Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
"SCIENZE BIOMEDICHE E BIOTECNOLOGICHE"

CICLO XXXI

COORDINATORE Prof. Paolo Pinton

_____

Novel RNAs in cancer: large scale analysis of Affymetrix Human Exon chips and Next Generation Sequencing

_____

Settore Scientifico Disciplinare BIO/17

Dottoranda
Dott.ssa Linda Minotti

Tutore
Prof. Stefano Volinia

_____ _____

(firma) *(firma)*

anni 2015/2018

INDEX

1 Introduction

1.1 Long non-coding RNAs classification, regulation and function

In the last years with the development of new molecular technologies, such as deep sequencing and DNA tiling arrays, the human transcriptome has been studied and characterized in its entirety. These techniques found that up to 70% of the genome is normally transcribed but only up to 2% serves as blueprints of protein genes [1]. In accordance with the latest version of Gencode (Version 28:November 2017,GRCh38-Ensembl 92,93; http://www.gencodegenes.org) only 34% of transcribed human genome sequences are composed of genes that encode for proteins. The remaining 66% are transcribed as non-coding RNAs (ncRNAs). Among the portion of the ncRNAs, 15779 are lncRNAs, which represent the major part of non-coding RNA regions ,Figure1.

Figure 1 Classification of human genes.

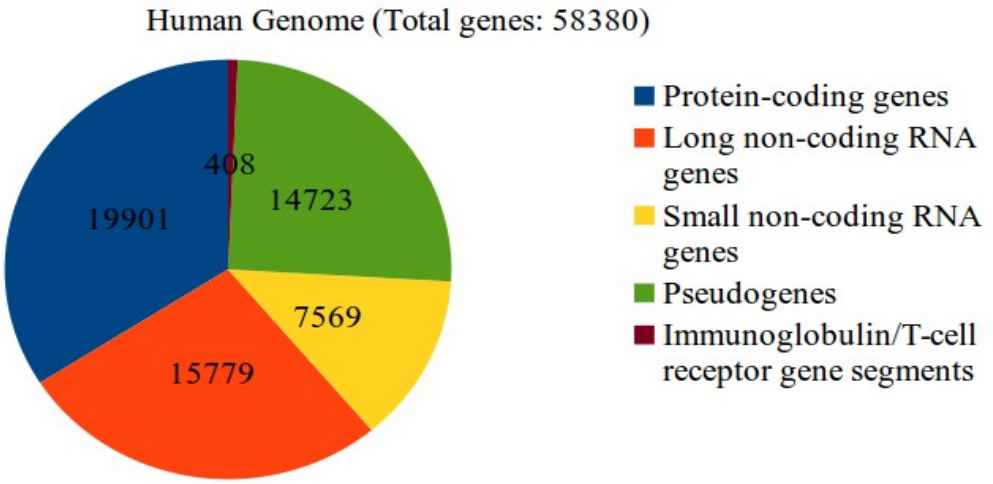


Figure 1 *Piechart of Gencode (V28) classification for the annotated 58380 genes of the entire human genome.*

LncRNAs are defined as more than 200 nt long and lacking an open reading frame of significant length, ie less than 100 aminoacids. LncRNAs were classified in 5 categories as described in the reference human gene annotation for the Encode project [2]. Their classification, showed in Figure 2, is based on their positions compared to host coding genes transcripts: i) sense overlapping, when are transcribed by the exons of coding genes in the same direction; ii)

antisense, transcribed in opposite strand of the known genes; iii) intronic, derived from the introns of host transcripts; iv) bi-directional, when their expression start in close genomic proximity of the near coding transcript expression start, situated on the opposite strand; v) intergenic, when thety are independent units within the genomic sequence between two genes.

Figure 2 Classification of long non-coding RNAs based on genomic position

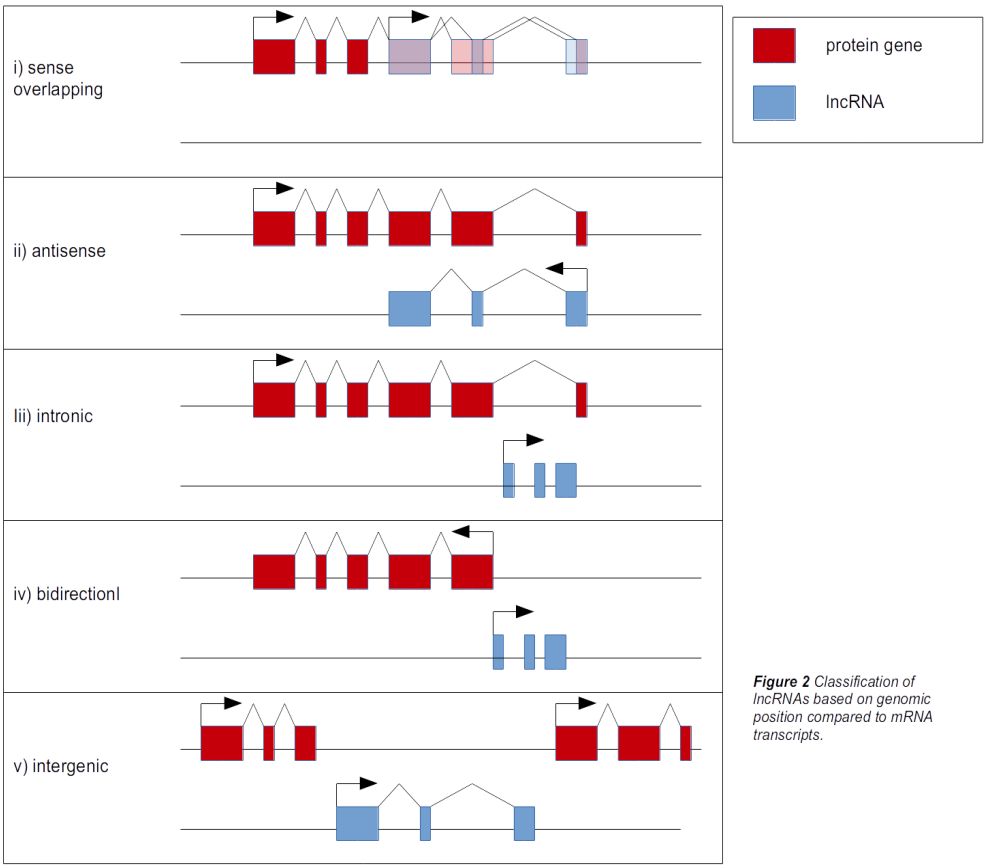


Figure 2 *Classification of lncRNAs based on genomic position compared to mRNA transcripts.*

The long intergenic non-coding RNAs (lincRNAs) are the largest group of lncRNAs in human genes according to GENCODE v28. LincRNAs expression was found to be closely associated with tissue-specificity compared to coding genes and they resulted ordinarily co-expressed with their neighbor protein genes [3]. LincRNAs don't overlap any known genes but they are highly conserved in mammals [4]. The second prevalent group of lncRNAs are the antisense RNAs which are transcribed from the opposite strand of the protein coding genes. They were also called natural antisense transcripts (NATs) that resulted partially overlapped with genes, primarily around the promoter or

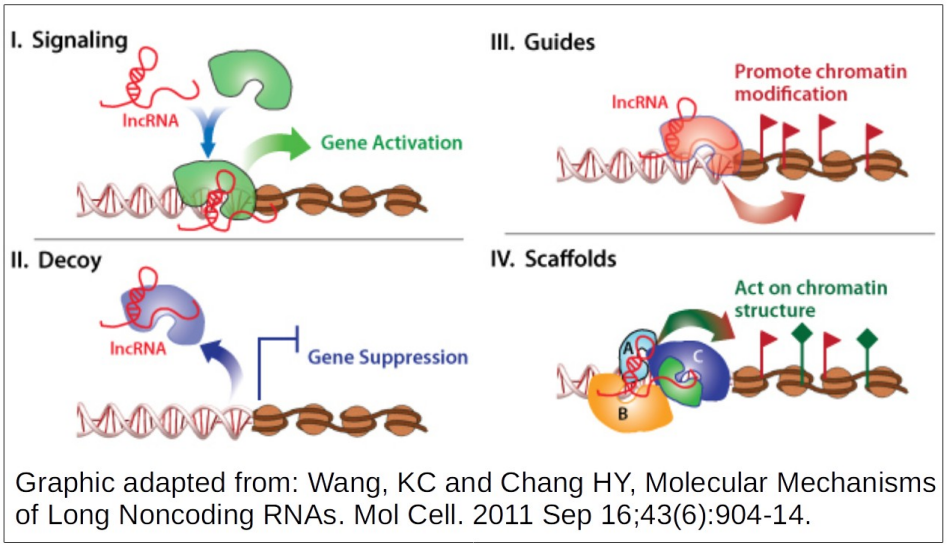
terminator sites [5]. Often NATs modulate the expression of the sense overlapped transcripts by the formation of duplex sense-antisense. It has been found that the sense-antisense pairs are typically co-expressed together and that they have a common pattern of evolutionary conservation [6]. The third class of lncRNAs includes the sense lncRNAs that are located on the same strand of the overlap genes and they are transcribed in the same direction but within intronic regions. This class is less common and to date poorly characterized when compared with other classes. Our recent studies are focused on a lncRNA situated in the first intron of the gene of the Transglutaminase 2 (TGM2), this lncRNA has been found correlated with its host gene [7]. Fourth class comprises a recently discovered type of lncRNAs, circular RNAs (circRNAs). CircRNAs form a covalently closed continuous loop without 5' or 3' polarities. The majority of circRNAs are abundant widespread, conserved and stable across different species, and show tissue and developmental stage specific features. They are generated through an alternative type of splicing, when a downstream splice donor is merged to an upstream splice acceptor through splice skipping or direct splice, a process named "back-splicing" [8]. The function of lncRNAs is difficult to study because it is implicated in a large number of post-transcriptional processes [9]. Nevertheless, several lncRNAs have been characterized and many patterns of action have been proposed including functions such as signaling, decoy, guides and scaffold Figure 3, furthermore enhancer RNAs and independent short peptides [10].

I. Signal lncRNAs act as a molecular signal to direct the transcription in presences of many stimuli and hence their presence could be interpreted as a measure of transcription dynamics [11].

II. Decoy lncRNAs regulate transcription by sequestering, through binding site regions, factors including transcription factors, catalytic proteins or their subunits, entire or parts of chromatin modifying complexes, as well as miRNAs [12].

III. Guide lncRNAs interact with ribonucleo-proteins complexes (RNPs) and targeting their enzymatic activities to the specific sequences of the genome resultingin them being essential for the proper localization of RNPs [13].

Figure 3 Classes of lncRNAs based on functional activity.



Graphic adapted from: Wang, KC and Chang HY, Molecular Mechanisms of Long Noncoding RNAs. Mol Cell. 2011 Sep 16;43(6):904-14.

IV. LncRNAs included in the scaffold class, play a structural role in providing an assembled platform for RNPs [14]. Depending on the nature of the proteins of the RNPs they could act both as transcriptional activators or repressors [15,16].

In addition, lncRNAs translated from enhanced regions, are considered as enhancer RNAs (eRNAs) that can influence the 3-dimensional organization of DNA (chromatin), acting as a connector for the protein of target enhancer regions [17]. Moreover, it was observed that lncRNAs encode for stable, functional short peptides (micropeptides), which may have their own specific functions [18].

1.2 Cancer type and lncRNAs related

Approximately 90% of **gastric** cancers (GC) are adenocarcinomas, which developed from the glands of the mucosa, the most superficial layer, of the stomach [19]. However, there are other types of cancer which arise from other types of tissue of the stomach, that in our dataset were classified as **stomach** [19]**.** The stomach group includes mucous-associated lymphomas, which originated from the lymphoid tissue of the stomach, leiomyosarcomas that arised from the muscles surrounding the mucosa and gastrointestinal stromal tumors, or carcinoid tumors. With about 1 million of GC diagnosis in worldwide population, of which about 700.000 died from this disease, GC was the 4th most common incident and the 2nd most common cause of cancer death [20]. GC was also considered as one of the highest cancer burdens, as measured by disability-adjusted life years

lost [21]. The incidence rate of GC varies between men and women and comparing different countries [22]. Helicobacter pylori infection has been considered as the most well-established risk factor of the GC [23]. Even though most gastric cancers are sporadic, approximately 10% show familial aggregation, that becomes a strong risk factor [24]. The radical treatment of GC is surgery, followed by traditional radiotherapy, chemotherapy, and the implementation of neoadjuvant therapy, that increase the 5-year survival rate of GC early diagnosed at 95% [25]. Comparing the expression levels of lncRNA AK023391, resulted in significant upregulation in gastric cancer samples then to adjacent normal tissues, and resulted correlated with poor survival in patients with gastric cancer [26]. With more than 1.3 million people diagnosed, **colorectal cancer (CRC)** is the third most common cancer in the worldwide population [27]. In 75%-80% of cases, CRC occurs sporadically as a result of the accumulation of both mutations and epigenetic modifications such as DNA methylation of several genes [28]. The stage at diagnosis is associated with the prognosis of CRC, which presents a 5-year survival rate of 90% at early diagnosis and a 5-year survival rate less than 10%, when are present distant metastases [29]. Accurate diagnosis and evaluation of treatment was based commonly on a colonoscopy, in which tissue specimens are collected and sent to a pathologist for diagnosis [30]. The correct diagnosis relies on the site of sample collection as well as examination of the pathology of tissue specimens using a detection strategy such as virtual colonoscopy [31], tests for DNA methylation markers in stool and fecal occult blood test [32]. Diagnostic, prognostic, biomarkers genes signature have been developed but are yet to be implemented into clinical trials because of several challenges [33]. In addition to genes associated with well-known pathways that are altered in CRC, microRNAs show potential as biomarkers of the disease [34]. Also lncRNAs were studied in CRC such as lncRNAs Breast Cancer Anti-Estrogen Resistance 4 (BCAR4), it increases colon cancer progression by enhancing cell proliferation and inhibiting apoptosis via BCAR4/β-catenin axis. Thus indicating that BCAR4 may be a useful new target for treatment of patients with colon cancer [35].

Pancreatic ductal adenocarcinoma (PDAC) is one the most deadly of all the solid malignancies tumors and is difficult to diagnose at its early stage [36]. The neoplastic glands have much differentiation between each other, this results in a difficulty to distinguish between a reactive non-neoplastic gland and a gland of

invasive adenocarcinoma [37]. Furthermore there is no effective therapy to PDAC, hence, it is very useful to discover diagnostic and prognostic molecular markers as they could be lncRNAs, such as MALAT1 [38]. **Hepatocellular carcinoma (HCC)** is a common cancer, its most prevalent risk factor remains as liver diseases such as chronic hepatitis virus infections and cirrhosis, caused by excessive alcohol consumption [39]. The incidences of mortality continue to rise, in spite of the development of prevention techniques, screening, and new technologies both in diagnosis and treatments [40]. Thus lncRNAs could be the new frontiers of the HCC studies as suggested by lncRNA-SVUGP2, that was found down-regulated in HCC samples and correlated with a better prognosis in patients with HCC [41]. **Lung** cancer with 80–85% being **non-small cell lung carcinoma (NSCLC)** is the leading cause of cancer-related death in both men and women with a very poor overall survival rate [42]. Historically, NSCLC was considered as one entity, and so standard care resulted on platinum-based therapy. However, with the discovery of the EGFR mutation and ALK rearrangements, the landscape of treatment has become more personalized [43]. With the development of new researches, an increasing number of roles that lncRNAs play in NSCLC have been found, and more and more evidence shows that lncRNAs have a close relationship with the patients' response to radiochemotherapy or molecular therapy [44]. The most frequent subtype of cancer that originates from **kidney** parenchyma is **clear cell renal cell carcinoma (ccRCC)** accounting for 70–80% of all malignancies [45]. The primary treatment of ccRCC today remains as the surgical removal of the affected kidney by complete or partial nephrectomy [46]. Lung cancer-associated transcript 1 (LUCAT1), a lncRNA was found upregulated in ccRCC tissues and renal cancer cell lines, and significantly correlated with the malignant stage and poor prognosis in ccRCC. Furthermore, LUCAT1 promoted proliferation and invasion in ccRCC cells through the AKT/GSK-3β signaling pathway and was induced by chemokine CXCL2. These findings indicate that the CXCL2/LUCAT1/AKT/GSK-3β axis is a potential therapeutic target and molecular biomarker for ccRCC [47]. **Serous epithelial ovarian cancer (EOC)** accounts for ~75% of subtypes that have no identifiable risk factors or precursor lesions, and only a few effective screening tools for early diagnosis are currently available [48]. EOC commonly originates from the ovarian surface epithelium (OSE) and/or ovarian inclusion cysts [49]. However, recently a novel hypothesis was proposed,

that high-grade serous ovarian cancer (HGSC), the most common histological subtype of EOC, develops from the Fallopian tubes [50]. A large number of lncRNAs presented a dysregulated expression in ovarian cancer and presented potential clinical implications [51]. **Prostate** cancer is one of the leading causes of mortality among males, the global burden of prostate cancer is rising but modification of lifestyle such as smoking cessation, exercise and specific diet offer opportunities to decrease the risk of developing prostate cancer [52]. Based on prostate-specific antigen the screening for prostate cancer remains controversial, owing to the high rate of over diagnosis and relative unnecessary prostate biopsies, despite evidence that it reduces mortality [53]. The common medical treatment approaches, both as monotherapy and in multimodal, include surgery, radiation therapy, chemotherapy, hormonal therapy and cryosurgery [54]. There is an urgent need for the development of sensitive and specific biomarkers for the early detection of prostate cancer to reduce overtreatment and accompanying morbidity. Regarding this, six differentially expressed lncRNAs in prostate cancer cell lines and patient samples compared to normal tissue were used. These markers were also successfully detected in patient urine samples and were found to be up-regulated when compared with normal (healthy) urine [55]. **Acute myeloid leukemia (AML)** is a heterogeneous disease caused by clonal expansion of myeloid progenitors (blasts) in the bone marrow and peripheral blood that is now cured in approximately 35%–40% of patients younger than the age of 60 years old [56]. However most patients with acute myeloid leukaemia (AML) die from the progressive disease after relapse; using deep sequencing techniques on primary and relapsed tumors, a phenomenon called clonal evolution has been characterized with both founding clones and novel subclones, impacting the therapeutic approach [57]. Many studies depict a landscape of important lncRNAs in AML and provide novel potential therapeutic targets and prognostic markers for AML treatment [58]. Furthermore, lncRNAs expression profiling can provide valuable information for improved risk stratification of AML patients [59]. **Breast cancer (BC)** represents the most common cancer affecting women worldwide. Research on BC management might benefit the development of new drugs, because it presents a heterogeneous nature [60]. BC represents about 25% of all female cancers, with nearly 1.7 million new cases diagnosed in 2012 with an incidence rate that has been predicted to rise further do to behavioral and

environmental changes that greatly affect the main risk factors [61]. The overexpression of lncRNAs CDKN2B-AS1, DSCAM-S1 and H19 in BC cells validate the relevance of the dysregulation pattern in cancer cells due to the presence of lncRNAs. The study further opens a new scope for experimental analysis to confirm the aberrant expression pattern of these lncRNAs which may act as potential bio-markers for the diagnosis and early detection of breast cancer [62]. The most common cancer of the **central nervous system** is the glioma that on the basis of its morphological appearance was classified as astrocytoma, oligodendroglioma, or ependymoma [63]. Astrocytomas express glial fibrillary acidic protein, an intermediate filament found in astrocytes that is routinely used as an aid in classifying a glioma as an astrocytoma [64]. The prognosis of glioma patients remains poor caused by the lack of effective means for diagnosing and treating glioma. Therefore, understanding the molecular mechanism of glioma progression is essential for effective treatments. Recent evidence indicates that lncRNAs may play important roles in regulating the progression of glioma, giving us the prospects of lncRNAs as diagnostic and prognostic biomarkers and therapeutic targets for glioma [65]. **Cutaneous melanoma** is the more aggressive form of the disease, occurring when melanocytes undergo changes and become malignant; this is less common than basal cell and squamous cell carcinomas, but is the most dangerous skin cancer [66]. Early diagnosis is very important because surgical excision is curative in approximately 99% of patients that have early stage of melanoma, before invading the dermis. The main treatment for melanoma currently is surgery, followed by radiotherapy and chemotherapy. Surgery can provide efficient tumor resection, if there are no metastases. Radiation therapy is used in severe cases and in conjunction with surgery, to increase the efficiency of treatment. Over the last 30 years, no drug or combination of drugs have demonstrated significant impact in improving patient survival which makes melanoma a challenge to treat [67]. Six differentially expressed lncRNAs were identified in different melanoma stages (I, II, III and IV) resulting in prognostic signature for risk-stratifying patients with melanoma [68]. More than 90% of tumors in the **head and neck** are **squamous cell carcinomas** which can arise in several places, often preventable, and if diagnosed early are usually curable. Unfortunately, patients often present an advanced stage of the disease that is incurable or requires aggressive treatment, which leaves them functionally

disabled [69]. Stopping smoking and drinking less alcohol are the primary preventions and are the most effective way to reduce mortality. In spite, of early detection should be a priority, given the excellent prognosis of early stage disease compared with the poor results in advanced stages, screening is the most cost effective when targeted at high risk groups such as heavy drinkers and smokers. Diagnosis is confirmed by biopsy of the primary site and fine needle aspiration of any enlarged lymph nodes. An increasing number of studies have shown that head and neck cancer is a genetic disease in which many oncogenes and tumor suppressor genes participate in synergistic processes involving many stages and pathways [70]. LncRNAs associated with head and neck cancer play important roles in prognosis, treatment, and prevention, making lncRNAs a good prognostic marker and treatment target to increase the expectancy and quality life of patients with head and neck cancer [71].

1.3 Microarray technology

Microarray is one of the high throughput technologies which is used to measure the expression of thousands of genes at the same time, also called genome chip, DNA chip or gene array. It is commonly composed of microscopic spots printed in specific positions of a solid surface such as plastic, glass or silicon biochip, showed in Figure 4. Each spot contains thousands of identical and specific sequences of DNA, that act as probes to detect specific gene or exon loci. To perform expression analysis, the RNA is extracted from the cells or tissue, fragmented and then copied to cDNA by reverse transcription. The fragments of cDNA are labeled with fluorescent molecules and add to the ordered array, then the fluorescence intensity is measured across the array, indicating the abundance of a predetermined set of sequences. Microarrays can be constructed with different approaches, depending on the type of scientific question that must be answered, affecting the number of probes under examination, costs, and the possibility of customization. Commonly microarrays for commercial use present a number of probes between 10 up to 5 million or more. Many companies produce microarray chips, of these Affymetrix is one of the most known. Canonical arrays contain probe designs on the few hundred bases proximal to the 3’ end of the known genes, and use expression at the 3’ end to approximate expression of the entire gene, assuming that the 3’ end of each gene is specifically defined, that

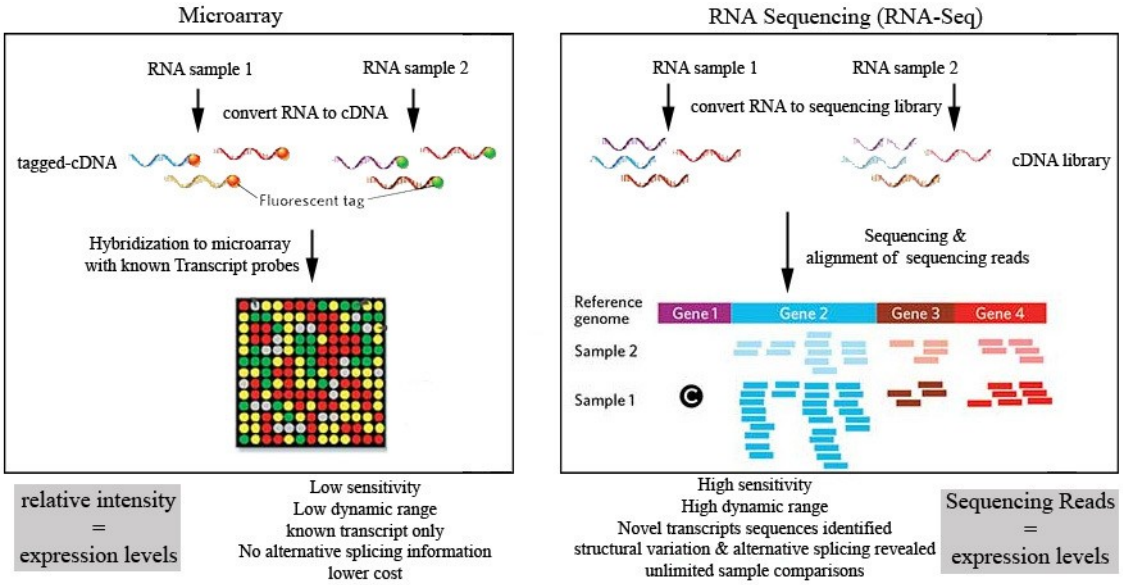
each transcripts present intact poly-A tail and that the entire length of the gene is expressed as a single unit [72]. More than 60% of genes present alternative splicing that produce thousands of transcript variants with potential distinct functions. About 50% of somatic point mutations disease-related cause changes in the pattern of splicing and 20% of cancer-related mutation cause the exon skipping events. Classical 3'-end expression microarrays cannot discriminate between the alternative spliced transcript that may have identical 3' ends. Furthermore, 3' arrays cannot detect transcripts that lack 3' exon, poly-A tail or that present genomic deletion or insertion. In the whole transcript array, the probes are designed along the entire length of the genes, to detect the differences between transcript variant [Affymetrix Application Note, Whole-transcript Expression Analysis]. Today, the densest array designed for profiling gene expression is Affymetrix Human Exon chip [73]. In this array probes targeting each exon are supported by two main sources. One are probes containing sequences from high confidence mRNA from RefSeq and GenBank databases, and from sequence tag evidence from dbEST. The other, are probes with a lower confidence set of purely computational predicted using software such as GENESCAN, TWINSCAN and Exoniphy [73,74]. The aim of the design was to include probesets for every potential exon on the entire human genome. Across the genome more than one million exon clusters were identified and grouped into transcript clusters. The project was intended to take alternative 5' or 3' splice site into consideration [75]. The Human Exon array contains over 6.5 million of probes clustered into 1.4 million of probesets that overlapped exon of genes. A probesets is therefore a Probe Selection Region (PSR) that represents a region of the genome predicted to act as an integral and coherent unit of transcriptional behavior. The median size of PRSs is 123 bp with a minimum of 25 bp and about 90% of them are represented by 4 probes. This redundancy allows the estimation of the presence of signal, relative expression and existence of alternative splicing [76]. With approximately 50% of probesets that target loci outside of Ensembl-defined protein coding exon, Exon Human array share the potential advantages of RNA-sequencing technology (RNA-seq) to detect transcription outside well-characterized loci [77].

1.4 RNA-seq technology

With Next Generation Sequencing technology (NGS) also called high-throughput sequencing, an entire human genome can be sequenced in a single day, revolutionizing the way that genomic researches advance [78]. In addition to the static genome, transcriptomics is the study of the complete set of transcripts found in a specific cell, tissue or organism for a given developmental stage or physiological condition [79].State of the art NGS has recently been utilized to study the dynamic transcriptome and the resulted technology is called RNA sequencing (RNA-seq), it is free from many of the limitations of other transcriptomic approaches, such as microarray and tag-based sequencing method [80]. With RNA-seq we would be able to measure the abundance of RNA and the resulting data could be interpreted in terms of transcriptional activity and nucleo acid stability. In the RNA-seq assays a population of RNA from a single sample, such as fresh or frozen tissue, Paired Formalin-Fixed Paraffin-Embedded (FFPE) tissue or cells lines, is converted to a library of cDNA fragments, see Figure 4, with adaptors attached to one end (single-end sequencing) or both ends (paired-end sequencing) [81]. The two main RNA-seq approaches for gene quantification in clinical RNA-seq are polyA+ selected or ribo-minus rRNA depletion that respectively reflect the mRNA pool or total RNA [82]. Eachof the fragments of the library of cDNA is sequenced in a high-troughput manner to obtain short molecules from single-end or paired-end, that are sequenced, called reads. The reads are 30-400 base-pair (bp) typically in length, depending on the sequencing machine used, in the long RNA-seq experiments the average reads size were greater than 200 bases. Albeit genomes are composed as linearly ordered sequencing of the nucleoacids, the transcriptome reorganized the information by splicing together non-contiguous exons to create mature transcripts [83]. The detection and characterization of these spliced RNAs was a specific crucial focus to the RNA-seq analysis, that needed the connectivity information about the full extent of splicing. To reconstruct the mature mRNA, the reads of the spliced RNA must have join together, these alignment challenges are further hindered by the presence of multiple copies of identical related genomic sequences. To combat these challenges various sequence alignments have been developed such as: Bowtie2 [84], BWA [85], TopHat2 [86] and STAR [87]. Nevertheless, the application of these algorithms negatively effect some areas of mapping accuracy, like sensitivity

and precision and of computational resources such run time and disk space usage [88].

Figure 4 Representation of Microarray and RNA-seq technology



1.5 Public datasets and projects

The National Center for Biotechnology Information (NCBI) contains the Gene Expression Omnibus (GEO), a public archive where there are stored big volumes of published high-throughput functional genomic data generated by the international research community [89]. As technologies advance, large-scale datasets become easier and cheaper to generate, but a major challenge remains upon how to integrate these datasets with each other [90]. The integrated Encyclopedia of DNA Elements (ENCODE) started in 2004 as a Pilot Project on 1% of the entire human genome and in 2007 the attempt was extended to the whole-genome assay [91]. ENCODE is thought to create a comprehensive catalog of all functional elements in the human genome, to enable scientific and medical communities to interpret and apply the knowledge in order to understand human biology and improve human health [92]. A functional element is defined as a discrete genome segment that encodes a defined product such as protein or non-coding RNA, and displays a reproducible biochemical signature, such as a protein-binding domain or a specific chromatin structure [93]. The ENCODE Consortium combine groups that share collective multiple technologies and approaches to

discover and classify the functional elements such as genes, transcripts and transcriptional regulatory regions, together with chromatin status and DNA methylation pattern [94]. All the data produced from the ENCODE Consortium are available at the ENCODE portal (www.encodeproject.org), curated, processed and validated by ENCODE Data Coordinator Center (DCC) [95]. The ENCODE data was also shared through the UCSC genome browser, GEO and Ensembl and adopted as common resource by the scientific community [96].

Reference Sequence (RefSeq) is an NCBI database that collects a curated, non-redundant set of reference sequences including genomic DNA contigs, mRNAs and proteins for known genes, and entire chromosomes. RefSeq provides a foundation for uniting sequence data with genetic and functional information and presents a synthesis of information integrated across multiple sources at a given time, having a role similar to a review article. The RefSeq collection is available without restriction and can be retrieved in several different ways, such as by searching or following available links in NCBI resources, including PubMed, Nucleotide, Protein, Gene, and Map Viewer, or searching with a sequence via BLAST, and downloading from the RefSeq FTP site [97]. Furthermore RefSeq could be visualized in GenomeBrowser [98].

The University of California Santa Cruz (UCSC); Known Genes track is a set of gene predictions based on data from RefSeq, GenBank, CCDS, Rfam, and the tRNA Genes track. Both protein-coding genes and non-coding RNA genes are included in UCSC tracks. For the transcripts of protein-coding genes they require at least two supports from RefSeq RNA or GenBank RNA sequences as well as one additional line of evidence. This is the same of the transcripts of non-coding RNA genes that require the support of one Rfam or tRNA prediction. This is a more conservative set of predictions compared to RefSeq. UCSC set generally has about 10% more protein-coding genes than RefsEQ and approximately four times as many putative non-coding genes [99].

The GENCODE Project: Encyclopedia of genes and gene variants is a consortium that aims to identify all gene features in the human genome using a combination of computational analysis, manual annotation, and experimental validation [100]. Few new protein-coding loci have been added since the first public release of this annotated data set, yet the number of alternative splicing transcripts annotated and ncRNAs has steadily increased [2]. With the new long-read sequencing

technologies this allows an improvement of current annotation, opening the way towards a complete annotation of lncRNAs expressed throughout a human lifetime [101].

The aim of the Genotype - Tissue Expression (GTEx) Project is to improve the health care of future generations by increasing the understanding of how changes in our genes contribute to common human diseases. Reagrding this GTEx researchers are now studying genes in different tissues obtained from many different people [102].

1.6 Tools and pipeline commonly used to discover novel lncRNAs

In the last decade canonical tools were used to search novel lncRNA based on RNA-seq data, summarized in Table 1. These were characterized based on 4 features: Protein-Coding Potential Predictor, lncRNA predictor, alignment-based or alignment-free.

Among the Protein Coding Predictors based on alignment we found CPC and COME. Coding Potential Calculator (CPC) can predict lncRNAs by investigating the open reading frame (ORF) information and alignment results [103]. COME can identify and characterize novel lncRNAs employing a decompose-compose method that integrates multiple sequence-derived and experiment-based features [104]. Between the Protein Coding Potential Predictors alignment-free there was Coding Potential Assesment Tool (CPAT) that can discriminate from coding protein genes and non-coding with high accuracy calculating Fickett TESTCODE score [105] and CPC2, an upgrade of CPC [106]. From the group of the lncRNA Predictors we can distinguish LncRNA_ID and lncRScan_SVM that are alignment-dependent. LncRNA_ID uses random forest (a machine learning model) based on various features like sequence characteristics of putative open reading frames, translation scores based on ribosomal coverage, and conservation against characterized protein families [107]. Using a support vector machine (SVM) lncRScan_SVM aims to classifying lncRNAs, using multiple features derived from Gencode datasets: gene structure, transcript sequence, potential codon sequence and conservation [108]. The lncRNA Predictors alignment-free are CNCI and PLEK. Coding Non-Coding Index (CNCI) is constructed on a matrix of adjoining nucleotide triplets (ANTs) and unequal distribution of codon (codon bias) that is useful for classifying incomplete transcripts and sense-antisense pairs. [109].

Table 1 Tools and pipeline related to lncRNAs

Tools/ Pipeline	link tools/pipeline	Type of data	type of cancer	link dataset	PMID/doi
CNCI	http://www.bioinfo.org/software/cnci	RNA-seq	not specific	not specific	23892401
COME	https://github.com/lulab/COME	RNA-seq	not specific	not specific	27608726
CPAT	http://rna-cpat.sourceforge.net/	RNA-seq	not specific	not specific	23335781
CPC	https://github.com/biocoder/cpc	RNA-seq	not specific	not specific	17631615
CPC2	http://cpc2.cbi.pku.edu.cn	RNA-seq	not specific	not specific	28521017
DeepLNC	http://bioserver.iiita.ac.in/deeplnc	RNA-seq	not specific	not specific	doi.org/10.1007/s13721-016-0129-2
FuncPred	www.funcpred.com	RNA-seq	not specific	GTEX projects	28361701
LncFinder	https://CRAN.R-project.org/package=LncFinder		not specific	not specific	30084867
LncmiRSRN	showed in the paper	Afymetrix Human Exon chip	prostate; glioblastoma; ovarian; lung	GEO: GSE21034; TCGA	29955818
LncRNA-ID	https://github.com/zhangy72/LncRNA-ID	RNA-seq	not specific	not specific	26315901
lncRNA-screen	https://github.com/NYU-BFX/lncRNA-screen	RNA-seq	not specific	TCGA	28583068
LncRNAnet	http://data.snu.ac.kr/pub/lncRNAnet	RNA-seq	not specific	not specific	29850775
lncRScan-SVM	http://sourceforge.net/projects/lncrscansvm/?source=directory	RNA-seq	not specific	not specific	26437338
PLEK	https://sourceforge.net/projects/plek/files/	RNA-seq	not specific	not specific	25239089
UClncR	http://bioinformaticstools.mayo.edu/research/UClncR	RNA-seq	not specific	not specific	29079769

Table 1 *List of the tools and pipeline useful to discover lncRNAs*

PLEK is a tool that distinguishes long non-coding from messenger RNAs using a computational pipeline based on improved k-mer scheme and a support vector machine (SVM) algorithm in the absence of genomic annotations [110].

Ultrafast and Comprehensive lncRNA detection pipeline takes standard RNA-seq alignment file, and performs transcript assembly, predicts lncRNA candidates, quantifies and annotates both known and novel lncRNA candidates, and generates a convenient report for downstream analysis. The pipeline works on both un-stranded and stranded RNA-seq thus it is useful to the prediction of lncRNAs that overlap other genes [111]. There is a study that integrates heterologous features in order to discover novel lncRNA. This study designed a new platform named LncFinder that extracts information not only from sequence-intrinsic composition but also from secondary structure and physicochemical properties [112]. In 2018 the capability of RNA-seq was demonstrated in the discovery of 5 lncRNA involved in colorectal cancer. This set of research used a whole transcriptome analysis to search for novel lncRNA and used TCGA public dataset to validate the results [113]. An open source comprehensive pipeline (*lncRNA-screen*) for lncRNA computationally discovered putative lncRNA transcripts over large multimodal datasets and was recently published. LncRNA-screen provides an automated solution which performed all the stages of data analysis from RNA-seq: data download, alignment, assembly, quality assessment, transcript filtration, novel lncRNA identification [114]. Among the hypothesis of ceRNAs, a new method was proposed to construct lncRNA related miRNA sponge regulatory networks (LncmiRSRNs) through the integration of matched lncRNA and mRNA expression profiles with clinical stage information and putative miRNA-target interactions. They proposed again publically available array-based data to extract expression data of lncRNAs and to link its expression to clinical stage. They re-annotated the Affymetrix Human Exon chip probes, to discard those that mapped protein coding genes and retain those that uniquely mapped lncRNAs [115]. A group investigated the correlation between gene expression profile and long non coding RNA expression, supposing that if a correlation existed they would probably present the same function. They exploited the knowledge of protein function onto non coding genes using the principle of association. Using different types of public data they can predict putative annotations of lncRNAs for cancer [116]. DeepLNC is a lncRNAs fast and accurate alternative prediction tool based on deep neural

network (DNN). DeepLNC is based on a unique feature for DNN classifier, k-mer pattern stores information previously manually annotated training datasets from LNCipedia and RefSeq. The k-mer information content generated on the basis of Shannon entropy function has resulted in an improved classifier accuracy [117]. A deep learning-based approach, lncRNAnet, is designed to identify lncRNAs that incorporate recurrent neural networks for RNA sequence modeling and convolutional neural networks for detecting stop codons to obtain an open reading frame indicator [118]. As we see in Table 1 the major parts of the tools applied to discover novel lncRNAs are based on RNA-seq. But RNA-seq had a bigger cost in terms of money and time to analyse, compared to the microarray approach. Thus, we thought of a method to discover novel lncRNAs taking advantage of Affymetrix microarray Human Exon chip and more of the available public dataset.

2 Thesis objective

My PhD was focused on the discovery of new lncRNAs implicated within cancer pathways. The difficulties of lncRNAs discoveries are that they show post transcriptional modification similar to the coding genes, like poly A tail. Furthermore, they are tissue specific and conserved as coding genes, but they are less transcripted. LncRNAs control multiple cellular processes and they are involved in many cancer pathways. Thanks to the development of new technologies, for the study of human genome and transcriptome, a large quantity of data is now public on the web. These data were produced from different laboratories worldwide. Is it possible to take advantage of this amount of free data to discover novel lncRNAs involved in cancer?

The objective of this thesis was to answer that question, therefore questioning whether public datasets of microarray are useful to the study of cancer, and in particular to discovery of novel lncRNAs involved.

3 Materials and methods

3.1 R

R is a language and environment for statistical computing and graphics, belonging to GNU Project software. The GNU operating system (https://www.gnu.org/home.it.html) consists of programs specifically released as free software packages both by the GNU Project and by third parties. The term "environment" means to characterize it as a completely planned and coherent system. R is an integrated suite of software facilities for data manipulation, calculation and graphical display. Indeed R provides a wide variety of statistical analysis such as linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering and more. Furthermore, R furnish a lot of graphical techniques to summarise and show the resulted data in an easy way. It includes an effective data handling and storage facility through the definition of variables. It provides a series of packages that contain operators for calculations on arrays, in particular matrices and a large, coherent, integrated collection of intermediate tools for data analysis. It supplies graphical facilities for data analysis and displays it either on-screen or saves a copy on a local machine. It is used through a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

3.1.1 Oligo Package

Oligo is a Bioconductor package used for preprocessing oligonucleotide microarrays, starting from the native format of the file output produced by the Affymetrix. Bioconductor offers tools for the high-throughput analysis and classification of the genomics data (http://www.bioconductor.org). It is based on R statistical programming languages, free open source and open development. The Oligo package allows users to preprocess microarray data using R, this is useful approach to combine tools already designed in R for downstream analysis and visualization. The Oligo software is implemented to support the analysis of large datasets, providing parallel execution of common tasks such as background subtraction, normalization and summarization [119]. In accordance with the Oligo user guide we used Affymetrix CEL files of all the samples of the dataset that we

wanted included in the analysis, in order to normalize the entire dataset at once. A CEL file is a data file created by Affymetrix DNA microarray software that contains the raw expression intensity values of the "probes" of an Affymetrix Chip and stores thousands of data points, which may make it a large file size. See R script for the command line used for this normalization. First we download the raw.data from GEO and stored it in a common folder. With the command “list.celfiles” we listed the entire number of samples of the common folder. Oligo also required the package “pd.huex.1.0.st.v2” that was used to annotate the probes on the exons probesets. Then the command “read.celfiles” was used to import and read all the CEL files with the annotated packages in order to create an object class named affyExonFS. AffyExonFS is an ExonFeatureSet object that was implemented by Oligo for the exon expression classes. Robust multi-array average algorithm (RMA) was used to create the expression matrix from the raw intensity values [120]. RMA performed background correction, log2 transformation and quantile normalization. Next, a linear model was fitted to the normalized data to obtain an expression measure for each probe set on each array, which is summarized in an expression matrix. Working with ExonFeatureSet object we processed the raw data of the probes to the probesets level providing the expression summaries at the exon level with the command “target=probeset”). For Exon Arrays, Affymetrix supplied an additional annotation file that could be used to summarized the raw data to the gene level. The probesets design in Affymetrix Human Exon chip are classified in three groups (core, extended, and full) depending on the the level of confidence of the sources used to generate the annotations.

For the GeneChip Human Exon 1.0 ST Array, the annotation sources for each level are:

Core Gene Annotation sources
• RefSeq alignments
• Genbank alignments of ‘complete CDS’ transcripts

Extended Gene Annotation sources
• cDNA alignments
• Ensembl annotations (Hubbard, T. et al.)

• Mapped syntenic mRNA from rat and mouse
• microRNA annotations
• Mitomap annotations
• Vegagene (The HAVANA group, Hillier et al., Heilig et al.)
• VegaPseudogene (The HAVANA group, Hillier et al., Heilig et al.)

Full Gene Annotations
• Geneid (Grup de Recerca en Informàtica Biomèdica)
• Genscan (Burge, C. et al.)
• GENSCAN Suboptimal (Burge, C. et al.)
• Exoniphy (Siepel et al.)
• RNAgene (Sean Eddy Lab)
• SgpGene (Grup de Recerca en Informàtica Biomèdica)
• TWINSCAN (Korf, I. et al.)

Core types probesets were named so because of the annotation was established on the validate annotation of the known genes. The extended types probesets derived its name from the sense that their annotation would extend boundaries of the core annotation of genes. The idea behind the name of the full group was that it could contain all putative genes, also those derived from predictive database. To summarize, the raw data to the gene level we used the command “rma” on affyExonFS object with “target = core” to obtain expression values of the known validated genes. Then with command “exprs” we extrapolated the expression values from the objects probesetsSummaries and geneSummaries and saved it as R dataframe. The R data.frame is a matrix contained in the column of the single samples and in the rows the normalized log2 expression values of single probeset or gene.

R script to preformed Oligo preprocessing microarray

```
source("https://bioconductor.org/biocLite.R")
biocLite("oligo")
biocLite("pd.huex.1.0.st.v2")
library(oligo)
library(pd.huex.1.0.st.v2)

celFiles <- list.celfiles()
affyExonFS <- read.celfiles(celFiles,pkgname="pd.huex.1.0.st.v2")
probesetSummaries <-rma(affyExonFS, target = "probeset")
probeData <-exprs(probesetSummaries)
probeData <- as.data.frame(probeData)
write.table(probeData,file = "probeData_Expression.txt", quote = FALSE, sep="\
t")

geneSummaries <-rma(affyExonFS, target = "core")
geneData <-exprs(geneSummaries)
geneData <- as.data.frame(geneData)
write.table(geneData,file = "geneData_Expression.txt", quote = FALSE, sep="\t")
```

3.1.2 Aroma package

Aroma package needed a specific structure of folder to perform the analysis [Henrik Bengtsson, Ken Simpson, James Bullard, Kasper Hansen “aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory [121]. As showed in Figure 5 the main folder named aroma, contains two other folders, called annotationData and rawData, respectively. The name of these three folders is specific and obligatory to be recognized from aroma. Aroma knows that in rawData it can find the raw.CEL file and in annotationData can search for cdf file and storage output. Furthermore, in the annotationData folder it stores the temporary file to save the progression of the analysis. The temporary file was then automatically deleted when the final output file was produced.

For the exon array analysis carried out here we need to be able to map transcript cluster IDs to exon IDs. For this reason, we cannot use the default CDF provided by Affymetrix , as we did with Oligo, which only has information on exon IDs but not on transcripts. Instead, we used custom CDFs that map transcript cluster IDs to exon IDs according to Affymetrix's definition of 'transcript clusters' [122]. The

custom CDF was downloaded from Affymetrix.aroma web page (http://aroma-project.org/), each unit corresponds to a transcript cluster and each group within a unit corresponds to an exon/probeset. Aroma offers different types of cdf depending on the level of analysis:

- HuEx-1_0-st-v2,coreR3,A20071112,EP.CDF - Core probesets: 18,708 units/ transcript clusters, 284,258 groups/probesets, and 1,082,385 probes.

- HuEx-1_0-st-v2,extendedR3,A20071112,EP.CDF - Extended + Core probesets: 147,476 units/transcript clusters, 804,085 groups/probesets, and 3,095,094 probes.

- HuEx-1_0-st-v2,fullR3,A20071112,EP.CDF - Full + Extended + Core probesets: 297,051 units/transcript clusters and 1,381,294 groups/probesets.

- HuEx-1_0-st-v2,mainR3,A20071112,EP.CDF - All 'main' design probesets (so includes 'free' and 'ambiguous' probesets): 312,355 units/transcript clusters and 1,400,703 groups/probesets.

We used the mainR3 version to perform our analysis because we wanted to work with all the exon probesets present in the chip. Therefore in the annotationData folder we needed to create another folder with the name of chipTypes, which in turn, needed to contain the folder with the name of cdf used. In the other branch of the scheme we must have a rawData folder that contains in turn the folder with the name of the platform. Every platform folder must contain another folder with the named cdf used, and the raw.CEL files. This structure allowed the performance of a big number of different analysis with the same script, changing every time the name of the folder in which the rawData and cdf data were contained. Furthermore, this permits the specific combination of cdf and rawData to perform analysis, that is recommended to change the "tag" of results every time we switch CDFs.

Figure 5 Aroma Package Work Scheme Structure

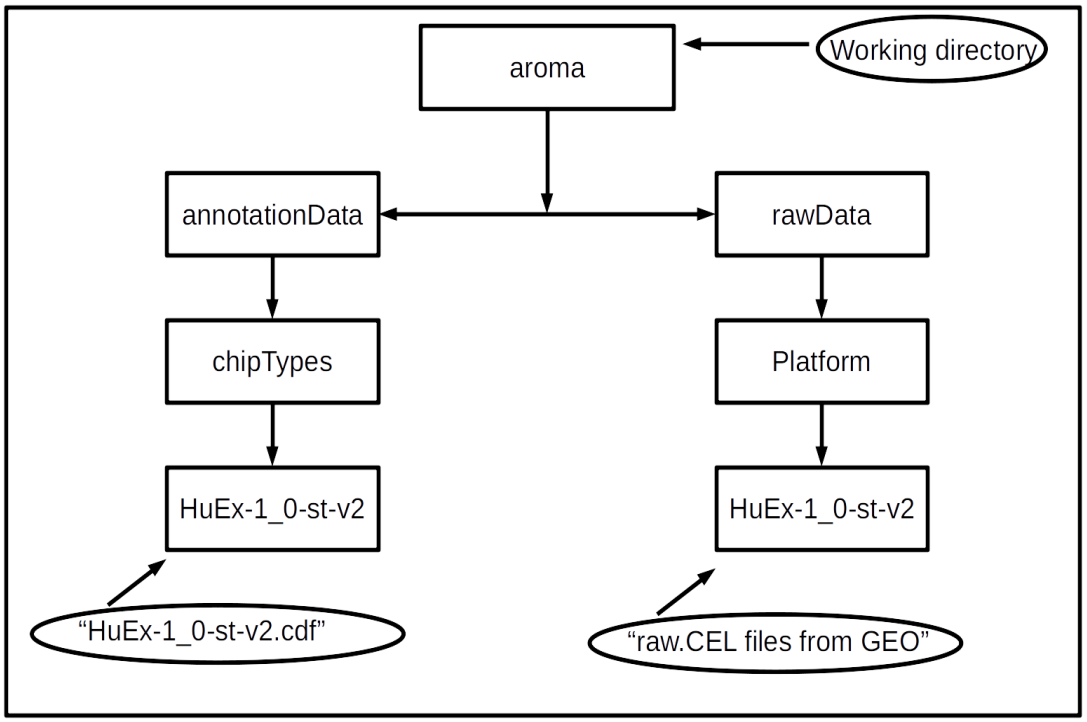


Figure 5 *Structure needed from Aroma to performed analysis. Rectangular shapes represented the folder, the oval forms displayed explication notes.*

Rscript used to performed preprocessing analysis of data are show in the APPENDIX A at the end of thesis. After we called “aroma.affymetrix” package in R we set the multiprocessing analysis to use each of the 8 CPU cores of the Linux server. Then we set up the annotation data defined by the custom cdf, named “cdf” with the specific “tag” of each output files. Next we setup “cs” the raw.CEL set and associated the “cdf” to the “cs”. Then we performed background adjustment and quantile normalization. These processes took approximately half an hour per arrays. Then we performed summarization fit exon-by-exon, a process that took around one hour per array. For the tumor/paired dataset we also performed the summarization to fit the entire transcript that estimated the overall expression of the “core” probesets of each of the known transcripts.

3.1.3 Other R package used

In the first part of pipeline we used a **genefilter** package of Bioconductor to perform the IQR analysis on the tumor/normal dataset [123]. Always for the first part of the pipeline we used **limma**, a Bioconductor software package that

provides integrated facilities for reading, normalizing and exploring data from gene expression experiments [124]. We used limma to discover the differential expression probesets between tumor and normal samples [125]. In the pipeline we used limma with paired dataset. For paired dataset we intended that the tumor tissue and normal adjacent tissue came from the same patient. Thus, when we compare the two groups we must compare every tumor tissue sample with the specific paired normal tissue sample. To do this we performed the analysis in blocks of two sizes. In this case a paired t-test was associated to statistical significance. Hence we had to create two factors; one with the information of the paired and the other with the information about the status (tumor or normal). Then we combined the two factors in a model matrix that was used by limma to perform the differential analysis. This passage was not necessary when we performed limma analysis in the unpaired dataset. For both of the two type of analysis, paired and unpaired we used lmFit to estimate the fold changes and standard errors by fitting a linear model for each of the probesets. The design matrix indicates the status of the group of arrays (tumor or normal) and in the case of paired analysis the arrays couples. The command ebayes were used to apply empirical Bayes smoothing to the standard errors. Next with the command topTable we extracted the output file that contains 6 columns: logFC value, AveExpr, t ,p.value, adjusted p.value, B. LogFC represents a log2-fold change (FC) between the two status groups. The AveExpr column is the average log2-expression level for that probesets across all the samples. Column t gives the moderated t-statistic. Column P.Value is the associated p-value and adj.P.Value is the p-value adjusted for multiple testing. In our case the adjustment was based on Benjamini and Hochberg's method "BH" which controls the false discovery rate. The B-statistic (lods or B) is the log-odds giving the probability that the probesets are differentially expressed. A B-statistic of zero corresponds to a 50-50 chance that the probesets is differentially expressed. The B-statistic is automatically adjusted for multiple testing by assuming that 1% of the probesets, are expected to be differentially expressed.

Harrell Miscellaneous (**Hmisc**) package contains many functions useful for data analysis, high-level graphics, utility operations, functions for computing sample size and power, advanced table making, variable clustering and character string

manipulation. We used its function rcorr to perform correlation analysis between Census Cancer Genes and orphan validated probesets.

3.2 Python

Python is an open source programming language and it is good for beginners. It is freely accessible in Python Software Foundation web page (https://www.python.org/), where documentation, tutorials, and guides are constantly evolving. Python is an interpreted language, meaning that we can change the code and quickly see the results, making it easy-to-read and powerful. It is a high-level language, which means that we could focus on what to do instead of how to do it, furthermore writing programs in Python takes less time than in some other languages such as C++, Java or Perl [126,127].

We used it to construct the third part of pipeline to discover orphan probesets cancer related. We make a Pearson correlation analysis with SciPy between orphan probesets and flanked regions. ScyPy is a collection of mathematical algorithms and convenience functions built on the Numpy extension of Python [128].

3.3 Dataset

3.3.1 Paired cancer dataset

For the analysis a dataset of human cancer and normal paired samples were constructed and analyzed with Human Exon1_0ST Affymetrix Chip obtained from public Gene Expression Omnibus Database (GEO). Table 2 describes the paired dataset which consist of 179 tumor and 179 normal adjacent tissue from 9 different types of cancer. In the group named “Stomach” all cancer that can be found in the stomach wall, including lymphomas, gastro intestinal stromal tumors and carcinoma tumors, were clustered together. The term “Gastric” is referred to only Adenocarcinomas arising from gastric epithelium. Raw. CEL files (ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSEnnnnn/) were used to create a normalized expression matrix with program language R. Aroma.package was applied to perform normalization, background correction, quantile normalization and robust multi-array average (RMA) approach. The annotation for cluster

transcripts is provided by Affymetrix (https://www.affymetrix.com/support/technical/byproduct.affx?product=huexon-st).

Table 2 Samples of Paired cancer dataset

Tissue	Type Tumor	Series accession GEO dataset	Number of samples
gastric	gastric cancer with different grade (mucosa)	GSE27342 GSE33335 GSE63089	70
stomach*	lymphomas, gastrointestinal stromal tumors, and carcinoid tumors	GSE30727	60
colon	colorectal cancer (CRC)	GSE21962 GSE77434	46
pancreas	pancreatic adenocarcinoma	GSE60646	20
liver	hepatocellular carcinoma (HCC)	GSE12941	20
lung	non-small cell lung carcinoma (NSCLC)	GSE12236 GSE22862	70
kidney	clear cell renal cell carcinoma (ccRCC)	GSE47032	20
ovarian	epithelial ovarian cancer (EOC) with high-grade serous carcinoma (HGSC)	GSE69429	12
prostate	prostate cancer	GSE71781	40
total			358

Table 2 *Dataset description of the paired samples obtained of normal and tumor tissues analyzed with HumanExon1_0-ST Affymetrix. *Stomach: The types of cancer that can be found in the stomach include lymphomas, gastrointestinal stromal tumors, and carcinoid tumors.*

3.3.2 Validation dataset

Unpaired dataset with 751 samples analyzed with Human Exon1_0ST Affymetrix Chip was used to validate the differently expressed orphan probesets in another cancer and normal dataset. In Table 3 there were 751 samples of dataset with series GEO number accession and informations.

Table 3 Samples of Unpaired dataset

Tissue	Type Tumor	Series accession GEO dataset	Number of samples
PBMC	peripheral blood mononucleated cell samples collected from living healthy donors	GSE30453	80
leukemia	acute myeloid leukemia	GSE30285	93
tumor cell line	breast cancer cell line	GSE16732	41
9 type of tumor cell line	Breast, Central Nervous System, Colon, Leukemia, Melanoma, Non-Small Cell Lung, Ovarian, Prostate, Renal	GSE29682	163
breast	breast cancer	GSE16534	84
colon	colorectal cancer	GSE16534 GSE69182	103
head and neck	Head and Neck Squamous Cell Carcinoma	GSE33205	44
lung	non-small cell lung cancers	GSE16534	43
pancreas	pancreatic ductal adenocarcinoma	GSE23397	14
prostate	prostate cancer	GSE12378 GSE29079	86
TOT			751

Table 3 *Dataset description of the samples obtained from tumor tissues, cancer cell lines and PBMC analyzed with HumanExon1_0-ST Affymetrix.*

We considered peripheral blood mononucleated cell (PBMC) as normal samples compared to the cancer group formed by tumor tissue and cancer cell lines. The entire dataset represented data of 11 different types of cancer. Some of them are in common with the paired dataset (colon, pancreas, lung, ovarian and prostate). The other cancer types are novel for our analysis (acute myeloid leukemia, breast cancer, central nervous system tumor, melanoma, renal and head and neck squamous cell carcinoma).

Raw.CEL files (ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSEnnnnn/) were used to create a normalized expression matrix with program language R. Aroma.package was applied to normalize background correction, quantile normalization and robust multi-array average (RMA). The annotation for cluster transcripts is

provided by Affymetrix (https://www.affymetrix.com/support/technical/byproduct.affx?product=huexon-st).

3.3.3 Encode validation dataset

As a validation dataset we used public samples of RNA-seq data from Encyclopedia of DNA Elements Encode (https://www.encodeproject.org). The dataset consisted of 173 samples described in Supplementary Table 1 divided in 84 normal tissues, 2 embryonic stem cells (ESC), 2 induced pluripotent stem cells (iPSC), 19 cellular fraction of K562 cells and 66 cell lines of which 38 were tumor cell lines. The Encode dataset was constructed from fastq file, the reads of the RNA-seq data were aligned with Star [87].

3.4 Pipeline

1) Detection of differentially expressed probesets between tumor/normal tissue

The matrix of the normalized expression values of probesets for each exon was used as an input in this analysis. Calculating the Inter Quartile Range (IQR) with R genefilter package excluded lowly expressed probesets and those probesets that exhibited little variation across all the samples. Furthermore, the probesets with expression value major of 8 in at least 20 samples (5%) were preserved. Comparison between tumor and normal tissue was performed using the limma package with R. Differentially expressed probesets were defined as those probesets with Benjamini-Hochberg corrected p value less than 0.05 [124].

2) Steps to discover orphan probesets

Were constructed coordinates' files (bed) of the probesets differentially expressed between tumor and normal tissue, that have been taken from Affymetrix Human Exon annotation (https://www.affymetrix.com/support/technical/byproduct.affx?product=huexon-st). The files.bed of RefSeq Genes, UCSC Genes, GenecodeGeneV24lift37Comprehensive were downloaded from GenomeBrowser, both whole transcripts and exons coordinates (https://genome-euro.ucsc.edu/cgi-bin/hgGatewayredirect=manual&source=genome.ucsc.edu). The intersect analysis

was performed using Bedtools on a Linux machine (http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html).

3) Validation of probesets with correlation analysis with flanked regions

The correlation analysis between predicted orphan probesets and overlapped or flanked coding regions was performed with Python pipeline. The correlation analysis was based on Speraman correlation with SciPy, giving a Rho value and relative p.value. The threshold p.value of 0.05 was selected to discriminate the probesets correlated (p.value < 0.05) from the probesets and not correlated (p.value > 0.05) with flanked coding regions.

The entire pipeline was reported as APPENDIX B at the end of the thesis.

3.5 Validation in Unpaired Affymetrix Dataset

To validate the orphan probesets resulted from the pipeline we performed the same analysis of the first part of the pipeline in the Unpaired dataset of tumor cell line, tumor samples and PBMC as show in R script. As for Paired dataset we performed an IQR analysis and we retained those probesets that were present a value greater than 8 in at least 5% of the samples, corresponding to 38 on 751 samples. Then we made the diferentially expression analysis with limma, comparing tumor cell line and tumor tissue together to PBMC, here considered as normal samples.

```
##################### Validation in Unpaired Dataset R script ###############
#upload uncoupled matrix and PhenoDataUncoupled
unco <- read.delim("UnMatrix_Encode.txt",sep="\t",row.names = 1)#9168
PDU <- read.delim("PD_Uncoupled.txt",row.names = 1)

#cut samples with PhenoData
uncoU <- unco[c(row.names(PDU)[1:751])]
write.table(uncoU, file="UnCoupled_PD_751.txt", quote = FALSE, sep = "\t")

#transform dataframe in matrix
uncoU <- as.matrix(uncoU) #9168

#call genfilter
library(genefilter)
IQR05U<- varFilter(uncoU, var.func=IQR, var.cutoff=0.5, filterByQuantile=FALSE)
#trasformo da matrice a dataframe
IQR05U <- as.data.frame(IQR05U) #8012
# subset at least 5% (47) of all samples with value > 8
m8su38U <- IQR05U[rowSums(IQR05U > 8) >= 38, ]#7779
#stessa cosa con altro comando per controllo
uncoA <- as.matrix(IQR05>8)#crea matrice con true false
count <- as.data.frame(rowSums(uncoA)) #conta quanti true per riga confermati 8230
write.table(IQR05U,file = "UnCoupledIQR.txt", quote = FALSE, sep = "\t")
write.table(m8su38U,file = "UnCoupledIQR_m8su38.txt", quote = FALSE, sep = "\t")

source("https://bioconductor.org/biocLite.R")
biocLite("limma")
library(limma)

#Limma for unpaired data
#definire colonna di interesse dei phenoData
design <- model.matrix(~factor(PDU$STATE))
colnames(design) <- c("Intercept","Tumor-Normal")
fit <- lmFit(m8su38U, design)
ebayes <- eBayes(fit)
Limma<- topTable(ebayes, number = 7779, coef = "Tumor-Normal", sort.by = "logFC")
write.table(Limma, file="LimmaUncoupled.txt", quote = FALSE, sep = "\t")

#subset significative Limma
LimmaS <- subset(Limma,adj.P.Val<0.05)#5843
write.table(LimmaS, file="LimmaUncoupled_Sign.txt", quote = FALSE, sep = "\t")
```

3.6 Validation on RNA-seq dataset

The General Transfer Format (GTF) file for annotation was constructed using the coordinates of the Human Genome Version 38 (HG38) of orphan probesets to validate them in Encode dataset. The ultrafast universal RNA-seq aligner, STAR,

was used with custom GTF to re-map the readings on the HG38 [87]. The normalization was performed whilst taking into account the sequencing depth and probesets length, Fragments Per Kilobase per Million mapped reads (FPKM). To calculate the FPKM values, the sum of the total reads in a sample were calculated and the value divided by 1,000,000 – this is the “per million” scaling factor. Then the read counts were divided by the “per million” scaling factor, obtaining reads per million (RPM), a normalization for sequencing depth. The RPM values were divided by the length of the probesets, in kilobases for paired-end RNA-seq, resulting in a matrix with log2 FPKM values, this takes into account that two reads can map to one fragment (and so it doesn’t count this fragment twice) [129].

The log2 FPKM matrix was loaded into BRB-tools, Spot Intensity filter was used as a threshold at the minimum value of 1 and Minimum Fold Change gene filter was applied.

The probesets that in less than 1% of the expression data values, having at least 1,5 fold change in either direction from the median value, were retained as validated.

3.7 Classification with the last version of Gencode V29

We re-analyzed and classified the 4,024 probesets with the lastest version of Gencode, V29. We used the python script of the third part of pipeline to correlate the probesets with flanked regions in Encode RNA-seq data. Furthermore, we added an analysis of clusters, to group those probesets that resulted inside the same gene, concordant and correlated between each other.

3.8 Conservation analysis

We used GenomeBrowser tool to see the conservation of the 4,024 orphan probesets [98]. Genome browser is a web open source tool which allows easy and rapid display of any portion of the genome at any scale, together with several dozen aligned annotation tracks (http://genome.ucsc.edu). This browser displays assembly contigs and gaps, mRNA and expressed sequence tag alignments, multiple gene predictions, cross-species homologous, single nucleotide polymorphisms, sequence-tagged sites, radiation hybrid data, transposon repeats, and more as a stack of co-registered tracks. On the tracks of GenomeBrowser we found two types of conservation analysis, 100 vertebrates and 7 vertebrates, see

example in Figure 6. We used the tool table track of Genome Browser to recover the coordinates of the sequences resulted conservative in the Vertebrates.

Figure 6 Example of the tracks for conservation analysis on Genome Browser.

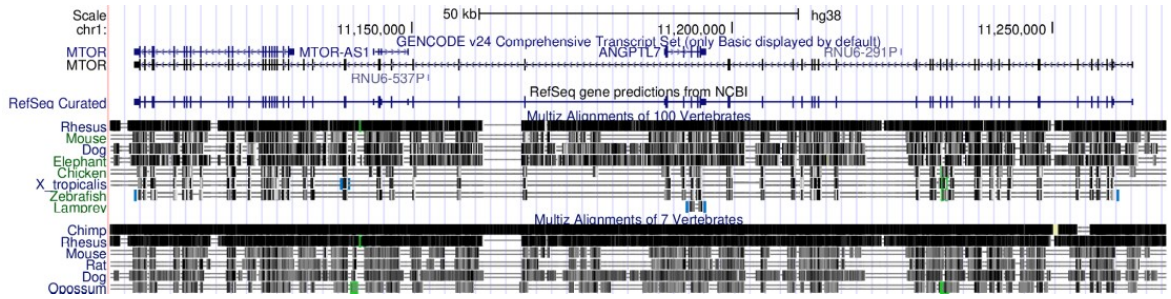


Figure 6 *Example of the tracks for conservation of MTOR gene on Genome Browser tool.*

The track shows the measurements of the evolutionary conservation of multiple alignments of vertebrate species using phastCons methods. PhastCons is sensitive to "runs" of conserved sites, and is therefore effective for picking out conserved elements [130]. We used two types of conservation tracks. The first was composed of about 100
vertebrates and second (7 vertebrates) a subgroup of the mammals. The two phylogeny trees of vertebrates used for the analysis are shown in Figure 7 .

Figure 7 Phylogeny tree of vertebrates used for the conservation analysis.

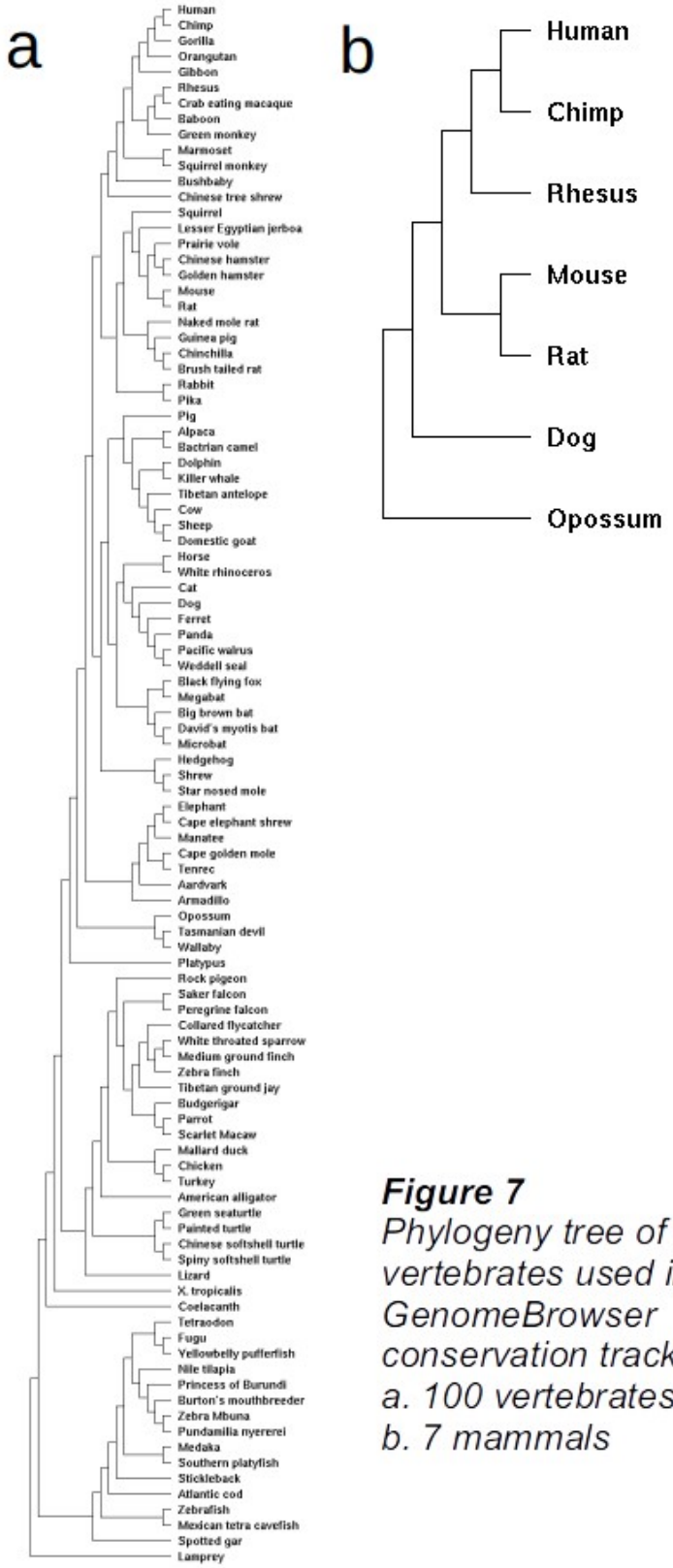


Figure 7
Phylogeny tree of the vertebrates used in GenomeBrowser conservation tracks
a. 100 vertebrates
b. 7 mammals

3.9 Coding potential analysis

Using genes coordinates on Human Genome version 19 of Genome Browser Tools (hg19) recovered the FASTA sequence of 4,024 orphan probesets. The FASTA input file was used with Coding-Potential Assessment Tool (CPAT), an alignment-free program, which uses logistic regression to distinguish between coding and non-coding transcripts on the basis of four sequence features. The first feature is Open Reading Frame (ORF) length that is one of the most useful to discriminate between non coding RNA sequence and messenger RNA. It is unlikely that a long putative ORF will be observed randomly in non-coding sequences [131]. The feature is based on putative maximum ORF. The ORF coverage, defined as the ratio of ORF to transcript length, was the second feature that CPAT considered to performed discrimination analysis. The ORF coverage is complementary to and independent of ORF length and presents an improvement of classification power respect to the ORF length. Large known long non-codingRNAs may contain putative long ORF by random chance and cannot classify well by ORF length alone. Those long ncRNAs usually present much lower ORF coverage than protein-coding RNAs [3].

The third feature calculated from CPAT is the Fickett score (FC). FC identifies protein-coding RNA and non-coding RNA based on a combination of nucleotide composition and the DNA code of codons, supported by linguistic studies property [132]. The last feature is hexamer usage bias, named Hexamer Score (XS) in the CPAT output. This analysis is based on the dependence between adjacent amino acids in proteins. Hexamer score calculates the specific degree of hexamer in a relative sequence. XS > 0 indicates a coding sequence, XS < 0 shows a non-coding sequence evidence [133]. The optimum cut-off to decide if a transcript could have coding probability is decided from a ROC curve. Coding Probability (CP) >= of 0.364 determines coding sequences, CP < 0.364 indicates non-coding sequences [105].

Python script to performed cpat analysis on the orphanprobesets

```
cpat.py -g orphan_fasta -d Human_logitModel.RData -x Human_Hexamer.tsv -o cpat_orphan
```

3.10 Correlation with Census Gene Cancer

The list of genes was download, which contained mutations that have been causally implicated in cancer from catalog Cancer Gene Census (CGC) [134]. The expression values of 708 CGCs was recovered from Encode datasets. The correlation analysis was performed with function rcorr of Hmisc package of R, that gave us an output with a rho e p.value. A p.value less than 0.001 was considered as a significant value. Rho near a value of 1 is indicative of positive correlation and rho near -1 indicates negative correlation. A rho value greater than 0.5 and value less than -0.5 was recognized as a strong correlation. We performed a specific functional analysis on the list of genes results correlated with our orphan probesets both (positive and negative correlation) with Panther, an example of usage shown in Figure 8 [135].

Figure 8 Panther web tool.

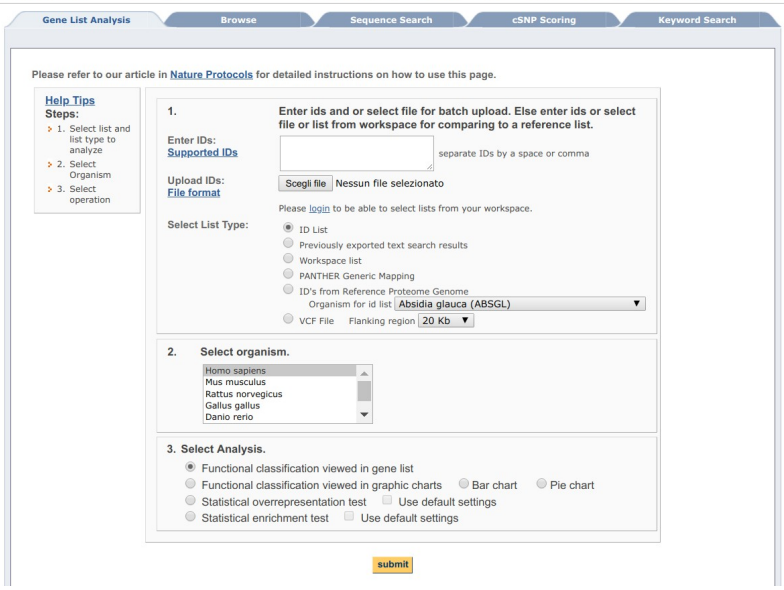


Figure 8 *Example of Panther usage on web tool.*

R script to performed correlation analysis with cancer census genes

```
#Upload expression matrix of Orphan Validated Porbesets and Cancer Census Gene
CG <- read.delim("CG_Encode_RPM_708",row.names = 1)
PB <- read.delim("Val_RPM_4024",row.names = 1)
#call hmisc package
library("Hmisc")
#function to create matrix with correlation results: rho and p-value
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor  =(cormat)[ut],
    p = pmat[ut]
  )
}
#use the function flattenCorrMatrix to recover data from the output analysis
res <- rcorr(t(PB),t(CG))
CorAll<-flattenCorrMatrix(res$r, res$P) #11193546

#the rcoor function of hmisc correlates the column of the matrices each other
#give it two matrix correlates other the two matrices also the column of the matrices each other
#so we have more tham 11 million of correlations
CorCG <- CorAll[- grep("Gen", CorAll$row),]
#subset 250278 correlation between census genes each other
CorCG_PB <- CorAll[- grep("Gen", CorAll$column),]
#subset 3099270 correlation between  census genes each other and census and  orphan probesets
CorCG_PB <- CorCG_PB[grep("Gen",CorCG_PB$row),]
#subset 2848992 correlation between census genes and  orphan probesets!!!
#analysis to recover significant results
CorCG_PB <- CorCG_PB[complete.cases(CorCG_PB),] #1614450 no nan
CorCG_PB_Sign05 <- subset(CorCG_PB, p<0.05) #299320
CorCG_PB_Sign001 <- subset(CorCG_PB, p<0.001) #93101
write.table(CorCG_PB_Sign001,file = "CorCG_PB_Sign001.txt", row.names = FALSE,quote =
FALSE,sep="\t")
#we retain the first more correlated gene for each probesets
#both positive and/or negative correlations
df <- subset(CorCG_PB_Sign001, cor>0)
dfminus <- subset(CorCG_PB_Sign001, cor<0)
df <- df[order(df$row, -(df$cor )), ] #sort for row and reverse cor
df <- df[ !duplicated(df$row), ] #Keep first and delete following duplicate
dfminus <- dfminus[order(dfminus$row, dfminus$cor ), ] #sort for row and cor
dfminus <- dfminus[ !duplicated(dfminus$row), ] #Keep first and delete following duplicate
CorCG_PB_Sign001_best <- rbind(df,dfminus)
write.table(CorCG_PB_Sign001_best,file = "CorCG_PB_Sign001_best.txt", sep="\t",quote =
FALSE,row.names = FALSE)
write.table(df,file = "CorCG_PB_Sign001_best_positive.txt", sep="\t",quote = FALSE,row.names
= FALSE)
write.table(dfminus,file = "CorCG_PB_Sign001_best_negative.txt", sep="\t",quote =
FALSE,row.names = FALSE)
```

4 Results and discussion

4.1 Preprocessing Affymetrix datasets

First of all we searched on GEO all the published datasets of cancer tissue or cell line and, where present, normal related tissue, analyzed with the AffymetrixHumanExon1.0ST array. On GEO every Affymetrix chip was classified as a platform with a specific identifier. AffyetrixHumanExon1.0ST in GEO was represented by two platforms, GPL5188 and GPL5175. The differences between the two platforms is based on the level (exon or gene) applied during preprocessing of the data. GPL5188 shows analysis at the exon level with total number of rows, representing exon probesets, of more than 1.4 million. GPL5175 exhibits analysis at gene level with a total number of probesets clustered in about 300,000 genes. For our analysis we directly used the original raw.CEL data which were the same for the two different platforms, corresponding to about 1.4 million exon probesets.

Platform GPL5188 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL5188) presents 207 series for a total of 9,358 samples, of these 4,637 are on cancer studies. Platform GPL5175 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL5175) shows 15,528 samples of which 5,642 are on cancer research. Then we analyzed in detail all the cancer related samples from both GPL5188 and GPL5175, around 9000 in total.

In 2010 a study was published in which gene expression signal performance of degraded RNA from fixed (FFPE) was compared with intact RNA from unfixed fresh-frozen (FF) specimens of paired tumor and normal samples. Both the raw probe and probeset level were evaluated for microarray signal dynamics. The results showed that, for expression signatures and differential expression analysis, neither FFPE nor FF produce significant differences to profile. But the results showed that probe performance was not uniform and decline sharply towards the 5′ end of genes. Additionally, RNA extracted from FFPE material was more degraded and fragmented than FF and that caused a reduction in dynamic range of expression signal. These results indicated that FFPE were not optimal samples to the studies of discovery, validation, and clinical application of microarray expression profiling [136]. We decided to discard the analysis of Formalin-Fixed Paraffin-Embedded (FFPE) samples because they presented low expression

signal that wasn't optimal for our type of discovery analysis. Next, we excluded from the analysis the data in which the original raw CEL files were not present. At this point we stayed with the 1109 samples that we split into two main datasets. The first dataset, used for pipeline analysis, was composed of those samples that presented tumor and paired normal tissue of the same patient. This dataset contained 358 paired (179 tumor, 179 normal) samples of 9 types of different cancer. We used the remaining ones to construct a validation dataset that contained 751 samples of tumor cell line, tumor tissue and PBMC. We started with Oligo working on raw CEL files to create a normalized expression matrix. We worked with 358 raw CEL files of about 1.4 million rows which weighed around 1.5 Gigabite each. We worked with a Linux server of 8 CPU cores, but we had some problems. The server continued to jump and as Oligo didn't save the progress it had made so far, we lost all data every time and it was necessary to start over from the beginning. Furthermore, observing the data procession we tried to calculate the time it would take for the entire analysis which resulted in it being months. So we tried to do the same analysis using Aroma that was known to be perform better for surveys with large volumes of data. With Aroma there was the possibility to use it in parallel with the 8 CPU cores for different parts of the analysis, reducing the length of time and energy expended. Aroma allowed us to perform background corrections, quantile normalization and RMA and gave us the matrix output in 20 days. We obtained two expression matrices with log2 normalized values of exon probesets, one of the paired tumor/normal tissue and other of cancer cell line, tumor tissue and PBMC. The first matrix of tumor/normal tissue was used for the pipeline analysis, while the other matrix was used as a validation dataset. In addition, for the paired tumor/normal dataset, we performed the normalization to the gene level, obtaining a third matrix with the log2 expression values of the probesets clustered in the known genes to use for correlation analysis.

4.2 Pipeline workflow

We established a computational pipeline with R to re-annotate the 1,400,703 probesets of the Affymetrix Human Exon 1.0 ST array, schematized in Figure 9. In Figure 9A we summarized the first part of pipeline performed in R. To identify those involved in cancer we first made an IQR analysis. With this analysis we discarded

the probesets that never change within the different samples. We obtained 581,149 probesets. Probesets with a very low value across the dataset were discarded, we only kept those with a value larger than 8 in at least 20 samples, that is 5% of all the dataset. On the remaining 164,600 probesets, we made a Limma analysis for paired data by comparing the samples of the tumor group against the samples of the normal group. The Limma analysis allowed the ability to determine which probesets were significantly differently expressed in the two groups, giving a logFC value that indicated in which groups the probesets were highly expressed and in which the results were lowly expressed. Taking into account the adjusted p.value, we obtained 76,713 significant probesets. We retained only probesets that were covered by at least four probes and discarded those that presented anomalies in the genome's coordinates. In order to classify the 67,585 significant probesets we use BEDtools to intersect the coordinates of probesets with the coordinates of known genes, in exon datasets as showed in Figure 9B. Proceeding to cascade we annotated as exonic probesets those that mapping within the exons of known genes in the same strand, first in RefSeq genes, then in UCSC genes and finally within GenecodeGeneV24lift37 genes, and we discarded them.

The same analysis was performed with the opposite strand between probesets and exons in order to retain the antisense probesets. We then intersected the coordinates of the remaining probesets not annotated with coordinates of the entire transcript of the datasets of known genes. Proceeding to cascade we classified as intergenic those that aren't within genes and considered as intronic those that are within genes. In order to confirm that the probesets classified as intergenic and intronic were really novel, we did a Pearson correlation analysis between the expression of probesets and the expression of known flanked coding regions.

As shown in Figure 9C, the intergenic probesets have been considered according to the criteria of our classification based on the strand. The intergenic_Ind are the probesets that have a strand opposite to both flanked genes so they can be defined as independent of these genes. The intergenic_Conc are the probesets that agree with both strands of flanked genes but they aren't correlated to either of them. The intergenic_Disc probesets are consistent with only one of the strands of

the two flanked genes but not correlated to them. Finally we performed the Pearson correlation of the intronic probesets with overlapped genes, Figure 9D.

Figure 9 Pipeline workflow to discover cancer related orphan HuEx probesets.

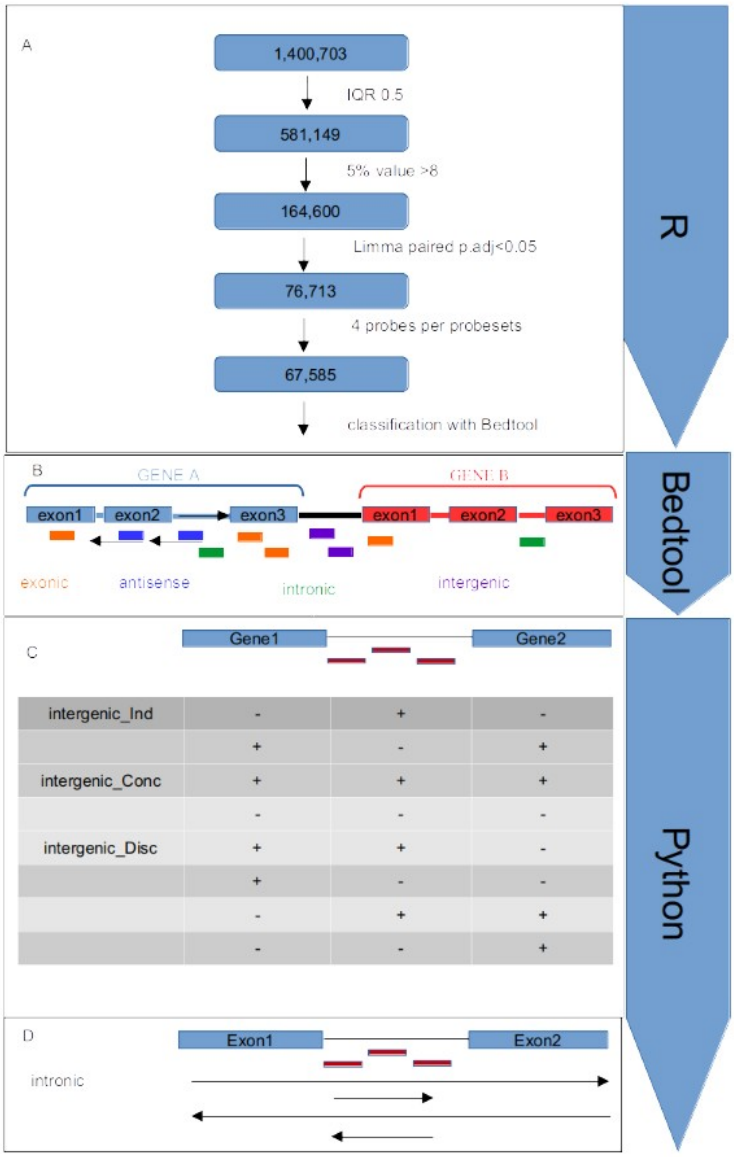


Figure 9 *Workflow of the pipeline to discover orphan probesets of Human Exon Affymetrix 1_0ST chip, cancer related.*
A) Analysis performed with R to retain the interesting probesets that results cancer related.
B) Classification of the probesets with Bedtool based on their position compared with known genes in human genome.
C) Correlation analysis between Intergenic probesets and flanked genes.
D) Correlation analysis between Intronic probesets and flanked exons.
C) and D) were performed using Python.

In Table 4 we summarized the results of the classification and correlation analysis. For the probesets that are classified with _Nan, it wasn't possible to perform correlation analysis with python, for probesets that are classified with _Miss flanked or overlapped gene expression information was not available.

For the next validation analysis we decided to retain the _Nan and the _Miss probesets, trying to validate them in RNA-seq. With bedtools and python correlation analysis we found 9,168 probesets differentially expressed between normal and tumor paired samples and not-overlapped to any known transcripts. We classified them based on their position compared to known transcripts as shown in the Figure 10. There were 3,829 antisense probesets that overlapped with known transcripts in the opposite strand. These could be novel antisense ncRNAs. We obtained 3,589 probesets that didn't overlap with any known transcripts, these could be independent transcripts and were therefore classified as novel. Finally, we found 1,750 probesets that overlapped intronic regions of known genes. These could be novel ncRNAs that are transcribed within genes.

Table 4 Classification of orphan probesets from pipeline

Classification Name	Number Probeset	Classification Results
intergenic_Nan	39	*
intergenic_Ind	1,337	independent transcript
intergenic_Disc	438	independent transcript
intergenic_Disc_Miss	985	**
intergenic_Conc	87	independent transcript
intergenic_Conc_Miss	703	**
intronic_Nan	8	*
intronic_Miss	911	**
intronic	831	independent transcript
antisense	3,829	independent antisense
Total	9,168	

Table 4 *New classification of probesets differentially expressed in normal/tumor sample*

** Nan: not possible correlation analysis*

*** Miss: not available flanked or overlapped gene expression information*

Figure 10 Pie chart of the pipeline classification

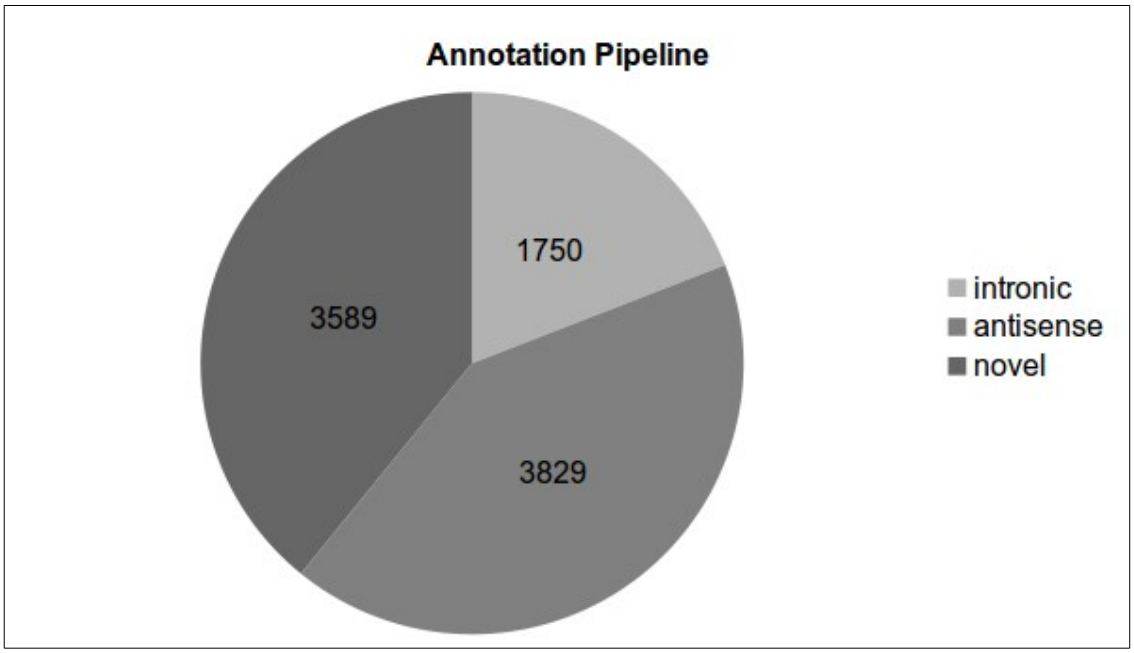


Figure 10 *The pie chart represent the counts of the 9168 orphan probesets that resulted classified from the Pipeline. Three classes: intronic, antisense e novel probesets.*

4.3 Validation on Unpaired Affymetrix dataset

To amplify the number of samples in the study we performed the same analysis of the first part of the pipeline in the Unpaired Affymetrix dataset, Figure 11. Unpaired dataset was composed of tumor cell line, cancer tissue and PBMC. Here we considered PBMC as normal samples compared to the other tumor group. We worked on the 9,168 probesets that resulted as orphan in previous pipeline. First, an IQR analysis, to discarded the probesets that never change within all the samples, was made. We obtained 8,012 probesets. To discard probesets with a very low value across the dataset, we only kept those having a value greater than 8 in at least 38 samples, that is 5% of all dataset. On the remaining 7,779 probesets, we made a Limma analysis by comparing the samples of the tumor group against the samples of the PBMC normal group. The Limma analysis allowed us to determine which probesets were significantly differently expressed in the two groups, giving a logFC value that indicated in which group the probesets

are highly expressed and in which are lowly expressed. Taking into account the adjusted p.value we validated 5,843 significant probesets.

Figure 11 Unpaired validation workflow

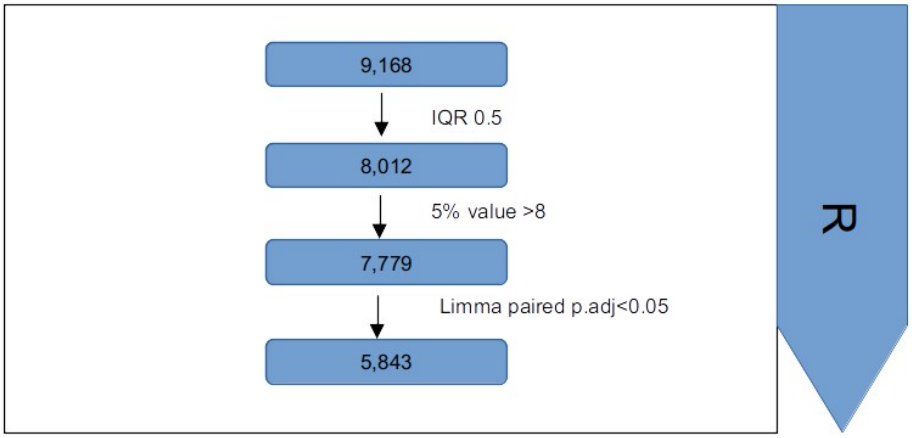


Figure 11 *Analysis performed with R to retain the validation in Unpaired dataset, orphan probesets that resulted cancer related.*

4.4 Validation on RNA-seq dataset

The Encode dataset, composed of normal tissue, ESC, iPSC, cellular fraction of K562 cell and 66 cell lines of which 38 tumor cell lines, was used to validated those probesets classified with the pipeline. In order to confirm a significant expression, probesets that present in 1% of the samples a fold-change of 1.5 were retained validated, resulted in 4,024 probesets, showed in Figure 12. We chose the threshold based on the composition of the dataset, we had two samples of embryonic stem cells and two samples of the induced pluripotent stem cells. The choice of the threshold was 1%, that corresponded to 2 samples of the 173 total samples, this allowed us to not exclude the probesets that were differentially expressed between stem cell and other samples. Of the 3,829 antisense probesets 1,517 were validated, corresponding to 40%. On the class of the 1,750 intronic probesets, 1,441 resulted validated, the 82%. Of the 3,589 intergenic probesets, 1,066 were validated, corresponding to the 30%. In the result of the 82% of the intronic sense probesets we have hypothesized to have a lot of false positives. We expected that many of the probesets are novel exons of host genes. The other two classes were in accordance with our expectations. Many of the

probesets were excluded because they were not significantly expressed in RNA-seq. We may suppos that they were background noise of the Affymetrix chip.

Figure 12 Pie chart of the probesets validated in RNA-seq

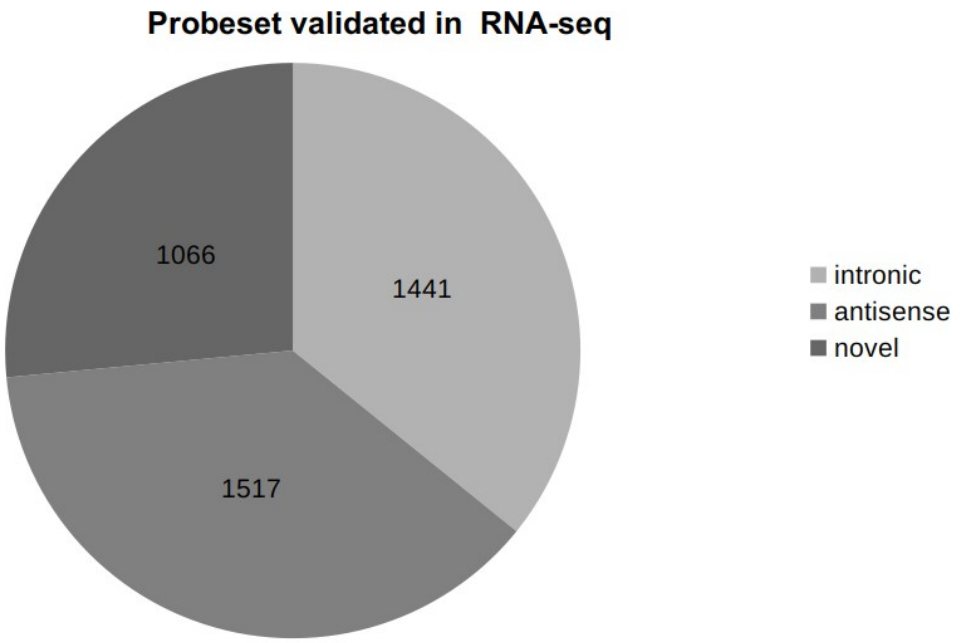


Figure 12 *The pie chart represents the 4,024 valid results of orphan probesets s in RNA-seq. Three classes: intronic, antisense e novel probesets.*

4.5 Classification with the last version of Gencode V29

In the piechart of Figure 13 we observed that compared to the precedent version V28 only the group of lncRNAs had a significant increase, more than 1,000 new lncRNA genes.

We decide to perform a new annotation and re-analysis of the correlation with the flanked regions of the 4,024 probesets validated in RNA-seq.

We re-utilized the same script of the third part of pipeline, applied on the RNA-seq Encode data. With a new classification we obtained 3,501 intragenic probesets, divided into three groups: exonicSense, intronicSense and antisense, Figure 14

Figure 13 Piechart of GencodeV29 classification of human genome.

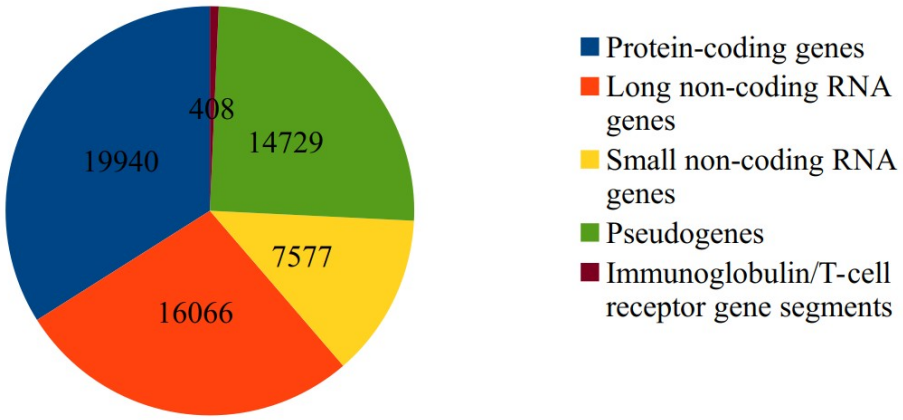


Figure 13 *Piechart of GencodeV29 classification of the 58720 genes of the entire human genome.*

Figure 14 Classification on Gencodev29 annotation.

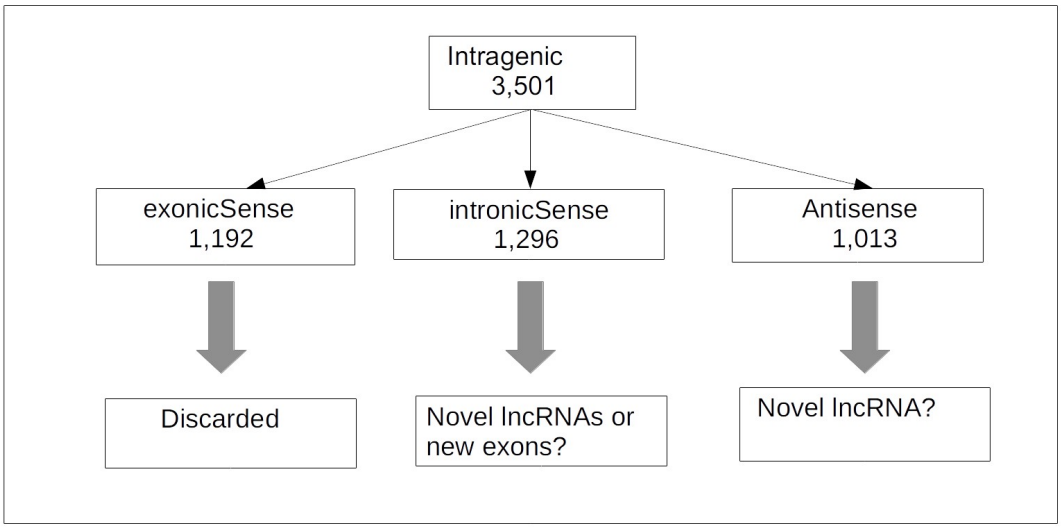


Figure 14 *Classification of intragenic probesets whit Gencodev29 annotation in three classes: exonixSense, intronicSense and antisense.*

To compare the differences between the previous classification and the last classification on Gencodev29, we performed analysis on GenomeBrowser, Figure 15.

Figure 15 GenomeBrowser analysis of exonicSense probesets.

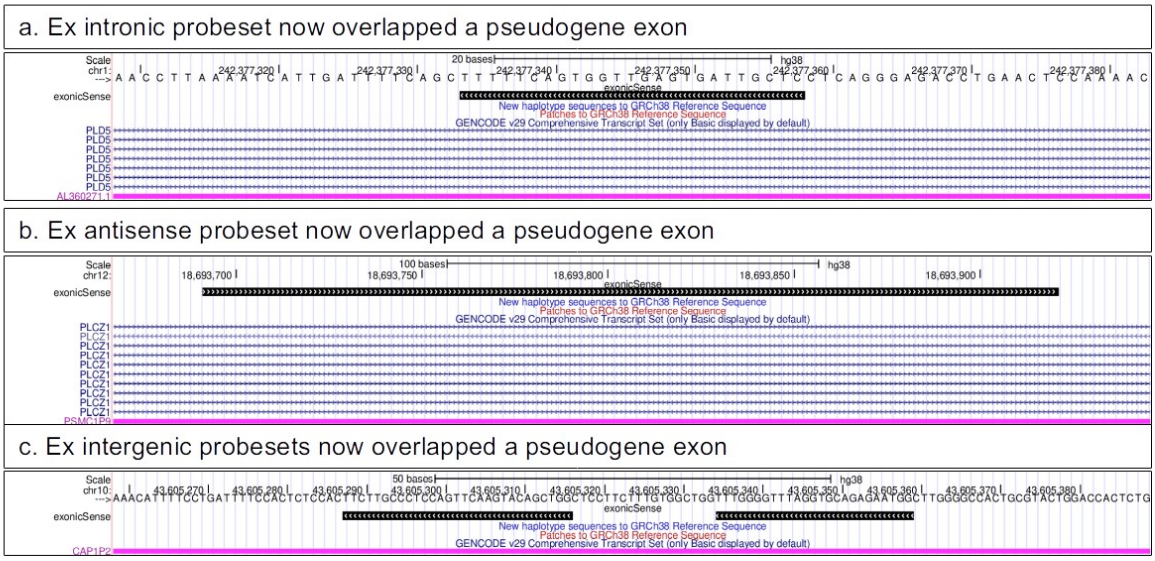


Figure 15 *In black the custom track of probesets. The symbols < or > inside it show the strand direction. In pink, novel pseudogenes in GencodeV29 and in blue the old annotation gene (Gencode v28 and earlier).*

In Figure 15a we observe a probeset previously classified as intronic of the gene PLD5 and now resulted as exonic in pseudogene AL360271.1. We conducted this analysis in a cascade manner. This means that we first searched for those probesets that resulted in exonic sense and classified it and then we reanalyzed the remaining probesets. So in this case the probesets overlapped two transcripts but we only considered the first exonic overlap. Proceeding with the analysis we found probesets that were classified as antisense, but now are exonic of a pseudo gene. We see an example of this case in Figure 15b where probeset previously consider antisense of PLCZ1 was now considered as an exonic probeset of the pseudogene PSMC1P9. In the last case depicted in Figure 15c we can observe probesets, that previously were retained as intergenic, with the new annotation on Gencodev29 fall in the exonic region of the pseudogene CAP1P2.

In general we observed that they effectively fell in the exon of known genes so we could discard them. For the group of 1,296 intronic sense RNAs, Figure 16, we obtained correlation values that allowed us to divide it into four sub-groups.

Figure 16 Summary analysis of intronic Sense probesets.

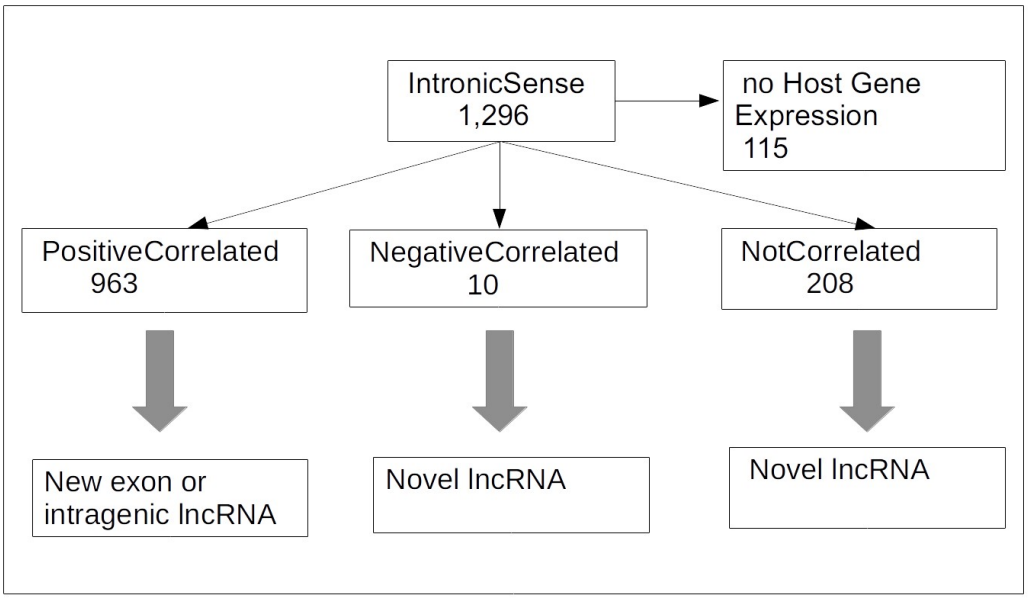


Figure 16 *Classification of intronic Sense probesets with Gencodev29 annotation in four classes based on the correlation with host genes.*

First of all we obtained 963 probesets that resulted in a positive correlation with the host genes. We thought that they probably were new exons. But in our previous study on lncRNA LOC107987281 situated in the first intron of the TGM2 gene, we found that it also correlated positively with TGM2. This indicated that we couldn't disregard these probesets, but that other types of analysis' were needed. Next, we obtained 10 probesets that resulted in a negative correlation, this made them a point of interest because not only could they be a novel independent transcript but they may have a role in the transcription of their host genes. The 208 probesets, that were surely not correlated with the host gene, could be novel lncRNAs. Lastly, we obtained 115 probesets in which there isn't an expression of the host gene. The same division was performed within the antisense group, Figure 17, where the considerations were different.

Figure 17 Summary analysis of antisense probesets.

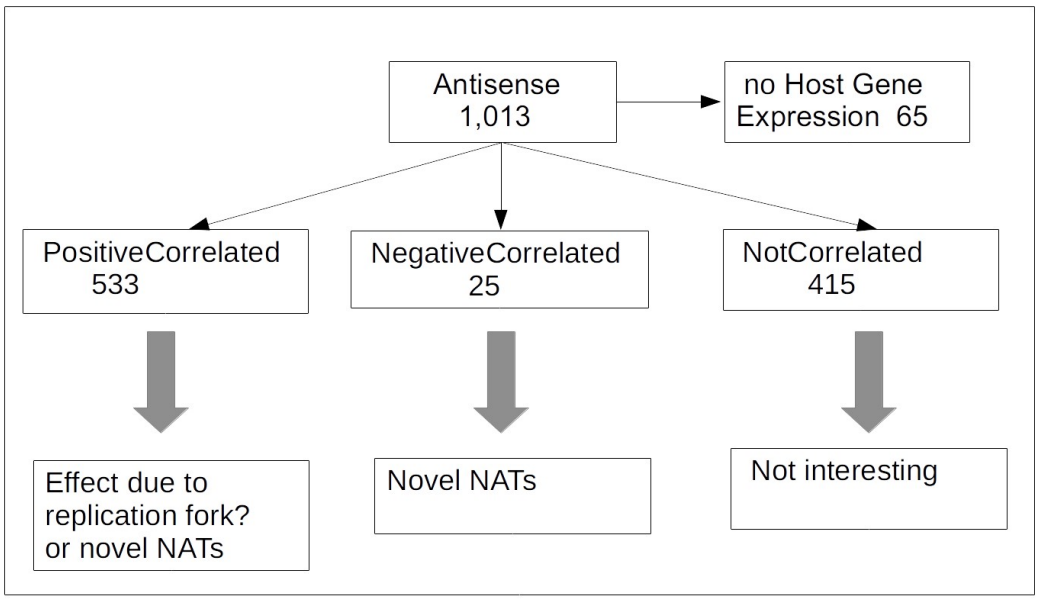


Figure 17 *Classification of antisense probesets with Gencodev29 annotation in four classes based on the correlation with host genes.*

For the subgroups of the probesets that were not correlated with any genes, 415, or there was not an expression of the Host gene, 65, we could consider them as novel NATs but they do not seem to produce any changes in host gene expression, so it is needed a correlation analysis with other genes. A different consideration could be applied to the antisense probesets that resulted having either positive and negative correlation with host genes. We obtained a large number, 533 probesets positively correlated to host genes compared to the small group of 25 of negatively correlated. We could suppose that this result is an effect of the replication fork that is not strand dependent. But we know that NATs could be both positively or negatively correlated with other genes, as showed in Figure 18 [137].

Thus, for the 25 negative correlations we are sure that the probesets are NATs with an opposite expression compared to host genes, so we could suppose that they are involved in the control of gene expression. However, for the positive correlated probesets we could only supposed that they are NATs with concordant expression with host genes and further analysis is needed.

Figure 18 Function of NATs.

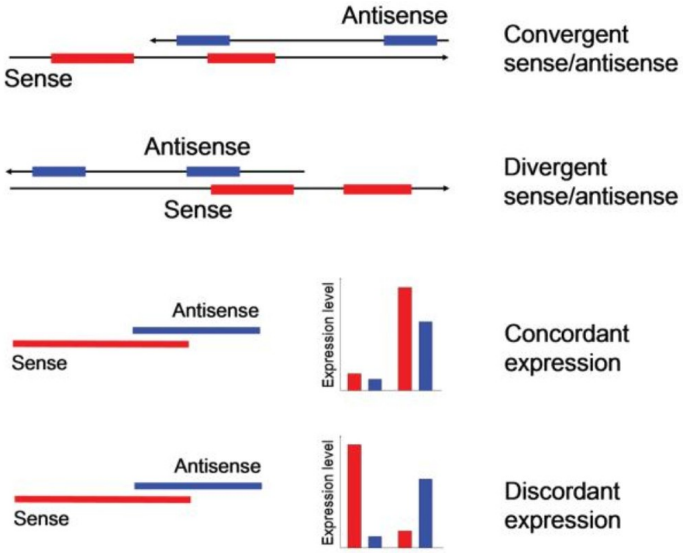


Image obtained from: Wight M. and Werner A. "The functions of natural antisense transcripts" Essays Biochem. 2013; 54: 91–101. doi: 10.1042/bse0540091

Lastly, we obtained 523 probesets that resulted in intergenic regions. For these probesets, we performed correlation analysis for the flanked same strand regions. As showed in Figure 19 we divided them in three subgroups.

Figure 19 Summary analysis of intragenic probesets.

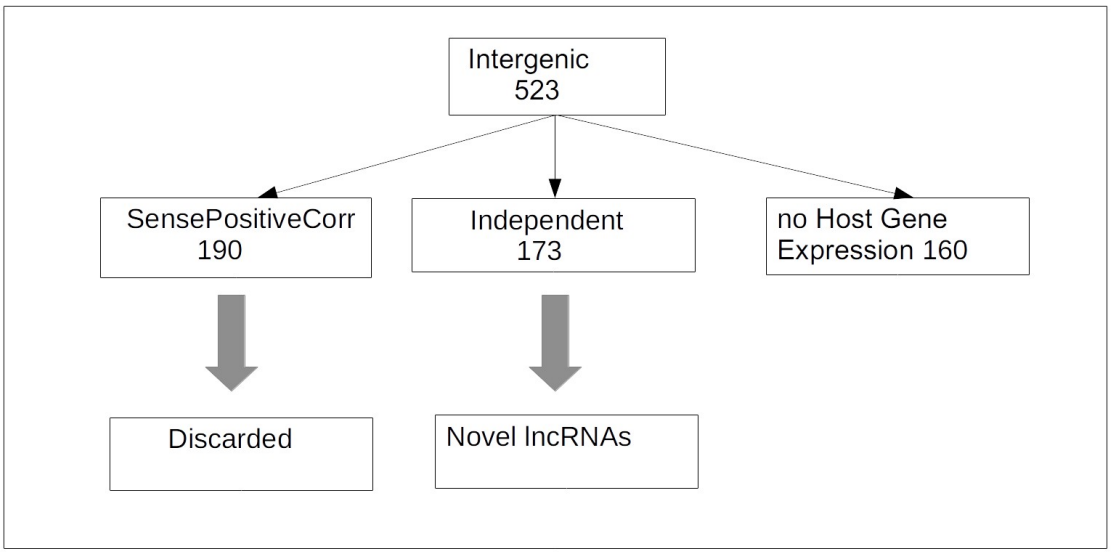


Figure 19 *Classification of intergenic probesets whit Gencodev29 annotation in four classes based on the correlation with flanked genes.*

The first group contains 190 probesets that resulted in positive correlatation to the flanked gene (both 5' or 3') with the same strand. We considered them to be novel exons or untranslated regions (UTR) of that gene, thus we discarded them. In the second group, we see 173 probesets that resulted completely independently compared to flanked regions. In this group we found probesets that didn't correlated to flanked genes with the same strand, or those probesets that presented an opposite strand compared to the flanked genes. So we were able to define them as independent transcripts and could be novel lncRNAs. In the third group, we found probesets for which there was not the expression of the host gene. We can validate that they are intergenic regions but couldn't validate them with the correlation analysis of flanked genes.

Furthermore, we clustered the intragenic probesets based on their positions. We considered the probesets that are in the same host gene as a cluster, resulting in concordance and correlation with each other, Figure 20.

Figure 20 Cluster analysis of intragenic probesets.

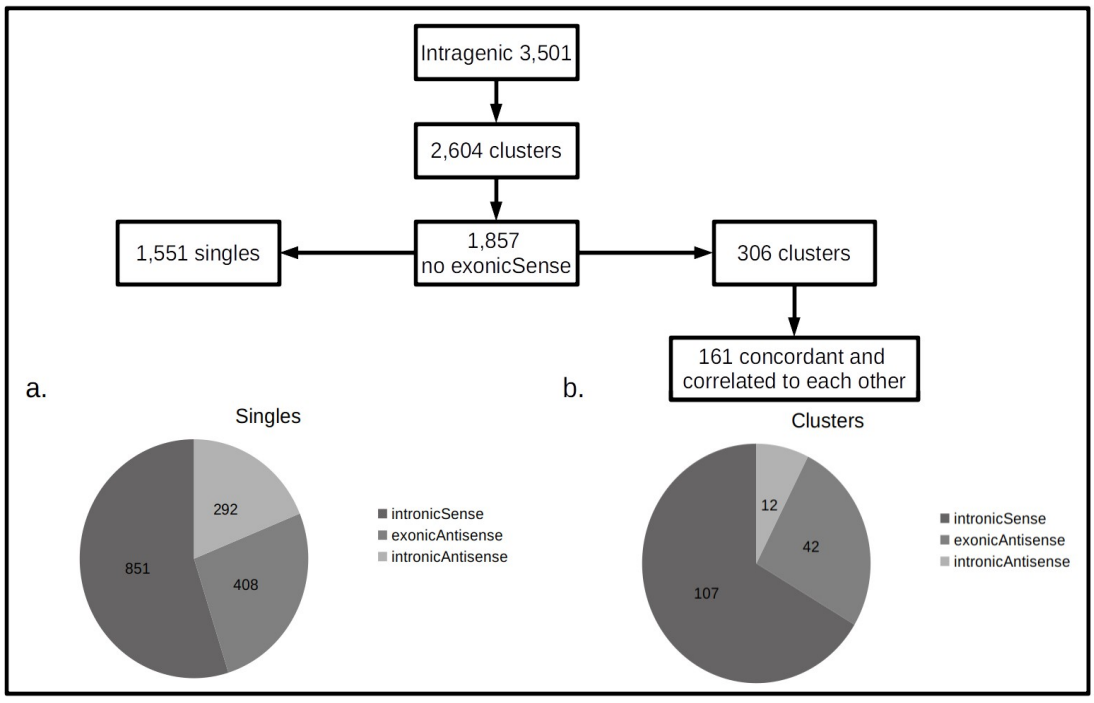


Figure 20 *Cluster analysis of intragenic probesets to group probesets based on the same host gene.*

Take into consideration, the 3,501 intragenic probesets were grouped in 2,604 clusters. From the 2,604 clusters we first discarded the exonix Sense RNAs. The remaining 1,857 clusters, were composed of 1,551 singles probesets, representing the probesets that are alone in the host gene. The other 306 clusters, were composed of probesets that are in the same host gene. Of these 306, 161 clusters presented probesets that are concordant and correlated to each other. The Figure 19 a and b showed the pie charts the classification of the single probesets and clusters respectively, in intronicSense, exonicAntisense and intronicAntisense classes.

4.6 Conservation analysis

We performed a conservation analysis on the 4,024 probesets using two tracks of Genome Browser tool. In the Table 5 are the summarized counts of the probesets that resulted in all the classification classes. The probesets in “All” column resulted in conservation of both the 100 vertebrates and the 7 vertebrates tracks. The “Not” column contained the probesets that did not result in any conservation of the two used tracks. In the “V7” and “V100” columns the results showed those probesets conserved in the 7 vertebrates and the 100 vertebrates tracks respectively.

Table 5 Summary of the conservation analysis.

Class	All	Not	V7	V100	Tot
exonic_Sense	974	63	9	146	1192
intronic_Sense	227	903	11	155	1296
exonic_Antisense	351	128	7	93	579
intronic_Antisense	101	257	11	65	434
intergenic	109	349	10	55	523
Tot	1762	1700	48	514	4024

Table 5 *In All: conserved both in 100 and 7 vertebrates track; Not: not conserved in any of the two tracks; V7: conserved only in 7 vertebrates track and V100: conserved only in the 100 vertebrates dataset.*

We observed that the major part of the probesets resulted in the conservation of both the 7 and 100 vertebrates track, and a small proportion showed a conservation in only one of the two tracks, with very low values in 7 vertebrate track.

Figure 21 Panel with summary of the conservation analysis.

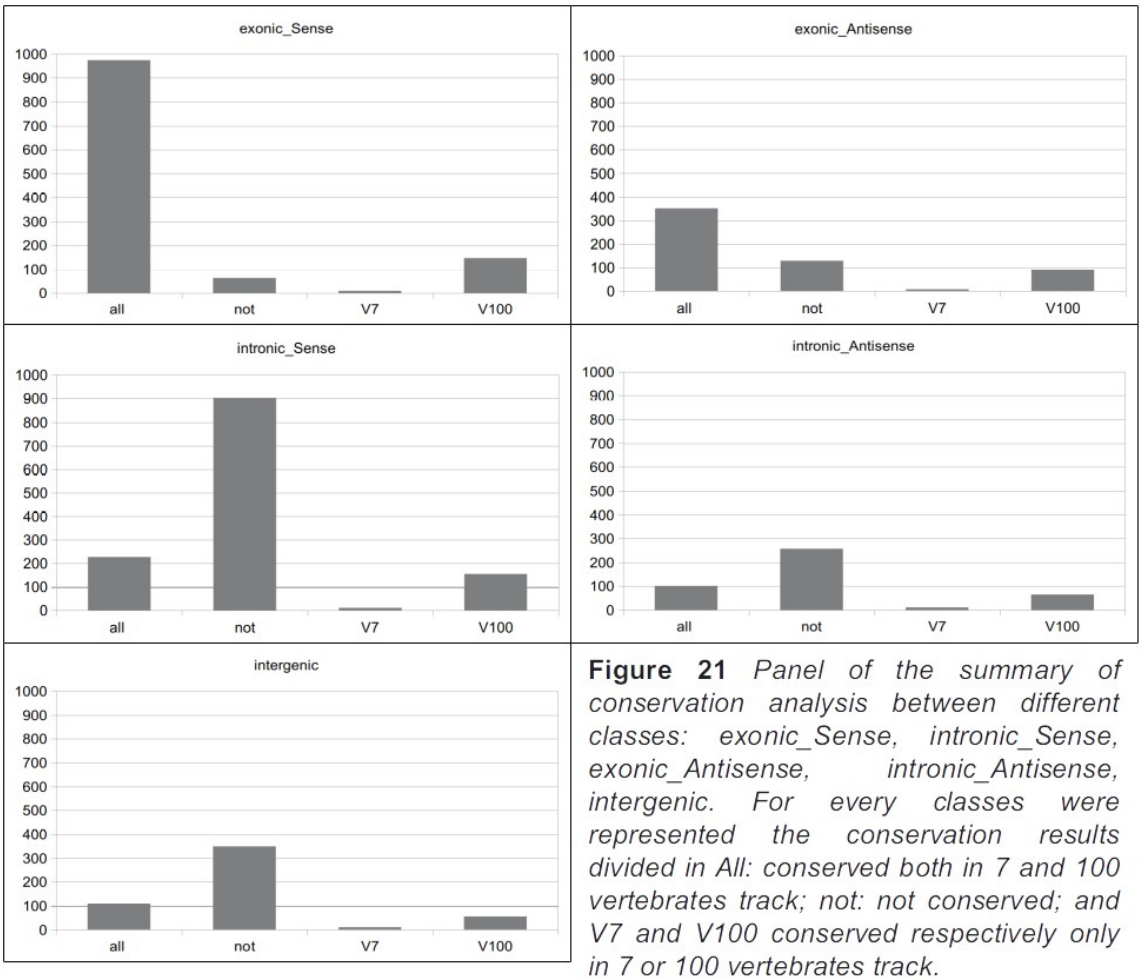


Figure 21 *Panel of the summary of conservation analysis between different classes: exonic_Sense, intronic_Sense, exonic_Antisense, intronic_Antisense, intergenic. For every classes were represented the conservation results divided in All: conserved both in 7 and 100 vertebrates track; not: not conserved; and V7 and V100 conserved respectively only in 7 or 100 vertebrates track.*

When we compared the results between various classes, see panel Figure 21, we observed that the exonic_Sense resulted in the best conservation of probesets, with 947 conserved in all the two tracks, corresponding to 82%. These results confirmed the previous data classification, because we anticipated that the probesets classified as exonic_Sense probably belonged to an exon of known genes. Instead, in the group of intronic_Sense, 903 probesets were not conserved, this information is useful to reduce the number of false positives and to focus the next analysis on the 227 conserved, corresponding to 18%. The antisense classes showed opposite results between the exonic and intronic groups. The high number of 351 probesets, that were more conserved in exonic_Antisense group could confirm the hypothesis of an effect of a replication fork. Finally, we could reduce

the number of interesting intergenic probesets to the 21%, corresponding to the 109 intergenic probesets conserved.

4.7 Coding potential analysis

We performed a coding potential analysis with CPAT tool based on the sequences of the 4,024 validated probesets. As reported in the literature we retained probable coding sequences, those that presented a CP > or = of 0.364 and non-coding regions those that have a CP value less than 0.364.

In Table 6 we reported the counts of the probable coding regions in the classes of re-annotation analysis with GencodeV28. We observed that the major part of the probesets resulted in predictive non-coding regions. This result could be positive but we observed a big proportion, 73%, of the predictive non-coding regions also in the exonic_Sense class and this is discordant with our expectation.

Though CPAT presented a sensitivity of 97% and a specificity of 97% both calculated on a 4,000 non-coding genes test, we supposed that CPAT analysis could not optimal because the probesets presented a small length size, that could falsifies the ORF length and ORF coverage calculations.

Table 6 ***Coding-Potential analysis***

Classification_GencodeV29	tot	non_coding	coding
exonicSense_ExonicSense	1192	872	320
intronicSense_positiveCorr	963	871	92
intronicSense_negativeCorr	10	8	2
intronicSense_notCorr	208	188	20
intronicSense_noHostGene	115	99	16
exonicAntisense_positiveCorr	339	266	73
exonicAntisense_negativeCorr	11	10	1
exonicAntisense_notCorr	189	152	37
exonicAntisense_notHostGene	39	32	7
intronicAntisense_positiveCorr	169	154	15
intronicAntisense_negativeCorr	14	12	2
intronicAntisense_notCorr	226	204	22
intronicAntisense_notHostGene	25	22	3
OutGen_SenseCorr	190	164	26
OutGen_Independent	173	153	20
Out_Gen_noHostGene	160	150	10

Table 6 *CPAT analysis on the 4,024 validated orphan probesets*

4.8 Cancer Genes Census correlation

We performed a correlation analysis between Cancer Genes Census (CGC) and the validated 4,024 orphan probesets. About 93,101 showed significant correlation between CGC and probesets, we retained the first more correlated gene for each probeset, both positive and/or negative correlations.

These resulted in 1,868 probesets positively correlated with 219 CGC genes and 1,701 probesets negatively correlated with 186 CGC genes. In order to give a functional biological predictive sense of our orphan probesets, we performed a functional analysis of the CGC genes probesets correlated with Panther.

In Figure 22 we summarized, in two pie charts, the molecular functions of the positive and negative correlated CGC with probesets. We observed that the molecular functions are the same, except from the channel regulatory activity.

Figure 22 Pie chart of molecular function classification of CGC with Panther.

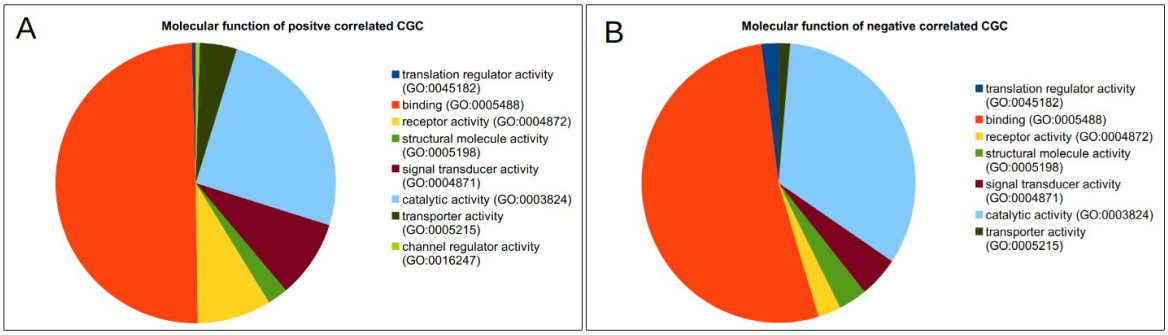


Figure 22 *Piechart of molecular function classification of CGC with Panther. A: CGC genes positive regulated with orphan probesets; B: CGC negative correlated with orphan probesets.*

All the significant results are showed in the tables of APPENDIX C.

5 Conclusion

In a previous study we used data from cancer analyzed with Affymetrix Human Exon Chip1_0ST to study the a lncRNA embedded within the TGM2 gene [7]. The Affymetrix Human Exon Chip allowed us to validate the presence of lncRNAs in an intron of the gene. Thus, following this successful study we decide to extend the method and use it to analyze the entire human genomic space in search of novel lncRNAs. We discovered that in our Affymetrix Human Exon dataset there were a large number of probesets than did not overlap with any type of known transcript in the current genome annotation. In fact, searching through literature we found that approximately 50% of Affymetrix Human Exon probesets target sequences outside Ensembl-defined exons. Therefore this Exon Human array to some extent shares the potential advantages of RNA-sequencing technology in detecting transcription outside well-characterized loci [77]. There are many studies that demonstrate that previously built microarrays designed for protein coding genes, contain probe sequences that overlap lncRNAs [138,139]. We investigated other studies in depth and found that there are many investigators who re-annotated the Human Exon chip probesets to discover novel long non-coding involved in cancer. For example, a study used the Human Exon chip to compare the conservation of antisense transcripts in mammals[140]. Other reports used this platform to investigate isoform variants involved in other diseases [141–143] or to detect gene fusion, somatic mutation patterns and pathway alterations causing cancer [144,145]. In a precedent study, the Human Exon chip was used to identify novel exons involved in glial brain tumor [146]. In a recent study, gene probe and transcript probe, from Affymetrix Gene Chip, were updated on the most recent knowledge of current genome and transcriptome. The single probes and the probe groups were reassigned into new genes and transcripts, based on functional region such as UTR, individual exon, CDS [147]. It was demonstrated that, among the type of Affymetrix expression chips, Human Exon Chip is the best to study novel ncRNAs. It contains more than 46% of not-annotated probesets compared with the about 12%, 21% and 6% of the HG U133A, HG 133 Plus 2.0 and Human Gene 1.0 respectively [148]. In a previous study, public data of gastric cancer were used for the re-annotation of over 6.5 million probes of the Affymetrix Human Exon 1.0 ST array to study long non-coding RNA. The study found that 136,053 probes

uniquely overlapped 9,294 lncRNAs at the gene level [149]. In another study, a computational pipeline was designed to re-annotate the probes from different Affymetrix array types. Among these, Affymetrix Human Exon 1.0 ST array demonstrated the most comprehensive coverage of the annotated human lncRNA [150]. A group of studies used the unique designed of Affymetrix Human Exon Array to discover specific novel lncRNAs in prostate cancer. They searched for un-annotated transcript clusters of probesets, which are differentially expressed in prostate cancer. Of 334 candidates 15 were validated in RT-PCR, confirming the usefulness of Affymetrix Human Exon in novel lncRNAs discovery [151].

At this stage the main advantage of using Affymetrix datasets over RNAseq cohorts is the much larger sample size of the chip-based cohorts. Thus the much higher statistical power in detecting true positive novel lncRNAs. Additionally, all the previous studies I cited above worked on a specific type of cancer. So, we decided to apply a pan-cancer approach, pooling datasets from different cancer types. Hence, we searched on GEO all published datasets of cancer tissue, tumor cell line and normal tissue, analyzed with the AffyetrixHumanExon1.0ST array and obtained about 9,000 samples. After exclusion from of datasets for which the original raw CEL files were not available or the original samples were from FFPE, we selected 1299 samples. These samples were split into two datasets, one used for pipeline analysis and one for the validation. The analysis dataset was composed of those samples that presented paired tumor and normal tissues from the same patient. The validation dataset contained 941 samples of tumor cell line, tumor tissue (11 different type of cancer) and white blood cells (PBMC), used as normal samples. The pipeline consisted of three different parts, based on three different considerations and three different bioinformatic tools. The first part was designed to assemble and normalize public data from different groups, to discover the differentially expressed probesets between tumor and normal paired tissue. First, R was used to construct and normalize the value of expression of the tumor and normal tissue sample, analyzed with Affymetrix Human Exon chip. R was also used with its package Limma to output the probesets related to cancer. Limma gave us the logFC and p.value that showed which probesets were significantly differentially expressed between the two groups (normal and tumor). The second part of pipeline was designed to discover non-annotated probesets, defined as orphan RNAs. Using Bedtools, the genome coordinate bed files for differentially

expressed probesets were intersected with genome coordinates.bed files of the 3 datasets of known genes with the exons from whole transcript (RefSeq, UCSC and GencodeV24). The antisense probesets were considered as predicted novel transcripts because being on the opposite strands of genes they did not really overlap known transcripts. Instead the intergenic probesets and the intronic probesets were used in the third and last part of the pipeline, written in Python. To validate the intronic and intergenic orphan probesets, must be excluded the correlation between these probesets and their overlapped or flanked genes with the same strand. We obtained more than 9,000 orphan probesets. Then, we re-utilized the first part of pipeline to validate them in the second dataset of Affymetrix HumanExonChip1.0ST. We can validate about 8,000 probesets that are significantly expressed in other 751 samples, increasing the total number of samples. Furthermore, with limma analysis we validated that 2/3 of the 9,000 probesets which were shown to be differentially expressed between tumor samples (cell and tissue) and PBMC considered as normal.

Then, we decided to switch to RNA-seq data to validate them. To perform this we used the public Encode dataset of the stem cell, induced pluripotent stem cell, tumor cell line and normal tissue. The re-mapping of probesets on this dataset allowed us to validate their expression, based on the presence or absence of reads in the probeset regions. We validate 4,024 orphan probesets that also expressed in RNA-seq data. Then, we classified the validated orphan probesets with the last version of Gencode, V29. We obtained about 500 probesets that could be novel lncRNAs. We obtained in depth intronicSense probesets that were negatively correlated (10) or not correlated (208) with the host genes. Between these, we are sure that they are independent, compared to host genes. Next, 25 probesets resulted as antisenses with a negative correlation with other genes so we supposed that they are novel NATs. Finally, we discovered 173 intragenic independent probesets that could be new transcripts of lncRNAs. Furthermore, we clustered the probesets based on their position. If more probesets resulted in the same host genes, concordant and correlated to each other, were grouped in a cluster of probeset that may cover the same lncRNA.

Through conservation analysis, with 100 vertebrates and 7 mammals on GenomeBrowser tracks, we have validated and discarded the exonic_Sense probesets, that resulted as expected, the best conserved. Furthermore, we have

reduced the number of probesets in the classes of intronic_Sense, antisense and intergenic probesets. This could indicate that these conserved probesets could effectively be novel transcripts of lncRNAs.

The coding potential analysis, probably, wasn't significant because of some of the probesets presented a small length size and this could falsifies the ORF length and ORF coverage calculations. Indeed we observed an unexpected phenomenon, that the probesets classified as exonic region presented a big number of non-coding regions. The CPAT analysis results were not good for our type of data so, other methods are needed to perform a classification of the probesets based on their coding potential. The correlation with CGC validated some of the orphan probesets as cancer related. It would be appropriate to deeply study this connection, one by one, to hypothesize a specific function of probesets.

Considering that all our samples are of total RNA, it would be interesting to see which probesets fall into poly-A sequences. Further future studies will be interesting, to deeply research these novel transcripts, passing to the *in vitro* analysis. The first step will be to validate the expression of this novel transcripts with q-PCT in a consistent number of samples, such as tumor cancer cell line or cancer fresh tissue. The next step will be the analysis of their function.

In conclusion with our pipeline, we used a pan-cancer dataset of microarrays and validation in RNAseq to discover novel candidate RNAs. These studies are an important starting point for the identification of novel RNAs *in vitro* and *in vivo*.

References

1. Kapranov, P.; Cheng, J.; Dike, S.; Nix, D.A.; Duttagupta, R.; Willingham, A.T.; Stadler, P.F.; Hertel, J.; Hackermüller, J.; Hofacker, I.L.; et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **2007**, *316*, 1484–1488, doi:10.1126/science.1138341.
2. Harrow, J.; Frankish, A.; Gonzalez, J.M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B.L.; Barrell, D.; Zadissa, A.; Searle, S.; et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **2012**, *22*, 1760–1774, doi:10.1101/gr.135350.111.
3. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927, doi:10.1101/gad.17446611.
4. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009**, *458*, 223–227, doi:10.1038/nature07672.
5. Faghihi, M.A.; Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 637–643, doi:10.1038/nrm2738.
6. Chen, J.; Sun, M.; Hurst, L.D.; Carmichael, G.G.; Rowley, J.D. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends Genet.* **2005**, *21*, 326–329, doi:10.1016/j.tig.2005.04.006.
7. Minotti, L.; Baldassari, F.; Galasso, M.; Volinia, S.; Bergamini, C.M.; Bianchi, N. A long non-coding RNA inside the type 2 transglutaminase gene tightly correlates with the expression of its transcriptional variants. *Amino Acids* **2018**, *50*, 421–438, doi:10.1007/s00726-017-2528-9.
8. Hou, L.-D.; Zhang, J. Circular RNAs: An emerging type of RNA in cancer. *Int J Immunopathol Pharmacol* **2017**, *30*, 1–6, doi:10.1177/0394632016686985.
9. Mercer, T.R.; Dinger, M.E.; Mattick, J.S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **2009**, *10*, 155–159, doi:10.1038/nrg2521.
10. Moran, V.A.; Perera, R.J.; Khalil, A.M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res.* **2012**, *40*, 6391–6400, doi:10.1093/nar/gks296.
11. Lin, A.; Li, C.; Xing, Z.; Hu, Q.; Liang, K.; Han, L.; Wang, C.; Hawke, D.H.; Wang, S.; Zhang, Y.; et al. The LINK-A lncRNA activates normoxic HIF1α signalling in triple-negative breast cancer. *Nat. Cell Biol.* **2016**, *18*, 213–224, doi:10.1038/ncb3295.
12. Fan, J.; Xing, Y.; Wen, X.; Jia, R.; Ni, H.; He, J.; Ding, X.; Pan, H.; Qian, G.; Ge, S.; et al. Long non-coding RNA ROR decoys gene-specific histone methylation to promote tumorigenesis. *Genome Biol.* **2015**, *16*, 139, doi:10.1186/s13059-015-0705-2.
13. Hajjari, M.; Salavaty, A. HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol Med* **2015**, *12*, 1–9, doi:10.7497/j.issn.2095-3941.2015.0006.
14. Ribeiro, D.M.; Zanzoni, A.; Cipriano, A.; Delli Ponti, R.; Spinelli, L.; Ballarino, M.; Bozzoni, I.; Tartaglia, G.G.; Brun, C. Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. *Nucleic Acids Res.* **2018**, *46*, 917–928, doi:10.1093/nar/gkx1169.
15. Aguilo, F.; Zhou, M.-M.; Walsh, M.J. Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer Res.* **2011**, *71*, 5365–5369, doi:10.1158/0008-5472.CAN-10-4379.
16. Kotake, Y.; Nakagawa, T.; Kitagawa, K.; Suzuki, S.; Liu, N.; Kitagawa, M.; Xiong, Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of

p15(INK4B) tumor suppressor gene. *Oncogene* **2011**, *30*, 1956–1962, doi:10.1038/onc.2010.568.
17. Wu, H.; Hu, Y.; Liu, X.; Song, W.; Gong, P.; Zhang, K.; Chen, Z.; Zhou, M.; Shen, X.; Qian, Y.; et al. LncRNA TRERNA1 Function as an Enhancer of SNAI1 Promotes Gastric Cancer Metastasis by Regulating Epithelial-Mesenchymal Transition. *Mol Ther Nucleic Acids* **2017**, *8*, 291–299, doi:10.1016/j.omtn.2017.06.021.
18. Huang, J.-Z.; Chen, M.; Chen, D.; Gao, X.-C.; Zhu, S.; Huang, H.; Hu, M.; Zhu, H.; Yan, G.-R. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* **2017**, *68*, 171-184.e6, doi:10.1016/j.molcel.2017.09.015.
19. Karimi, P.; Islami, F.; Anandasabapathy, S.; Freedman, N.D.; Kamangar, F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol. Biomarkers Prev.* **2014**, *23*, 700–713, doi:10.1158/1055-9965.EPI-13-1057.
20. Jemal, A.; Center, M.M.; DeSantis, C.; Ward, E.M. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol. Biomarkers Prev.* **2010**, *19*, 1893–1907, doi:10.1158/1055-9965.EPI-10-0437.
21. Soerjomataram, I.; Lortet-Tieulent, J.; Parkin, D.M.; Ferlay, J.; Mathers, C.; Forman, D.; Bray, F. Global burden of cancer in 2008: a systematic analysis of disability-adjusted life-years in 12 world regions. *Lancet* **2012**, *380*, 1840–1850, doi:10.1016/S0140-6736(12)60919-2.
22. Ferlay, J.; Shin, H.-R.; Bray, F.; Forman, D.; Mathers, C.; Parkin, D.M. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **2010**, *127*, 2893–2917, doi:10.1002/ijc.25516.
23. Suzuki, H.; Iwasaki, E.; Hibi, T. Helicobacter pylori and gastric cancer. *Gastric Cancer* **2009**, *12*, 79–87, doi:10.1007/s10120-009-0507-x.
24. Oliveira, C.; Pinheiro, H.; Figueiredo, J.; Seruca, R.; Carneiro, F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. *Lancet Oncol.* **2015**, *16*, e60-70, doi:10.1016/S1470-2045(14)71016-2.
25. Song, Z.; Wu, Y.; Yang, J.; Yang, D.; Fang, X. Progress in the treatment of advanced gastric cancer. *Tumour Biol.* **2017**, *39*, 1010428317714626, doi:10.1177/1010428317714626.
26. Huang, Y.; Zhang, J.; Hou, L.; Wang, G.; Liu, H.; Zhang, R.; Chen, X.; Zhu, J. LncRNA AK023391 promotes tumorigenesis and invasion of gastric cancer through activation of the PI3K/Akt signaling pathway. *J. Exp. Clin. Cancer Res.* **2017**, *36*, 194, doi:10.1186/s13046-017-0666-2.
27. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2015**, *136*, E359-386, doi:10.1002/ijc.29210.
28. Migliore, L.; Migheli, F.; Spisni, R.; Coppedè, F. Genetics, cytogenetics, and epigenetics of colorectal cancer. *J. Biomed. Biotechnol.* **2011**, *2011*, 792362, doi:10.1155/2011/792362.
29. Coppedè, F.; Lopomo, A.; Spisni, R.; Migliore, L. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *World J. Gastroenterol.* **2014**, *20*, 943–956, doi:10.3748/wjg.v20.i4.943.
30. Kekelidze, M.; D'Errico, L.; Pansini, M.; Tyndall, A.; Hohmann, J. Colorectal cancer: current imaging methods and future perspectives for the diagnosis, staging and therapeutic response evaluation. *World J. Gastroenterol.* **2013**, *19*, 8502–8514, doi:10.3748/wjg.v19.i46.8502.
31. Philip, A.K.; Lubner, M.G.; Harms, B. Computed tomographic colonography. *Surg Clin North Am* **2011**, *91*, 127–139, doi:10.1016/j.suc.2010.10.010.

32. Ip, S.; Sokoro, A.A.H.; Kaita, L.; Ruiz, C.; McIntyre, E.; Singh, H. Use of fecal occult blood testing in hospitalized patients: results of an audit. *Can J Gastroenterol Hepatol* **2014**, *28*, 489–494.
33. Van Schaeybroeck, S.; Allen, W.L.; Turkington, R.C.; Johnston, P.G. Implementing prognostic and predictive biomarkers in CRC clinical trials. *Nat Rev Clin Oncol* **2011**, *8*, 222–232, doi:10.1038/nrclinonc.2011.15.
34. Orang, A.V.; Barzegari, A. MicroRNAs in colorectal cancer: from diagnosis to targeted therapy. *Asian Pac. J. Cancer Prev.* **2014**, *15*, 6989–6999.
35. Ouyang, S.; Zheng, X.; Zhou, X.; Chen, Z.; Yang, X.; Xie, M. LncRNA BCAR4 promotes colon cancer progression via activating Wnt/β-catenin signaling. *Oncotarget* **2017**, *8*, 92815–92826, doi:10.18632/oncotarget.21590.
36. Duguang, L.; Jin, H.; Xiaowei, Q.; Peng, X.; Xiaodong, W.; Zhennan, L.; Jianjun, Q.; Jie, Y. The involvement of lncRNAs in the development and progression of pancreatic cancer. *Cancer Biol Ther* **2017**, *18*, 927–936, doi:10.1080/15384047.2017.1385682.
37. Hruban, R.H.; Klimstra, D.S. Adenocarcinoma of the pancreas. *Semin Diagn Pathol* **2014**, *31*, 443–451, doi:10.1053/j.semdp.2014.08.004.
38. Liu, J.-H.; Chen, G.; Dang, Y.-W.; Li, C.-J.; Luo, D.-Z. Expression and prognostic significance of lncRNA MALAT1 in pancreatic cancer tissues. *Asian Pac. J. Cancer Prev.* **2014**, *15*, 2971–2977.
39. Davis, G.L.; Dempster, J.; Meler, J.D.; Orr, D.W.; Walberg, M.W.; Brown, B.; Berger, B.D.; O'Connor, J.K.; Goldstein, R.M. Hepatocellular carcinoma: management of an increasingly common problem. *Proc (Bayl Univ Med Cent)* **2008**, *21*, 266–280.
40. Balogh, J.; Victor, D.; Asham, E.H.; Burroughs, S.G.; Boktour, M.; Saharia, A.; Li, X.; Ghobrial, R.M.; Monsour, H.P. Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma* **2016**, *3*, 41–53, doi:10.2147/JHC.S61146.
41. Hu, J.; Song, C.; Duan, B.; Zhang, X.; Li, D.; Zhu, L.; Gao, H. LncRNA-SVUGP2 suppresses progression of hepatocellular carcinoma. *Oncotarget* **2017**, *8*, 97835–97850, doi:10.18632/oncotarget.18279.
42. Molina, J.R.; Yang, P.; Cassivi, S.D.; Schild, S.E.; Adjei, A.A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* **2008**, *83*, 584–594, doi:10.4065/83.5.584.
43. NSCLC: An Update of Driver Mutations, Their Role in Pathogenesis and Clinical Significance. - PubMed - NCBI Available online: https://www.ncbi.nlm.nih.gov/pubmed/26422542 (accessed on Nov 23, 2018).
44. Zhan, Y.; Zang, H.; Feng, J.; Lu, J.; Chen, L.; Fan, S. Long non-coding RNAs associated with non-small cell lung cancer. *Oncotarget* **2017**, *8*, 69174–69184, doi:10.18632/oncotarget.20088.
45. Cohen, H.T.; McGovern, F.J. Renal-cell carcinoma. *N. Engl. J. Med.* **2005**, *353*, 2477–2490, doi:10.1056/NEJMra043172.
46. Van Poppel, H.; Becker, F.; Cadeddu, J.A.; Gill, I.S.; Janetschek, G.; Jewett, M.A.S.; Laguna, M.P.; Marberger, M.; Montorsi, F.; Polascik, T.J.; et al. Treatment of localised renal cell carcinoma. *Eur. Urol.* **2011**, *60*, 662–672, doi:10.1016/j.eururo.2011.06.040.
47. Zheng, Z.; Zhao, F.; Zhu, D.; Han, J.; Chen, H.; Cai, Y.; Chen, Z.; Xie, W. Long Non-Coding RNA LUCAT1 Promotes Proliferation and Invasion in Clear Cell Renal Cell Carcinoma Through AKT/GSK-3β Signaling Pathway. *Cell. Physiol. Biochem.* **2018**, *48*, 891–904, doi:10.1159/000491957.
48. Kim, J.; Coffey, D.M.; Creighton, C.J.; Yu, Z.; Hawkins, S.M.; Matzuk, M.M. High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3921–3926, doi:10.1073/pnas.1117135109.

49. Scully, R.E. Classification of human ovarian tumors. *Environ Health Perspect* **1987**, *73*, 15–24.
50. Lessard-Anderson, C.R.; Handlogten, K.S.; Molitor, R.J.; Dowdy, S.C.; Cliby, W.A.; Weaver, A.L.; Sauver, J.S.; Bakkum-Gamez, J.N. Effect of tubal sterilization technique on risk of serous epithelial ovarian and primary peritoneal carcinoma. *Gynecol. Oncol.* **2014**, *135*, 423–427, doi:10.1016/j.ygyno.2014.10.005.
51. Zhan, L.; Li, J.; Wei, B. Long non-coding RNAs in ovarian cancer. *J Exp Clin Cancer Res* **2018**, *37*, doi:10.1186/s13046-018-0793-4.
52. Cuzick, J.; Thorat, M.A.; Andriole, G.; Brawley, O.W.; Brown, P.H.; Culig, Z.; Eeles, R.A.; Ford, L.G.; Hamdy, F.C.; Holmberg, L.; et al. Prevention and early detection of prostate cancer. *Lancet Oncol.* **2014**, *15*, e484-492, doi:10.1016/S1470-2045(14)70211-6.
53. Attard, G.; Parker, C.; Eeles, R.A.; Schröder, F.; Tomlins, S.A.; Tannock, I.; Drake, C.G.; de Bono, J.S. Prostate cancer. *Lancet* **2016**, *387*, 70–82, doi:10.1016/S0140-6736(14)61947-4.
54. Chen, F.-Z.; Zhao, X.-K. Prostate cancer: current treatment and prevention strategies. *Iran Red Crescent Med J* **2013**, *15*, 279–284, doi:10.5812/ircmj.6499.
55. Lee, B.; Mazar, J.; Aftab, M.N.; Qi, F.; Shelley, J.; Li, J.-L.; Govindarajan, S.; Valerio, F.; Rivera, I.; Thurn, T.; et al. Long noncoding RNAs as putative biomarkers for prostate cancer detection. *J Mol Diagn* **2014**, *16*, 615–626, doi:10.1016/j.jmoldx.2014.06.009.
56. Döhner, H.; Weisdorf, D.J.; Bloomfield, C.D. Acute Myeloid Leukemia. *N. Engl. J. Med.* **2015**, *373*, 1136–1152, doi:10.1056/NEJMra1406184.
57. Ding, L.; Ley, T.J.; Larson, D.E.; Miller, C.A.; Koboldt, D.C.; Welch, J.S.; Ritchey, J.K.; Young, M.A.; Lamprecht, T.; McLellan, M.D.; et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **2012**, *481*, 506–510, doi:10.1038/nature10738.
58. Lei, L.; Xia, S.; Liu, D.; Li, X.; Feng, J.; Zhu, Y.; Hu, J.; Xia, L.; Guo, L.; Chen, F.; et al. Genome-wide characterization of lncRNAs in acute myeloid leukemia. *Brief. Bioinformatics* **2018**, *19*, 627–635, doi:10.1093/bib/bbx007.
59. Mer, A.S.; Lindberg, J.; Nilsson, C.; Klevebring, D.; Wang, M.; Grönberg, H.; Lehmann, S.; Rantalainen, M. Expression levels of long non-coding RNAs are prognostic for AML outcome. *J Hematol Oncol* **2018**, *11*, 52, doi:10.1186/s13045-018-0596-2.
60. Baldassari, F.; Zerbinati, C.; Galasso, M.; Corrà, F.; Minotti, L.; Agnoletto, C.; Previati, M.; Croce, C.M.; Volinia, S. Screen for MicroRNA and Drug Interactions in Breast Cancer Cell Lines Points to miR-126 as a Modulator of CDK4/6 and PIK3CA Inhibitors. *Front Genet* **2018**, *9*, 174, doi:10.3389/fgene.2018.00174.
61. Arnold, M.; Karim-Kos, H.E.; Coebergh, J.W.; Byrnes, G.; Antilla, A.; Ferlay, J.; Renehan, A.G.; Forman, D.; Soerjomataram, I. Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory. *Eur. J. Cancer* **2015**, *51*, 1164–1187, doi:10.1016/j.ejca.2013.09.002.
62. Tripathi, R.; Soni, A.; Varadwaj, P.K. Integrated analysis of dysregulated lncRNA expression in breast cancer cell identified by RNA-seq study. *Non-coding RNA Research* **2016**, *1*, 35–42, doi:10.1016/j.ncrna.2016.09.002.
63. Gladson, C.L.; Prayson, R.A.; Liu, W.M. The pathobiology of glioma tumors. *Annu Rev Pathol* **2010**, *5*, 33–50, doi:10.1146/annurev-pathol-121808-102109.
64. Wen, P.Y.; Kesari, S. Malignant gliomas in adults. *N. Engl. J. Med.* **2008**, *359*, 492–507, doi:10.1056/NEJMra0708126.
65. Xi, J.; Sun, Q.; Ma, L.; Kang, J. Long non-coding RNAs in glioma progression. *Cancer Letters* **2018**, *419*, 203–209, doi:10.1016/j.canlet.2018.01.041.
66. Foletto, M.C.; Haas, S.E. Cutaneous melanoma: new advances in treatment. *Anais Brasileiros de Dermatologia* **2014**, *89*, 301, doi:10.1590/abd1806-4841.20142540.

67. Tsao, H.; Atkins, M.B.; Sober, A.J. Management of cutaneous melanoma. *N. Engl. J. Med.* **2004**, *351*, 998–1012, doi:10.1056/NEJMra041245.
68. Yang, S.; Xu, J.; Zeng, X. A six-long non-coding RNA signature predicts prognosis in melanoma patients. *Int J Oncol* **2018**, *52*, 1178–1188, doi:10.3892/ijo.2018.4268.
69. Sanderson, R.J.; Ironside, J.A.D. Squamous cell carcinomas of the head and neck. *BMJ* **2002**, *325*, 822–827.
70. Bonnin, N.; Armandy, E.; Carras, J.; Ferrandon, S.; Battiston-Montagne, P.; Aubry, M.; Guihard, S.; Meyronet, D.; Foy, J.-P.; Saintigny, P.; et al. MiR-422a promotes loco-regional recurrence by targeting NT5E/CD73 in head and neck squamous cell carcinoma. *Oncotarget* **2016**, *7*, 44023–44038, doi:10.18632/oncotarget.9829.
71. Ma, X.; Sheng, S.; Wu, J.; Jiang, Y.; Gao, X.; Cen, X.; Wu, J.; Wang, S.; Tang, Y.; Tang, Y.; et al. LncRNAs as an intermediate in HPV16 promoting myeloid-derived suppressor cell recruitment of head and neck squamous cell carcinoma. *Oncotarget* **2017**, *8*, 42061–42075, doi:10.18632/oncotarget.14939.
72. Lockstone, H.E. Exon array data analysis using Affymetrix power tools and R statistical software. *Brief. Bioinformatics* **2011**, *12*, 634–644, doi:10.1093/bib/bbq086.
73. Kapur, K.; Xing, Y.; Ouyang, Z.; Wong, W.H. Exon arrays provide accurate assessments of gene expression. *Genome Biol.* **2007**, *8*, R82, doi:10.1186/gb-2007-8-5-r82.
74. Kwan, T.; Benovoy, D.; Dias, C.; Gurd, S.; Serre, D.; Zuzan, H.; Clark, T.A.; Schweitzer, A.; Staples, M.K.; Wang, H.; et al. Heritability of alternative splicing in the human genome. *Genome Res.* **2007**, *17*, 1210–1218, doi:10.1101/gr.6281007.
75. Clark, T.A.; Schweitzer, A.C.; Chen, T.X.; Staples, M.K.; Lu, G.; Wang, H.; Williams, A.; Blume, J.E. Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.* **2007**, *8*, R64, doi:10.1186/gb-2007-8-4-r64.
76. Gardina, P.J.; Clark, T.A.; Shimada, B.; Staples, M.K.; Yang, Q.; Veitch, J.; Schweitzer, A.; Awad, T.; Sugnet, C.; Dee, S.; et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **2006**, *7*, 325, doi:10.1186/1471-2164-7-325.
77. Yates, T.; Okoniewski, M.J.; Miller, C.J. X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. *Nucleic Acids Res.* **2008**, *36*, D780-786, doi:10.1093/nar/gkm779.
78. Behjati, S.; Tarpey, P.S. What is next generation sequencing? *Arch Dis Child Educ Pract Ed* **2013**, *98*, 236–238, doi:10.1136/archdischild-2013-304340.
79. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63, doi:10.1038/nrg2484.
80. Qian, X.; Ba, Y.; Zhuang, Q.; Zhong, G. RNA-Seq technology and its application in fish transcriptomics. *OMICS* **2014**, *18*, 98–110, doi:10.1089/omi.2013.0110.
81. Esteve-Codina, A.; Arpi, O.; Martinez-García, M.; Pineda, E.; Mallo, M.; Gut, M.; Carrato, C.; Rovira, A.; Lopez, R.; Tortosa, A.; et al. A Comparison of RNA-Seq Results from Paired Formalin-Fixed Paraffin-Embedded and Fresh-Frozen Glioblastoma Tissue Samples. *PLoS One* **2017**, *12*, doi:10.1371/journal.pone.0170632.
82. Cui, P.; Lin, Q.; Ding, F.; Xin, C.; Gong, W.; Zhang, L.; Geng, J.; Zhang, B.; Yu, X.; Yang, J.; et al. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **2010**, *96*, 259–265, doi:10.1016/j.ygeno.2010.07.010.
83. Hastings, M.L.; Krainer, A.R. Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **2001**, *13*, 302–309.
84. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359, doi:10.1038/nmeth.1923.
85. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760, doi:10.1093/bioinformatics/btp324.

86. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S.L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **2013**, *14*, R36, doi:10.1186/gb-2013-14-4-r36.
87. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21, doi:10.1093/bioinformatics/bts635.
88. Grant, G.R.; Farkas, M.H.; Pizarro, A.D.; Lahens, N.F.; Schug, J.; Brunk, B.P.; Stoeckert, C.J.; Hogenesch, J.B.; Pierce, E.A. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics* **2011**, *27*, 2518–2528, doi:10.1093/bioinformatics/btr427.
89. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210.
90. Barrett, T.; Troup, D.B.; Wilhite, S.E.; Ledoux, P.; Rudnev, D.; Evangelista, C.; Kim, I.F.; Soboleva, A.; Tomashevsky, M.; Marshall, K.A.; et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* **2009**, *37*, D885-890, doi:10.1093/nar/gkn764.
91. ENCODE Project Consortium The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **2004**, *306*, 636–640, doi:10.1126/science.1105136.
92. ENCODE Project Consortium; Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigó, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T.; et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **2007**, *447*, 799–816, doi:10.1038/nature05874.
93. ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74, doi:10.1038/nature11247.
94. ENCODE Project Consortium A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **2011**, *9*, e1001046, doi:10.1371/journal.pbio.1001046.
95. Sloan, C.A.; Chan, E.T.; Davidson, J.M.; Malladi, V.S.; Strattan, J.S.; Hitz, B.C.; Gabdank, I.; Narayanan, A.K.; Ho, M.; Lee, B.T.; et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **2016**, *44*, D726-732, doi:10.1093/nar/gkv1160.
96. Rosenbloom, K.R.; Sloan, C.A.; Malladi, V.S.; Dreszer, T.R.; Learned, K.; Kirkup, V.M.; Wong, M.C.; Maddren, M.; Fang, R.; Heitner, S.G.; et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* **2013**, *41*, D56-63, doi:10.1093/nar/gks1172.
97. Pruitt, K.D.; Maglott, D.R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **2001**, *29*, 137–140.
98. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006, doi:10.1101/gr.229102.
99. Hsu, F.; Kent, W.J.; Clawson, H.; Kuhn, R.M.; Diekhans, M.; Haussler, D. The UCSC Known Genes. *Bioinformatics* **2006**, *22*, 1036–1046, doi:10.1093/bioinformatics/btl048.
100. Harrow, J.; Denoeud, F.; Frankish, A.; Reymond, A.; Chen, C.-K.; Chrast, J.; Lagarde, J.; Gilbert, J.G.R.; Storey, R.; Swarbreck, D.; et al. GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **2006**, *7 Suppl 1*, S4.1-9, doi:10.1186/gb-2006-7-s1-s4.
101. Uszczynska-Ratajczak, B.; Lagarde, J.; Frankish, A.; Guigó, R.; Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **2018**, *19*, 535–548, doi:10.1038/s41576-018-0017-y.
102. GTEx Consortium The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585, doi:10.1038/ng.2653.

103. Kong, L.; Zhang, Y.; Ye, Z.-Q.; Liu, X.-Q.; Zhao, S.-Q.; Wei, L.; Gao, G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **2007**, *35*, W345-349, doi:10.1093/nar/gkm391.
104. Hu, L.; Xu, Z.; Hu, B.; Lu, Z.J. COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Res.* **2017**, *45*, e2, doi:10.1093/nar/gkw798.
105. Wang, L.; Park, H.J.; Dasari, S.; Wang, S.; Kocher, J.-P.; Li, W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **2013**, *41*, e74, doi:10.1093/nar/gkt006.
106. Kang, Y.-J.; Yang, D.-C.; Kong, L.; Hou, M.; Meng, Y.-Q.; Wei, L.; Gao, G. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16, doi:10.1093/nar/gkx428.
107. Achawanantakun, R.; Chen, J.; Sun, Y.; Zhang, Y. LncRNA-ID: Long non-coding RNA IDentification using balanced random forests. *Bioinformatics* **2015**, *31*, 3897–3905, doi:10.1093/bioinformatics/btv480.
108. Sun, L.; Liu, H.; Zhang, L.; Meng, J. lncRScan-SVM: A Tool for Predicting Long Non-Coding RNAs Using Support Vector Machine. *PLoS ONE* **2015**, *10*, e0139654, doi:10.1371/journal.pone.0139654.
109. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166, doi:10.1093/nar/gkt646.
110. Li, A.; Zhang, J.; Zhou, Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* **2014**, *15*, 311, doi:10.1186/1471-2105-15-311.
111. Sun, Z.; Nair, A.; Chen, X.; Prodduturi, N.; Wang, J.; Kocher, J.-P. UClncR: Ultrafast and comprehensive long non-coding RNA detection from RNA-seq. *Sci Rep* **2017**, *7*, 14196, doi:10.1038/s41598-017-14595-3.
112. Han, S.; Liang, Y.; Ma, Q.; Xu, Y.; Zhang, Y.; Du, W.; Wang, C.; Li, Y. LncFinder: an integrated platform for long non-coding RNA identification utilizing sequence intrinsic composition, structural information and physicochemical property. *Brief. Bioinformatics* **2018**, doi:10.1093/bib/bby065.
113. Yamada, A.; Yu, P.; Lin, W.; Okugawa, Y.; Boland, C.R.; Goel, A. A RNA-Sequencing approach for the identification of novel long non-coding RNA biomarkers in colorectal cancer. *Sci Rep* **2018**, *8*, 575, doi:10.1038/s41598-017-18407-6.
114. Gong, Y.; Huang, H.-T.; Liang, Y.; Trimarchi, T.; Aifantis, I.; Tsirigos, A. lncRNA-screen: an interactive platform for computationally screening long non-coding RNAs in large genomics datasets. *BMC Genomics* **2017**, *18*, 434, doi:10.1186/s12864-017-3817-0.
115. Zhang, J.; Liu, L.; Li, J.; Le, T.D. LncmiRSRN: identification and analysis of long non-coding RNA related miRNA sponge regulatory network in human cancer. *Bioinformatics* **2018**, doi:10.1093/bioinformatics/bty525.
116. Perron, U.; Provero, P.; Molineris, I. In silico prediction of lncRNA function using tissue specific and evolutionary conserved expression. *BMC Bioinformatics* **2017**, *18*, 144, doi:10.1186/s12859-017-1535-x.
117. Tripathi, R.; Patel, S.; Kumari, V.; Chakraborty, P.; Varadwaj, P.K. DeepLNC, a long non-coding RNA prediction tool using deep neural network. *Netw Model Anal Health Inform Bioinforma* **2016**, *5*, 21, doi:10.1007/s13721-016-0129-2.
118. Baek, J.; Lee, B.; Kwon, S.; Yoon, S. LncRNAnet: long non-coding RNA identification using deep learning. *Bioinformatics* **2018**, *34*, 3889–3897, doi:10.1093/bioinformatics/bty418.

119. Carvalho, B.S.; Irizarry, R.A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **2010**, *26*, 2363–2367, doi:10.1093/bioinformatics/btq431.
120. Irizarry, R.A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y.D.; Antonellis, K.J.; Scherf, U.; Speed, T.P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **2003**, *4*, 249–264, doi:10.1093/biostatistics/4.2.249.
121. Purdom, E.; Simpson, K.M.; Robinson, M.D.; Conboy, J.G.; Lapuk, A.V.; Speed, T.P. FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics* **2008**, *24*, 1707–1714, doi:10.1093/bioinformatics/btn284.
122. exon_probeset_trans_clust_whitepaper.pdf.
123. genefilter: methods for filtering genes from high-throughput experiments version 1.64.0 from Bioconductor Available online: https://rdrr.io/bioc/genefilter/ (accessed on Nov 24, 2018).
124. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47, doi:10.1093/nar/gkv007.
125. ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5373812/ (accessed on Nov 24, 2018).
126. Lundh, F. Academic citation for Python 2002.
127. Oliphant, T.E. Python for Scientific Computing. *Computing in Science Engineering* **2007**, *9*, 10–20, doi:10.1109/MCSE.2007.58.
128. SciPy: Open source scientific tools for Python | BibSonomy Available online: https://www.bibsonomy.org/bibtex/21b37d2cc741af879d7958f2f7c23c420/microcuts (accessed on Nov 24, 2018).
129. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515, doi:10.1038/nbt.1621.
130. Combining phylogenetic and hidden Markov models in biosequence analysis. - PubMed - NCBI Available online: https://www.ncbi.nlm.nih.gov/pubmed/15285899 (accessed on Nov 24, 2018).
131. Frith, M.C.; Bailey, T.L.; Kasukawa, T.; Mignone, F.; Kummerfeld, S.K.; Madera, M.; Sunkara, S.; Furuno, M.; Bult, C.J.; Quackenbush, J.; et al. Discrimination of non-protein-coding transcripts from protein-coding mRNA. *RNA Biol* **2006**, *3*, 40–48.
132. Fickett, J.W. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res.* **1982**, *10*, 5303–5318.
133. Fickett, J.W.; Tung, C.S. Assessment of protein coding measures. *Nucleic Acids Res.* **1992**, *20*, 6441–6450.
134. Futreal, P.A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M.R. A census of human cancer genes. *Nat. Rev. Cancer* **2004**, *4*, 177–183, doi:10.1038/nrc1299.
135. Mi, H.; Muruganujan, A.; Casagrande, J.T.; Thomas, P.D. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* **2013**, *8*, 1551–1566, doi:10.1038/nprot.2013.092.
136. Abdueva, D.; Wing, M.; Schaub, B.; Triche, T.; Davicioni, E. Quantitative expression profiling in formalin-fixed paraffin-embedded samples by affymetrix microarrays. *J Mol Diagn* **2010**, *12*, 409–417, doi:10.2353/jmoldx.2010.090155.
137. Wight, M.; Werner, A. The functions of natural antisense transcripts. *Essays Biochem* **2013**, *54*, 91–101, doi:10.1042/bse0540091.

138. Gellert, P.; Ponomareva, Y.; Braun, T.; Uchida, S. Noncoder: a web interface for exon array-based detection of long non-coding RNAs. *Nucleic Acids Res.* **2013**, *41*, e20, doi:10.1093/nar/gks877.
139. Liao, Q.; Xiao, H.; Bu, D.; Xie, C.; Miao, R.; Luo, H.; Zhao, G.; Yu, K.; Zhao, H.; Skogerbø, G.; et al. ncFANs: a web server for functional annotation of long non-coding RNAs. *Nucleic Acids Res.* **2011**, *39*, W118-124, doi:10.1093/nar/gkr432.
140. Ling, M.H.; Ban, Y.; Wen, H.; Wang, S.M.; Ge, S.X. Conserved expression of natural antisense transcripts in mammals. *BMC Genomics* **2013**, *14*, 243, doi:10.1186/1471-2164-14-243.
141. Bemmo, A.; Benovoy, D.; Kwan, T.; Gaffney, D.J.; Jensen, R.V.; Majewski, J. Gene expression and isoform variation analysis using Affymetrix Exon Arrays. *BMC Genomics* **2008**, *9*, 529, doi:10.1186/1471-2164-9-529.
142. Yamamoto, M.L.; Clark, T.A.; Gee, S.L.; Kang, J.-A.; Schweitzer, A.C.; Wickrema, A.; Conboy, J.G. Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis. *Blood* **2009**, *113*, 3363–3370, doi:10.1182/blood-2008-05-160325.
143. Risueño, A.; Roson-Burgo, B.; Dolnik, A.; Hernandez-Rivas, J.M.; Bullinger, L.; De, J.L.R. A robust estimation of exon expression to identify alternative spliced genes applied to human tissues and cancer samples. *BMC Genomics* **2014**, *15*, 879–879, doi:10.1186/1471-2164-15-879.
144. Lin, E.; Li, L.; Guan, Y.; Soriano, R.; Rivers, C.S.; Mohan, S.; Pandita, A.; Tang, J.; Modrusan, Z. Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. *Mol. Cancer Res.* **2009**, *7*, 1466–1476, doi:10.1158/1541-7786.MCR-08-0522.
145. Kan, Z.; Jaiswal, B.S.; Stinson, J.; Janakiraman, V.; Bhatt, D.; Stern, H.M.; Yue, P.; Haverty, P.M.; Bourgon, R.; Zheng, J.; et al. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **2010**, *466*, 869–873, doi:10.1038/nature09208.
146. French, P.J.; Peeters, J.; Horsman, S.; Duijm, E.; Siccama, I.; Van, M. den B.; Luider, T.M.; Kros, J.M.; Van, P. der S.; Sillevis, P.S. Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. *Cancer Res* **2007**, *67*, 5635–5642, doi:10.1158/0008-5472.CAN-06-2869.
147. Saka, E.; Harrison, B.J.; West, K.; Petruska, J.C.; Rouchka, E.C. Framework for reanalysis of publicly available Affymetrix® GeneChip® data sets based on functional regions of interest. BMC Genomics **2017**, 18.
148. Risueño, A.; Fontanillo, C.; Dinger, M.E.; De Las Rivas, J. GATExplorer: genomic and transcriptomic explorer; mapping expression probes to gene loci, transcripts, exons and ncRNAs. *BMC Bioinformatics* **2010**, *11*, 221, doi:10.1186/1471-2105-11-221.
149. Cao, W.-J.; Wu, H.-L.; He, B.-S.; Zhang, Y.-S.; Zhang, Z.-Y. Analysis of long non-coding RNA expression profiles in gastric cancer. *World J. Gastroenterol.* **2013**, *19*, 3658–3664, doi:10.3748/wjg.v19.i23.3658.
150. Du, Z.; Fei, T.; Verhaak, R.G.W.; Su, Z.; Zhang, Y.; Brown, M.; Chen, Y.; Liu, X.S. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat. Struct. Mol. Biol.* **2013**, *20*, 908–913, doi:10.1038/nsmb.2591.
151. Böttcher, R.; Hoogland, A.M.; Dits, N.; Verhoef, E.I.; Kweldam, C.; Waranecki, P.; Bangma, C.H.; van Leenders, G.J.L.H.; Jenster, G. Novel long non-coding RNAs are specific diagnostic and prognostic markers for prostate cancer. *Oncotarget* **2015**, *6*, 4036–4050, doi:10.18632/oncotarget.2879.

APPENDIX A : Rscript Aroma

```
library("aroma.affymetrix")
library(future)
plan(multiprocess)
verbose <- Arguments$getVerbose(-8, timestamp=TRUE)
print(verbose)
setwd("~/aroma")

#customCDF binary version downloaded from aroma-project.org
chipType<-"HuEx-1_0-st-v2"
cdf <- AffymetrixCdfFile$byChipType("HuEx-1_0-st-v2",tags = "mainR3,A20071112,EP")
print(cdf)
#AffymetrixCdfFile:
#Path: annotationData/chipTypes/HuEx-1_0-st-v2
#Filename: HuEx-1_0-st-v2,mainR3,A20071112,EP.cdf
#File size: 207.11 MiB (217167633 bytes)
#Chip type: HuEx-1_0-st-v2,mainR3,A20071112,EP
#RAM: 0.00MB
#File format: v4 (binary; XDA)
#Dimension: 2560x2560
#Number of cells: 6553600
#Number of units: 312355
#Cells per unit: 20.98
#Number of QC units: 1

cs <- AffymetrixCelSet$byName("Coupled", cdf=cdf)
print(cs)
#AffymetrixCelSet:
#Name: paired
#Tags:
#Path: rawData/Coupled/HuEx-1_0-st-v2
#Platform: Affymetrix
#Chip type: HuEx-1_0-st-v2,mainR3,A20071112,EP
#Number of arrays: 807
#Names: GSM1143384_PS41_N, GSM1143385_PS41_T, GSM1143386_PS43_N, ...,
GSM881210_ec_NM6_TS_Vi_nm_51 [807]
#Time period: [not reported if more than 500 arrays]
#Total file size: 73401.54MB
#RAM: 0.85MB

#comand used to associated cdf with dataset, could be used in each time of the analysis to
change cdf
#e.g. from "main cdf" to "core cdf"
setCdf(cs, cdf)

#Background Adjustment and Normalization

#RMA background correction
bc <- RmaBackgroundCorrection(cs,tags="*,mainR3")
```

```
csBC <- process(bc,verbose=verbose)

#quantile normalization
qn <- QuantileNormalization(csBC, typesToUpdate="pm")
print(qn)
#QuantileNormalization:
#  Data set: paired
#Input tags: RBC,mainR3
#User tags: *
#  Asterisk ('*') tags: QN
#Output tags: RBC,mainR3,QN
#Number of files: 807 (51190.37MB)
#Platform: Affymetrix
#Chip type: HuEx-1_0-st-v2,mainR3,A20071112,EP
#Algorithm parameters: {subsetToUpdate: NULL, typesToUpdate: chr "pm", subsetToAvg:
NULL, typesToAvg: chr "pm", .targetDistribution: NULL}
#Output path: probeData/Coupled,RBC,mainR3,QN/HuEx-1_0-st-v2
#Is done: FALSE
#RAM: 0.00MB

csN <- process(qn, verbose=verbose)
print(csN)
#AffymetrixCelSet:
#  Name: paired
#Tags: RBC,mainR3,QN
#Path: probeData/Coupled,RBC,mainR3,QN/HuEx-1_0-st-v2
#Platform: Affymetrix
#Chip type: HuEx-1_0-st-v2,mainR3,A20071112,EP
#Number of arrays: 807
#Names: GSM1143384_PS41_N, GSM1143385_PS41_T, GSM1143386_PS43_N, ...,
GSM881210_ec_NM6_TS_Vi_nm_51 [807]
#Time period: [not reported if more than 500 arrays]
#Total file size: 51190.37MB
#RAM: 0.85MB

#summarization : to fit exon-by-exon
plmEx <- ExonRmaPlm(csN, mergeGroups=FALSE)
print(plmEx)
#ExonRmaPlm:
#  Data set: paired
#Chip type: HuEx-1_0-st-v2,mainR3,A20071112,EP
#Input tags: RBC,mainR3,QN
#Output tags: RBC,mainR3,QN,RMA
#Parameters: {probeModel: chr "pm", shift: num 0, flavor: chr "affyPLM", treatNAsAs: chr
"weights", mergeGroups: logi FALSE}
#Path: plmData/Coupled,RBC,mainR3,QN,RMA/HuEx-1_0-st-v2
#RAM: 0.00MB

setOption(aromaSettings, "memory/ram", 10.0)
```

```
#extract normalized expression matrix
cesEx <- getChipEffectSet(plmEx)
CoupledMatrix <- extractDataFrame(cesEx, units=NULL, addNames=TRUE)
CoupledMatrix <- round(CoupledMatrix, digits = 2)
write.table(CoupledMatrix,file = "CoupledMatrix.txt", quote = FALSE, sep ="\t")

#summarization : to fit a summary of the entire transcript
plmTr <- ExonRmaPlm(csN,mergeGroups = TRUE)
print(plmTr)
#ExonRmaPlm:
#  Data set: prova
#Chip type: HuEx-1_0-st-v2,mainR3,A20071112,EP
#Input tags: RBC,mainR3,QN
#Output tags: RBC,mainR3,QN,RMA,merged
#Parameters: {probeModel: chr "pm", shift: num 0, flavor: chr "affyPLM", treatNAsAs: chr
"weights", mergeGroups: logi TRUE}
#Path: plmData/prova,RBC,mainR3,QN,RMA,merged/HuEx-1_0-st-v2
#RAM: 0.00MB

setOption(aromaSettings, "memory/ram", 1.0)
fit(plmTr,verbose = verbose)

#extract normalized expression of transcipt matrix
cesTr <- getChipEffectSet(plmTr)
trFit <- extractDataFrame(cesTr, units=NULL, addNames=TRUE)
```

APPENDIX B : Rscript pipeline

```
###perform IQR analysis###
source("https://bioconductor.org/biocLite.R")
biocLite("genefilter")
library(genefilter)
IQR05<- varFilter(Coupled, var.func=IQR, var.cutoff=0.5, filterByQuantile=FALSE)#581149
IQR05 <- round(IQR05, digits = 2)
write.table(IQR05,file = "CoupledIQR.txt", quote = FALSE, sep = "\t")

###subset of 5% of samples(20)must have at least value > 8###
IQR05 <- IQR05[rowSums(IQR05 > 8) >= 20, ]#164600

###perfomed limma differentiation analysis###
source("https://bioconductor.org/biocLite.R")
biocLite("limma")
library(limma)

#upload informations about the samples data
PhenoDataCoupled <- read.delim(PhenoDataCoupled)

#performed limma analysis for paired tumor/normal samples
paired <- factor(PhenoDataCoupled$Coupled)
status <- factor(PhenoDataCoupled$Status)
design <- model.matrix(~paired+status)
fit <- lmFit(IQR05A, design)
fit <- eBayes(fit)
LimmaCoupled<- topTable(fit, adjust.method="BH",number = 437403, coef = "statustumor",
sort.by = "logFC")
write.table(LimmaCoupled, file="LimmaCoupled.txt", quote = FALSE, sep = "\t")

##################################Bedtools#####################################
#work in linux terminal
bedtools intersect -a Limma.bed -b RefSeq.bed -wao -s > int IrefseqSS.txt
#-wao retain all the probesets with overlap bp number also zero
#- s same strand
#- S opposite strand
#proceding to cascade
bedtools intersect -a noIrefseqSS.bed -b RefSeq.bed -wao -S > int IrefseqOS.txt
bedtools intersect -a noIrefseq.bed -b ucsc.bed -wao -s > int IucscSS.txt
bedtools intersect -a noIucscSS.bed -b ucsc.bed -wao -S > int IucscOS.txt
bedtools intersect -a noIucsc.bed -b GencodeV24lift27.bed -wao -s > int IgencodeSS.txt
bedtools intersect -a noIgencode.bed -b GencodeV24lift27.bed -wao -S > int IgencodeOS.txt

##################################Python script################################

#############################################################################
        ######correlation analysis between intergenic probesets and flanked genes#####
#############################################################################
import csv
```

```
import sys
from collections import defaultdict

#fileOut1 = open('GeneDX.Corr.b.txt', 'wb')
#fileOut2 = open('GeneSX.Corr.b.txt', 'wb')
#fileOut3 = open('polyTranscript.DX.b.txt', 'w')
#fileOut4 = open('polyTranscript.SX.b.txt', 'w')
fileOut5 = open('complete19_5_17.txt', 'w')

#upload i file probesest and transcripts
#ifile = open('probe.19.csv', 'rb')
#fileb = open('gene.19.csv', 'rb')

ifile = open('intergenic.csv', 'rb')
fileb = open('transciptCoordRefSeq.csv', 'rb')

probe = csv.DictReader(ifile)
gene = csv.DictReader(fileb)

for row in probe:
        print row
        break

#create ditionary with probesets and flanked genes rigth(DX) and left (SX)
GeneRight=defaultdict(str)
GeneLeft=defaultdict(str)
GeneIndep=defaultdict(str)

numProb=0
numGene=0

for rowP in probe:
        fileb.seek(0)
        numGene=0
        numProb=numProb+1
        print numProb
        for rowG in gene:
                numGene=numGene+1
                #print numProb, numGene
                if      rowP['cromP'] == rowG['cromG'] and \
                int(rowP['startP']) >= int(rowG['startI']) and int(rowP['endP']) <= int(rowG['endI']) :
                        if rowP['strandP'] == rowG['strandDX']:
                                GeneRight[rowP['Probe']]=rowG['geneDX']
                        if rowP['strandP'] == rowG['strandSX']:
                                GeneLeft[rowP['Probe']]=rowG['geneSX']
                        if rowP['strandP'] != rowG['strandSX'] and rowP['strandP'] !=
rowG['strandDX']:
                                GeneIndep[rowP['Probe']]=True

import scipy
```

```
from scipy import stats
print "GeneLeft",len (GeneLeft)
#print GeneLeft
print "GeneRight", len (GeneRight)
#print GeneRight
print "GeneIndep",len (GeneIndep)
#print GeneIndep


#upload correlation matrix of flanked genes
geneDX=defaultdict(list)
with open("ExpressionTranscript.csv","rb") as csvf:
    contentDX = csv.reader(csvf, delimiter=",")
    for row in contentDX:
        for value in row[1:]:
            geneDX[row[0]].append(float(value))
csvf.close

print "geneDX", len (geneDX)
geneSX=defaultdict(list)
with open("ExpressionTranscript.csv","rb") as csvf:
    contentSX = csv.reader(csvf, delimiter=",")
    for row in contentSX:
        for value in row[1:]:
            geneSX[row[0]].append(float(value))
csvf.close
print "geneSX", len (geneSX)

iprobe=defaultdict(list)
with open("ExpressionProbe .csv","rb") as csvf:
    contentI = csv.reader(csvf, delimiter=",")
    for row in contentI:
        for value in row[1:]:
            iprobe[row[0]].append(float(value))
csvf.close
print "iprobe", len (iprobe)

#perfomed correlation analysis of probesets and specific flanked genes
#(rho, pvalue)
TotalProbes=set()
for probe in GeneRight:
        TotalProbes.add(probe)
for probe in GeneLeft:
        TotalProbes.add(probe)

print len (TotalProbes)
#print TotalProbes

#cluster near probesets
polyTranscriptR=defaultdict(list)
```

```
for probe in GeneRight:
	polyTranscriptR[GeneRight[probe]].append(probe)

polyTranscriptL=defaultdict(list)
for probe in GeneLeft:

	polyTranscriptL[GeneLeft[probe]].append(probe)

for probe in TotalProbes:
	if geneDX[GeneRight[probe]] is not None:
		try:
			(Rho, pValue)=scipy.stats.spearmanr (iprobe[probe],
geneDX[GeneRight[probe]])
			print >> fileOut5, "%s\t%s\t%3.2f\t%3.2g" %
(probe,GeneRight[probe],Rho,pValue),
		except:
			print >> fileOut5, "%s\t%s\t%s\t%s" % (probe,GeneRight[probe],'not
possible Correlation','na'),
	else:
		print fileOut5, "%s\t%s\t%s\t%s" % (probe,GeneRight[probe],
		'not concordant with geneDX','na'),
	if polyTranscriptR[GeneRight[probe]] is not None:
		print >> fileOut5, "\t%s" % (polyTranscriptR[GeneRight[probe]]),
	else:
		print >> fileOut5, "\t%s" % ('na'),
		#sys.exit(1)

	if geneSX[GeneLeft[probe]] is not None:
		try:
			(Rho, pValue)=scipy.stats.spearmanr (iprobe[probe],
geneSX[GeneLeft[probe]])
			print >> fileOut5, "\t%s\t%s\t%3.2f\t%3.2g" %
(probe,GeneLeft[probe],Rho,pValue),
		except:
			print >> fileOut5, "\t%s\t%s\t%s\t%s" % (probe, GeneLeft[probe], 'not
possible Correlation','na'),
	else:
		print  >> fileOut5, "\t%s\t%s\t%s\t%s" % (probe, GeneLeft[probe],
		'not concordant with geneSX','na'),

	if polyTranscriptL[GeneLeft[probe]] is not None:
		print >> fileOut5, "\t%s" % (polyTranscriptL[GeneLeft[probe]])
	else:
		print >> fileOut5, "\t%s" % ('na')

fileOut5.close()
##########################################################################
############Correlation analysis between intragenic probesets and hostgenes###########
##########################################################################
import csv
```

```
import scipy
from scipy import stats
import time
start_time = time.time()

#We have, right now, 2 datasets with the same
#array of probeset ids. We can map'em together to find where are the probes positions

#opening the matrix with the tissue expression and the probe sets related
#to the expression genes

matrix1 = open("/home/kalpof/Desktop/workbioinfo/intragenicprobes
/MatrixTrascript.txt","rb")
matrix2 = open("/home/kalpof/Desktop/workbioinfo/intragenicprobes
/IntragenicExpressionMatrix.csv","rb")
#opening the matrix with the tissue expression and the probe sets related
#to the trascript genes

file1 = open("/home/kalpof/Desktop/workbioinfo/intragenicprobes/Transcript.txt","rb")
#opening a gRange set, with start, stop, string, ecc.. from the transcript genes

file2 = open("/home/kalpof/Desktop/workbioinfo/intragenicprobes/TrueIntragenic.txt","rb")
#opening a gRange set, with start, stop, string, ecc.. from the intergenic sequences

#parser from file to list. Creates a list containing (usually) 5 values (but can differ
#between the type of input). Here the lists contains 5 values, the probeset ID, the chr
#the start, stop and the direction of the string (forward/reverse). This is not creating a dict
def parser(a_file, n):
  array= []
   read = csv.reader(a_file,delimiter=",")
   for row in read:
      array.append([element.split() for element in row])
      #split of the elements into an array

   new_list = []
   for e in array:
      new_list.append(e)
  return new_list

# a function which parses the CSV matrix into a dictionary with key
# equals to the probe set ID and the related value equals to a list
# containing the tissue sample and the expression value in double format
# the 4S at the end means for statistics, cause it's usefull for
# calculate Rho and P-value elements
def parserMatrixCSV4S(a_file):
   array=[]
    array2 = []
     read = csv.reader(a_file,delimiter=",")
  for row in read:
      for element in row:
```

```
#print element
 #element = element.split()
      try:
                  array2.append(float(element))
           except:
                  array2.append(element)

           array.append(array2)
           array2 = []

    c = {}
    for e in array:
           c[e[0]] = e[1:]
    return c

# a function which parses the txt matrix into a dictionary with key
# equals to the probe set ID and the related value equals to a list
# containing the tissue sample and the expression value in double format
# The 4S at the end means for statistics, cause it's usefull for
# calculate Rho and P-value elements
def parserMatrixTXT4S(a_file):
    array=[]
    array2 = []
    read = csv.reader(a_file,delimiter=",")
    for row in read:
           for element in row[0].split():
                  try:
                         array2.append(float(element))
                  except:
                         array2.append(element)

           array.append(array2)
           array2 = []

    c = {}
    for e in array:
           c[e[0]]= e[1:]
    return c

# a function used to find Related elements in 2 sequences between transcript
# products and intragenic products
def findRelates(list1, list2):
    #the left branch is the transcript, and the right one is the intragenic
    list3= []
    for e1 in list1:
           for e2 in list2:
                  if(e1[0][1] == e2[0][1] and
                  int(e2[0][2]) >= int(e1[0][2]) and int(e2[0][3]) <= int(e1[0][3])):
                  list3.append([float(e1[0][0]), float(e2[0][0]), e1[0][4], e2[0][4],
                         e1[0][2], e2[0][2], e1[0][3], e2[0][3], e1[0][1]])
```

```
                return list3

# the same function as above, but usefull to find the number of probes
# where the string sense is equal to the trascript one
def findRelatesSameString(list1, list2):
  #the left branch is the transcript, and the right one is the intragenic
   list3= []
   for e1 in list1:
        for e2 in list2:
            if(e1[0][1] == e2[0][1] and e1[0][4] == e2[0][4] and
            int(e2[0][2]) >= int(e1[0][2]) and int(e2[0][3]) <= int(e1[0][3])):
                list3.append([float(e1[0][0]), float(e2[0][0])])
        return list3

print("loading matrices...")
m1 = parserMatrixTXT4S(matrix1)
m2 = parserMatrixCSV4S(matrix2)
print("done!")

print("")

print("loading intragenic and trascription arrays...")
list1 = parser(file1,5) #list of the probesedt Ids from the trascript file
list2 = parser(file2,5) #list of the probes Ids from the Intragenic file
print("done!")

print("")

print("finding related sequences...")
list3 = findRelates(list1,list2)
print("done!")

print("")

i=0
for e in list3:
   i=i+1

print("On the total of " +str(len(list2)) + " probes, the total related probes are " + str(i))
last_file = open("/home/kalpof/Desktop/workbioinfo/result6.txt","w")
for e in list3:
   try:
     (Rho, pValue)= scipy.stats.spearmanr(m1[e[0]], m2[e[1]])
        last_file.write(str(e[0])+"\t"+str(e[1])+"\t"+str(e[8])+"\t"+str(e[5])+"\t"+str(e[7])+"\
t"+str(Rho)+"\t"+str(e[2])
        +"\t"+str(e[3])+"\t"+str(pValue)+"\n")
   except:
        last_file.write(str(e[0])+"\tMismatch\tNaN\tNaN\n")

print("--- %s seconds ---" % (time.time() - start_time))
```

APPENDIX C : Significant results

Intronic Sense Negative RNAs Correlated and Not Correlated

Classification_Pipeline	Classification_GencodeV29	hg38Coordinates	strand	ConservationAnalysis	hosting Gene Symbol	Probe_ID	length BP	level	logFC_Limma	adj.P.Val_Limma	coding_prob_CPAT	CGC_pos	CGC_pos_corr	CGC_pos_p	CGC_neg	CGC_neg_corr	CGC_neg_p
InGen_SS_641	intronicSense_negativeCorr	chr17:35597075-35597099	+	Cons7_no_Cons100_no	AP2B1	3718691	24	extended	-0.1	0.0	0.000	DCAF12L2	0.4	0.0	MAP2K4	-0.5	0.0
InGen_SS_Miss_168	intronicSense_negativeCorr	chr2:230725357-230725428	+	Cons7_yes_Cons100_yes	CAB39	2531531	71	extended	0.3	0.0	0.013	ZNRF3	0.5	0.0	ARHGAP5	-0.5	0.0
InGen_SS_663	intronicSense_negativeCorr	chr17:1441779-1441867	-	Cons7_no_Cons100_no	CRK	3740194	88	extended	-0.1	0.0	0.000	ZNRF3	0.5	0.0	STAG1	-0.5	0.0
InGen_SS_Miss_322	intronicSense_negativeCorr	chr6:5123875-5123906	-	Cons7_no_Cons100_no	LYRM4	2939902	31	full	-0.1	0.0	0.000	ZNRF3	0.7	0.0	ARHGAP5	-0.4	0.0
OutGen_Disc_298	intronicSense_negativeCorr	chr1:19485975-19486004	+	Cons7_no_Cons100_yes	MINOS1	2323728	29	free	-0.1	0.0	0.000	GLI1	0.4	0.0	TET2	-0.5	0.0
InGen_SS_224	intronicSense_negativeCorr	chr5:134225562-134225651	-	Cons7_yes_Cons100_yes	PPP2CA	2876073	89	full	-0.1	0.0	0.000	POLE	0.4	0.0	HOOK3	-0.4	0.0
InGen_SS_770	intronicSense_negativeCorr	chr21:44864472-44864576	-	Cons7_no_Cons100_no	PTTG1IP	3934709	104	extended	-0.1	0.0	0.000	KMT2D	0.4	0.0	RB1	-0.5	0.0
InGen_SS_218	intronicSense_negativeCorr	chr5:112900918-112900976	-	Cons7_no_Cons100_no	REEP5	2871218	58	extended	-0.1	0.0	0.083	ZNRF3	0.6	0.0	STAG1	-0.4	0.0
InGen_OS_1746	intronicSense_negativeCorr	chr8:73295042-73295084	-	Cons7_yes_Cons100_yes	RPL7	3140592	42	free	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_294	intronicSense_negativeCorr	chr6:11124989-11125174	+	Cons7_no_Cons100_no	SMIM13	2894892	185	extended	-0.1	0.0	0.431	ZNRF3	0.9	0.0	CDKN1B	-0.3	0.0
InGen_SS_91	intronicSense_notCorr	chr2:69487914-69487953	-	Cons7_no_Cons100_no	AAK1	2558177	39	extended	-0.1	0.0	0.073	NA	NA	NA	NA	NA	NA
InGen_SS_752	intronicSense_notCorr	chr20:25390289-25390321	-	Cons7_no_Cons100_no	ABHD12	3901927	32	full	-0.1	0.0	0.000	NUTM1	0.6	0.0	HOOK3	-0.4	0.0
InGen_SS_390	intronicSense_notCorr	chr9:130755777-130755807	+	Cons7_no_Cons100_no	ABL1	3191737	30	extended	-0.1	0.0	0.000	ZNRF3	0.5	0.0	ARHGAP5	-0.6	0.0
InGen_OS_1081	intronicSense_notCorr	chr5:172768563-172769038	+	Cons7_yes_Cons100_yes	AC022217.3	2841157	475	extended	-0.2	0.0	0.810	TP63	0.3	0.0	ABI1	-0.3	0.0
InGen_OS_1082	intronicSense_notCorr	chr5:172770752-172771086	+	Cons7_yes_Cons100_yes	AC022217.3	2841167	334	extended	-0.3	0.0	0.641	NA	NA	NA	NA	NA	NA
InGen_OS_1041	intronicSense_notCorr	chr5:132494637-132494662	+	Cons7_no_Cons100_no	AC116366.3	2828600	25	full	-0.1	0.0	0.000	TP63	0.3	0.0	ABI1	-0.3	0.0
InGen_SS_Miss_559	intronicSense_notCorr	chr12:93083329-93083356	-	Cons7_yes_Cons100_no	AC138123.1	3465758	27	extended	0.2	0.0	0.000	IL21R	0.4	0.0	NA	NA	NA
InGen_SS_709	intronicSense_notCorr	chr19:38727104-38727135	+	Cons7_yes_Cons100_yes	ACTN4	3832712	31	extended	-0.2	0.0	0.000	KMT2D	0.4	0.0	CRNKL1	-0.5	0.0
InGen_SS_62	intronicSense_notCorr	chr2:65248097-65248126	+	Cons7_no_Cons100_no	ACTR2	2485799	29	extended	-0.1	0.0	0.000	ZNRF3	0.9	0.0	NCOA2	-0.3	0.0
InGen_SS_Miss_290	intronicSense_notCorr	chr5:179208493-179208524	-	Cons7_no_Cons100_no	ADAMTS2	2889981	31	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_223	intronicSense_notCorr	chr5:132963092-132963181	-	Cons7_no_Cons100_no	AFF4	2875613	89	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	STAG1	-0.5	0.0

InGen_SS_61	intronicSense_notCorr	chr2:64565383-64565448	+	Cons7_no_Cons100_no	AFTPH	2485502	65	extended	-0.1	0.0	0.000	ZNRF3	0.9	0.0	NA	NA	NA
InGen_SS_59	intronicSense_notCorr	chr2:27067025-27067052	+	Cons7_no_Cons100_no	AGBL5	2474216	27	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1218	intronicSense_notCorr	chr6:11250001-11250027	+	Cons7_no_Cons100_no	AL139807.1	2894943	26	full	-0.2	0.0	0.000	NA	NA	NA	TMEM127	-0.4	0.0
InGen_SS_548	intronicSense_notCorr	chr13:100539854-100539887	-	Cons7_no_Cons100_yes	AL356966.1	3523144	33	extended	0.3	0.0	0.000	NA	NA	NA	CREB3L2	-0.3	0.0
OutGen_Ind_156	intronicSense_notCorr	chr1:192246253-192246568	-	Cons7_yes_Cons100_yes	AL390957.1	2448887	315	extended	0.4	0.0	0.002	CNBD1	0.7	0.0	PRDM2	-0.4	0.0
InGen_SS_Miss_348	intronicSense_notCorr	chr7:92346834-92346858	+	Cons7_no_Cons100_no	ANKIB1	3012572	24	extended	0.2	0.0	0.000	ZNRF3	0.5	0.0	GNA11	-0.5	0.0
InGen_SS_184	intronicSense_notCorr	chr4:73099049-73099079	-	Cons7_yes_Cons100_no	ANKRD17	2773055	30	extended	0.7	0.0	0.000	ZNRF3	0.5	0.0	MAP2K4	-0.5	0.0
InGen_SS_460	intronicSense_notCorr	chr11:83209316-83209398	+	Cons7_no_Cons100_yes	ANKRD42	3342581	82	full	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_538	intronicSense_notCorr	chr12:45216422-45216545	+	Cons7_no_Cons100_no	ANO6	3412575	123	full	-0.1	0.0	0.000	SETD1B	0.3	0.0	MLLT11	-0.3	0.0
InGen_SS_Miss_515	intronicSense_notCorr	chr11:112284783-112284815	+	Cons7_no_Cons100_no	AP002884.1	3349067	32	full	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_713	intronicSense_notCorr	chr19:49771259-49771291	+	Cons7_no_Cons100_no	AP2A1	3838936	32	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_447	intronicSense_notCorr	chr11:972888-972950	+	Cons7_no_Cons100_no	AP2A2	3316482	62	full	-0.1	0.0	0.000	SLC34A2	0.6	0.0	ARHGAP5	-0.3	0.0
InGen_SS_711	intronicSense_notCorr	chr19:41903936-41904010	+	Cons7_no_Cons100_no	ARHGEF1	3834558	74	full	-0.1	0.0	0.000	DCAF12L2	0.6	0.0	ARHGAP5	-0.3	0.0
InGen_SS_Miss_736	intronicSense_notCorr	chr17:46288325-46288436	-	Cons7_no_Cons100_no	ARL17B	3760311	111	extended	-0.1	0.0	0.027	GLI1	0.3	0.0	TET2	-0.3	0.0
InGen_SS_49	intronicSense_notCorr	chr1:155425854-155425880	-	Cons7_no_Cons100_no	ASH1L	2437460	26	extended	-0.1	0.0	0.006	NA	NA	NA	NA	NA	NA
InGen_SS_83	intronicSense_notCorr	chr2:25792677-25792720	-	Cons7_no_Cons100_no	ASXL2	2544993	43	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	STAG1	-0.4	0.0
InGen_SS_757	intronicSense_notCorr	chr20:59032083-59032109	-	Cons7_no_Cons100_yes	ATP5F1E	3911808	26	extended	-0.1	0.0	0.000	DGCR8	0.5	0.0	SMAD4	-0.5	0.0
InGen_OS_666	intronicSense_notCorr	chr3:183278921-183278949	+	Cons7_no_Cons100_no	B3GNT5	2654985	28	full	-0.1	0.0	0.076	NA	NA	NA	NA	NA	NA
InGen_SS_756	intronicSense_notCorr	chr20:49677321-49677347	-	Cons7_no_Cons100_no	B4GALT5	3908981	26	full	-0.1	0.0	0.001	ZNRF3	0.6	0.0	CRNKL1	-0.4	0.0
InGen_SS_424	intronicSense_notCorr	chr10:119667789-119667820	+	Cons7_no_Cons100_no	BAG3	3267364	31	full	0.1	0.0	0.000	DCAF12L2	0.6	0.0	ARHGAP5	-0.4	0.0
InGen_SS_658	intronicSense_notCorr	chr17:81452129-81452155	+	Cons7_no_Cons100_no	BAHCC1	3737896	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_44	intronicSense_notCorr	chr1:93764166-93764195	-	Cons7_no_Cons100_no	BCAR3	2423576	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_648	intronicSense_notCorr	chr17:60701878-60701907	+	Cons7_no_Cons100_no	BCAS3	3729764	29	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_804	intronicSense_notCorr	chrX:15551862-15551923	+	Cons7_yes_Cons100_yes	BMX	3969839	61	extended	0.2	0.0	0.037	CCR7	0.7	0.0	NA	NA	NA
InGen_SS_331	intronicSense_notCorr	chr8:26408716-26408810	+	Cons7_no_Cons100_no	BNIP3L	3091032	94	extended	-0.1	0.0	0.041	ZNRF3	0.6	0.0	RB1	-0.3	0.0
InGen_SS_Miss_298	intronicSense_notCorr	chr6:36225365-36225397	+	Cons7_yes_Cons100_yes	BRPF3	2905011	32	full	-0.1	0.0	0.000	DCAF12L2	0.6	0.0	NA	NA	NA
InGen_SS_573	intronicSense_notCorr	chr14:93325195-93325267	-	Cons7_no_Cons100_no	BTBD7	3577336	72	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	STAG1	-0.5	0.0
InGen_SS_	intronicSense_notCorr	chr21:32866333-	+	Cons7_no_Cons100_	C21orf62-	3918314	29	full	-0.5	0.0	0.000	NA	NA	NA	NA	NA	NA

Miss_819		32866362		no	AS1												
InGen_SS_356	intronicSense_notCorr	chr8:85220183-85220218	-	Cons7_no_Cons100_no	C8orf59	3142951	35	extended	0.1	0.0	0.000	ZNRF3	0.5	0.0	FOXO3	-0.5	0.0
InGen_SS_443	intronicSense_notCorr	chr10:73851389-73851435	-	Cons7_no_Cons100_no	CAMK2G	3294924	46	full	-0.3	0.0	0.000	DGCR8	0.5	0.0	STAG1	-0.5	0.0
InGen_SS_35	intronicSense_notCorr	chr1:19343331-19343364	-	Cons7_no_Cons100_no	CAPZB	2399782	33	extended	-0.1	0.0	0.000	KMT2D	0.3	0.0	PRKAR1A	-0.3	0.0
InGen_SS_455	intronicSense_notCorr	chr11:64353888-64353915	+	Cons7_no_Cons100_yes	CCDC88B	3334558	27	full	0.1	0.0	0.424	NA	NA	NA	NA	NA	NA
InGen_SS_520	intronicSense_notCorr	chr12:55728214-55728250	-	Cons7_no_Cons100_no	CD63	3457178	36	full	-0.1	0.0	0.000	DGCR8	0.4	0.0	ARHGAP5	-0.5	0.0
InGen_SS_589	intronicSense_notCorr	chr15:42726450-42726479	-	Cons7_no_Cons100_yes	CDAN1	3620753	29	extended	-0.3	0.0	0.000	DGCR8	0.5	0.0	ARHGAP5	-0.5	0.0
InGen_SS_540	intronicSense_notCorr	chr13:114251945-114252183	+	Cons7_no_Cons100_yes	CDC16	3503192	238	full	0.2	0.0	0.000	ZNRF3	0.6	0.0	STAG1	-0.5	0.0
InGen_SS_737	intronicSense_notCorr	chr20:5163836-5164100	+	Cons7_no_Cons100_no	CDS2	3874906	264	extended	-0.1	0.0	0.012	ZNRF3	0.6	0.0	GOLGA5	-0.3	0.0
InGen_SS_633	intronicSense_notCorr	chr16:80633351-80633378	-	Cons7_no_Cons100_no	CDYL2	3701312	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_129	intronicSense_notCorr	chr3:134489771-134489807	+	Cons7_no_Cons100_yes	CEP63	2643539	36	extended	-0.1	0.0	0.002	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_655	intronicSense_notCorr	chr16:11060818-11060853	+	Cons7_no_Cons100_no	CLEC16A	3648140	35	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_413	intronicSense_notCorr	chr10:35149767-35149793	+	Cons7_no_Cons100_no	CREM	3242364	26	full	-0.1	0.0	0.000	ZNRF3	0.9	0.0	STAG1	-0.3	0.0
InGen_OS_2972	intronicSense_notCorr	chr17:7252136-7252162	-	Cons7_no_Cons100_yes	CTDNEP1	3743521	26	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_349	intronicSense_notCorr	chr7:102282668-102282701	+	Cons7_no_Cons100_yes	CUX1	3016519	33	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_609	intronicSense_notCorr	chr15:22914691-22914723	-	Cons7_yes_Cons100_yes	CYFIP1	3583684	32	extended	-0.1	0.0	0.000	DCAF12L2	0.5	0.0	ARHGAP5	-0.5	0.0
InGen_SS_697	intronicSense_notCorr	chr19:1422719-1422979	+	Cons7_no_Cons100_no	DAZAP1	3815885	260	extended	0.2	0.0	0.000	FUS	0.5	0.0	STAG1	-0.6	0.0
InGen_SS_403	intronicSense_notCorr	chr9:123483363-123483400	-	Cons7_no_Cons100_no	DENND1A	3224718	37	full	-0.1	0.0	0.000	DCAF12L2	0.6	0.0	RB1	-0.3	0.0
InGen_SS_77	intronicSense_notCorr	chr2:233403407-233403486	+	Cons7_no_Cons100_no	DGKD	2532988	79	extended	-0.1	0.0	0.005	NA	NA	NA	NA	NA	NA
InGen_SS_215	intronicSense_notCorr	chr5:80651074-80651098	-	Cons7_yes_Cons100_yes	DHFR	2864632	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_437	intronicSense_notCorr	chr10:21872135-21872209	-	Cons7_no_Cons100_no	DNAJC1	3280932	74	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
InGen_SS_640	intronicSense_notCorr	chr17:17977760-17977789	+	Cons7_yes_Cons100_yes	DRC3	3712855	29	full	-0.2	0.0	0.001	NA	NA	NA	NA	NA	NA
InGen_SS_220	intronicSense_notCorr	chr5:118973885-118974133	-	Cons7_no_Cons100_yes	DTWD2	2872494	248	extended	0.2	0.0	0.000	DCAF12L2	0.6	0.0	RB1	-0.3	0.0
InGen_SS_744	intronicSense_notCorr	chr20:34535712-34535742	+	Cons7_no_Cons100_no	DYNLRB1	3882979	30	full	-0.1	0.0	0.000	ZNRF3	0.3	0.0	GNAQ	-0.3	0.0
InGen_OS_1255	intronicSense_notCorr	chr6:52500557-52500586	+	Cons7_no_Cons100_no	EFHC1	2910283	29	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_94	intronicSense_notCorr	chr2:88626731-88626841	-	Cons7_no_Cons100_no	EIF2AK3	2563701	110	full	-0.1	0.0	0.000	MDM4	0.3	0.0	ITGAV	-0.3	0.0
InGen_SS_80	intronicSense_notCorr	chr2:3283249-3283292	-	Cons7_no_Cons100_no	EIPR1	2538543	43	full	-0.1	0.0	0.000	CCR7	0.3	0.0	NA	NA	NA
InGen_SS_67	intronicSense_notCorr	chr2:120015177-120015207	+	Cons7_no_Cons100_no	EPB41L5	2503191	30	extended	-0.1	0.0	0.000	ELF3	0.3	0.0	MLLT6	-0.4	0.0

InGen_SS_204	intronicSense_notCorr	chr5:172834922-172834949	+	Cons7_no_Cons100_yes	ERGIC1	2841186	27	extended	-0.1	0.0	0.000	NA	NA	NA	ARID2	-0.3	0.0
InGen_SS_Miss_319	intronicSense_notCorr	chr6:170307084-170307133	+	Cons7_no_Cons100_no	FAM120B	2937867	49	full	-0.1	0.0	0.000	DCAF12L2	0.4	0.0	RB1	-0.5	0.0
InGen_SS_Miss_251	intronicSense_notCorr	chr4:186160609-186160637	+	Cons7_no_Cons100_no	FAM149A	2755015	28	extended	-0.2	0.0	0.000	NA	NA	NA	PTEN	-0.3	0.0
InGen_SS_Miss_586	intronicSense_notCorr	chr14:55305510-55305586	+	Cons7_no_Cons100_no	FBXO34	3536796	76	extended	-0.1	0.0	0.000	ZNRF3	0.5	0.0	CRNKL1	-0.3	0.0
InGen_SS_Miss_588	intronicSense_notCorr	chr14:55299354-55299378	+	Cons7_no_Cons100_yes	FBXO34	3536838	24	extended	0.2	0.0	0.106	NA	NA	NA	NA	NA	NA
InGen_SS_462	intronicSense_notCorr	chr11:110461344-110461431	+	Cons7_no_Cons100_no	FDX1	3348201	87	extended	-0.1	0.0	0.003	NA	NA	NA	PPM1D	-0.3	0.0
InGen_SS_Miss_336	intronicSense_notCorr	chr7:4683679-4683703	+	Cons7_no_Cons100_no	FOXK1	2988249	24	full	-0.1	0.0	0.474	DCAF12L2	0.4	0.0	STAG1	-0.5	0.0
InGen_SS_Miss_337	intronicSense_notCorr	chr7:4761316-4761340	+	Cons7_no_Cons100_no	FOXK1	2988295	24	full	-0.3	0.0	0.000	DCAF12L2	0.5	0.0	ARHGAP5	-0.4	0.0
InGen_SS_Miss_14	intronicSense_notCorr	chr1:41366306-41366405	+	Cons7_yes_Cons100_yes	FOXO6	2332255	99	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_253	intronicSense_notCorr	chr4:48779319-48779358	-	Cons7_no_Cons100_no	FRYL	2768618	39	full	-0.1	0.0	0.000	KMT2D	0.3	0.0	SDHD	-0.3	0.0
InGen_SS_Miss_790	intronicSense_notCorr	chr19:19429457-19429489	+	Cons7_no_Cons100_no	GATAD2A	3825794	32	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
InGen_SS_255	intronicSense_notCorr	chr6:13397402-13397434	-	Cons7_no_Cons100_no	GFOD1	2942388	32	full	-0.1	0.0	0.001	NA	NA	NA	NA	NA	NA
InGen_SS_89	intronicSense_notCorr	chr2:69356063-69356159	-	Cons7_no_Cons100_no	GFPT1	2558090	96	extended	-0.1	0.0	0.009	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
InGen_SS_771	intronicSense_notCorr	chr22:23088424-23088454	+	Cons7_no_Cons100_yes	GNAZ	3939130	30	full	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_158	intronicSense_notCorr	chr3:179451239-179451273	-	Cons7_no_Cons100_no	GNB4	2706960	34	full	-0.1	0.0	0.000	KMT2D	0.3	0.0	NA	NA	NA
InGen_SS_397	intronicSense_notCorr	chr9:36274187-36274222	-	Cons7_no_Cons100_no	GNE	3205145	35	extended	-0.1	0.0	0.000	NA	NA	NA	ARID2	-0.4	0.0
InGen_SS_Miss_421	intronicSense_notCorr	chr9:130054243-130054277	+	Cons7_no_Cons100_no	GPR107	3191276	34	full	-0.2	0.0	0.000	DCAF12L2	0.5	0.0	ARHGAP5	-0.4	0.0
InGen_SS_Miss_422	intronicSense_notCorr	chr9:130119040-130119068	+	Cons7_no_Cons100_no	GPR107	3191311	28	extended	-0.1	0.0	0.000	DGCR8	0.5	0.0	ARHGAP5	-0.5	0.0
InGen_SS_Miss_357	intronicSense_notCorr	chr7:6529079-6529103	-	Cons7_no_Cons100_no	GRID2IP	3037478	24	extended	-0.2	0.0	0.001	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_446	intronicSense_notCorr	chr16:85248140-85248171	+	Cons7_no_Cons100_no	GSE1	3672196	31	full	-0.1	0.0	0.000	SLC34A2	0.5	0.0	ARHGAP5	-0.4	0.0
InGen_SS_Miss_562	intronicSense_notCorr	chr12:112202278-112202416	-	Cons7_no_Cons100_no	HECTD4	3471938	138	extended	-0.1	0.0	0.005	ZNRF3	0.5	0.0	NSD1	-0.4	0.0
InGen_SS_Miss_563	intronicSense_notCorr	chr12:112296460-112296648	-	Cons7_no_Cons100_no	HECTD4	3472047	188	full	-0.1	0.0	0.056	ZNRF3	0.7	0.0	CDKN1B	-0.5	0.0
InGen_SS_Miss_77	intronicSense_notCorr	chr1:91387435-91387505	-	Cons7_no_Cons100_no	HFM1	2422684	70	extended	-0.1	0.0	0.000	ZNRF3	0.5	0.0	ARHGAP5	-0.5	0.0
InGen_SS_234	intronicSense_notCorr	chr6:26164725-26164749	+	Cons7_no_Cons100_yes	HIST1H2BD	2899187	24	full	0.3	0.0	0.007	ZNRF3	1.0	0.0	CDKN1B	-0.3	0.0
InGen_OS_2514	intronicSense_notCorr	chr14:64539555-64539628	+	Cons7_no_Cons100_no	HSPA2	3540154	73	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_430	intronicSense_notCorr	chr10:132745838-132745863	+	Cons7_no_Cons100_no	INPP5A	3272304	25	full	-0.1	0.0	0.761	NA	NA	NA	PTEN	-0.3	0.0
InGen_SS_431	intronicSense_notCorr	chr10:132757824-132757923	+	Cons7_no_Cons100_no	INPP5A	3272325	99	full	0.1	0.0	0.000	NA	NA	NA	PALB2	-0.3	0.0
InGen_SS_432	intronicSense_notCorr	chr10:13264581	+	Cons7_no_Cons100_	INPP5A	3272240	34	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA

		6-132645850		no													
lnGen_SS_482	intronicSense_notCorr	chr11:77886417-77886462	-	Cons7_no_Cons100_no	INTS4	3383086	45	full	-0.2	0.0	0.005	NA	NA	NA	NA	NA	NA
lnGen_SS_Miss_98	intronicSense_notCorr	chr1:201740301-201740336	-	Cons7_no_Cons100_yes	IPO9-AS1	2450978	35	full	-0.2	0.0	0.001	NA	NA	NA	NA	NA	NA
lnGen_SS_Miss_474	intronicSense_notCorr	chr10:132137305-132137329	+	Cons7_no_Cons100_no	JAKMIP3	3271850	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_SS_181	intronicSense_notCorr	chr4:20771780-20771839	-	Cons7_no_Cons100_no	KCNIP4	2763061	59	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_SS_164	intronicSense_notCorr	chr4:6849695-6849722	+	Cons7_no_Cons100_no	KIAA0232	2717148	27	extended	-0.1	0.0	0.000	ZNRF3	0.8	0.0	CDKN1B	-0.4	0.0
lnGen_SS_51	intronicSense_notCorr	chr1:202903559-202903614	-	Cons7_no_Cons100_no	KLHL12	2451439	55	extended	-0.1	0.0	0.006	NA	NA	NA	NA	NA	NA
lnGen_SS_201	intronicSense_notCorr	chr5:154740199-154740355	+	Cons7_no_Cons100_no	LARP1	2836841	156	extended	-0.1	0.0	0.005	NA	NA	NA	NA	NA	NA
lnGen_SS_Nan_1	intronicSense_notCorr	chr3:156929802-156929980	+	Cons7_yes_Cons100_yes	LEKR1	2649217	178	extended	0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_SS_Nan_2	intronicSense_notCorr	chr3:156930062-156930089	+	Cons7_yes_Cons100_yes	LEKR1	2649218	27	extended	0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_SS_777	intronicSense_notCorr	chr22:31237953-31237995	+	Cons7_no_Cons100_no	LIMK2	3942903	42	extended	-0.1	0.0	0.005	NA	NA	NA	NA	NA	NA
lnGen_SS_Miss_802	intronicSense_notCorr	chr19:27747068-27747092	-	Cons7_no_Cons100_no	LINC00662	3857363	24	full	0.2	0.0	0.000	ZNRF3	0.7	0.0	GNA11	-0.4	0.0
lnGen_SS_Miss_494	intronicSense_notCorr	chr10:132965147-132965177	-	Cons7_no_Cons100_no	LINC01166	3314813	30	full	-0.2	0.0	0.000	TP63	0.3	0.0	ABI1	-0.3	0.0
lnGen_SS_Miss_450	intronicSense_notCorr	chr9:103204150-103204815	-	Cons7_no_Cons100_no	LINC01492	3218337	665	extended	0.3	0.0	0.015	CCR7	0.5	0.0	KMT2C	-0.3	0.0
lnGen_SS_Miss_202	intronicSense_notCorr	chr3:107431696-107431746	+	Cons7_no_Cons100_no	LINC01990	2634942	50	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_SS_Miss_592	intronicSense_notCorr	chr14:77064615-77064646	+	Cons7_no_Cons100_no	LINC02288	3545297	31	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_SS_295	intronicSense_notCorr	chr7:98204192-98204229	+	Cons7_yes_Cons100_yes	LMTK2	3014212	37	full	-0.1	0.0	0.000	ZNRF3	1.0	0.0	CDKN1B	-0.3	0.0
lnGen_SS_137	intronicSense_notCorr	chr3:37108270-37108303	-	Cons7_no_Cons100_no	LRRFIP2	2669239	33	full	-0.1	0.0	0.001	ZNRF3	0.5	0.0	CNBP	-0.5	0.0
lnGen_SS_653	intronicSense_notCorr	chr17:63687790-63687912	+	Cons7_no_Cons100_no	MAP3K3	3730850	122	extended	-0.1	0.0	0.037	ZNRF3	0.7	0.0	STAG1	-0.4	0.0
lnGen_SS_244	intronicSense_notCorr	chr6:161079071-161079209	+	Cons7_no_Cons100_no	MAP3K4	2934827	138	extended	-0.1	0.0	0.004	ZNRF3	0.6	0.0	CDKN1B	-0.5	0.0
lnGen_SS_685	intronicSense_notCorr	chr18:35007042-35007070	+	Cons7_no_Cons100_no	MAPRE2	3784424	28	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_SS_195	intronicSense_notCorr	chr5:66741795-66741825	+	Cons7_no_Cons100_no	MAST4	2812715	30	full	-0.1	0.0	0.001	NA	NA	NA	NA	NA	NA
lnGen_SS_634	intronicSense_notCorr	chr16:84066456-84066481	-	Cons7_no_Cons100_no	MBTPS1	3702319	25	extended	0.1	0.0	0.000	ZNRF3	0.4	0.0	RHOA	-0.4	0.0
lnGen_OS_506	intronicSense_notCorr	chr2:111117870-111117983	-	Cons7_no_Cons100_no	MIR4435-2HG	2570813	113	full	-0.1	0.0	0.000	NA	NA	NA	ABI1	-0.3	0.0
lnGen_OS_507	intronicSense_notCorr	chr2:111121550-111121579	-	Cons7_no_Cons100_no	MIR4435-2HG	2570827	29	ambiguous	-0.2	0.0	0.000	TCL1A	0.5	0.0	NA	NA	NA
lnGen_SS_612	intronicSense_notCorr	chr16:22009367-22009397	+	Cons7_yes_Cons100_no	MOSMO	3652295	30	full	-0.1	0.0	0.000	MPL	0.4	0.0	SMAD4	-0.3	0.0
lnGen_SS_21	intronicSense_notCorr	chr1:167733454-167733484	+	Cons7_no_Cons100_no	MPZL1	2365963	30	extended	-0.1	0.0	0.001	ZNRF3	0.7	0.0	STAG1	-0.5	0.0
lnGen_OS_652	intronicSense_notCorr	chr3:139044402-139044426	+	Cons7_yes_Cons100_yes	MRPS22	2644825	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA

InGen_SS_340	intronicSense_notCorr	chr8:97676708-97676789	+	Cons7_no_Cons100_no	MTDH	3108451	81	extended	-0.2	0.0	0.011	ZNRF3	0.6	0.0	PIK3R1	-0.4	0.0
InGen_SS_Miss_321	intronicSense_notCorr	chr6:2690187-2690220	-	Cons7_no_Cons100_no	MYLK4	2938871	33	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_606	intronicSense_notCorr	chr16:3471752-3471777	+	Cons7_no_Cons100_yes	NAA60	3645920	25	full	-0.1	0.0	0.000	DCAF12L2	0.5	0.0	ARHGAP5	-0.4	0.0
InGen_SS_Miss_896	intronicSense_notCorr	chr1:1759477-1759502	-	Cons7_no_Cons100_no	NADK	4044727	25	full	-0.1	0.0	0.000	DGCR8	0.5	0.0	ARHGAP5	-0.6	0.0
InGen_SS_550	intronicSense_notCorr	chr13:101304578-101304751	-	Cons7_no_Cons100_no	NALCN	3523426	173	extended	-0.1	0.0	0.006	NA	NA	NA	PTEN	-0.3	0.0
InGen_SS_Miss_457	intronicSense_notCorr	chr10:46030474-46030524	-	Cons7_no_Cons100_yes	NCOA4	3246380	50	full	-0.1	0.0	0.233	DCAF12L2	0.5	0.0	STAG1	-0.4	0.0
InGen_SS_611	intronicSense_notCorr	chr16:15650180-15650208	+	Cons7_no_Cons100_yes	NDE1	3649814	28	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_108	intronicSense_notCorr	chr2:239965551-239965609	-	Cons7_no_Cons100_no	NDUFA10	2606594	58	full	-0.1	0.0	0.789	FUS	0.4	0.0	CRNKL1	-0.4	0.0
InGen_SS_Miss_437	intronicSense_notCorr	chr9:14349626-14349651	-	Cons7_no_Cons100_no	NFIB	3199361	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_806	intronicSense_notCorr	chrX:17682675-17682714	+	Cons7_no_Cons100_no	NHS	3970401	39	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_81	intronicSense_notCorr	chr2:10676989-10677100	-	Cons7_no_Cons100_no	NOL10	2540279	111	extended	-0.1	0.0	0.008	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
InGen_SS_Miss_698	intronicSense_notCorr	chr16:21425853-21425956	-	Cons7_no_Cons100_no	NPIPB3	3684179	103	extended	0.2	0.0	0.001	C15orf65	0.4	0.0	GNA11	-0.3	0.0
InGen_SS_Miss_700	intronicSense_notCorr	chr16:21426224-21426311	-	Cons7_no_Cons100_no	NPIPB3	3684182	87	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_701	intronicSense_notCorr	chr16:21426778-21427156	-	Cons7_no_Cons100_no	NPIPB3	3684183	378	extended	0.3	0.0	0.006	ZNRF3	0.7	0.0	CDKN1B	-0.4	0.0
InGen_SS_Miss_702	intronicSense_notCorr	chr16:21427530-21428076	-	Cons7_no_Cons100_no	NPIPB3	3684184	546	extended	0.2	0.0	0.039	TP63	0.3	0.0	ABI1	-0.3	0.0
InGen_SS_Miss_703	intronicSense_notCorr	chr16:21436517-21436760	-	Cons7_no_Cons100_no	NPIPB3	3684192	243	extended	0.2	0.0	0.002	ZNRF3	0.9	0.0	CDKN1B	-0.3	0.0
InGen_SS_Miss_598	intronicSense_notCorr	chr14:105323836-105323917	+	Cons7_no_Cons100_no	PACS2	3554655	81	ambiguous	-0.1	0.0	0.000	RGPD3	0.6	0.0	PTEN	-0.3	0.0
InGen_SS_794	intronicSense_notCorr	chr22:42900089-42900139	-	Cons7_no_Cons100_no	PACSIN2	3962696	50	full	-0.1	0.0	0.000	ZNRF3	0.4	0.0	CRNKL1	-0.4	0.0
InGen_SS_Miss_12	intronicSense_notCorr	chr1:28369954-28370071	+	Cons7_yes_Cons100_yes	PHACTR4	2327442	117	extended	-0.1	0.0	0.000	DCAF12L2	0.5	0.0	STAG1	-0.5	0.0
InGen_SS_674	intronicSense_notCorr	chr17:38799050-38799095	-	Cons7_yes_Cons100_yes	PIP4K2B	3755393	45	free	-0.1	0.0	0.000	DGCR8	0.5	0.0	ARHGAP5	-0.5	0.0
InGen_SS_341	intronicSense_notCorr	chr8:109471879-109471939	+	Cons7_yes_Cons100_yes	PKHD1L1	3111634	60	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_395	intronicSense_notCorr	chr9:5419120-5419147	-	Cons7_no_Cons100_no	PLGRKT	3197572	27	full	-0.2	0.0	0.000	C15orf65	0.4	0.0	HIP1	-0.3	0.0
InGen_SS_786	intronicSense_notCorr	chr22:50411338-50411367	+	Cons7_no_Cons100_no	PPP6R2	3950816	29	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	CRNKL1	-0.4	0.0
InGen_SS_554	intronicSense_notCorr	chr14:58249978-58250100	+	Cons7_no_Cons100_no	PSMA3	3537865	122	extended	-0.1	0.0	0.003	ZNRF3	0.5	0.0	RB1	-0.5	0.0
InGen_SS_561	intronicSense_notCorr	chr14:90256840-90256868	+	Cons7_no_Cons100_no	PSMC1	3548282	28	full	-0.1	0.0	0.000	MYOD1	0.5	0.0	NA	NA	NA
InGen_SS_86	intronicSense_notCorr	chr2:53965741-53965933	-	Cons7_no_Cons100_no	PSME4	2553365	192	extended	-0.1	0.0	0.013	ZNRF3	0.5	0.0	RB1	-0.5	0.0
InGen_SS_746	intronicSense_notCorr	chr20:50552583-50552610	+	Cons7_no_Cons100_no	PTPN1	3888781	27	extended	-0.1	0.0	0.000	ZNRF3	0.4	0.0	JAK1	-0.5	0.0
InGen_SS_106	intronicSense_notCorr	chr2:219294008-	-	Cons7_yes_Cons100	PTPRN	2600098	94	full	-0.1	0.0	0.013	NA	NA	NA	NA	NA	NA

		219294102		_yes													
InGen_SS_552	intronicSense_notCorr	chr13:114031663-114031692	-	Cons7_no_Cons100_no	RASA3	3526878	29	full	0.1	0.0	0.000	NA	NA	NA	PTEN	-0.5	0.0
InGen_OS_2417	intronicSense_notCorr	chr13:48532859-48532990	+	Cons7_no_Cons100_no	RB1	3489114	131	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_475	intronicSense_notCorr	chr11:65658242-65658282	-	Cons7_no_Cons100_no	RELA	3377801	40	full	-0.1	0.0	0.000	DCAF12L2	0.5	0.0	TRIP11	-0.6	0.0
InGen_SS_25	intronicSense_notCorr	chr1:183710230-183710270	+	Cons7_no_Cons100_yes	RGL1	2371359	40	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_602	intronicSense_notCorr	chr14:49586781-49586837	-	Cons7_yes_Cons100_yes	RPS29	3563334	56	extended	-0.2	0.0	0.000	NA	NA	NA	TBX3	-0.3	0.0
InGen_SS_Miss_51	intronicSense_notCorr	chr1:218314839-218314978	+	Cons7_no_Cons100_no	RRP15	2380581	139	extended	-0.1	0.0	0.004	ZNRF3	0.7	0.0	MALT1	-0.3	0.0
InGen_SS_481	intronicSense_notCorr	chr11:77820385-77820453	-	Cons7_no_Cons100_no	RSF1	3383039	68	full	-0.1	0.0	0.000	CNOT3	0.4	0.0	MAP2K4	-0.4	0.0
InGen_SS_29	intronicSense_notCorr	chr1:237603395-237603431	+	Cons7_no_Cons100_no	RYR2	2387273	36	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1029	intronicSense_notCorr	chr5:116574704-116574823	+	Cons7_yes_Cons100_yes	SEMA6A-AS2	2825178	119	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1030	intronicSense_notCorr	chr5:116574824-116574855	+	Cons7_yes_Cons100_yes	SEMA6A-AS2	2825179	31	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2837	intronicSense_notCorr	chr16:30395211-30395236	-	Cons7_no_Cons100_no	SEPT1	3687781	25	free	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_79	intronicSense_notCorr	chr2:241340791-241341093	+	Cons7_no_Cons100_no	SEPT2	2536344	302	extended	0.2	0.0	0.020	ZNRF3	0.5	0.0	STAG1	-0.5	0.0
InGen_SS_574	intronicSense_notCorr	chr14:99468264-99468293	-	Cons7_no_Cons100_no	SETD3	3579250	29	full	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3404	intronicSense_notCorr	chr21:45507472-45507502	-	Cons7_no_Cons100_no	SLC19A1	3935073	30	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_719	intronicSense_notCorr	chr19:6443232-6443257	-	Cons7_no_Cons100_no	SLC25A23	3847884	25	extended	-0.1	0.0	0.000	ZNRF3	0.8	0.0	RB1	-0.3	0.0
InGen_SS_34	intronicSense_notCorr	chr1:9071553-9071664	-	Cons7_no_Cons100_no	SLC2A5	2395605	111	full	-0.1	0.0	0.036	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_4	intronicSense_notCorr	chr1:8330942-8330990	+	Cons7_no_Cons100_no	SLC45A1	2318963	48	full	-0.1	0.0	0.013	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_213	intronicSense_notCorr	chr3:112183465-112183926	-	Cons7_yes_Cons100_yes	SLC9C1	2688645	461	extended	0.4	0.0	0.030	NA	NA	NA	NA	NA	NA
InGen_SS_484	intronicSense_notCorr	chr11:93538126-93538227	-	Cons7_no_Cons100_no	SMCO4	3386759	101	full	-0.1	0.0	0.024	DCAF12L2	0.4	0.0	ARHGAP5	-0.3	0.0
InGen_SS_666	intronicSense_notCorr	chr17:2228207-2228234	-	Cons7_no_Cons100_no	SMG6	3740920	27	extended	-0.1	0.0	0.001	NA	NA	NA	NA	NA	NA
InGen_SS_297	intronicSense_notCorr	chr7:128031523-128031655	+	Cons7_no_Cons100_yes	SND1	3022625	132	extended	-0.2	0.0	0.000	DCAF12L2	0.3	0.0	KRAS	-0.3	0.0
InGen_SS_Miss_899	intronicSense_notCorr	chr2:1283306-1283336	+	Cons7_no_Cons100_no	SNTG2	2466517	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_194	intronicSense_notCorr	chr5:54524935-54524963	+	Cons7_no_Cons100_no	SNX18	2809640	28	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	STAG1	-0.3	0.0
InGen_SS_166	intronicSense_notCorr	chr4:7630201-7630227	+	Cons7_no_Cons100_no	SORCS2	2717321	26	full	-0.1	0.0	0.023	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_248	intronicSense_notCorr	chr4:123024993-123025033	+	Cons7_no_Cons100_yes	SPATA5	2742158	40	full	0.2	0.0	0.014	NA	NA	NA	NA	NA	NA
InGen_SS_553	intronicSense_notCorr	chr14:35021493-35021591	+	Cons7_no_Cons100_no	SRP54	3532337	98	extended	-0.1	0.0	0.000	ZNRF3	0.4	0.0	STAG1	-0.4	0.0
InGen_SS_636	intronicSense_notCorr	chr17:2314430-2314718	+	Cons7_no_Cons100_no	SRR	3706231	288	extended	-0.1	0.0	0.004	NA	NA	NA	ARID2	-0.3	0.0

InGen_SS_527	intronicSense_notCorr	chr12:130819893-130819924	-	Cons7_no_Cons100_no	STX2	3478475	31	full	-0.1	0.0	0.013	ZNRF3	0.7	0.0	ARHGAP5	-0.4	0.0
InGen_SS_646	intronicSense_notCorr	chr17:55146554-55146657	+	Cons7_no_Cons100_no	STXBP4	3727557	103	extended	-0.1	0.0	0.009	NA	NA	NA	NA	NA	NA
InGen_SS_545	intronicSense_notCorr	chr13:48001137-48001172	-	Cons7_no_Cons100_no	SUCLA2	3513355	35	full	0.1	0.0	0.000	TERT	0.4	0.0	ARHGAP5	-0.3	0.0
InGen_SS_600	intronicSense_notCorr	chr15:101678085-101678148	-	Cons7_no_Cons100_no	TARSL2	3642411	63	extended	-0.1	0.0	0.040	NA	NA	NA	NA	NA	NA
InGen_SS_784	intronicSense_notCorr	chr22:46977374-46977400	+	Cons7_no_Cons100_no	TBC1D22A	3949297	26	full	0.3	0.0	0.000	C15orf65	0.4	0.0	CTNND1	-0.3	0.0
InGen_SS_135	intronicSense_notCorr	chr3:17175918-17175944	-	Cons7_no_Cons100_no	TBC1D5	2664899	26	full	-0.1	0.0	0.000	ZNRF3	0.8	0.0	CDKN1B	-0.4	0.0
InGen_SS_157	intronicSense_notCorr	chr3:177156447-177156476	-	Cons7_no_Cons100_no	TBL1XR1	2706412	29	extended	0.1	0.0	0.001	ZNRF3	0.6	0.0	KIF5B	-0.4	0.0
InGen_SS_241	intronicSense_notCorr	chr6:133891667-133891692	+	Cons7_no_Cons100_yes	TCF21	2926451	25	full	-0.2	0.0	0.000	NA	NA	NA	MAX	-0.3	0.0
InGen_SS_Miss_456	intronicSense_notCorr	chr10:49972205-49972265	+	Cons7_no_Cons100_no	TIMM23B	3246345	60	full	0.1	0.0	0.000	ZNRF3	0.8	0.0	CDKN1B	-0.4	0.0
InGen_SS_792	intronicSense_notCorr	chr22:38246101-38246132	-	Cons7_yes_Cons100_yes	TMEM184B	3960461	31	full	-0.2	0.0	0.000	DGCR8	0.4	0.0	NIN	-0.4	0.0
InGen_SS_793	intronicSense_notCorr	chr22:38225358-38225395	-	Cons7_no_Cons100_yes	TMEM184B	3960448	37	extended	-0.2	0.0	0.158	NA	NA	NA	NA	NA	NA
InGen_OS_2813	intronicSense_notCorr	chr16:8782309-8782760	-	Cons7_no_Cons100_yes	TMEM186	3679498	451	extended	-0.4	0.0	0.015	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_533	intronicSense_notCorr	chr11:125111478-125111509	-	Cons7_no_Cons100_no	TMEM218	3396350	31	extended	-0.1	0.0	0.000	ZNRF3	1.0	0.0	CDKN1B	-0.3	0.0
InGen_SS_Miss_334	intronicSense_notCorr	chr6:157310590-157310619	-	Cons7_no_Cons100_no	TMEM242	4048792	29	extended	-0.1	0.0	0.358	ZNRF3	0.5	0.0	ARHGAP5	-0.6	0.0
InGen_SS_779	intronicSense_notCorr	chr22:40217826-40217946	+	Cons7_no_Cons100_no	TNRC6B	3946255	120	extended	-0.1	0.0	0.000	ZNRF3	0.3	0.0	GNAQ	-0.3	0.0
InGen_SS_571	intronicSense_notCorr	chr14:92026834-92026866	-	Cons7_yes_Cons100_yes	TRIP11	3576858	32	full	0.3	0.0	0.000	ZNRF3	0.9	0.0	CDKN1B	-0.3	0.0
InGen_SS_572	intronicSense_notCorr	chr14:92026565-92026609	-	Cons7_no_Cons100_yes	TRIP11	3576856	44	full	0.2	0.0	0.000	ZNRF3	1.0	0.0	CDKN1B	-0.3	0.0
InGen_SS_724	intronicSense_notCorr	chr19:12731308-12731338	-	Cons7_yes_Cons100_yes	TRIR	3851707	30	extended	-0.1	0.0	0.000	DGCR8	0.4	0.0	ARHGAP5	-0.5	0.0
InGen_SS_800	intronicSense_notCorr	chr22:50234115-50234141	-	Cons7_no_Cons100_no	TUBGCP6	3965682	26	full	0.1	0.0	0.105	ZNRF3	1.0	0.0	CDKN1B	-0.3	0.0
InGen_OS_205	intronicSense_notCorr	chr1:52414823-52414853	-	Cons7_yes_Cons100_yes	TUT4	2412838	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_626	intronicSense_notCorr	chr16:11698425-11698479	-	Cons7_no_Cons100_no	TXNDC11	3680500	54	full	-0.1	0.0	0.000	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
InGen_SS_168	intronicSense_notCorr	chr4:39731585-39731643	+	Cons7_no_Cons100_no	UBE2K	2724518	58	extended	-0.2	0.0	0.000	ZNRF3	0.5	0.0	RB1	-0.4	0.0
InGen_SS_172	intronicSense_notCorr	chr4:119237367-119237423	+	Cons7_no_Cons100_no	USP53	2741252	56	extended	-0.1	0.0	0.006	ZNRF3	0.9	0.0	JAK1	-0.3	0.0
InGen_SS_179	intronicSense_notCorr	chr4:10087656-10087683	-	Cons7_no_Cons100_no	WDR1	2760397	27	full	-0.1	0.0	0.000	DGCR8	0.4	0.0	SMAD4	-0.5	0.0
InGen_SS_180	intronicSense_notCorr	chr4:10099555-10099584	-	Cons7_no_Cons100_no	WDR1	2760417	29	full	-0.1	0.0	0.000	KMT2D	0.4	0.0	RABEP1	-0.5	0.0
InGen_SS_721	intronicSense_notCorr	chr19:10926515-10926653	-	Cons7_no_Cons100_no	YIPF2	3850593	138	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	ARHGAP5	-0.6	0.0
InGen_SS_Miss_848	intronicSense_notCorr	chr22:37646487-37646514	-	Cons7_no_Cons100_no	Z83844.2	3960194	27	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_SS_141	intronicSense_notCorr	chr3:44937900-	-	Cons7_yes_Cons100	ZDHHC3	2671681	37	extended	0.1	0.0	0.085	ZNRF3	0.6	0.0	ARHGAP	-0.5	0.0

		44937937		_yes											5			
InGen_SS_669	intronicSense_notCorr	chr17:5122895-5122920	-	Cons7_no_Cons100_no	ZNF232	3742581	25	full	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_621	intronicSense_notCorr	chr15:64593327-64593373	+	Cons7_yes_Cons100_yes	ZNF609	3598118	46	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_262	intronicSense_notCorr	chr3:40505929-40505956	+	Cons7_no_Cons100_no	ZNF620	2618704	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
InGen_SS_Miss_269	intronicSense_notCorr	chr5:61375994-61376053	+	Cons7_yes_Cons100_yes	ZSWIM6	2811453	59	extended	0.1	0.0	0.002	NA	NA	NA	NA	NA	NA	NA

Antisense Positive RNAs Correlated and Negative Correlated

Classification_Pipeline	Classification_GencodeV29	hg38Coordinates	strand	ConservationAnalysis	hostingGeneSymbol	Probe_ID	length_BP	level	logFC_Limma	adj.P.Val_Limma	coding_prob_CPAT	CGC_pos	CGC_pos_corr	CGC_pos_p	CGC_neg	CGC_neg_corr	CGC_neg_p
InGen_OS_2971	exonicAntisense_negativeCorr	chr17:7224344-7224452	-	Cons7_yes_Cons100_yes	ACADVL	3743438	108	extended	-0.2	0.0	0.127	ERCC2	0.4	0.0	ARHGAP5	-0.3	0.0
InGen_OS_3189	exonicAntisense_negativeCorr	chr19:48807979-48808099	+	Cons7_no_Cons100_no	BCAT2	3837997	120	extended	-0.1	0.0	0.025	ESR1	0.4	0.0	NA	NA	NA
InGen_OS_2548	exonicAntisense_negativeCorr	chr14:103519974-103520021	+	Cons7_yes_Cons100_yes	CKB	3553763	47	extended	-0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1606	exonicAntisense_negativeCorr	chr8:11842902-11842997	+	Cons7_no_Cons100_no	CTSB	3086257	95	extended	0.1	0.0	0.000	ZNRF3	0.6	0.0	ARHGAP5	-0.6	0.0
InGen_OS_1607	exonicAntisense_negativeCorr	chr8:11845100-11845198	+	Cons7_yes_Cons100_yes	CTSB	3086263	98	extended	0.3	0.0	0.000	DGCR8	0.5	0.0	ARHGAP5	-0.5	0.0
InGen_OS_1480	exonicAntisense_negativeCorr	chr7:135170382-135170411	+	Cons7_no_Cons100_no	CYREN	3025763	29	extended	-0.1	0.0	0.000	TP63	0.3	0.0	ABI1	-0.3	0.0
InGen_OS_1481	exonicAntisense_negativeCorr	chr7:135170435-135170504	+	Cons7_no_Cons100_no	CYREN	3025764	69	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1189	exonicAntisense_negativeCorr	chr5:151028124-151028294	-	Cons7_yes_Cons100_yes	GPX3	2881665	170	extended	-0.5	0.0	0.000	ZNRF3	0.3	0.0	NCOA2	-0.4	0.0
InGen_OS_1725	exonicAntisense_negativeCorr	chr8:47736238-47736741	-	Cons7_no_Cons100_yes	SPIDR	3134021	503	extended	-0.3	0.0	0.002	PIK3R1	0.6	0.0	WHSC1	-0.6	0.0
InGen_OS_1240	exonicAntisense_negativeCorr	chr6:33299945-33300191	+	Cons7_no_Cons100_no	TAPBP	2903608	246	extended	0.3	0.0	0.007	DGCR8	0.6	0.0	MALT1	-0.5	0.0
InGen_OS_918	exonicAntisense_negativeCorr	chr4:55415994-55416023	-	Cons7_no_Cons100_no	TMEM165	2769940	29	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2766	exonicAntisense_positiveCorr	chr16:2328631-2328708	+	Cons7_no_Cons100_yes	ABCA3	3644712	77	full	-0.2	0.0	0.011	NA	NA	NA	NA	NA	NA
InGen_OS_2937	exonicAntisense_positiveCorr	chr17:62966539-62966586	+	Cons7_yes_Cons100_yes	AC005972.3	3730440	47	extended	-0.1	0.0	0.000	CRTC1	0.5	0.0	SDHD	-0.4	0.0
InGen_OS_3014	exonicAntisense_positiveCorr	chr17:43932447-43932530	-	Cons7_no_Cons100_yes	AC007993.2	3758759	83	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3112	exonicAntisense_positiveCorr	chr19:1876038-1876064	+	Cons7_no_Cons100_no	AC012615.1	3816104	26	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_890	exonicAntisense_positiveCorr	chr4:1713229-1713264	-	Cons7_no_Cons100_no	AC016773.1	2757340	35	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2299	exonicAntisense_positiveCorr	chr12:127145957-127146010	+	Cons7_no_Cons100_no	AC079949.1	3437078	53	full	-0.1	0.0	0.000	TNFRSF17	0.4	0.0	NA	NA	NA
InGen_OS_2932	exonicAntisense_positiveCorr	chr17:48646696-	+	Cons7_no_Cons100_yes	AC103702.2	3725328	127	full	-0.1	0.0	0.000	NA	NA	NA	ARID2	-0.3	0.0

		48646823																
lnGen_OS_883	exonicAntisense_positiveCorr	chr4:183504460-183504500	+	Cons7_no_Cons100_no	AC107214.1	2753950	40	extended	-0.1	0.0	0.000	KLK2	0.7	0.0	NA	NA	NA
lnGen_OS_1406	exonicAntisense_positiveCorr	chr7:5527284-5527883	+	Cons7_yes_Cons100_yes	ACTB	2988699	599	extended	0.1	0.0	0.625	CDH1	0.3	0.0	FOXP1	-0.3	0.0
lnGen_OS_1407	exonicAntisense_positiveCorr	chr7:5528656-5528696	+	Cons7_yes_Cons100_yes	ACTB	2988701	40	extended	0.2	0.0	0.006	CD79A	0.5	0.0	BTG1	-0.3	0.0
lnGen_OS_2952	exonicAntisense_positiveCorr	chr17:81511075-81511109	+	Cons7_yes_Cons100_yes	ACTG1	3737987	34	extended	0.2	0.0	0.014	NUTM2B	0.4	0.0	KMT2C	-0.3	0.0
lnGen_OS_2655	exonicAntisense_positiveCorr	chr15:58640796-58640920	+	Cons7_yes_Cons100_yes	ADAM10	3595810	124	extended	0.2	0.0	0.020	NA	NA	NA	NA	NA	NA
lnGen_OS_2526	exonicAntisense_positiveCorr	chr14:75427650-75427681	+	Cons7_no_Cons100_no	AF111167.2	3544570	31	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2105	exonicAntisense_positiveCorr	chr11:62516532-62520320	+	Cons7_yes_Cons100_yes	AHNAK	3333505	3788	extended	-0.1	0.0	0.982	FLNA	0.4	0.0	FANCA	-0.3	0.0
lnGen_OS_2106	exonicAntisense_positiveCorr	chr11:62522887-62522920	+	Cons7_yes_Cons100_yes	AHNAK	3333506	33	extended	-0.2	0.0	0.019	NA	NA	NA	NA	NA	NA
lnGen_OS_2550	exonicAntisense_positiveCorr	chr14:104950600-104950634	+	Cons7_no_Cons100_yes	AHNAK2	3554520	34	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3298	exonicAntisense_positiveCorr	chr20:53504088-53504158	+	Cons7_no_Cons100_no	AL354993.2	3889616	70	extended	0.1	0.0	0.000	RGPD3	0.6	0.0	PTEN	-0.3	0.0
lnGen_OS_55	exonicAntisense_positiveCorr	chr1:84077595-84077628	+	Cons7_no_Cons100_no	AL359504.2	2344395	33	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2468	exonicAntisense_positiveCorr	chr13:78596167-78596196	-	Cons7_no_Cons100_no	AL445209.1	3518927	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2469	exonicAntisense_positiveCorr	chr13:78596304-78596372	-	Cons7_no_Cons100_no	AL445209.1	3518929	68	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1802	exonicAntisense_positiveCorr	chr9:62837450-62837566	+	Cons7_no_Cons100_no	AL512625.2	3172297	116	full	-0.1	0.0	0.184	NA	NA	NA	NA	NA	NA
lnGen_OS_1818	exonicAntisense_positiveCorr	chr9:101425523-101425612	+	Cons7_yes_Cons100_yes	ALDOB	3182449	89	extended	-0.5	0.0	0.002	NA	NA	NA	NA	NA	NA
lnGen_OS_1820	exonicAntisense_positiveCorr	chr9:101427564-101427593	+	Cons7_yes_Cons100_yes	ALDOB	3182452	29	extended	-0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3664	exonicAntisense_positiveCorr	chr17:81891806-81891955	-	Cons7_yes_Cons100_yes	ANAPC11	4041434	149	full	-0.1	0.0	0.000	RGPD3	0.7	0.0	PTEN	-0.3	0.0
lnGen_OS_2351	exonicAntisense_positiveCorr	chr12:45216376-45216619	-	Cons7_no_Cons100_yes	ANO6	3452007	243	extended	-0.1	0.0	0.000	CNBD1	0.4	0.0	NA	NA	NA
lnGen_OS_2686	exonicAntisense_positiveCorr	chr15:89805376-89805439	+	Cons7_yes_Cons100_yes	ANPEP	3607833	63	full	-0.4	0.0	0.174	NA	NA	NA	NA	NA	NA
lnGen_OS_2657	exonicAntisense_positiveCorr	chr15:60347470-60347584	+	Cons7_no_Cons100_yes	ANXA2	3596392	114	extended	0.3	0.0	0.006	NA	NA	NA	NA	NA	NA
lnGen_OS_2219	exonicAntisense_positiveCorr	chr11:134762686-134762767	-	Cons7_no_Cons100_no	AP001999.1	3399826	81	full	-0.1	0.0	0.000	MALAT1bis	0.4	0.0	NA	NA	NA
lnGen_OS_297	exonicAntisense_positiveCorr	chr1:228082678-228082718	-	Cons7_yes_Cons100_yes	ARF1	2459404	40	full	-0.1	0.0	0.176	NUTM1	0.5	0.0	SMAD4	-0.3	0.0
lnGen_OS_2772	exonicAntisense_positiveCorr	chr16:18793180-18793304	+	Cons7_yes_Cons100_yes	ARL6IP1	3650722	124	extended	0.2	0.0	0.005	NA	NA	NA	NSD1	-0.5	0.0
lnGen_OS_1899	exonicAntisense_positiveCorr	chr9:124877230-124877265	-	Cons7_no_Cons100_yes	ARPC5L	3225226	35	full	-0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2652	exonicAntisense_positiveCorr	chr15:52548981-52549078	+	Cons7_no_Cons100_no	ARPP19	3594282	97	extended	0.2	0.0	0.044	NA	NA	NA	NSD1	-0.5	0.0
lnGen_OS_117	exonicAntisense_positiveCorr	chr1:201127118-201127231	+	Cons7_yes_Cons100_yes	ASCL5	2374535	113	full	-0.1	0.0	0.000	CNBD1	0.7	0.0	PRDM2	-0.4	0.0
lnGen_OS_2109	exonicAntisense_positiveCorr	chr11:63671570-63671605	+	Cons7_yes_Cons100_yes	ATL3	3333939	35	full	-0.1	0.0	0.000	MDM2	0.5	0.0	NA	NA	NA

lnGen_OS_487	exonicAntisense_positiveCorr	chr2:85754123-85754170	-	Cons7_no_Cons100_yes	ATOH8	2562493	47	full	-0.1	0.0	0.000	CCR7	0.4	0.0	KMT2C	-0.3	0.0
lnGen_OS_2385	exonicAntisense_positiveCorr	chr12:110326407-110326460	-	Cons7_yes_Cons100_yes	ATP2A2	3471179	53	free	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1456	exonicAntisense_positiveCorr	chr7:99459176-99459200	+	Cons7_no_Cons100_yes	ATP5MF	3014807	24	extended	0.2	0.0	0.004	NA	NA	NA	NA	NA	NA
lnGen_OS_2283	exonicAntisense_positiveCorr	chr12:111597959-111597991	+	Cons7_yes_Cons100_yes	ATXN2	3431986	32	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2286	exonicAntisense_positiveCorr	chr12:111598803-111598945	+	Cons7_yes_Cons100_yes	ATXN2	3432001	142	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1457	exonicAntisense_positiveCorr	chr7:99968155-99968179	+	Cons7_no_Cons100_yes	AZGP1	3015102	24	extended	-0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1107	exonicAntisense_positiveCorr	chr5:17275862-17276332	-	Cons7_yes_Cons100_yes	BASP1	2850308	470	extended	0.2	0.0	0.004	NA	NA	NA	NA	NA	NA
lnGen_OS_2571	exonicAntisense_positiveCorr	chr14:23310977-23311003	-	Cons7_no_Cons100_yes	BCL2L2	3557341	26	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2396	exonicAntisense_positiveCorr	chr12:122021965-122022034	-	Cons7_yes_Cons100_yes	BCL7A	3475454	69	extended	-0.1	0.0	0.000	GLI1	0.3	0.0	CTNND1	-0.4	0.0
lnGen_OS_679	exonicAntisense_positiveCorr	chr3:4984397-4984526	-	Cons7_yes_Cons100_yes	BHLHE40	2661208	129	extended	0.2	0.0	0.008	NA	NA	NA	NA	NA	NA
lnGen_OS_1207	exonicAntisense_positiveCorr	chr5:181052964-181053081	-	Cons7_no_Cons100_no	BTNL9	2890953	117	extended	-0.1	0.0	0.000	GLI1	0.4	0.0	TET2	-0.3	0.0
lnGen_OS_6	exonicAntisense_positiveCorr	chr1:1081980-1082009	+	Cons7_no_Cons100_no	C1orf159	2315490	29	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_277	exonicAntisense_positiveCorr	chr1:184622873-184622934	-	Cons7_no_Cons100_no	C1orf21	2447778	61	full	-0.1	0.0	0.000	ESR1	0.5	0.0	RALGDS	-0.3	0.0
lnGen_OS_3459	exonicAntisense_positiveCorr	chr22:49657201-49657237	+	Cons7_no_Cons100_no	C22orf34	3950268	36	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1580	exonicAntisense_positiveCorr	chr7:134933042-134933171	-	Cons7_yes_Cons100_yes	CALD1	3074136	129	extended	0.2	0.0	0.212	KLK2	0.9	0.0	POLE	-0.3	0.0
lnGen_OS_3210	exonicAntisense_positiveCorr	chr19:12940087-12940149	-	Cons7_yes_Cons100_yes	CALR	3851902	62	extended	0.2	0.0	0.009	NA	NA	NA	NA	NA	NA
lnGen_OS_3211	exonicAntisense_positiveCorr	chr19:12940264-12940346	-	Cons7_yes_Cons100_yes	CALR	3851903	82	extended	0.2	0.0	0.000	MDM2	0.4	0.0	NA	NA	NA
lnGen_OS_1577	exonicAntisense_positiveCorr	chr7:128769647-128770933	-	Cons7_yes_Cons100_yes	CALU	3071857	1286	extended	0.3	0.0	0.026	KLK2	0.6	0.0	ARID1B	-0.3	0.0
lnGen_OS_1205	exonicAntisense_positiveCorr	chr5:179729459-179730403	-	Cons7_yes_Cons100_yes	CANX	2890215	944	extended	0.2	0.0	0.003	FOXA1	0.4	0.0	JAK2	-0.4	0.0
lnGen_OS_1571	exonicAntisense_positiveCorr	chr7:116526549-116526686	-	Cons7_yes_Cons100_yes	CAV1	3069183	137	extended	-0.2	0.0	0.001	NA	NA	NA	NA	NA	NA
lnGen_OS_2954	exonicAntisense_positiveCorr	chr17:82104945-82104972	+	Cons7_no_Cons100_no	CCDC57	3738546	27	free	-0.1	0.0	0.000	GLI1	0.4	0.0	TET2	-0.4	0.0
lnGen_OS_2190	exonicAntisense_positiveCorr	chr11:69648042-69648119	-	Cons7_yes_Cons100_yes	CCND1	3380052	77	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2973	exonicAntisense_positiveCorr	chr17:7580086-7580262	-	Cons7_yes_Cons100_yes	CD68	3743834	176	extended	0.3	0.0	0.065	ZNRF3	0.9	0.0	CDKN1B	-0.3	0.0
lnGen_OS_1071	exonicAntisense_positiveCorr	chr5:150407288-150407317	+	Cons7_yes_Cons100_yes	CD74	2835507	29	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3191	exonicAntisense_positiveCorr	chr19:54465228-54465259	+	Cons7_yes_Cons100_yes	CDC42EP5	3841379	31	full	-0.1	0.0	0.773	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
lnGen_OS_2297	exonicAntisense_positiveCorr	chr12:123265219-123265307	+	Cons7_yes_Cons100_yes	CDK2AP1	3435787	88	extended	0.1	0.0	0.031	MNX1	0.3	0.0	NA	NA	NA
lnGen_OS_2100	exonicAntisense_positiveCorr	chr11:47469909-47469992	+	Cons7_yes_Cons100_yes	CELF1	3329856	83	extended	-0.1	0.0	0.008	NA	NA	NA	NA	NA	NA
lnGen_OS_150	exonicAntisense_positiveCorr	chr1:243255078-	+	Cons7_yes_Cons100_yes	CEP170	2388524	40	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA

		243255118																
lnGen_OS_2201	exonicAntisense_positiveCorr	chr11:93661604-93661689	-	Cons7_yes_Cons100_yes	CEP295	3386810	85	full	-0.1	0.0	0.060	NA	NA	NA	NA	NA	NA
lnGen_OS_1278	exonicAntisense_positiveCorr	chr6:118651429-118651476	+	Cons7_yes_Cons100_yes	CEP85L	2923322	47	full	-0.1	0.0	0.937	NA	NA	NA	NA	NA	NA
lnGen_OS_1279	exonicAntisense_positiveCorr	chr6:118651496-118651523	+	Cons7_yes_Cons100_no	CEP85L	2923323	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2730	exonicAntisense_positiveCorr	chr15:57548131-57548162	-	Cons7_no_Cons100_no	CGNL1	3626182	31	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2844	exonicAntisense_positiveCorr	chr16:53156059-53156368	-	Cons7_yes_Cons100_yes	CHD9	3691901	309	extended	-0.2	0.0	0.035	FBXW7	0.3	0.0	SMAD3	-0.3	0.0
lnGen_OS_667	exonicAntisense_positiveCorr	chr3:190307382-190307538	+	Cons7_yes_Cons100_yes	CLDN1	2657795	156	extended	0.4	0.0	0.094	POU2AF1	0.7	0.0	PRDM2	-0.4	0.0
lnGen_OS_1239	exonicAntisense_positiveCorr	chr6:31736293-31736460	+	Cons7_yes_Cons100_yes	CLIC1	2902630	167	extended	0.3	0.0	0.011	KLK2	0.6	0.0	KMT2C	-0.4	0.0
lnGen_OS_1544	exonicAntisense_positiveCorr	chr7:74289572-74289607	-	Cons7_no_Cons100_no	CLIP2	3056442	35	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1615	exonicAntisense_positiveCorr	chr8:27599776-27599915	+	Cons7_yes_Cons100_yes	CLU	3091474	139	extended	-0.5	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1373	exonicAntisense_positiveCorr	chr6:122996359-122996385	-	Cons7_no_Cons100_yes	CLVS2	2972400	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_295	exonicAntisense_positiveCorr	chr1:224616417-224616531	-	Cons7_no_Cons100_no	CNIH3	2458170	114	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1262	exonicAntisense_positiveCorr	chr6:75085990-75086238	+	Cons7_yes_Cons100_yes	COL12A1	2913877	248	extended	0.3	0.0	0.003	CD28	0.5	0.0	NA	NA	NA
lnGen_OS_2935	exonicAntisense_positiveCorr	chr17:50185778-50185831	+	Cons7_yes_Cons100_yes	COL1A1	3726289	53	free	1.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_556	exonicAntisense_positiveCorr	chr2:189009073-189009187	-	Cons7_yes_Cons100_yes	COL3A1	2591635	114	extended	0.7	0.0	0.010	GRM3	0.3	0.0	IKBKB	-0.3	0.0
lnGen_OS_2958	exonicAntisense_positiveCorr	chr17:82273384-82273413	+	Cons7_yes_Cons100_yes	CSNK1D	3738731	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3385	exonicAntisense_positiveCorr	chr21:43776197-43776308	+	Cons7_no_Cons100_yes	CSTB	3923313	111	extended	0.1	0.0	0.407	NA	NA	NA	NA	NA	NA
lnGen_OS_2867	exonicAntisense_positiveCorr	chr16:88712691-88712716	-	Cons7_yes_Cons100_yes	CTU2	3704338	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_372	exonicAntisense_positiveCorr	chr2:72130840-72130905	+	Cons7_no_Cons100_no	CYP26B1	2488506	65	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_49	exonicAntisense_positiveCorr	chr1:57423997-57424061	+	Cons7_yes_Cons100_yes	DAB1	2337987	64	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_698	exonicAntisense_positiveCorr	chr3:49470271-49470367	-	Cons7_yes_Cons100_yes	DAG1	2674547	96	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2266	exonicAntisense_positiveCorr	chr12:91146081-91146229	+	Cons7_yes_Cons100_yes	DCN	3425654	148	extended	-0.2	0.0	0.002	NA	NA	NA	NA	NA	NA
lnGen_OS_2267	exonicAntisense_positiveCorr	chr12:91178373-91178507	+	Cons7_yes_Cons100_yes	DCN	3425658	134	extended	-0.3	0.0	0.000	FLNA	0.3	0.0	NA	NA	NA
lnGen_OS_373	exonicAntisense_positiveCorr	chr2:74361499-74361528	+	Cons7_yes_Cons100_yes	DCTN1	2489203	29	full	0.2	0.0	0.003	NA	NA	NA	NA	NA	NA
lnGen_OS_1419	exonicAntisense_positiveCorr	chr7:35037786-35037861	+	Cons7_no_Cons100_no	DPY19L1	2996868	75	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3068	exonicAntisense_positiveCorr	chr18:31042617-31042646	+	Cons7_no_Cons100_no	DSC3	3783357	29	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_726	exonicAntisense_positiveCorr	chr3:107240670-107240696	-	Cons7_no_Cons100_yes	DUBR	2687545	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2725	exonicAntisense_positiveCorr	chr15:45152479-45152505	-	Cons7_yes_Cons100_yes	DUOX1	3622277	26	full	-0.1	0.0	0.471	DCAF12L2	0.4	0.0	RB1	-0.3	0.0

lnGen_OS_2392	exonicAntisense_positiveCorr	chr12:120496177-120496211	-	Cons7_no_Cons100_no	DYNLL1	3474574	34	extended	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3335	exonicAntisense_positiveCorr	chr20:34540696-34540798	-	Cons7_no_Cons100_yes	DYNLRB1	3903476	102	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3285	exonicAntisense_positiveCorr	chr20:33686371-33686500	+	Cons7_yes_Cons100_yes	E2F1	3882646	129	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1323	exonicAntisense_positiveCorr	chr6:20402285-20402315	-	Cons7_yes_Cons100_yes	E2F3	2944600	30	full	-0.1	0.0	0.000	RGPD3	0.6	0.0	PTEN	-0.3	0.0
lnGen_OS_1261	exonicAntisense_positiveCorr	chr6:73517872-73517901	+	Cons7_yes_Cons100_yes	EEF1A1	2913626	29	extended	-0.2	0.0	0.012	FOXA1	0.3	0.0	NSD1	-0.3	0.0
lnGen_OS_782	exonicAntisense_positiveCorr	chr3:184321541-184321682	-	Cons7_yes_Cons100_yes	EIF4G1	2708447	141	extended	0.3	0.0	0.057	CNBD1	0.7	0.0	PRDM2	-0.4	0.0
lnGen_OS_1025	exonicAntisense_positiveCorr	chr5:95961586-95961655	+	Cons7_yes_Cons100_yes	ELL2	2821091	69	extended	-0.3	0.0	0.684	NA	NA	NA	NA	NA	NA
lnGen_OS_749	exonicAntisense_positiveCorr	chr3:134795612-134795642	-	Cons7_yes_Cons100_yes	EPHB1	2696494	30	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3122	exonicAntisense_positiveCorr	chr19:11383202-11383234	+	Cons7_yes_Cons100_yes	EPOR	3821196	32	extended	-0.1	0.0	0.000	C15orf65	0.4	0.0	GNA11	-0.3	0.0
lnGen_OS_40	exonicAntisense_positiveCorr	chr1:44355026-44355059	+	Cons7_yes_Cons100_yes	ERI3	2333896	33	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3485	exonicAntisense_positiveCorr	chr22:29268216-29268271	-	Cons7_yes_Cons100_yes	EWSR1	3956748	55	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_735	exonicAntisense_positiveCorr	chr3:122384211-122384265	-	Cons7_yes_Cons100_yes	FAM162A	2691963	54	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2257	exonicAntisense_positiveCorr	chr12:62191242-62191272	+	Cons7_yes_Cons100_yes	FAM19A2	3419134	30	full	-0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1990	exonicAntisense_positiveCorr	chr10:5685140-5685201	-	Cons7_no_Cons100_yes	FAM208B	3275101	61	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3188	exonicAntisense_positiveCorr	chr19:48601125-48601150	+	Cons7_no_Cons100_yes	FAM83E	3837881	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2838	exonicAntisense_positiveCorr	chr16:30924221-30924266	-	Cons7_yes_Cons100_yes	FBXL19	3688087	45	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1965	exonicAntisense_positiveCorr	chr10:101695176-101695247	+	Cons7_yes_Cons100_yes	FBXW4	3261312	71	full	-0.1	0.0	0.265	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
lnGen_OS_1623	exonicAntisense_positiveCorr	chr8:38466492-38466561	+	Cons7_no_Cons100_no	FGFR1	3094699	69	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_428	exonicAntisense_positiveCorr	chr2:215365528-215365604	+	Cons7_yes_Cons100_yes	FN1	2526813	76	extended	0.4	0.0	0.004	KLK2	0.6	0.0	SPEN	-0.3	0.0
lnGen_OS_429	exonicAntisense_positiveCorr	chr2:215371952-215371982	+	Cons7_yes_Cons100_yes	FN1	2526818	30	extended	0.5	0.0	0.008	KLK2	0.8	0.0	SPEN	-0.3	0.0
lnGen_OS_430	exonicAntisense_positiveCorr	chr2:215384062-215384321	+	Cons7_yes_Cons100_yes	FN1	2526824	259	extended	0.5	0.0	0.007	KLK2	1.0	0.0	NA	NA	NA
lnGen_OS_431	exonicAntisense_positiveCorr	chr2:215384458-215384980	+	Cons7_yes_Cons100_yes	FN1	2526825	522	extended	0.4	0.0	0.001	KLK2	0.6	0.0	AFF1	-0.3	0.0
lnGen_OS_2322	exonicAntisense_positiveCorr	chr12:6536714-6536769	-	Cons7_yes_Cons100_yes	GAPDH	3441985	55	full	0.3	0.0	0.079	FOXA1	0.4	0.0	KMT2C	-0.4	0.0
lnGen_OS_2323	exonicAntisense_positiveCorr	chr12:6537630-6537655	-	Cons7_yes_Cons100_yes	GAPDH	3441987	25	extended	0.3	0.0	0.000	NUTM2B	0.3	0.0	NSD1	-0.4	0.0
lnGen_OS_1190	exonicAntisense_positiveCorr	chr5:151268788-151269131	-	Cons7_no_Cons100_no	GM2A	2881916	343	extended	0.1	0.0	0.038	MNX1	0.3	0.0	NA	NA	NA
lnGen_OS_2181	exonicAntisense_positiveCorr	chr11:64285144-64285243	-	Cons7_no_Cons100_yes	GPR137	3376847	99	extended	-0.1	0.0	0.056	CCR7	0.3	0.0	CUX1	-0.3	0.0
lnGen_OS_2713	exonicAntisense_positiveCorr	chr15:32732936-32733153	-	Cons7_yes_Cons100_yes	GREM1	3616793	217	extended	0.7	0.0	0.002	CCR4	0.6	0.0	NA	NA	NA
lnGen_OS_1910	exonicAntisense_positiveCorr	chr9:137165017-	-	Cons7_yes_Cons100_no	GRIN1	3230863	28	full	-0.1	0.0	0.000	NA	NA	NA	ARID2	-0.3	0.0

		137165045															
lnGen_OS_2186	exonicAntisense_positiveCorr	chr11:67584710-67584771	-	Cons7_yes_Cons100_yes	GSTP1	3378981	61	extended	0.2	0.0	0.015	NA	NA	NA	NA	NA	NA
lnGen_OS_2187	exonicAntisense_positiveCorr	chr11:67585170-67585221	-	Cons7_yes_Cons100_yes	GSTP1	3378982	51	extended	0.3	0.0	0.106	NA	NA	NA	NA	NA	NA
lnGen_OS_1617	exonicAntisense_positiveCorr	chr8:30657845-30658065	+	Cons7_no_Cons100_yes	GTF2E2	3092592	220	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2087	exonicAntisense_positiveCorr	chr11:5226610-5226742	+	Cons7_yes_Cons100_yes	HBB	3318304	132	extended	-0.5	0.0	0.023	NA	NA	NA	NA	NA	NA
lnGen_OS_3562	exonicAntisense_positiveCorr	chrX:153971247-153971291	+	Cons7_yes_Cons100_yes	HCFC1	3996084	44	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1832	exonicAntisense_positiveCorr	chr9:113376174-113376227	+	Cons7_no_Cons100_no	HDHD3	3185609	53	extended	-0.1	0.0	0.123	NA	NA	NA	NA	NA	NA
lnGen_OS_3047	exonicAntisense_positiveCorr	chr17:82418728-82418813	-	Cons7_no_Cons100_no	HEXDC	3775028	85	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1329	exonicAntisense_positiveCorr	chr6:27133040-27133148	-	Cons7_yes_Cons100_yes	HIST1H2AG	2946713	108	full	0.4	0.0	0.000	MTCP1	0.4	0.0	NA	NA	NA
lnGen_OS_1233	exonicAntisense_positiveCorr	chr6:27814450-27814482	+	Cons7_yes_Cons100_yes	HIST1H2AJ	2900057	32	extended	0.6	0.0	0.001	NA	NA	NA	NA	NA	NA
lnGen_OS_103	exonicAntisense_positiveCorr	chr1:149842295-149842700	+	Cons7_yes_Cons100_yes	HIST2H2AA3	2357914	405	extended	0.3	0.0	0.039	FOXA1	0.3	0.0	NSD1	-0.3	0.0
lnGen_OS_1330	exonicAntisense_positiveCorr	chr6:29945477-29945810	-	Cons7_no_Cons100_no	HLA-A	2948152	333	extended	0.2	0.0	0.019	KLK2	0.5	0.0	BRD3	-0.3	0.0
lnGen_OS_1238	exonicAntisense_positiveCorr	chr6:31354489-31354517	+	Cons7_no_Cons100_yes	HLA-B	2902269	28	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1705	exonicAntisense_positiveCorr	chr8:29064405-29064435	-	Cons7_yes_Cons100_yes	HMBOX1	3129591	30	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1340	exonicAntisense_positiveCorr	chr6:34240780-34240888	-	Cons7_yes_Cons100_yes	HMGA1	2951059	108	full	0.5	0.0	0.000	MYB	0.5	0.0	MALT1	-0.4	0.0
lnGen_OS_1341	exonicAntisense_positiveCorr	chr6:34244924-34244961	-	Cons7_yes_Cons100_yes	HMGA1	2951076	37	extended	0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_880	exonicAntisense_positiveCorr	chr4:173334658-173334682	+	Cons7_yes_Cons100_yes	HMGB2	2752003	24	ambiguous	-0.2	0.0	0.553	GLI1	0.4	0.0	TET2	-0.3	0.0
lnGen_OS_2491	exonicAntisense_positiveCorr	chr14:21269323-21269352	+	Cons7_yes_Cons100_yes	HNRNPC	3528026	29	full	0.1	0.0	0.000	CD28	0.5	0.0	NA	NA	NA
lnGen_OS_1415	exonicAntisense_positiveCorr	chr7:27102117-27102194	+	Cons7_yes_Cons100_yes	HOXA2	2994049	77	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_394	exonicAntisense_positiveCorr	chr2:128318070-128318096	+	Cons7_yes_Cons100_yes	HS6ST1	2505005	26	extended	-0.1	0.0	0.000	CSF3R	0.3	0.0	MLLT6	-0.3	0.0
lnGen_OS_1353	exonicAntisense_positiveCorr	chr6:44252003-44252055	-	Cons7_yes_Cons100_yes	HSP90AB1	2955060	52	extended	0.2	0.0	0.000	NA	NA	NA	NSD1	-0.5	0.0
lnGen_OS_2377	exonicAntisense_positiveCorr	chr12:103943565-103943822	-	Cons7_yes_Cons100_yes	HSP90B1	3468855	257	extended	0.2	0.0	0.009	KLK2	0.5	0.0	CREBBP	-0.3	0.0
lnGen_OS_1837	exonicAntisense_positiveCorr	chr9:125236938-125237041	+	Cons7_yes_Cons100_yes	HSPA5	3189105	103	extended	0.2	0.0	0.000	GPC5	0.3	0.0	NA	NA	NA
lnGen_OS_1838	exonicAntisense_positiveCorr	chr9:125238664-125239175	+	Cons7_yes_Cons100_yes	HSPA5	3189107	511	extended	0.2	0.0	0.251	GPC5	0.3	0.0	NA	NA	NA
lnGen_OS_1409	exonicAntisense_positiveCorr	chr7:8261948-8262167	+	Cons7_yes_Cons100_yes	ICA1	2989716	219	full	-0.1	0.0	0.000	NUTM1	0.5	0.0	NSD1	-0.3	0.0
lnGen_OS_26	exonicAntisense_positiveCorr	chr1:23559236-23559405	+	Cons7_yes_Cons100_yes	ID3	2325147	169	extended	-0.2	0.0	0.220	NA	NA	NA	NA	NA	NA
lnGen_OS_2142	exonicAntisense_positiveCorr	chr11:309262-309293	-	Cons7_no_Cons100_no	IFITM2	3357770	31	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2554	exonicAntisense_positiveCorr	chr14:105707833-105708034	+	Cons7_yes_Cons100_yes	IGHA1	3554938	201	extended	-0.5	0.0	0.009	POU5F1	0.5	0.0	FOXP1	-0.4	0.0

InGen_OS_2555	exonicAntisense_positiveCorr	chr14:105708378-105708611	+	Cons7_no_Cons100_yes	IGHA1	3554939	233	extended	-0.5	0.0	0.045	CD79B	0.4	0.0	FOXP1	-0.3	0.0
InGen_OS_380	exonicAntisense_positiveCorr	chr2:88860888-88860916	+	Cons7_yes_Cons100_yes	IGKJ4	2493068	28	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3476	exonicAntisense_positiveCorr	chr22:22904993-22905024	-	Cons7_yes_Cons100_yes	IGLJ3	3954509	31	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2193	exonicAntisense_positiveCorr	chr11:72002545-72002574	-	Cons7_no_Cons100_no	IL18BP	3380904	29	extended	0.3	0.0	0.001	DCAF12L2	0.5	0.0	ARHGAP5	-0.4	0.0
InGen_OS_1594	exonicAntisense_positiveCorr	chr7:155297809-155298110	-	Cons7_no_Cons100_no	INSIG1	3081023	301	extended	-0.1	0.0	0.148	CCR7	0.4	0.0	ALDH2	-0.2	0.0
InGen_OS_1508	exonicAntisense_positiveCorr	chr7:2559022-2559132	-	Cons7_no_Cons100_yes	IQCE	3035836	110	full	-0.1	0.0	0.041	NA	NA	NA	ARID2	-0.3	0.0
InGen_OS_1932	exonicAntisense_positiveCorr	chr10:32901371-32901618	+	Cons7_yes_Cons100_yes	ITGB1	3241730	247	extended	0.3	0.0	0.007	KLK2	0.6	0.0	FUS	-0.3	0.0
InGen_OS_1167	exonicAntisense_positiveCorr	chr5:134526642-134526670	-	Cons7_yes_Cons100_yes	JADE2	2876255	28	full	-0.1	0.0	0.000	GLI1	0.3	0.0	TET2	-0.3	0.0
InGen_OS_2927	exonicAntisense_positiveCorr	chr17:46193073-46193231	+	Cons7_yes_Cons100_yes	KANSL1	3723799	158	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2410	exonicAntisense_positiveCorr	chr13:30306921-30306946	+	Cons7_no_Cons100_yes	KATNAL1	3483895	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_384	exonicAntisense_positiveCorr	chr2:98823340-98823368	+	Cons7_no_Cons100_yes	KIAA1211L	2495662	28	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1489	exonicAntisense_positiveCorr	chr7:138981083-138981112	+	Cons7_yes_Cons100_yes	KIAA1549	3026778	29	free	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2918	exonicAntisense_positiveCorr	chr17:41586757-41586793	+	Cons7_yes_Cons100_yes	KRT14	3721379	36	extended	-0.2	0.0	0.578	NA	NA	NA	NA	NA	NA
InGen_OS_2249	exonicAntisense_positiveCorr	chr12:52897340-52897496	+	Cons7_yes_Cons100_yes	KRT8	3415570	156	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2250	exonicAntisense_positiveCorr	chr12:52901981-52902016	+	Cons7_yes_Cons100_yes	KRT8	3415572	35	extended	0.4	0.0	0.038	NA	NA	NA	NA	NA	NA
InGen_OS_2251	exonicAntisense_positiveCorr	chr12:52904879-52905023	+	Cons7_yes_Cons100_yes	KRT8	3415573	144	extended	0.1	0.0	0.186	CD79A	0.6	0.0	SPEN	-0.4	0.0
InGen_OS_30	exonicAntisense_positiveCorr	chr1:30732829-30733664	+	Cons7_no_Cons100_no	LAPTM5	2328141	835	extended	0.3	0.0	0.121	NUP214	0.7	0.0	MALT1	-0.7	0.0
InGen_OS_2416	exonicAntisense_positiveCorr	chr13:46126722-46126925	+	Cons7_no_Cons100_yes	LCP1	3488476	203	extended	0.3	0.0	0.128	CD79B	0.4	0.0	FOXP1	-0.3	0.0
InGen_OS_2162	exonicAntisense_positiveCorr	chr11:18407176-18407576	-	Cons7_yes_Cons100_yes	LDHA	3365432	400	extended	0.4	0.0	0.045	FOXA1	0.4	0.0	JAK2	-0.4	0.0
InGen_OS_695	exonicAntisense_positiveCorr	chr3:45594408-45594475	-	Cons7_no_Cons100_no	LIMD1	2671823	67	extended	-0.1	0.0	0.000	CD79B	0.3	0.0	FOXP1	-0.3	0.0
InGen_OS_3403	exonicAntisense_positiveCorr	chr21:45288074-45288123	-	Cons7_no_Cons100_no	LINC00205	3934901	49	full	-0.2	0.0	0.036	NA	NA	NA	NA	NA	NA
InGen_OS_2710	exonicAntisense_positiveCorr	chr15:32536928-32536985	-	Cons7_no_Cons100_no	LINC02256	3616715	57	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2618	exonicAntisense_positiveCorr	chr14:105093709-105093895	-	Cons7_no_Cons100_no	LINC02298	3581435	186	full	-0.1	0.0	0.000	TAL2	0.3	0.0	NIN	-0.3	0.0
InGen_OS_3824	exonicAntisense_positiveCorr	chr1:919527-919559	+	Cons7_no_Cons100_no	LINC02593	4053405	32	full	-0.1	0.0	0.000	CCR7	0.7	0.0	NA	NA	NA
InGen_OS_1829	exonicAntisense_positiveCorr	chr9:111037943-111038035	+	Cons7_yes_Cons100_yes	LPAR1	3184816	92	full	-0.1	0.0	0.000	CCR7	0.4	0.0	KMT2C	-0.3	0.0
InGen_OS_789	exonicAntisense_positiveCorr	chr3:188880312-188880336	-	Cons7_no_Cons100_no	LPP	2710220	24	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_588	exonicAntisense_positiveCorr	chr2:237763072-237763382	-	Cons7_no_Cons100_yes	LRRFIP1	2605580	310	extended	0.2	0.0	0.000	FNBP1	0.3	0.0	NA	NA	NA
InGen_OS_1771	exonicAntisense_positiveCorr	chr8:143020964-	-	Cons7_no_Cons100_yes	LY6E	3157275	27	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA

		143020991															
lnGen_OS_1842	exonicAntisense_positiveCorr	chr9:125707075-125707105	+	Cons7_no_Cons100_yes	MAPKAP1	3189301	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_32	exonicAntisense_positiveCorr	chr1:32334193-32334583	+	Cons7_yes_Cons100_yes	MARCKSL1	2328901	390	extended	0.5	0.0	0.012	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
lnGen_OS_2873	exonicAntisense_positiveCorr	chr16:89914376-89914405	-	Cons7_no_Cons100_no	MC1R	3705114	29	extended	-0.1	0.0	0.003	NA	NA	NA	NA	NA	NA
lnGen_OS_2874	exonicAntisense_positiveCorr	chr16:89914731-89914760	-	Cons7_no_Cons100_yes	MC1R	3705115	29	extended	-0.1	0.0	0.002	NA	NA	NA	NA	NA	NA
lnGen_OS_762	exonicAntisense_positiveCorr	chr3:151085876-151085918	-	Cons7_yes_Cons100_yes	MED12L	2701002	42	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_763	exonicAntisense_positiveCorr	chr3:151086724-151086771	-	Cons7_yes_Cons100_yes	MED12L	2701005	47	full	-0.1	0.0	0.136	NA	NA	NA	NA	NA	NA
lnGen_OS_1565	exonicAntisense_positiveCorr	chr7:100430174-100430447	-	Cons7_yes_Cons100_yes	MEPCE	3064003	273	extended	-0.1	0.0	0.389	NA	NA	NA	NA	NA	NA
lnGen_OS_2827	exonicAntisense_positiveCorr	chr16:15409288-15409328	-	Cons7_no_Cons100_no	MPV17L	3681922	40	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1121	exonicAntisense_positiveCorr	chr5:69217939-69218023	-	Cons7_no_Cons100_yes	MRPS36	2860595	84	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3659	exonicAntisense_positiveCorr	chrM:4107-4230	-	Cons7_yes_Cons100_yes	MT-ND1	4037705	123	extended	-0.2	0.0	0.001	PTPN13	0.3	0.0	PIK3CB	-0.3	0.0
lnGen_OS_3656	exonicAntisense_positiveCorr	chrM:14206-14477	+	Cons7_yes_Cons100_yes	MT-ND6	4037676	271	ambiguous	-0.1	0.0	0.004	CBLB	0.4	0.0	SRSF3	-0.6	0.0
lnGen_OS_3658	exonicAntisense_positiveCorr	chrM:2772-2851	-	Cons7_yes_Cons100_yes	MT-RNR2	4037702	79	extended	-0.2	0.0	0.000	ACKR3	0.3	0.0	PIK3CB	-0.3	0.0
lnGen_OS_3655	exonicAntisense_positiveCorr	chrM:14674-14741	+	Cons7_yes_Cons100_yes	MT-TE	4037674	67	extended	-0.2	0.0	0.002	MALAT1	0.3	0.0	KEAP1	-0.4	0.0
lnGen_OS_2081	exonicAntisense_positiveCorr	chr11:1017781-1017929	+	Cons7_no_Cons100_yes	MUC6	3316566	148	extended	-0.5	0.0	0.105	NA	NA	NA	NA	NA	NA
lnGen_OS_3520	exonicAntisense_positiveCorr	chrX:3346516-3346546	+	Cons7_no_Cons100_no	MXRA5	3967092	30	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3012	exonicAntisense_positiveCorr	chr17:42536571-42536766	-	Cons7_yes_Cons100_yes	NAGLU	3757948	195	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3814	exonicAntisense_positiveCorr	chr9:137215662-137215688	-	Cons7_no_Cons100_yes	NDOR1	4050533	26	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2183	exonicAntisense_positiveCorr	chr11:65428333-65428757	-	Cons7_no_Cons100_no	NEAT1	3377619	424	extended	0.2	0.0	0.132	ZNRF3	1.0	0.0	CDKN1B	-0.3	0.0
lnGen_OS_2184	exonicAntisense_positiveCorr	chr11:65430392-65431203	-	Cons7_no_Cons100_no	NEAT1	3377621	811	extended	0.2	0.0	0.011	ZNRF3	0.4	0.0	NA	NA	NA
lnGen_OS_2856	exonicAntisense_positiveCorr	chr16:69565988-69566012	-	Cons7_yes_Cons100_yes	NFAT5	3696652	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2993	exonicAntisense_positiveCorr	chr17:28043004-28043033	-	Cons7_yes_Cons100_yes	NLK	3750485	29	full	-0.1	0.0	0.259	NA	NA	NA	NA	NA	NA
lnGen_OS_3026	exonicAntisense_positiveCorr	chr17:56594070-56594095	-	Cons7_yes_Cons100_yes	NOG	3763548	25	full	-0.2	0.0	0.376	NA	NA	NA	NA	NA	NA
lnGen_OS_2637	exonicAntisense_positiveCorr	chr15:34343026-34343066	+	Cons7_yes_Cons100_yes	NOP10	3588128	40	full	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2270	exonicAntisense_positiveCorr	chr12:95790350-95790378	+	Cons7_no_Cons100_yes	NTN4	3427008	28	full	-0.2	0.0	0.000	RGPD3	0.7	0.0	PTEN	-0.3	0.0
lnGen_OS_127	exonicAntisense_positiveCorr	chr1:205717201-205717565	+	Cons7_yes_Cons100_yes	NUCKS1	2376638	364	extended	0.2	0.0	0.008	CDX2	0.5	0.0	TET2	-0.3	0.0
lnGen_OS_128	exonicAntisense_positiveCorr	chr1:205727736-205727798	+	Cons7_yes_Cons100_yes	NUCKS1	2376644	62	full	0.2	0.0	0.000	CDX2	0.5	0.0	TET2	-0.4	0.0
lnGen_OS_1656	exonicAntisense_positiveCorr	chr8:109333814-109333911	+	Cons7_yes_Cons100_yes	NUDCD1	3111538	97	extended	-0.1	0.0	0.000	TAL2	0.3	0.0	NA	NA	NA
lnGen_OS_1527	exonicAntisense_positiveCorr	chr7:44606625-	-	Cons7_yes_Cons100_yes	OGDH	3048797	83	full	-0.1	0.0	0.001	NA	NA	NA	NA	NA	NA

		44606708																
lnGen_OS_2394	exonicAntisense_positiveCorr	chr12:12162664 9-121626744	-	Cons7_yes_Cons100_yes	ORAI1	3475273	95	full	-0.1	0.0	0.146	NA	NA	NA	NA	NA	NA
lnGen_OS_1646	exonicAntisense_positiveCorr	chr8:100709519-100709646	+	Cons7_yes_Cons100_yes	PABPC1	3109395	127	extended	0.3	0.0	0.007	NA	NA	NA	NSD1	-0.3	0.0
lnGen_OS_1116	exonicAntisense_positiveCorr	chr5:50667094-50667167	-	Cons7_yes_Cons100_yes	PARP8	2856127	73	extended	0.3	0.0	0.000	NFKB2	0.4	0.0	NCOA1	-0.3	0.0
lnGen_OS_1185	exonicAntisense_positiveCorr	chr5:141494808-141494838	-	Cons7_yes_Cons100_yes	PCDHGC5	2878647	30	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3364	exonicAntisense_positiveCorr	chr20:64259961-64259989	-	Cons7_yes_Cons100_yes	PCMTD2	3914386	28	extended	-0.1	0.0	0.020	NA	NA	NA	NA	NA	NA
lnGen_OS_1870	exonicAntisense_positiveCorr	chr9:75891252-75891282	-	Cons7_yes_Cons100_yes	PCSK5	3210332	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_326	exonicAntisense_positiveCorr	chr2:10790795-10790819	+	Cons7_yes_Cons100_yes	PDIA6	2469589	24	extended	0.2	0.0	0.000	NUTM2B	0.4	0.0	PIM1	-0.3	0.0
lnGen_OS_441	exonicAntisense_positiveCorr	chr2:238288468-238288500	+	Cons7_no_Cons100_no	PER2	2534806	32	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2885	exonicAntisense_positiveCorr	chr17:4945705-4945907	+	Cons7_yes_Cons100_yes	PFN1	3707382	202	extended	0.1	0.0	0.008	CDH1	0.3	0.0	KLF6	-0.2	0.0
lnGen_OS_2141	exonicAntisense_positiveCorr	chr11:289133-289200	-	Cons7_no_Cons100_no	PGGHG	3357763	67	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3616	exonicAntisense_positiveCorr	chrX:78125853-78125982	-	Cons7_yes_Cons100_yes	PGK1	4013431	129	extended	0.5	0.0	0.008	NUTM2B	0.5	0.0	LIFR	-0.3	0.0
lnGen_OS_3350	exonicAntisense_positiveCorr	chr20:59747735-59747759	-	Cons7_yes_Cons100_yes	PHACTR3	3912056	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3106	exonicAntisense_positiveCorr	chr18:62716549-62716574	-	Cons7_no_Cons100_no	PHLPP1	3811282	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_945	exonicAntisense_positiveCorr	chr4:88008281-88008328	-	Cons7_yes_Cons100_yes	PKD2	2777264	47	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3343	exonicAntisense_positiveCorr	chr20:44531782-44531809	-	Cons7_no_Cons100_yes	PKIG	3906985	27	ambiguous	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2665	exonicAntisense_positiveCorr	chr15:72199217-72199533	+	Cons7_no_Cons100_yes	PKM	3600661	316	extended	0.6	0.0	0.026	KLK2	0.5	0.0	NSD1	-0.3	0.0
lnGen_OS_2666	exonicAntisense_positiveCorr	chr15:72208777-72208806	+	Cons7_yes_Cons100_yes	PKM	3600665	29	extended	0.6	0.0	0.304	KLK2	0.7	0.0	SPEN	-0.3	0.0
lnGen_OS_1682	exonicAntisense_positiveCorr	chr8:143917300-143917477	+	Cons7_yes_Cons100_yes	PLEC	3119928	177	extended	0.2	0.0	0.662	CCR7	0.4	0.0	KMT2C	-0.3	0.0
lnGen_OS_2057	exonicAntisense_positiveCorr	chr10:122374702-122374734	-	Cons7_yes_Cons100_yes	PLEKHA1	3310600	32	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1397	exonicAntisense_positiveCorr	chr6:160718786-160718815	-	Cons7_no_Cons100_yes	PLG	2982890	29	extended	-0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1366	exonicAntisense_positiveCorr	chr6:98835360-98835387	-	Cons7_yes_Cons100_yes	POU3F2	2966062	27	extended	-0.1	0.0	0.093	NA	NA	NA	NA	NA	NA
lnGen_OS_1528	exonicAntisense_positiveCorr	chr7:44801299-44801598	-	Cons7_yes_Cons100_yes	PPIA	3048864	299	extended	0.2	0.0	0.358	RPN1	0.5	0.0	MALAT1	-0.5	0.0
lnGen_OS_2132	exonicAntisense_positiveCorr	chr11:111765938-111766036	+	Cons7_no_Cons100_yes	PPP2R1B	3348727	98	full	-0.1	0.0	0.039	NA	NA	NA	NA	NA	NA
lnGen_OS_819	exonicAntisense_positiveCorr	chr4:16083395-16083453	+	Cons7_no_Cons100_yes	PROM1	2719759	58	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2411	exonicAntisense_positiveCorr	chr13:39037834-39037863	+	Cons7_yes_Cons100_yes	PROSER1	3486191	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_269	exonicAntisense_positiveCorr	chr1:171541577-171542152	-	Cons7_yes_Cons100_yes	PRRC2C	2443925	575	extended	0.2	0.0	0.012	NA	NA	NA	NA	NA	NA
lnGen_OS_3175	exonicAntisense_positiveCorr	chr19:42868754-42868789	+	Cons7_no_Cons100_yes	PSG1	3834986	35	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA

lnGen_OS_250	exonicAntisense_positiveCorr	chr1:151400484-151400574	-	Cons7_yes_Cons100_yes	PSMB4	2435042	90	extended	0.2	0.0	0.464	NA	NA	NA	NA	NA	NA
lnGen_OS_251	exonicAntisense_positiveCorr	chr1:151400780-151400823	-	Cons7_yes_Cons100_yes	PSMB4	2435043	43	extended	0.3	0.0	0.001	NA	NA	NA	NA	NA	NA
lnGen_OS_3229	exonicAntisense_positiveCorr	chr19:38379242-38379275	-	Cons7_yes_Cons100_yes	PSMD8	3861349	33	free	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3013	exonicAntisense_positiveCorr	chr17:42842057-42842245	-	Cons7_no_Cons100_yes	PSME3	3758194	188	extended	0.3	0.0	0.001	NA	NA	NA	NA	NA	NA
lnGen_OS_579	exonicAntisense_positiveCorr	chr2:231708618-231708650	-	Cons7_no_Cons100_yes	PTMA	2603590	32	extended	0.2	0.0	0.000	CDX2	0.5	0.0	TET2	-0.3	0.0
lnGen_OS_580	exonicAntisense_positiveCorr	chr2:231708709-231708750	-	Cons7_yes_Cons100_yes	PTMA	2603591	41	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_581	exonicAntisense_positiveCorr	chr2:231711392-231711419	-	Cons7_yes_Cons100_yes	PTMA	2603593	27	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_582	exonicAntisense_positiveCorr	chr2:231711912-231711943	-	Cons7_yes_Cons100_yes	PTMA	2603594	31	extended	0.3	0.0	0.000	CCR7	0.6	0.0	CIC	-0.3	0.0
lnGen_OS_583	exonicAntisense_positiveCorr	chr2:231712466-231712505	-	Cons7_yes_Cons100_yes	PTMA	2603595	39	extended	0.3	0.0	0.000	BRCA2	0.5	0.0	SKI	-0.4	0.0
lnGen_OS_584	exonicAntisense_positiveCorr	chr2:231712910-231713048	-	Cons7_yes_Cons100_yes	PTMA	2603596	138	extended	0.3	0.0	0.011	FOXA1	0.4	0.0	NSD1	-0.4	0.0
lnGen_OS_2324	exonicAntisense_positiveCorr	chr12:6770621-6770651	-	Cons7_no_Cons100_yes	PTMS	3442308	30	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3386	exonicAntisense_positiveCorr	chr21:44849756-44851211	+	Cons7_no_Cons100_yes	PTTG1IP	3923950	1455	extended	0.1	0.0	0.012	NUTM2B	0.3	0.0	KMT2C	-0.3	0.0
lnGen_OS_1147	exonicAntisense_positiveCorr	chr5:80960915-80960942	-	Cons7_yes_Cons100_yes	RASGRF2	2864692	27	extended	-0.1	0.0	0.002	NA	NA	NA	NA	NA	NA
lnGen_OS_2831	exonicAntisense_positiveCorr	chr16:24540107-24540144	-	Cons7_yes_Cons100_yes	RBBP6	3685539	37	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_910	exonicAntisense_positiveCorr	chr4:26319764-26319799	-	Cons7_yes_Cons100_yes	RBPJ	2764426	35	full	-0.2	0.0	0.000	CD79A	0.6	0.0	NIN	-0.4	0.0
lnGen_OS_3132	exonicAntisense_positiveCorr	chr19:14006988-14007130	+	Cons7_no_Cons100_yes	RFX1	3822538	142	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2692	exonicAntisense_positiveCorr	chr15:93088990-93089015	+	Cons7_no_Cons100_yes	RGMA	3609316	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_615	exonicAntisense_positiveCorr	chr3:49359224-49360052	+	Cons7_yes_Cons100_yes	RHOA	2622081	828	extended	0.1	0.0	0.004	FOXA1	0.4	0.0	JAK2	-0.3	0.0
lnGen_OS_452	exonicAntisense_positiveCorr	chr2:20447521-20447591	-	Cons7_yes_Cons100_yes	RHOB	2542953	70	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_74	exonicAntisense_positiveCorr	chr1:112703005-112703061	+	Cons7_yes_Cons100_yes	RHOC	2352324	56	extended	0.2	0.0	0.045	CCR7	0.3	0.0	CUX1	-0.3	0.0
lnGen_OS_2498	exonicAntisense_positiveCorr	chr14:49862597-49862768	+	Cons7_yes_Cons100_yes	RN7SL2	3534863	171	extended	-0.2	0.0	0.127	FBXW7	0.5	0.0	BRD4	-0.5	0.0
lnGen_OS_3220	exonicAntisense_positiveCorr	chr19:17862920-17862973	-	Cons7_yes_Cons100_yes	RPL18A	3854670	53	extended	0.1	0.0	0.028	NUTM2B	0.4	0.0	NSD1	-0.4	0.0
lnGen_OS_3267	exonicAntisense_positiveCorr	chr19:55386394-55386428	-	Cons7_yes_Cons100_yes	RPL28	3871439	34	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3268	exonicAntisense_positiveCorr	chr19:55386582-55386661	-	Cons7_yes_Cons100_yes	RPL28	3871440	79	extended	0.3	0.0	0.000	KLK2	0.9	0.0	POLE	-0.3	0.0
lnGen_OS_3440	exonicAntisense_positiveCorr	chr22:39315511-39315535	+	Cons7_yes_Cons100_yes	RPL3	3945768	24	extended	-0.2	0.0	0.000	ROBO2	0.3	0.0	NSD1	-0.5	0.0
lnGen_OS_3441	exonicAntisense_positiveCorr	chr22:39317569-39317608	+	Cons7_yes_Cons100_yes	RPL3	3945771	39	extended	-0.2	0.0	0.051	CDX2	0.6	0.0	KMT2C	-0.4	0.0
lnGen_OS_3622	exonicAntisense_positiveCorr	chrX:101391475-101391563	-	Cons7_yes_Cons100_yes	RPL36A	4015757	88	extended	0.2	0.0	0.034	NA	NA	NA	NA	NA	NA
lnGen_OS_3623	exonicAntisense_positiveCorr	chrX:101395368-	-	Cons7_yes_Cons100_yes	RPL36A	4015759	24	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA

		101395392																
lnGen_OS_3549	exonicAntisense_positiveCorr	chrX:119789930-119789993	+	Cons7_yes_Cons100_yes	RPL39	3988971	63	extended	0.2	0.0	0.003	CCR7	0.3	0.0	CUX1	-0.3	0.0
lnGen_OS_1245	exonicAntisense_positiveCorr	chr6:34418370-34418424	+	Cons7_yes_Cons100_yes	RPS10	2904127	54	extended	0.1	0.0	0.000	KLK2	0.4	0.0	NSD1	-0.4	0.0
lnGen_OS_1246	exonicAntisense_positiveCorr	chr6:34425114-34425210	+	Cons7_yes_Cons100_yes	RPS10	2904131	96	extended	0.2	0.0	0.006	KLK2	0.6	0.0	STAT5B	-0.3	0.0
lnGen_OS_3167	exonicAntisense_positiveCorr	chr19:39433523-39433555	+	Cons7_yes_Cons100_yes	RPS16	3833033	32	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3242	exonicAntisense_positiveCorr	chr19:41861118-41861142	-	Cons7_yes_Cons100_yes	RPS19	3863277	24	extended	0.2	0.0	0.000	CNBD1	0.7	0.0	PRDM2	-0.3	0.0
lnGen_OS_2765	exonicAntisense_positiveCorr	chr16:1962737-1962769	+	Cons7_yes_Cons100_yes	RPS2	3644236	32	extended	0.1	0.0	0.000	CDH1	0.3	0.0	NSD1	-0.3	0.0
lnGen_OS_1630	exonicAntisense_positiveCorr	chr8:56074388-56074412	+	Cons7_yes_Cons100_yes	RPS20	3099070	24	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2024	exonicAntisense_positiveCorr	chr10:78035624-78035650	-	Cons7_yes_Cons100_yes	RPS24	3296570	26	extended	0.2	0.0	0.004	CDH1	0.4	0.0	KMT2C	-0.4	0.0
lnGen_OS_2194	exonicAntisense_positiveCorr	chr11:75404772-75404797	-	Cons7_yes_Cons100_yes	RPS3	3382292	25	extended	0.2	0.0	0.000	MNX1	0.3	0.0	NOTCH1	-0.3	0.0
lnGen_OS_3272	exonicAntisense_positiveCorr	chr19:58393379-58393426	-	Cons7_yes_Cons100_yes	RPS5	3872901	47	extended	0.2	0.0	0.281	CACNA1D	0.3	0.0	NSD1	-0.4	0.0
lnGen_OS_692	exonicAntisense_positiveCorr	chr3:39410782-39410812	-	Cons7_yes_Cons100_yes	RPSA	2670014	30	extended	0.2	0.0	0.000	NA	NA	NA	NSD1	-0.5	0.0
lnGen_OS_2261	exonicAntisense_positiveCorr	chr12:63821955-63821984	+	Cons7_yes_Cons100_yes	RXYLT1-AS1	3419630	29	ambiguous	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_980	exonicAntisense_positiveCorr	chr5:9545711-9545775	+	Cons7_no_Cons100_no	SEMA5A	2801328	64	full	-0.1	0.0	0.004	NA	NA	NA	NA	NA	NA
lnGen_OS_2539	exonicAntisense_positiveCorr	chr14:94378458-94378636	+	Cons7_yes_Cons100_yes	SERPINA1	3549673	178	extended	0.3	0.0	0.183	NA	NA	NA	NA	NA	NA
lnGen_OS_2540	exonicAntisense_positiveCorr	chr14:94380887-94381107	+	Cons7_yes_Cons100_yes	SERPINA1	3549675	220	extended	0.2	0.0	0.407	GRM3	0.5	0.0	NA	NA	NA
lnGen_OS_2541	exonicAntisense_positiveCorr	chr14:94382668-94382966	+	Cons7_yes_Cons100_yes	SERPINA1	3549676	298	extended	0.2	0.0	0.328	ESR1	0.3	0.0	NA	NA	NA
lnGen_OS_1954	exonicAntisense_positiveCorr	chr10:79557071-79557188	+	Cons7_no_Cons100_no	SFTPA2	3253936	117	extended	-0.2	0.0	0.021	NA	NA	NA	NA	NA	NA
lnGen_OS_3101	exonicAntisense_positiveCorr	chr18:45838975-45839044	-	Cons7_yes_Cons100_yes	SIGLEC15	3806121	69	full	-0.1	0.0	0.045	CCR7	0.3	0.0	CUX1	-0.3	0.0
lnGen_OS_1455	exonicAntisense_positiveCorr	chr7:96322112-96322179	+	Cons7_yes_Cons100_yes	SLC25A13	3013743	67	full	-0.1	0.0	0.057	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
lnGen_OS_2757	exonicAntisense_positiveCorr	chr15:91854198-91854243	-	Cons7_yes_Cons100_yes	SLCO3A1	3639266	45	full	-0.2	0.0	0.000	C15orf65	0.3	0.0	IL6ST	-0.3	0.0
lnGen_OS_2042	exonicAntisense_positiveCorr	chr10:100913150-100913188	-	Cons7_yes_Cons100_yes	SLF2	3303635	38	extended	-0.1	0.0	0.011	TCL1A	0.3	0.0	NA	NA	NA
lnGen_OS_2470	exonicAntisense_positiveCorr	chr13:87672581-87672645	-	Cons7_no_Cons100_yes	SLITRK5	3519988	64	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1678	exonicAntisense_positiveCorr	chr8:142741174-142741284	+	Cons7_yes_Cons100_yes	SLURP1	3119261	110	full	-0.1	0.0	0.566	NA	NA	NA	NA	NA	NA
lnGen_OS_1495	exonicAntisense_positiveCorr	chr7:151245588-151245675	+	Cons7_yes_Cons100_yes	SMARCD3	3031993	87	full	-0.1	0.0	0.364	KLK2	0.9	0.0	NA	NA	NA
lnGen_OS_1165	exonicAntisense_positiveCorr	chr5:132814176-132814508	-	Cons7_yes_Cons100_yes	SOWAHA	2875488	332	extended	-0.1	0.0	0.089	NA	NA	NA	NA	NA	NA
lnGen_OS_1413	exonicAntisense_positiveCorr	chr7:20784279-20784404	+	Cons7_yes_Cons100_yes	SP8	2992080	125	free	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3337	exonicAntisense_positiveCorr	chr20:35616100-35616185	-	Cons7_yes_Cons100_yes	SPAG4	3904116	85	full	-0.1	0.0	0.000	HOXD11	0.5	0.0	HOOK3	-0.3	0.0

lnGen_OS_1073	exonicAntisense_positiveCorr	chr5:151662562-151663028	+	Cons7_yes_Cons100_yes	SPARC	2835934	466	extended	0.5	0.0	0.002	KLK2	0.7	0.0	SPEN	-0.3	0.0
lnGen_OS_1074	exonicAntisense_positiveCorr	chr5:151669709-151669774	+	Cons7_yes_Cons100_yes	SPARC	2835938	65	extended	0.6	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1075	exonicAntisense_positiveCorr	chr5:151671569-151671642	+	Cons7_yes_Cons100_yes	SPARC	2835939	73	extended	0.7	0.0	0.081	KLK2	0.9	0.0	BCORL1	-0.2	0.0
lnGen_OS_1549	exonicAntisense_positiveCorr	chr7:75493623-75493682	-	Cons7_no_Cons100_no	SPDYE5	3057344	59	extended	0.3	0.0	0.000	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
lnGen_OS_1550	exonicAntisense_positiveCorr	chr7:75494942-75495120	-	Cons7_no_Cons100_no	SPDYE5	3057346	178	extended	0.1	0.0	0.026	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
lnGen_OS_1347	exonicAntisense_positiveCorr	chr6:43171408-43171596	-	Cons7_yes_Cons100_yes	SRF	2954467	188	full	-0.1	0.0	0.000	KLK2	0.8	0.0	NA	NA	NA
lnGen_OS_2811	exonicAntisense_positiveCorr	chr16:2769071-2769098	-	Cons7_yes_Cons100_yes	SRRM2	3677161	27	full	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_376	exonicAntisense_positiveCorr	chr2:85888675-85888835	+	Cons7_yes_Cons100_yes	ST3GAL5	2491846	160	full	-0.1	0.0	0.000	NTRK3	0.4	0.0	NA	NA	NA
lnGen_OS_2034	exonicAntisense_positiveCorr	chr10:88880636-88880673	-	Cons7_no_Cons100_yes	STAMBPL1	3299496	37	full	-0.1	0.0	0.000	GLI1	0.4	0.0	TET2	-0.3	0.0
lnGen_OS_686	exonicAntisense_positiveCorr	chr3:31532735-31532824	-	Cons7_yes_Cons100_yes	STT3B	2667781	89	ambiguous	-0.2	0.0	0.003	NA	NA	NA	NA	NA	NA
lnGen_OS_1744	exonicAntisense_positiveCorr	chr8:69564011-69564100	-	Cons7_yes_Cons100_yes	SULF1	3139551	89	full	0.7	0.0	0.016	KLK2	0.8	0.0	NA	NA	NA
lnGen_OS_424	exonicAntisense_positiveCorr	chr2:202238466-202238492	+	Cons7_yes_Cons100_yes	SUMO1	2523146	26	full	-0.1	0.0	0.000	CDX2	0.4	0.0	JAK2	-0.4	0.0
lnGen_OS_1846	exonicAntisense_positiveCorr	chr9:133362971-133363751	+	Cons7_yes_Cons100_yes	SURF4	3192907	780	extended	0.2	0.0	0.218	CDX2	0.5	0.0	KMT2C	-0.5	0.0
lnGen_OS_3818	exonicAntisense_positiveCorr	chr9:133361964-133362441	+	Cons7_yes_Cons100_yes	SURF4	4050809	477	extended	0.2	0.0	0.036	HMGA2	0.4	0.0	PIK3R1	-0.4	0.0
lnGen_OS_1830	exonicAntisense_positiveCorr	chr9:112175234-112175278	+	Cons7_no_Cons100_no	SUSD1	3185144	44	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1988	exonicAntisense_positiveCorr	chr10:133569713-133569752	+	Cons7_no_Cons100_no	SYCE1	3273018	39	full	-0.1	0.0	0.000	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
lnGen_OS_956	exonicAntisense_positiveCorr	chr4:119030313-119030341	-	Cons7_no_Cons100_yes	SYNPO2	2783423	28	extended	-0.3	0.0	0.000	FNBP1	0.3	0.0	NA	NA	NA
lnGen_OS_957	exonicAntisense_positiveCorr	chr4:119030424-119030865	-	Cons7_yes_Cons100_yes	SYNPO2	2783424	441	extended	-0.3	0.0	0.030	KLK2	0.6	0.0	FUS	-0.3	0.0
lnGen_OS_2356	exonicAntisense_positiveCorr	chr12:49763989-49764470	-	Cons7_yes_Cons100_yes	TMBIM6	3454087	481	extended	-0.2	0.0	0.047	MNX1	0.3	0.0	NA	NA	NA
lnGen_OS_806	exonicAntisense_positiveCorr	chr4:1721093-1721130	+	Cons7_no_Cons100_no	TMEM129	2714956	37	full	-0.1	0.0	0.000	TAL2	0.5	0.0	SPECC1	-0.3	0.0
lnGen_OS_1127	exonicAntisense_positiveCorr	chr5:72816707-72816732	-	Cons7_yes_Cons100_yes	TNPO1	2862138	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1128	exonicAntisense_positiveCorr	chr5:72816747-72816775	-	Cons7_yes_Cons100_yes	TNPO1	2862139	28	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3247	exonicAntisense_positiveCorr	chr19:44901059-44901104	-	Cons7_yes_Cons100_yes	TOMM40	3865063	45	free	0.5	0.0	0.001	NKX2-1	0.6	0.0	NA	NA	NA
lnGen_OS_142	exonicAntisense_positiveCorr	chr1:223845689-223845719	+	Cons7_yes_Cons100_no	TP53BP2	2382223	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2325	exonicAntisense_positiveCorr	chr12:6869289-6869390	-	Cons7_yes_Cons100_yes	TPI1	3442342	101	extended	0.4	0.0	0.035	CDX2	0.5	0.0	KMT2C	-0.4	0.0
lnGen_OS_3217	exonicAntisense_positiveCorr	chr19:16101654-16102696	-	Cons7_yes_Cons100_yes	TPM4	3853749	1042	extended	0.3	0.0	0.073	KLK2	0.5	0.0	KMT2C	-0.4	0.0
lnGen_OS_1581	exonicAntisense_positiveCorr	chr7:138554723-138554860	-	Cons7_yes_Cons100_yes	TRIM24	3075312	137	free	0.2	0.0	0.012	NA	NA	NA	NA	NA	NA
lnGen_OS_1475	exonicAntisense_positiveCorr	chr7:120857837-	+	Cons7_no_Cons100_yes	TSPAN12	3021094	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA

		120857861															
lnGen_OS_2646	exonicAntisense_positiveCorr	chr15:42920543-42920576	+	Cons7_yes_Cons100_yes	TTBK2	3591242	33	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3397	exonicAntisense_positiveCorr	chr21:37073325-37073359	-	Cons7_yes_Cons100_yes	TTC3	3931290	34	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_13	exonicAntisense_positiveCorr	chr1:2789591-2789623	+	Cons7_yes_Cons100_yes	TTC34	2316799	32	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3644	exonicAntisense_positiveCorr	chrY:19076944-19077009	+	Cons7_no_Cons100_no	TTTY14	4031005	65	full	-0.1	0.0	0.161	NA	NA	NA	NA	NA	NA
lnGen_OS_2247	exonicAntisense_positiveCorr	chr12:49185601-49185625	+	Cons7_yes_Cons100_yes	TUBA1A	3413800	24	extended	0.3	0.0	0.000	APC	0.4	0.0	MSN	-0.4	0.0
lnGen_OS_1331	exonicAntisense_positiveCorr	chr6:30720400-30720424	-	Cons7_no_Cons100_yes	TUBB	2948588	24	extended	0.2	0.0	0.030	CDX2	0.4	0.0	TET2	-0.3	0.0
lnGen_OS_1332	exonicAntisense_positiveCorr	chr6:30724657-30725198	-	Cons7_yes_Cons100_yes	TUBB	2948594	541	extended	0.3	0.0	0.002	FOXA1	0.5	0.0	NSD1	-0.4	0.0
lnGen_OS_3819	exonicAntisense_positiveCorr	chr9:137242699-137242755	-	Cons7_yes_Cons100_yes	TUBB4B	4051520	56	full	0.2	0.0	0.843	NA	NA	NA	NA	NA	NA
lnGen_OS_3488	exonicAntisense_positiveCorr	chr22:30969261-30969312	-	Cons7_yes_Cons100_yes	TUG1	3957636	51	full	-0.1	0.0	0.263	CD79A	0.6	0.0	NIN	-0.4	0.0
lnGen_OS_1791	exonicAntisense_positiveCorr	chr9:25678142-25678279	+	Cons7_yes_Cons100_yes	TUSC1	3165567	137	extended	-0.1	0.0	0.000	CRTC1	0.4	0.0	FANCD2	-0.3	0.0
lnGen_OS_245	exonicAntisense_positiveCorr	chr1:145995441-145995469	+	Cons7_yes_Cons100_yes	TXNIP	2432568	28	extended	-0.3	0.0	0.501	FLNA	0.3	0.0	NA	NA	NA
lnGen_OS_246	exonicAntisense_positiveCorr	chr1:145992835-145994763	+	Cons7_yes_Cons100_yes	TXNIP	2432570	1929	extended	-0.3	0.0	0.249	KLK2	0.4	0.0	NA	NA	NA
lnGen_OS_2378	exonicAntisense_positiveCorr	chr12:104287178-104287204	-	Cons7_no_Cons100_no	TXNRD1	3468968	26	full	-0.1	0.0	0.236	NA	NA	NA	NA	NA	NA
lnGen_OS_2379	exonicAntisense_positiveCorr	chr12:104288921-104289006	-	Cons7_no_Cons100_yes	TXNRD1	3468969	85	full	0.2	0.0	0.000	GPC5	0.3	0.0	NA	NA	NA
lnGen_OS_2380	exonicAntisense_positiveCorr	chr12:104311368-104311407	-	Cons7_yes_Cons100_yes	TXNRD1	3468972	39	full	0.2	0.0	0.025	CNBD1	0.6	0.0	PRDM2	-0.3	0.0
lnGen_OS_842	exonicAntisense_positiveCorr	chr4:67700986-67701152	+	Cons7_no_Cons100_yes	UBA6	2729707	166	full	-0.1	0.0	0.067	NA	NA	NA	NA	NA	NA
lnGen_OS_858	exonicAntisense_positiveCorr	chr4:102827535-102827630	+	Cons7_yes_Cons100_yes	UBE2D3	2737872	95	ambiguous	-0.1	0.0	0.000	KLK2	0.9	0.0	POLE	-0.3	0.0
lnGen_OS_3411	exonicAntisense_positiveCorr	chr22:19479121-19479146	+	Cons7_no_Cons100_no	UFD1	3936915	25	full	-0.1	0.0	0.006	NA	NA	NA	NA	NA	NA
lnGen_OS_2902	exonicAntisense_positiveCorr	chr17:19867610-19867635	+	Cons7_yes_Cons100_yes	ULK2	3714149	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3221	exonicAntisense_positiveCorr	chr19:18832118-18832185	-	Cons7_no_Cons100_yes	UPF1	3855252	67	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2485	exonicAntisense_positiveCorr	chr13:114301809-114301833	-	Cons7_no_Cons100_yes	UPF3A	3527014	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1350	exonicAntisense_positiveCorr	chr6:43785457-43785544	-	Cons7_no_Cons100_yes	VEGFA	2954892	87	extended	0.2	0.0	0.018	NA	NA	NA	NA	NA	NA
lnGen_OS_3031	exonicAntisense_positiveCorr	chr17:59839857-59840175	-	Cons7_yes_Cons100_yes	VMP1	3764922	318	extended	0.4	0.0	0.001	KLK2	0.4	0.0	PIM1	-0.3	0.0
lnGen_OS_697	exonicAntisense_positiveCorr	chr3:49015410-49015481	-	Cons7_no_Cons100_no	WDR6	2673832	71	extended	-0.1	0.0	0.057	DCAF12L2	0.6	0.0	ARHGAP5	-0.4	0.0
lnGen_OS_1649	exonicAntisense_positiveCorr	chr8:100920102-100920354	+	Cons7_yes_Cons100_yes	YWHAZ	3109448	252	extended	0.3	0.0	0.001	CTNNB1	0.4	0.0	NA	NA	NA
lnGen_OS_2368	exonicAntisense_positiveCorr	chr12:56118273-56118358	-	Cons7_yes_Cons100_yes	ZC3H10	3457436	85	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3158	exonicAntisense_positiveCorr	chr19:36418593-36418619	+	Cons7_no_Cons100_no	ZFP82	3831374	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA

InGen_OS_2227	exonicAntisense_positiveCorr	chr12:6673369-6673438	+	Cons7_yes_Cons100_yes	ZNF384	3402691	69	extended	0.3	0.0	0.031	NA	NA	NA	NA	NA	NA
InGen_OS_3161	exonicAntisense_positiveCorr	chr19:37793380-37793477	+	Cons7_no_Cons100_no	ZNF573	3832065	97	full	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_339	intronicAntisense_negativeCorr	chr2:38481810-38481835	+	Cons7_yes_Cons100_yes	AC011247.1	2477854	25	extended	0.4	0.0	0.000	MDS2	0.6	0.0	GNA11	-0.3	0.0
InGen_OS_1885	intronicAntisense_negativeCorr	chr9:94346542-94346611	-	Cons7_yes_Cons100_yes	AL691447.2	3215503	69	full	0.4	0.0	0.120	NA	NA	NA	NA	NA	NA
InGen_OS_2065	intronicAntisense_negativeCorr	chr10:12492704-124927068	-	Cons7_no_Cons100_no	AL731577.2	3311398	26	extended	-0.1	0.0	0.000	ZNRF3	0.9	0.0	CDKN1B	-0.3	0.0
InGen_OS_946	intronicAntisense_negativeCorr	chr4:90838564-90838630	-	Cons7_yes_Cons100_yes	CCSER1	2777882	66	core	0.2	0.0	0.013	CDX2	0.3	0.0	MED12	-0.3	0.0
InGen_OS_947	intronicAntisense_negativeCorr	chr4:90838634-90838705	-	Cons7_yes_Cons100_yes	CCSER1	2777883	71	core	0.2	0.0	0.054	TAL2	0.6	0.0	BTG1	-0.3	0.0
InGen_OS_2843	intronicAntisense_negativeCorr	chr16:53130570-53130603	-	Cons7_yes_Cons100_yes	CHD9	3691891	33	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_360	intronicAntisense_negativeCorr	chr2:47837223-47837254	+	Cons7_no_Cons100_no	FBXO11	2481192	31	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_904	intronicAntisense_negativeCorr	chr4:22726442-22726476	-	Cons7_no_Cons100_yes	GBA3	2763403	34	extended	0.2	0.0	0.000	ID3	0.4	0.0	PTPRB	-0.3	0.0
InGen_OS_2812	intronicAntisense_negativeCorr	chr16:4508967-4509058	-	Cons7_no_Cons100_no	HMOX2	3678189	91	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2125	intronicAntisense_negativeCorr	chr11:82689799-82689825	+	Cons7_no_Cons100_no	MIR4300HG	3342377	26	core	0.1	0.0	0.015	PIM1	0.6	0.0	ARHGAP26	-0.5	0.0
InGen_OS_143	intronicAntisense_negativeCorr	chr1:228208050-228208108	+	Cons7_yes_Cons100_yes	OBSCN-AS1	2383998	58	full	0.1	0.0	0.000	PDE4DIP	0.6	0.0	EML4	-0.5	0.0
InGen_OS_1968	intronicAntisense_negativeCorr	chr10:103628620-103628648	+	Cons7_no_Cons100_no	SH3PXD2A	3262341	28	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1137	intronicAntisense_negativeCorr	chr5:76241991-76242022	-	Cons7_yes_Cons100_yes	SV2C	2863245	31	extended	0.3	0.0	0.000	ECT2L	0.5	0.0	NA	NA	NA
InGen_OS_383	intronicAntisense_negativeCorr	chr2:97827977-97828181	+	Cons7_yes_Cons100_yes	TMEM131	2495244	204	extended	0.3	0.0	0.001	CHEK2	0.7	0.0	PIK3R1	-0.5	0.0
InGen_OS_2352	intronicAntisense_positiveCorr	chr12:46383718-46383742	-	Cons7_no_Cons100_no	AC008014.1	3452370	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2850	intronicAntisense_positiveCorr	chr16:65277300-65277351	-	Cons7_no_Cons100_no	AC009055.1	3694820	51	extended	0.2	0.0	0.010	NA	NA	NA	NA	NA	NA
InGen_OS_1102	intronicAntisense_positiveCorr	chr5:1596632-1596658	-	Cons7_no_Cons100_yes	AC026412.3	2846090	26	extended	-0.1	0.0	0.000	ZNRF3	0.5	0.0	B2M	-0.3	0.0
InGen_OS_593	intronicAntisense_positiveCorr	chr3:8580363-8580433	+	Cons7_no_Cons100_no	AC034187.1	2609423	70	extended	-0.1	0.0	0.000	NA	NA	NA	FANCC	-0.3	0.0
OutGen_Conc_Miss_106	intronicAntisense_positiveCorr	chr8:60652038-60652115	+	Cons7_yes_Cons100_yes	AC068389.1	3100242	77	extended	-0.2	0.0	0.004	NA	NA	NA	NA	NA	NA
InGen_OS_1609	intronicAntisense_positiveCorr	chr8:12601645-12601674	+	Cons7_no_Cons100_no	AC068587.4	3086662	29	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_58	intronicAntisense_positiveCorr	chr1:85577222-85577328	+	Cons7_no_Cons100_no	AC092807.3	2344884	106	full	-0.1	0.0	0.001	NA	NA	NA	QKI	-0.3	0.0
InGen_OS_1613	intronicAntisense_positiveCorr	chr8:22258325-22258364	+	Cons7_no_Cons100_no	AC105206.2	3089311	39	extended	-0.1	0.0	0.016	ZNRF3	1.0	0.0	NA	NA	NA
InGen_OS_3043	intronicAntisense_positiveCorr	chr17:81364944-81365012	-	Cons7_no_Cons100_no	AC110285.3	3773841	68	full	-0.1	0.0	0.033	NA	NA	NA	NA	NA	NA
InGen_OS_385	intronicAntisense_positiveCorr	chr2:99827909-99827933	+	Cons7_no_Cons100_no	AFF3	2496066	24	full	-0.3	0.0	0.071	NA	NA	NA	NA	NA	NA
InGen_OS_1671	intronicAntisense_positiveCorr	chr8:140598134-140598162	+	Cons7_no_Cons100_no	AGO2	3118487	28	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1351	intronicAntisense_positiveCorr	chr6:44075913-	-	Cons7_no_Cons100_no	AL109615.2	2954982	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA

		44075938																
lnGen_OS_1352	intronicAntisense_positiveCorr	chr6:44076655-44076713	-	Cons7_yes_Cons100_yes	AL109615.2	2954983	58	full	-0.1	0.0	0.066	NA	NA	NA	NA	NA	NA
lnGen_OS_3551	intronicAntisense_positiveCorr	chrX:123960083-123960107	+	Cons7_no_Cons100_yes	AL121601.1	3989716	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_302	intronicAntisense_positiveCorr	chr1:234714415-234714444	-	Cons7_no_Cons100_no	AL160408.6	2461583	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1509	intronicAntisense_positiveCorr	chr7:2718740-2718823	-	Cons7_no_Cons100_no	AMZ1	3035898	83	extended	-0.1	0.0	0.000	DCAF12L2	0.6	0.0	NA	NA	NA
lnGen_OS_475	intronicAntisense_positiveCorr	chr2:69741725-69741792	-	Cons7_no_Cons100_no	ANXA4	2558276	67	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2202	intronicAntisense_positiveCorr	chr11:95162802-95162831	-	Cons7_no_Cons100_no	AP000787.1	3387237	29	full	-0.1	0.0	0.006	NA	NA	NA	NA	NA	NA
lnGen_OS_1021	intronicAntisense_positiveCorr	chr5:78984831-78984865	+	Cons7_no_Cons100_no	ARSB	2817200	34	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1148	intronicAntisense_positiveCorr	chr5:82009435-82009538	-	Cons7_yes_Cons100_yes	ATG10	2865014	103	extended	0.5	0.0	0.001	MYOD1	0.4	0.0	NA	NA	NA
lnGen_OS_2108	intronicAntisense_positiveCorr	chr11:63670929-63671153	+	Cons7_no_Cons100_no	ATL3	3333937	224	full	-0.2	0.0	0.000	TP63	0.3	0.0	ABI1	-0.3	0.0
lnGen_OS_1652	intronicAntisense_positiveCorr	chr8:102863695-102863806	+	Cons7_yes_Cons100_yes	AZIN1	3110073	111	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3535	intronicAntisense_positiveCorr	chrX:40108585-40108618	+	Cons7_no_Cons100_yes	BCOR	3974427	33	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_870	intronicAntisense_positiveCorr	chr4:158671431-158671463	+	Cons7_no_Cons100_no	C4orf46	2749557	32	free	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2800	intronicAntisense_positiveCorr	chr16:88953283-88953333	+	Cons7_no_Cons100_no	CBFA2T3	3673776	50	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_731	intronicAntisense_positiveCorr	chr3:112333539-112333582	-	Cons7_no_Cons100_no	CD200	2688694	43	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_82	intronicAntisense_positiveCorr	chr1:116530494-116530519	+	Cons7_no_Cons100_no	CD58	2353582	25	extended	-0.1	0.0	0.000	ALK	0.5	0.0	NA	NA	NA
lnGen_OS_3352	intronicAntisense_positiveCorr	chr20:61433961-61434015	-	Cons7_no_Cons100_no	CDH4	3912456	54	extended	-0.1	0.0	0.323	NA	NA	NA	NA	NA	NA
lnGen_OS_2750	intronicAntisense_positiveCorr	chr15:80791807-80791837	-	Cons7_no_Cons100_no	CEMIP	3635419	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1806	intronicAntisense_positiveCorr	chr9:71769559-71769583	+	Cons7_no_Cons100_yes	CEMIP2	3174412	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2381	intronicAntisense_positiveCorr	chr12:104457973-104458151	-	Cons7_no_Cons100_yes	CHST11	3469027	178	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2687	intronicAntisense_positiveCorr	chr15:90232848-90232877	+	Cons7_no_Cons100_no	CIB1	3607986	29	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_70	intronicAntisense_positiveCorr	chr1:102933762-102933834	+	Cons7_no_Cons100_no	COL11A1	2349303	72	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_813	intronicAntisense_positiveCorr	chr4:5822680-5822826	+	Cons7_no_Cons100_no	CRMP1	2716864	146	full	-0.2	0.0	0.153	HMGN2P46	0.4	0.0	STAT5B	-0.3	0.0
lnGen_OS_1612	intronicAntisense_positiveCorr	chr8:19755940-19755972	+	Cons7_no_Cons100_yes	CSGALNACT1	3088402	32	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_802	intronicAntisense_positiveCorr	chr4:1245178-1245215	+	Cons7_no_Cons100_no	CTBP1	2714635	37	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2213	intronicAntisense_positiveCorr	chr11:126316701-126316816	-	Cons7_no_Cons100_no	DCPS	3396980	115	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_19	intronicAntisense_positiveCorr	chr1:12608382-12608421	+	Cons7_no_Cons100_no	DHRS3	2320948	39	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_547	intronicAntisense_positiveCorr	chr2:172106450-172106487	-	Cons7_no_Cons100_no	DLX2-DT	2587018	37	full	-0.1	0.0	0.151	NA	NA	NA	NA	NA	NA

InGen_OS_1313	intronicAntisense_positiveCorr	chr6:7542248-7542274	-	Cons7_no_Cons100_no	DSP	2940699	26	full	-0.1	0.0	0.000	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
InGen_OS_1040	intronicAntisense_positiveCorr	chr5:118974287-118974313	+	Cons7_no_Cons100_yes	DTWD2	2825510	26	extended	0.4	0.0	0.002	CDX2	0.7	0.0	KMT2C	-0.3	0.0
InGen_OS_1325	intronicAntisense_positiveCorr	chr6:20403895-20403924	-	Cons7_yes_Cons100_yes	E2F3	2944607	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3181	intronicAntisense_positiveCorr	chr19:45642877-45642902	+	Cons7_no_Cons100_no	EML2	3836396	25	extended	-0.1	0.0	0.000	MUC16	0.5	0.0	DICER1	-0.4	0.0
InGen_OS_2236	intronicAntisense_positiveCorr	chr12:15788981-15789008	+	Cons7_yes_Cons100_yes	EPS8	3406436	27	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_421	intronicAntisense_positiveCorr	chr2:201064344-201064380	+	Cons7_yes_Cons100_yes	FAM126B	2522594	36	extended	0.3	0.0	0.003	CCR7	0.5	0.0	MALT1	-0.4	0.0
InGen_OS_2630	intronicAntisense_positiveCorr	chr15:29570155-29570191	+	Cons7_no_Cons100_no	FAM189A1	3586071	36	full	-0.1	0.0	0.008	NA	NA	NA	NA	NA	NA
InGen_OS_3510	intronicAntisense_positiveCorr	chr22:48490324-48490397	-	Cons7_no_Cons100_no	FAM19A5	3964625	73	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1796	intronicAntisense_positiveCorr	chr9:37514753-37514843	+	Cons7_no_Cons100_no	FBXO10	3168970	90	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3555	intronicAntisense_positiveCorr	chrX:131792966-131792991	+	Cons7_no_Cons100_yes	FIRRE	3991073	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_220	intronicAntisense_positiveCorr	chr1:93448376-93448602	-	Cons7_no_Cons100_yes	FNBP1L	2423411	226	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_970	intronicAntisense_positiveCorr	chr4:173169910-173169950	-	Cons7_no_Cons100_no	GALNT7	2793902	40	full	-0.1	0.0	0.012	NA	NA	NA	NA	NA	NA
InGen_OS_3135	intronicAntisense_positiveCorr	chr19:14489644-14489684	+	Cons7_yes_Cons100_no	GIPC1	3822797	40	extended	0.3	0.0	0.000	CCR7	0.6	0.0	NA	NA	NA
InGen_OS_3625	intronicAntisense_positiveCorr	chrX:102712733-102712776	-	Cons7_no_Cons100_no	GPRASP2	4016245	43	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2711	intronicAntisense_positiveCorr	chr15:32719026-32719115	-	Cons7_no_Cons100_no	GREM1	3616773	89	full	0.1	0.0	0.000	KLK2	0.7	0.0	NA	NA	NA
InGen_OS_1757	intronicAntisense_positiveCorr	chr8:101493644-101493796	-	Cons7_no_Cons100_yes	GRHL2	3147130	152	full	-0.1	0.0	0.000	RGPD3	0.5	0.0	NA	NA	NA
InGen_OS_36	intronicAntisense_positiveCorr	chr1:36956542-36956576	+	Cons7_no_Cons100_no	GRIK3	2330595	34	full	-0.1	0.0	0.001	GLI1	0.4	0.0	TET2	-0.3	0.0
InGen_OS_1772	intronicAntisense_positiveCorr	chr8:143557322-143557368	-	Cons7_no_Cons100_no	GSDMD	3157527	46	extended	0.3	0.0	0.000	NF2	0.5	0.0	NCOA1	-0.5	0.0
InGen_OS_1045	intronicAntisense_positiveCorr	chr5:135398670-135398698	+	Cons7_no_Cons100_yes	H2AFY	2829764	28	full	-0.1	0.0	0.000	CCR7	0.3	0.0	CUX1	-0.3	0.0
InGen_OS_1327	intronicAntisense_positiveCorr	chr6:26125905-26125932	-	Cons7_no_Cons100_no	HIST1H2AC	2946286	27	extended	-0.1	0.0	0.000	DCAF12L2	0.5	0.0	NCOA2	-0.3	0.0
InGen_OS_1704	intronicAntisense_positiveCorr	chr8:29054775-29054800	-	Cons7_no_Cons100_no	HMBOX1	3129566	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3341	intronicAntisense_positiveCorr	chr20:44360306-44360484	-	Cons7_no_Cons100_no	HNF4A	3906893	178	extended	-0.4	0.0	0.002	ATP1A1	0.3	0.0	HIP1	-0.3	0.0
InGen_OS_1518	intronicAntisense_positiveCorr	chr7:27097034-27097063	-	Cons7_yes_Cons100_yes	HOTAIRM1	3042747	29	full	-0.2	0.0	0.020	NA	NA	NA	NA	NA	NA
InGen_OS_1416	intronicAntisense_positiveCorr	chr7:27174488-27174516	+	Cons7_yes_Cons100_yes	HOXA10	2994130	28	full	-0.1	0.0	0.000	CD79A	0.6	0.0	NIN	-0.4	0.0
InGen_OS_3022	intronicAntisense_positiveCorr	chr17:48633809-48633865	-	Cons7_yes_Cons100_yes	HOXB-AS4	3761439	56	extended	0.2	0.0	0.000	NUTM1	0.5	0.0	HOOK3	-0.4	0.0
InGen_OS_659	intronicAntisense_positiveCorr	chr3:160100763-160100794	+	Cons7_yes_Cons100_no	IL12A-AS1	2650116	31	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_622	intronicAntisense_positiveCorr	chr3:57164868-57164899	+	Cons7_no_Cons100_no	IL17RD	2625597	31	extended	-0.2	0.0	0.000	TAL2	0.6	0.0	NIN	-0.3	0.0
InGen_OS_996	intronicAntisense_positiveCorr	chr5:55944975-	+	Cons7_no_Cons100_no	IL6ST	2810100	27	extended	0.3	0.0	0.000	TPM4	0.4	0.0	MALAT1	-0.4	0.0

		55945002															
InGen_OS_2220	intronicAntisense_positiveCorr	chr12:2812532-2812561	+	Cons7_no_Cons100_yes	ITFG2-AS1	3401125	29	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1931	intronicAntisense_positiveCorr	chr10:30058942-30058991	+	Cons7_no_Cons100_yes	JCAD	3240896	49	full	-0.1	0.0	0.000	NA	NA	NA	PTEN	-0.3	0.0
InGen_OS_3114	intronicAntisense_positiveCorr	chr19:2264497-2264521	+	Cons7_no_Cons100_no	JSRP1	3816382	24	extended	-0.1	0.0	0.000	MUC16	0.5	0.0	DICER1	-0.3	0.0
InGen_OS_2530	intronicAntisense_positiveCorr	chr14:88323891-88323917	+	Cons7_no_Cons100_yes	KCNK10	3547518	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_456	intronicAntisense_positiveCorr	chr2:27444326-27444354	-	Cons7_no_Cons100_no	KRTCAP3	2545870	28	extended	-0.2	0.0	0.000	ZNRF3	1.0	0.0	CDKN1B	-0.3	0.0
InGen_OS_3087	intronicAntisense_positiveCorr	chr18:13494104-13494130	-	Cons7_no_Cons100_yes	LDLRAD4	3799977	26	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3088	intronicAntisense_positiveCorr	chr18:13499471-13499541	-	Cons7_no_Cons100_no	LDLRAD4	3799982	70	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_126	intronicAntisense_positiveCorr	chr1:205455712-205455748	+	Cons7_no_Cons100_yes	LEMD1	2376459	36	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2471	intronicAntisense_positiveCorr	chr13:89279316-89279346	-	Cons7_no_Cons100_no	LINC00440	3520153	30	full	-0.1	0.0	0.006	NA	NA	NA	NA	NA	NA
InGen_OS_1023	intronicAntisense_positiveCorr	chr5:88661164-88661213	+	Cons7_no_Cons100_yes	LINC00461	2819340	49	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_827	intronicAntisense_positiveCorr	chr4:41878858-41878893	+	Cons7_yes_Cons100_yes	LINC00682	2725310	35	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1083	intronicAntisense_positiveCorr	chr5:173735299-173735335	+	Cons7_no_Cons100_no	LINC01484	2841679	36	full	-0.1	0.0	0.332	NA	NA	NA	NA	NA	NA
InGen_OS_1810	intronicAntisense_positiveCorr	chr9:90414788-90415116	+	Cons7_no_Cons100_no	LINC01508	3178885	328	extended	0.1	0.0	0.003	CDX2	0.5	0.0	TET2	-0.3	0.0
InGen_OS_1619	intronicAntisense_positiveCorr	chr8:37547018-37547072	+	Cons7_yes_Cons100_yes	LINC01605	3094062	54	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_997	intronicAntisense_positiveCorr	chr5:56480940-56480967	+	Cons7_no_Cons100_no	LINC01948	2810232	27	full	-0.1	0.0	0.000	GLI1	0.4	0.0	TET2	-0.3	0.0
InGen_OS_2731	intronicAntisense_positiveCorr	chr15:58540922-58541057	-	Cons7_no_Cons100_no	LIPC	3626543	135	extended	-0.1	0.0	0.001	NA	NA	NA	NA	NA	NA
InGen_OS_992	intronicAntisense_positiveCorr	chr5:36151220-36151247	+	Cons7_no_Cons100_no	LMBRD2	2806550	27	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_537	intronicAntisense_positiveCorr	chr2:149039425-149039483	-	Cons7_no_Cons100_no	LYPD6B	2580534	58	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3250	intronicAntisense_positiveCorr	chr19:45160151-45160184	-	Cons7_no_Cons100_no	MARK4	3865188	33	full	-0.1	0.0	0.000	RALGDS	0.4	0.0	NA	NA	NA
InGen_OS_3251	intronicAntisense_positiveCorr	chr19:45162454-45162488	-	Cons7_no_Cons100_no	MARK4	3865195	34	full	-0.1	0.0	0.000	RALGDS	0.3	0.0	NA	NA	NA
InGen_OS_1271	intronicAntisense_positiveCorr	chr6:89721213-89721327	+	Cons7_no_Cons100_no	MDN1	2916930	114	extended	0.5	0.0	0.000	SLC34A2	0.4	0.0	NA	NA	NA
InGen_OS_1272	intronicAntisense_positiveCorr	chr6:89721586-89721616	+	Cons7_no_Cons100_no	MDN1	2916934	30	extended	0.4	0.0	0.016	MYB	0.5	0.0	MALT1	-0.4	0.0
InGen_OS_1273	intronicAntisense_positiveCorr	chr6:89721713-89721769	+	Cons7_no_Cons100_no	MDN1	2916935	56	extended	0.4	0.0	0.016	NA	NA	NA	NA	NA	NA
InGen_OS_473	intronicAntisense_positiveCorr	chr2:66446005-66446111	-	Cons7_no_Cons100_no	MEIS1	2557143	106	full	-0.1	0.0	0.000	TCL1A	0.6	0.0	NA	NA	NA
InGen_OS_3055	intronicAntisense_positiveCorr	chr17:83089926-83089952	-	Cons7_no_Cons100_no	METRNL	3775572	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_482	intronicAntisense_positiveCorr	chr2:60353181-60353205	+	Cons7_no_Cons100_no	MIR4432HG	2484075	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3509	intronicAntisense_positiveCorr	chr22:46104243-46104272	-	Cons7_no_Cons100_no	MIRLET7BHG	3963958	29	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA

InGen_OS_11	intronicAntisense_positiveCorr	chr1:2330370-2330424	+	Cons7_no_Cons100_no	MORN1	2316525	54	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2857	intronicAntisense_positiveCorr	chr16:69566808-69566855	-	Cons7_no_Cons100_yes	NFAT5	3696655	47	full	-0.1	0.0	0.122	NA	NA	NA	ABI1	-0.3	0.0
InGen_OS_3303	intronicAntisense_positiveCorr	chr20:63266732-63266956	+	Cons7_no_Cons100_no	NKAIN4	3893286	224	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_115	intronicAntisense_positiveCorr	chr1:183258109-183258135	+	Cons7_no_Cons100_no	NMNAT2	2371191	26	extended	0.3	0.0	0.130	CNBD1	0.7	0.0	PRDM2	-0.4	0.0
InGen_OS_1855	intronicAntisense_positiveCorr	chr9:136527721-136527759	+	Cons7_no_Cons100_no	NOTCH1	3194411	38	full	-0.1	0.0	0.000	NUTM1	0.6	0.0	HOOK3	-0.4	0.0
InGen_OS_387	intronicAntisense_positiveCorr	chr2:104852565-104852607	+	Cons7_yes_Cons100_yes	PANTR1	2497809	42	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_445	intronicAntisense_positiveCorr	chr2:241149080-241149243	+	Cons7_no_Cons100_no	PASK	2536182	163	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3618	intronicAntisense_positiveCorr	chrX:92100351-92100480	-	Cons7_no_Cons100_no	PCDH11X	4014653	129	extended	-0.2	0.0	0.000	CDX2	0.4	0.0	TET2	-0.4	0.0
InGen_OS_2698	intronicAntisense_positiveCorr	chr15:10148881 6-101488916	+	Cons7_no_Cons100_no	PCSK6	3611945	100	full	-0.2	0.0	0.044	NA	NA	NA	ACVR1	-0.3	0.0
InGen_OS_884	intronicAntisense_positiveCorr	chr4:185535161-185535190	+	Cons7_no_Cons100_no	PDLIM3	2754738	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1834	intronicAntisense_positiveCorr	chr9:120877494-120877537	+	Cons7_no_Cons100_no	PHF19	3187543	43	extended	-0.1	0.0	0.000	GLI1	0.4	0.0	TET2	-0.4	0.0
InGen_OS_162	intronicAntisense_positiveCorr	chr1:2455855-2455884	-	Cons7_no_Cons100_yes	PLCH2	2392501	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1479	intronicAntisense_positiveCorr	chr7:132424112-132424137	+	Cons7_no_Cons100_no	PLXNA4	3024728	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_363	intronicAntisense_positiveCorr	chr2:55693126-55693150	+	Cons7_yes_Cons100_yes	PNPT1	2483096	24	full	0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3444	intronicAntisense_positiveCorr	chr22:42601263-42601354	+	Cons7_no_Cons100_no	POLDIP3	3947500	91	extended	-0.1	0.0	0.000	KLK2	0.7	0.0	PER1	-0.4	0.0
InGen_OS_166	intronicAntisense_positiveCorr	chr1:3321071-3321099	-	Cons7_no_Cons100_no	PRDM16	2393183	28	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_167	intronicAntisense_positiveCorr	chr1:3329228-3329258	-	Cons7_no_Cons100_no	PRDM16	2393188	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2245	intronicAntisense_positiveCorr	chr12:42589077-42589157	+	Cons7_yes_Cons100_yes	PRICKLE1	3412136	80	full	-0.1	0.0	0.000	TCL1A	0.6	0.0	NA	NA	NA
InGen_OS_2920	intronicAntisense_positiveCorr	chr17:42965042-42965165	+	Cons7_no_Cons100_no	PTGES3L-AARSD1	3722279	123	extended	-0.1	0.0	0.000	CCR7	0.3	0.0	CUX1	-0.3	0.0
InGen_OS_141	intronicAntisense_positiveCorr	chr1:220254971-220255003	+	Cons7_yes_Cons100_no	RAB3GAP2	2381098	32	extended	0.2	0.0	0.000	CCR7	0.3	0.0	CUX1	-0.3	0.0
InGen_OS_742	intronicAntisense_positiveCorr	chr3:128726474-128726565	-	Cons7_no_Cons100_yes	RAB7A	2694450	91	extended	-0.1	0.0	0.014	CD79A	0.6	0.0	NIN	-0.4	0.0
InGen_OS_271	intronicAntisense_positiveCorr	chr1:174160041-174160066	-	Cons7_no_Cons100_yes	RABGAP1L	2444655	25	extended	-0.1	0.0	0.000	CD79B	0.4	0.0	FOXP1	-0.3	0.0
InGen_OS_2235	intronicAntisense_positiveCorr	chr12:15221465-15221501	+	Cons7_no_Cons100_yes	RERG	3406293	36	full	-0.1	0.0	0.090	NA	NA	NA	NA	NA	NA
InGen_OS_200	intronicAntisense_positiveCorr	chr1:44407135-44407229	-	Cons7_no_Cons100_no	RNF220	2409715	94	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_2876	intronicAntisense_positiveCorr	chr17:360653-360677	+	Cons7_no_Cons100_no	RPH3AL	3705398	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_1311	intronicAntisense_positiveCorr	chr6:7146389-7146421	-	Cons7_no_Cons100_yes	RREB1	2940465	32	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_27	intronicAntisense_positiveCorr	chr1:25247740-25247799	+	Cons7_no_Cons100_no	RSRP1	2325871	59	full	-0.1	0.0	0.000	MDM2	0.5	0.0	NA	NA	NA
InGen_OS_3374	intronicAntisense_positiveCorr	chr21:34889570-	+	Cons7_no_Cons100_no	RUNX1	3919424	29	free	-0.1	0.0	0.000	DCAF12L2	0.4	0.0	ARHGAP5	-0.3	0.0

		34889599															
lnGen_OS_3375	intronicAntisense_positiveCorr	chr21:34988022-34988049	+	Cons7_no_Cons100_no	RUNX1	3919457	27	full	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1028	intronicAntisense_positiveCorr	chr5:116573895-116573921	+	Cons7_no_Cons100_yes	SEMA6A	2825175	26	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1942	intronicAntisense_positiveCorr	chr10:50623401-50623426	+	Cons7_yes_Cons100_yes	SGMS1	3246714	25	free	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2975	intronicAntisense_positiveCorr	chr17:11402808-11402832	-	Cons7_no_Cons100_no	SHISA6	3745708	24	full	-0.5	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_975	intronicAntisense_positiveCorr	chr5:1110799-1110846	+	Cons7_no_Cons100_no	SLC12A7	2799008	47	extended	-0.1	0.0	0.025	NA	NA	NA	NA	NA	NA
lnGen_OS_1454	intronicAntisense_positiveCorr	chr7:96321719-96321873	+	Cons7_no_Cons100_yes	SLC25A13	3013741	154	extended	-0.1	0.0	0.000	HOXD11	0.4	0.0	PTEN	-0.3	0.0
lnGen_OS_3195	intronicAntisense_positiveCorr	chr19:58491614-58491643	+	Cons7_no_Cons100_no	SLC27A5	3844206	29	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_541	intronicAntisense_positiveCorr	chr2:161625570-161625605	-	Cons7_yes_Cons100_yes	SLC4A10	2584007	35	full	-0.2	0.0	0.014	NA	NA	NA	NA	NA	NA
lnGen_OS_2738	intronicAntisense_positiveCorr	chr15:67147870-67147899	-	Cons7_no_Cons100_no	SMAD3	3630431	29	free	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1858	intronicAntisense_positiveCorr	chr9:2016602-2016646	-	Cons7_yes_Cons100_yes	SMARCA2	3196468	44	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3309	intronicAntisense_positiveCorr	chr20:4172335-4172363	-	Cons7_no_Cons100_no	SMOX	3895882	28	full	-0.1	0.0	0.047	RGPD3	0.7	0.0	PTEN	-0.3	0.0
lnGen_OS_1924	intronicAntisense_positiveCorr	chr10:17453746-17453772	+	Cons7_no_Cons100_no	ST8SIA6	3237042	26	full	-0.1	0.0	0.000	MDS2	0.9	0.0	BMPR1A	-0.3	0.0
lnGen_OS_1925	intronicAntisense_positiveCorr	chr10:17453878-17453906	+	Cons7_no_Cons100_no	ST8SIA6	3237044	28	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1291	intronicAntisense_positiveCorr	chr6:147203749-147203889	+	Cons7_yes_Cons100_yes	STXBP5-AS1	2929876	140	core	-0.1	0.0	0.056	TSC2	0.5	0.0	PIK3R1	-0.3	0.0
lnGen_OS_1131	intronicAntisense_positiveCorr	chr5:76170335-76170489	-	Cons7_yes_Cons100_yes	SV2C	2863204	154	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3093	intronicAntisense_positiveCorr	chr18:26256381-26256408	-	Cons7_yes_Cons100_no	TAF4B	3802233	27	extended	0.7	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_913	intronicAntisense_positiveCorr	chr4:37977716-37977782	-	Cons7_no_Cons100_yes	TBC1D1	2765962	66	full	-0.1	0.0	0.336	NUTM1	0.4	0.0	TET2	-0.3	0.0
lnGen_OS_2899	intronicAntisense_positiveCorr	chr17:18651565-18651604	+	Cons7_yes_Cons100_yes	TBC1D28	3713467	39	extended	0.3	0.0	0.000	MLLT11	0.7	0.0	PER1	-0.5	0.0
lnGen_OS_604	intronicAntisense_positiveCorr	chr3:17299610-17299642	+	Cons7_no_Cons100_no	TBC1D5	2612943	32	extended	-0.1	0.0	0.002	NA	NA	NA	NA	NA	NA
lnGen_OS_3051	intronicAntisense_positiveCorr	chr17:82866520-82866544	-	Cons7_no_Cons100_no	TBCD	3775380	24	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_3054	intronicAntisense_positiveCorr	chr17:82914648-82914699	-	Cons7_no_Cons100_no	TBCD	3775416	51	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2389	intronicAntisense_positiveCorr	chr12:114409202-114409271	-	Cons7_yes_Cons100_yes	TBX5-AS1	3472684	69	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_412	intronicAntisense_positiveCorr	chr2:187556570-187556724	+	Cons7_no_Cons100_no	TFPI	2519406	154	extended	-0.1	0.0	0.000	NUTM1	0.4	0.0	NA	NA	NA
lnGen_OS_67	intronicAntisense_positiveCorr	chr1:91885649-91885771	+	Cons7_no_Cons100_yes	TGFBR3	2346571	122	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_594	intronicAntisense_positiveCorr	chr3:9396839-9396935	+	Cons7_yes_Cons100_yes	THUMPD3-AS1	2609605	96	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1526	intronicAntisense_positiveCorr	chr7:38353681-38353711	-	Cons7_yes_Cons100_yes	TRG-AS1	3046699	30	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2339	intronicAntisense_positiveCorr	chr12:30965371-30965396	-	Cons7_no_Cons100_no	TSPAN11	3449488	25	full	-0.2	0.0	0.047	NA	NA	NA	NA	NA	NA

lnGen_OS_3592	intronicAntisense_positiveCorr	chrX:38562009-38562035	-	Cons7_no_Cons100_no	TSPAN7	4005019	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3428	intronicAntisense_positiveCorr	chr22:28442348-28442374	+	Cons7_no_Cons100_yes	TTC28	3941587	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3667	intronicAntisense_positiveCorr	chr1:2648928-2648961	+	Cons7_no_Cons100_no	TTC34	4042452	33	full	-0.4	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3669	intronicAntisense_positiveCorr	chr1:2652131-2652206	+	Cons7_no_Cons100_no	TTC34	4042460	75	full	-0.4	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3671	intronicAntisense_positiveCorr	chr1:2654936-2654974	+	Cons7_no_Cons100_no	TTC34	4042463	38	extended	-0.3	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3674	intronicAntisense_positiveCorr	chr1:2652773-2652846	+	Cons7_no_Cons100_no	TTC34	4042467	73	full	-0.3	0.0	0.021	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3677	intronicAntisense_positiveCorr	chr1:2653981-2654046	+	Cons7_no_Cons100_no	TTC34	4042473	65	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3678	intronicAntisense_positiveCorr	chr1:2654174-2654198	+	Cons7_no_Cons100_no	TTC34	4042475	24	full	-0.3	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3682	intronicAntisense_positiveCorr	chr1:2654619-2654658	+	Cons7_no_Cons100_no	TTC34	4042481	39	full	-0.4	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3746	intronicAntisense_positiveCorr	chr1:2697692-2697851	+	Cons7_no_Cons100_no	TTC34	4042614	159	full	-0.3	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_1477	intronicAntisense_positiveCorr	chr7:129952353-129952382	+	Cons7_yes_Cons100_yes	UBE2H	3023694	29	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_1150	intronicAntisense_positiveCorr	chr5:83473090-83473147	-	Cons7_no_Cons100_yes	VCAN	2865279	57	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_1774	intronicAntisense_positiveCorr	chr9:2621473-2621652	+	Cons7_yes_Cons100_yes	VLDLR-AS1	3160176	179	free	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3541	intronicAntisense_positiveCorr	chrX:54358476-54358500	+	Cons7_yes_Cons100_yes	WNK3	3978408	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3074	intronicAntisense_positiveCorr	chr18:48136788-48136819	+	Cons7_no_Cons100_no	ZBTB7C	3787478	31	extended	-0.2	0.0	0.000	NUTM1	0.7	0.0	HOOK3	-0.3	0.0	
lnGen_OS_3076	intronicAntisense_positiveCorr	chr18:48408736-48408855	+	Cons7_yes_Cons100_yes	ZBTB7C	3787541	119	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3178	intronicAntisense_positiveCorr	chr19:43619400-43619476	+	Cons7_no_Cons100_no	ZNF428	3835155	76	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3194	intronicAntisense_positiveCorr	chr19:58362089-58362159	+	Cons7_no_Cons100_no	ZNF497	3843997	70	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_3271	intronicAntisense_positiveCorr	chr19:57822891-57822915	-	Cons7_no_Cons100_no	ZNF587B	3872465	24	full	-0.2	0.0	0.134	NA	NA	NA	NA	NA	NA	NA
lnGen_OS_1640	intronicAntisense_positiveCorr	chr8:80873373-80873402	+	Cons7_no_Cons100_no	ZNF704	3104834	29	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA	NA

Intergenic RNAs independent

Classification_Pipeline	hg38Coordinates	strand	ConservationAnalysis	hosting Gene Symbol	Probe_ID	length BP	level	logFC _Limma	adj.P.Val_Limma	coding_prob_ CPAT	CGC_ pos	CGC_ pos_ corr	CGC_ posi _p	CGC_ neg	CGC_ neg_ corr	CGC_ neg_ p
lnGen_OS_1141	chr5:77958115-77958173	-	Cons7_no_Cons100_yes	novel	2863717	58	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_1574	chr7:127652117-127652149	-	Cons7_yes_Cons100_yes	novel	3071333	32	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
lnGen_OS_2425	chr13:74134066-	+	Cons7_yes_Cons100_yes	novel	3493822	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA

	74134091															
InGen_OS_2481	chr13:112062850-112062973	-	Cons7_yes_Cons100_yes	novel	3526065	123	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_3286	chr20:35160678-35160706	+	Cons7_no_Cons100_no	novel	3883208	28	extended	-0.1	0.0	0.257	NA	NA	NA	NA	NA	NA
InGen_OS_3627	chrX:113505608-113505634	-	Cons7_no_Cons100_no	novel	4018547	26	extended	-0.1	0.0	0.140	HMGN2P46	0.4	0.0	NA	NA	NA
InGen_OS_3654	chrM:236-260	+	Cons7_no_Cons100_no	novel	4037570	24	ambiguous	-0.4	0.0	0.012	PER1	0.6	0.0	NRAS	-0.6	0.0
InGen_OS_562	chr2:202376459-202376490	-	Cons7_yes_Cons100_yes	novel	2595308	31	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_737	chr3:124033693-124033728	-	Cons7_yes_Cons100_yes	novel	2692665	35	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_798	chr3:197644175-197644208	-	Cons7_no_Cons100_no	novel	2713549	33	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
InGen_OS_818	chr4:15002179-15002296	+	Cons7_no_Cons100_no	novel	2719366	117	full	-0.1	0.0	0.079	NA	NA	NA	NA	NA	NA
InGen_OS_824	chr4:24471652-24471710	+	Cons7_yes_Cons100_yes	novel	2721552	58	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_43	chr6:6007817-6007841	-	Cons7_no_Cons100_no	novel	2940179	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_64	chr5:101042180-101042261	-	Cons7_no_Cons100_no	novel	2868966	81	extended	0.1	0.0	0.010	MUC16	0.4	0.0	NCOA2	-0.4	0.0
OutGen_Conc_71	chr10:24844220-24844296	-	Cons7_no_Cons100_no	novel	3281715	76	extended	-0.2	0.0	0.003	NA	NA	NA	NA	NA	NA
OutGen_Conc_81	chr11:21580467-21580493	+	Cons7_no_Cons100_no	novel	3323695	26	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_130	chr11:65559267-65559300	-	Cons7_no_Cons100_no	novel	3377720	33	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_158	chr16:5098612-5098674	-	Cons7_no_Cons100_no	novel	3678583	62	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_162	chr16:54290203-54290245	-	Cons7_no_Cons100_yes	novel	3692314	42	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_176	chr18:12071821-12072020	+	Cons7_no_Cons100_no	novel	3779414	199	extended	-0.1	0.0	0.004	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_183	chr19:2840937-2840967	+	Cons7_no_Cons100_no	novel	3816665	30	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_276	chr3:197640887-197641161	-	Cons7_no_Cons100_no	novel	2713545	274	extended	0.3	0.0	0.014	DGCR8	0.4	0.0	ARHGAP5	-0.4	0.0
OutGen_Conc_Miss_277	chr3:197642260-197642607	-	Cons7_no_Cons100_no	novel	2713547	347	extended	0.3	0.0	0.030	KMT2D	0.3	0.0	NCOA4	-0.3	0.0
OutGen_Conc_Miss_278	chr3:197643660-197643697	-	Cons7_no_Cons100_no	novel	2713548	37	extended	0.4	0.0	0.000	NA	NA	NA	ARID2	-0.3	0.0
OutGen_Conc_Miss_279	chr3:197650773-197650838	-	Cons7_no_Cons100_no	novel	2713553	65	extended	0.4	0.0	0.003	ZNRF3	0.6	0.0	ARHGAP5	-0.5	0.0
OutGen_Conc_Miss_280	chr3:197653667-197653708	-	Cons7_yes_Cons100_yes	novel	2713558	41	full	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_301	chr4:1602067-1602098	-	Cons7_no_Cons100_no	novel	2757256	31	full	0.3	0.0	0.019	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_329	chr5:107194903-107194931	-	Cons7_no_Cons100_no	novel	2869838	28	extended	0.1	0.0	0.000	CNBD1	0.7	0.0	PRDM2	-0.4	0.0
OutGen_Conc_Miss_355	chr7:42889060-42889085	-	Cons7_no_Cons100_no	novel	3047941	25	extended	-0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_377	chr9:136621243-136621276	+	Cons7_no_Cons100_no	novel	3194490	33	full	-0.1	0.0	0.000	DCAF12L2	0.5	0.0	ARHGAP5	-0.3	0.0
OutGen_Conc_Miss_378	chr9:136621861-136621887	+	Cons7_no_Cons100_no	novel	3194491	26	full	-0.1	0.0	0.000	DCAF12L2	0.5	0.0	ARHGAP5	-0.4	0.0
OutGen_Conc_Miss_379	chr9:136625170-136625211	+	Cons7_no_Cons100_no	novel	3194497	41	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA

OutGen_Conc_Miss_383	chr9:91078459-91078506	-	Cons7_no_Cons100_no	novel	3214345	47	full	-0.1	0.0	0.012	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_391	chr10:109457084-109457139	-	Cons7_no_Cons100_no	novel	3306233	55	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_401	chr12:40620839-40620937	+	Cons7_no_Cons100_no	novel	3411710	98	extended	-0.2	0.0	0.026	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_407	chr12:1937408-1937435	-	Cons7_no_Cons100_no	novel	3440191	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_409	chr12:49843145-49843184	-	Cons7_no_Cons100_yes	novel	3454156	39	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_432	chr14:49549001-49549027	-	Cons7_no_Cons100_no	novel	3563309	26	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_433	chr14:91264552-91264588	-	Cons7_no_Cons100_no	novel	3576371	36	full	-0.2	0.0	0.000	DCAF12L2	0.6	0.0	RB1	-0.3	0.0
OutGen_Conc_Miss_45	chr4:184504649-184504744	-	Cons7_no_Cons100_no	novel	2796455	95	full	-0.1	0.0	0.225	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_472	chr17:49902084-49902118	+	Cons7_yes_Cons100_yes	novel	3726078	34	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_496	chr20:58018664-58018748	+	Cons7_no_Cons100_no	novel	3890789	84	full	-0.1	0.0	0.000	ECT2L	0.6	0.0	NA	NA	NA
OutGen_Conc_Miss_498	chr21:34073097-34073123	+	Cons7_no_Cons100_no	novel	3918994	26	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_511	chr22:49305445-49305478	-	Cons7_no_Cons100_no	novel	3965067	33	full	-0.1	0.0	0.628	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_512	chr22:49333897-49333922	-	Cons7_no_Cons100_no	novel	3965074	25	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_520	chrX:40005789-40005868	-	Cons7_no_Cons100_no	novel	4005398	79	full	-0.1	0.0	0.000	CDX2	0.4	0.0	NA	NA	NA
OutGen_Conc_Miss_521	chrX:40013896-40013926	-	Cons7_no_Cons100_yes	novel	4005409	30	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_604	chr9:62830999-62831200	+	Cons7_no_Cons100_no	novel	3172296	201	full	-0.1	0.0	0.000	C15orf65	0.4	0.0	GNA11	-0.3	0.0
OutGen_Conc_Miss_623	chr10:120187118-120187178	+	Cons7_no_Cons100_no	novel	3267530	60	extended	-0.1	0.0	0.000	C15orf65	0.6	0.0	GNA11	-0.4	0.0
OutGen_Conc_Miss_643	chr13:109424165-109424270	+	Cons7_yes_Cons100_yes	novel	3501084	105	extended	-0.2	0.0	0.011	PPM1D	0.3	0.0	NSD1	-0.6	0.0
OutGen_Conc_Miss_703	chrX:4544518-4544604	-	Cons7_no_Cons100_no	novel	3998086	86	extended	0.7	0.0	0.000	RGPD3	0.7	0.0	PTEN	-0.3	0.0
OutGen_Conc_Miss_704	chrX:4542560-4542588	-	Cons7_no_Cons100_no	novel	3998084	28	extended	0.5	0.0	0.037	RGPD3	0.7	0.0	PTEN	-0.3	0.0
OutGen_Conc_Miss_705	chrX:4540654-4540958	-	Cons7_no_Cons100_no	novel	3998081	304	extended	0.3	0.0	0.001	RGPD3	0.7	0.0	PTEN	-0.4	0.0
OutGen_Conc_Miss_72	chr6:169814167-169814193	+	Cons7_no_Cons100_no	novel	2937642	26	full	-0.4	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_73	chr6:170029935-170029966	+	Cons7_no_Cons100_no	novel	2937673	31	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Conc_Miss_96	chr7:56374242-56374304	-	Cons7_no_Cons100_no	novel	3052024	62	extended	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_115	chr1:2317903-2317928	+	Cons7_no_Cons100_no	novel	2316494	25	full	0.1	0.0	0.000	TAL2	0.6	0.0	NIN	-0.3	0.0
OutGen_Disc_139	chr7:114083559-114083583	-	Cons7_no_Cons100_no	novel	3068725	24	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_148	chr5:180494291-180494385	-	Cons7_yes_Cons100_yes	novel	2890723	94	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_186	chr4:166716367-166716508	+	Cons7_no_Cons100_no	novel	2750883	141	extended	-0.1	0.0	0.012	CDX2	0.4	0.0	TET2	-0.4	0.0
OutGen_Disc_212	chr3:173396084-173396116	-	Cons7_yes_Cons100_yes	novel	2705875	32	full	-0.1	0.0	0.248	NA	NA	NA	NA	NA	NA
OutGen_Disc_219	chr9:32955854-32955888	-	Cons7_no_Cons100_no	novel	3203303	34	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA

OutGen_Disc_22	chr14:105076371-105076480	-	Cons7_no_Cons100_no	novel	3581428	109	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_24	chr12:120317379-120317417	+	Cons7_no_Cons100_no	novel	3434329	38	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_251	chr19:57483451-57483637	+	Cons7_no_Cons100_no	novel	3843281	186	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_262	chr11:125433582-125433609	-	Cons7_no_Cons100_no	novel	3396570	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_3	chr5:14011623-14011647	-	Cons7_no_Cons100_yes	novel	2849196	24	full	-0.2	0.0	0.000	RGPD3	0.6	0.0	PTEN	-0.3	0.0
OutGen_Disc_336	chr1:150981935-150981960	+	Cons7_yes_Cons100_yes	novel	2358616	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_338	chr16:8992144-8992168	+	Cons7_no_Cons100_no	novel	3647600	24	extended	-0.1	0.0	0.000	CCR7	0.5	0.0	CIC	-0.4	0.0
OutGen_Disc_365	chr17:39612752-39612792	+	Cons7_yes_Cons100_yes	novel	3720313	40	full	-0.1	0.0	0.662	NA	NA	NA	NA	NA	NA
OutGen_Disc_369	chr10:87449709-87449873	+	Cons7_no_Cons100_yes	novel	3256541	164	extended	0.1	0.0	0.020	NA	NA	NA	STAT3	-0.4	0.0
OutGen_Disc_380	chr5:122127084-122127118	+	Cons7_no_Cons100_no	novel	2826117	34	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_52	chr1:204045675-204045718	-	Cons7_yes_Cons100_yes	novel	2451837	43	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_96	chr9:111896485-111896571	-	Cons7_no_Cons100_no	novel	3220815	86	full	0.1	0.0	0.000	TAL2	0.5	0.0	NA	NA	NA
OutGen_Disc_Miss_151	chr13:61090846-61090916	+	Cons7_no_Cons100_no	novel	3492106	70	extended	0.3	0.0	0.000	CDX2	0.5	0.0	TET2	-0.3	0.0
OutGen_Disc_Miss_169	chr9:64724560-64724591	+	Cons7_no_Cons100_no	novel	3173283	31	extended	0.3	0.0	0.048	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_175	chr20:31587659-31587707	-	Cons7_no_Cons100_no	novel	3902478	48	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_255	chr11:15844446-15844471	+	Cons7_yes_Cons100_yes	novel	3321888	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_268	chr18:14458354-14458392	+	Cons7_no_Cons100_no	novel	3780579	38	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_29	chr5:82615604-82615630	+	Cons7_no_Cons100_no	novel	2818362	26	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_315	chr7:103280600-103280631	-	Cons7_no_Cons100_no	novel	3065618	31	extended	-0.1	0.0	0.961	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_355	chr13:106919875-106919907	+	Cons7_yes_Cons100_yes	novel	3500510	32	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_41	chr19:14071254-14071281	-	Cons7_no_Cons100_no	novel	3852488	27	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_421	chr12:123151942-123152006	+	Cons7_no_Cons100_no	novel	3435741	64	full	-0.1	0.0	0.001	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_424	chr12:51869166-51869233	-	Cons7_no_Cons100_no	novel	3455031	67	full	-0.1	0.0	0.000	TAL2	0.6	0.0	NIN	-0.3	0.0
OutGen_Disc_Miss_466	chr18:14458506-14458548	+	Cons7_no_Cons100_no	novel	3780580	42	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_490	chr4:182815260-182815417	+	Cons7_no_Cons100_no	novel	2753661	157	extended	0.4	0.0	0.000	RGPD3	0.6	0.0	PTEN	-0.4	0.0
OutGen_Disc_Miss_547	chr9:63038290-63038317	-	Cons7_no_Cons100_no	novel	3207568	27	extended	0.3	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_551	chr18:79955864-79955925	+	Cons7_no_Cons100_no	novel	3795431	61	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_6	chr20:19754103-19754138	-	Cons7_no_Cons100_no	novel	3899869	35	extended	0.5	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_606	chr5:149472056-149472085	-	Cons7_no_Cons100_no	novel	2880884	29	full	-0.3	0.0	0.858	NA	NA	NA	NA	NA	NA

OutGen_Disc_Miss_645	chr1:3484078-3484114	-	Cons7_no_Cons100_no	novel	2393336	36	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_763	chr6:28988665-28988693	-	Cons7_no_Cons100_no	novel	2947640	28	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_764	chr10:103855669-103855721	-	Cons7_yes_Cons100_yes	novel	3304982	52	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_771	chr12:5201268-5201302	-	Cons7_no_Cons100_no	novel	3441468	34	full	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_781	chr12:58328403-58328432	-	Cons7_no_Cons100_no	novel	3459033	29	core	0.2	0.0	0.006	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_802	chr2:91557034-91557059	-	Cons7_no_Cons100_no	novel	2564192	25	extended	0.3	0.0	0.000	NA	NA	NA	NSD1	-0.5	0.0
OutGen_Disc_Miss_810	chr2:128474059-128474114	+	Cons7_no_Cons100_no	novel	2505060	55	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_862	chr9:78032203-78032320	+	Cons7_no_Cons100_no	novel	3175882	117	full	-0.1	0.0	0.000	TP63	0.3	0.0	ABI1	-0.3	0.0
OutGen_Disc_Miss_917	chr2:128450159-128450199	+	Cons7_no_Cons100_no	novel	2505054	40	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Disc_Miss_971	chr8:23859154-23859189	+	Cons7_no_Cons100_no	novel	3090123	35	full	-0.1	0.0	0.000	C15orf65	0.6	0.0	GNA11	-0.4	0.0
OutGen_Disc_Miss_986	chr17:1267830-1268032	+	Cons7_no_Cons100_no	novel	3705738	202	extended	-0.1	0.0	0.004	NA	NA	NA	NA	NA	NA
OutGen_Ind_1000	chr4:139177760-139177785	+	Cons7_no_Cons100_no	novel	2744576	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1009	chr4:185128668-185128726	-	Cons7_no_Cons100_no	novel	2796777	58	full	-0.1	0.0	0.090	NA	NA	NA	NA	NA	NA
OutGen_Ind_1085	chr16:54374079-54374106	+	Cons7_no_Cons100_yes	novel	3661409	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1120	chr2:88764929-88765345	-	Cons7_no_Cons100_no	novel	2563760	416	full	-0.1	0.0	0.514	GLI1	0.5	0.0	TET2	-0.4	0.0
OutGen_Ind_1145	chr17:82273834-82273865	+	Cons7_no_Cons100_no	novel	3738735	31	full	-0.2	0.0	0.000	CCR7	0.4	0.0	B2M	-0.3	0.0
OutGen_Ind_1156	chr6:158231672-158231751	-	Cons7_no_Cons100_no	novel	2981729	79	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1172	chr6:28635231-28635260	+	Cons7_no_Cons100_no	novel	2900604	29	full	-0.1	0.0	0.003	NA	NA	NA	NA	NA	NA
OutGen_Ind_1179	chr3:39181114-39181189	+	Cons7_no_Cons100_no	novel	2618144	75	free	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1190	chr17:81341123-81341288	+	Cons7_yes_Cons100_yes	novel	3737859	165	extended	-0.1	0.0	0.009	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
OutGen_Ind_1247	chr17:19724761-19724951	+	Cons7_no_Cons100_no	novel	3714124	190	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1249	chr1:4439460-4439485	-	Cons7_no_Cons100_no	novel	2393988	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1269	chr13:108496053-108496080	-	Cons7_no_Cons100_no	novel	3525064	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1305	chrX:18872413-18872449	-	Cons7_no_Cons100_no	novel	4001536	36	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1309	chr19:36267279-36267318	-	Cons7_no_Cons100_no	novel	3860343	39	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_1333	chrX:129779858-129779883	-	Cons7_yes_Cons100_yes	novel	4021327	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_134	chr3:123457643-123457680	+	Cons7_no_Cons100_no	novel	2639401	37	full	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_157	chr7:155252299-155252330	-	Cons7_no_Cons100_no	novel	3080981	31	full	-0.3	0.0	0.000	ESR1	0.3	0.0	NA	NA	NA
OutGen_Ind_158	chr11:57086491-57086686	-	Cons7_yes_Cons100_yes	novel	3373603	195	extended	0.3	0.0	0.001	NA	NA	NA	NA	NA	NA
OutGen_Ind_194	chr16:353331-353475	+	Cons7_no_Cons100_no	novel	3642799	144	full	-0.1	0.0	0.000	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
OutGen_Ind_204	chr4:139179022-139179049	+	Cons7_no_Cons100_no	novel	2744581	27	extended	-0.1	0.0	0.000	HMGN2P46	0.4	0.0	NA	NA	NA

OutGen_Ind_234	chr3:128617398-128617466	+	Cons7_no_Cons100_no	novel	2641237	68	core	-0.1	0.0	0.055	NA	NA	NA	NA	NA	NA
OutGen_Ind_238	chr9:33289576-33289652	-	Cons7_no_Cons100_yes	novel	3203515	76	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_276	chr2:42975585-42975619	+	Cons7_no_Cons100_no	novel	2479148	34	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_285	chr9:136575694-136575760	-	Cons7_no_Cons100_no	novel	3230238	66	full	-0.1	0.0	0.000	TAL2	0.6	0.0	NIN	-0.3	0.0
OutGen_Ind_303	chrX:154516283-154516409	+	Cons7_no_Cons100_no	novel	3996544	126	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_31	chr19:18897248-18897274	+	Cons7_no_Cons100_no	novel	3825436	26	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_312	chr19:36682656-36682680	+	Cons7_no_Cons100_no	novel	3831512	24	extended	0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_321	chr6:4005619-4005701	-	Cons7_no_Cons100_no	novel	2939538	82	extended	-0.1	0.0	0.000	ZNRF3	0.6	0.0	GNA11	-0.4	0.0
OutGen_Ind_335	chr1:154999559-154999636	-	Cons7_yes_Cons100_yes	novel	2437024	77	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_343	chr2:67426-67454	+	Cons7_no_Cons100_no	novel	2466100	28	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_355	chr2:43170962-43170997	+	Cons7_no_Cons100_yes	novel	2479297	35	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_366	chr18:23113833-23113868	-	Cons7_no_Cons100_no	novel	3801229	35	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_375	chr1:234723198-234723225	-	Cons7_no_Cons100_no	novel	2461591	27	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_390	chr5:107672424-107672480	+	Cons7_yes_Cons100_yes	novel	2823071	56	full	-0.1	0.0	0.000	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
OutGen_Ind_391	chr2:45010553-45010650	+	Cons7_no_Cons100_no	novel	2480007	97	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_41	chr3:183428892-183428926	-	Cons7_no_Cons100_no	novel	2708044	34	full	-0.1	0.0	0.000	ACSL6	0.3	0.0	ARAF	-0.3	0.0
OutGen_Ind_429	chr2:76504177-76504230	+	Cons7_no_Cons100_no	novel	2489952	53	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_44	chr5:178075545-178075589	-	Cons7_no_Cons100_no	novel	2889421	44	extended	-0.4	0.0	0.002	NA	NA	NA	NA	NA	NA
OutGen_Ind_444	chr17:75858630-75858657	+	Cons7_no_Cons100_no	novel	3735300	27	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_467	chr6:37137379-37137404	-	Cons7_no_Cons100_no	novel	2952182	25	full	-0.2	0.0	0.000	ELF3	0.3	0.0	MLLT6	-0.4	0.0
OutGen_Ind_473	chr14:50668528-50668557	+	Cons7_no_Cons100_no	novel	3535241	29	full	-0.2	0.0	0.119	RGPD3	0.7	0.0	PTEN	-0.3	0.0
OutGen_Ind_475	chr22:23521808-23521950	+	Cons7_no_Cons100_yes	novel	3939346	142	extended	-0.1	0.0	0.051	NA	NA	NA	NA	NA	NA
OutGen_Ind_498	chr4:183737101-183737132	-	Cons7_no_Cons100_no	novel	2796118	31	extended	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_509	chr18:80147530-80147592	-	Cons7_no_Cons100_no	novel	3814788	62	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_533	chr2:240899546-240899597	+	Cons7_no_Cons100_no	novel	2536001	51	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_567	chr1:119828918-119828951	+	Cons7_no_Cons100_no	novel	2354718	33	free	0.2	0.0	0.076	NA	NA	NA	NA	NA	NA
OutGen_Ind_60	chr2:42794363-42794798	+	Cons7_no_Cons100_yes	novel	2479054	435	extended	-0.2	0.0	0.002	NCOR1	0.3	0.0	MAF	-0.3	0.0
OutGen_Ind_608	chr4:122493730-122493760	+	Cons7_no_Cons100_no	novel	2742037	30	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_614	chr17:57699811-57699864	+	Cons7_yes_Cons100_yes	novel	3728323	53	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_616	chr2:236306171-236306208	+	Cons7_no_Cons100_no	novel	2533958	37	full	-0.1	0.0	0.000	RGPD3	0.7	0.0	PTEN	-0.3	0.0

OutGen_Ind_629	chr1:30908948-30909030	+	Cons7_no_Cons100_no	novel	2328178	82	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_645	chr4:87936126-87936156	-	Cons7_no_Cons100_no	novel	2777241	30	extended	-0.1	0.0	0.000	CDX2	0.4	0.0	TET2	-0.4	0.0
OutGen_Ind_665	chr2:138278379-138278404	-	Cons7_no_Cons100_no	novel	2578530	25	extended	-0.1	0.0	0.000	RGPD3	0.7	0.0	PTEN	-0.3	0.0
OutGen_Ind_669	chr6:7051337-7051445	-	Cons7_no_Cons100_yes	novel	2940413	108	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_677	chr2:172082872-172082912	-	Cons7_no_Cons100_no	novel	2586980	40	full	-0.1	0.0	0.321	APC	0.3	0.0	STAT6	-0.4	0.0
OutGen_Ind_679	chr1:1247756-1247815	+	Cons7_no_Cons100_no	novel	2315650	59	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_690	chr11:118993235-118993302	-	Cons7_no_Cons100_no	novel	3394053	67	extended	-0.1	0.0	0.017	GLI1	0.4	0.0	TET2	-0.4	0.0
OutGen_Ind_691	chr18:28422212-28422269	+	Cons7_no_Cons100_no	novel	3783099	57	extended	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_732	chr12:49060038-49060062	+	Cons7_yes_Cons100_yes	novel	3413762	24	full	-0.2	0.0	0.017	NUTM1	0.7	0.0	HOOK3	-0.3	0.0
OutGen_Ind_74	chr6:14352034-14352069	+	Cons7_no_Cons100_no	novel	2895936	35	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_770	chr1:200739313-200739371	-	Cons7_yes_Cons100_no	novel	2450450	58	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_800	chr15:79431801-79432032	-	Cons7_yes_Cons100_no	novel	3635007	231	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_822	chr11:67447584-67447614	+	Cons7_no_Cons100_no	novel	3337107	30	extended	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_865	chr1:25568904-25568929	-	Cons7_no_Cons100_no	novel	2402347	25	full	-0.2	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_866	chr2:54456016-54456057	-	Cons7_no_Cons100_no	novel	2553473	41	free	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_88	chr17:82101331-82101403	+	Cons7_no_Cons100_no	novel	3738536	72	full	-0.1	0.0	0.101	RBM15	0.5	0.0	IL6ST	-0.4	0.0
OutGen_Ind_881	chr20:62952533-62952567	-	Cons7_no_Cons100_no	novel	3913625	34	extended	-0.1	0.0	0.000	C15orf65	0.6	0.0	HIP1	-0.4	0.0
OutGen_Ind_926	chr8:98974500-98974549	-	Cons7_no_Cons100_no	novel	3146328	49	full	-0.1	0.0	0.592	NA	NA	NA	NA	NA	NA
OutGen_Ind_968	chr10:132797287-132797312	+	Cons7_no_Cons100_no	novel	3272376	25	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Ind_986	chr11:102931653-102931679	+	Cons7_no_Cons100_no	novel	3346737	26	full	-0.1	0.0	0.000	NA	NA	NA	NA	NA	NA
OutGen_Nan_1	chr1:119862011-119862044	-	Cons7_no_Cons100_no	novel	2431208	33	ambiguous	0.2	0.0	0.000	NA	NA	NA	NA	NA	NA

Cluster RNA analysis

Cluster ID	hostCoordinates	host Strand	Cluster	average Rho	average pValue	Cluster Classification
17	chr6:34236873-34246231	+	['lnGen_OS_1340', 'lnGen_OS_1341']	0.2	0.003	['exonicAntisense', 'exonicAntisense']
32	chr1:212950520-212992037	+	['lnGen_SS_Miss_49', 'lnGen_SS_Miss_50']	0.5	0.000	['intronicSense', 'intronicSense']
47	chr20:3249305-3407625	-	['lnGen_SS_Miss_813', 'lnGen_SS_Miss_814']	0.4	0.000	['intronicSense', 'intronicSense']
56	chr4:10074339-10116949	-	['lnGen_SS_180', 'lnGen_SS_179']	0.6	0.000	['intronicSense', 'intronicSense']
77	chr8:141207166-141308305	-	['lnGen_SS_Miss_406', 'lnGen_SS_367']	0.4	0.000	['intronicSense', 'intronicSense']
96	chr17:78261349-78278492	-	['lnGen_SS_Miss_744', 'lnGen_SS_Miss_743']	0.7	0.000	['intronicSense', 'intronicSense']
102	chr3:156825481-157046129	+	['lnGen_SS_Nan_2', 'lnGen_SS_Nan_1']	0.2	0.002	['intronicSense', 'intronicSense']
103	chr16:1351923-1364113	+	['lnGen_SS_603', 'lnGen_SS_602']	0.4	0.000	['intronicSense', 'intronicSense']
106	chr2:101007617-101252866	-	['lnGen_SS_Miss_184', 'lnGen_SS_Miss_183']	0.6	0.000	['intronicSense', 'intronicSense']
109	chr11:62433542-62556235	-	['lnGen_OS_2106', 'lnGen_OS_2105']	0.4	0.000	['exonicAntisense', 'exonicAntisense']
129	chr6:30720201-30725426	+	['lnGen_OS_1331', 'lnGen_OS_1332']	0.3	0.000	['exonicAntisense', 'exonicAntisense']
157	chr5:151661096-151687165	-	['lnGen_OS_1074', 'lnGen_OS_1073', 'lnGen_OS_1075']	0.5	0.000	['exonicAntisense', 'exonicAntisense', 'exonicAntisense']
161	chr2:46916157-47076137	+	['lnGen_SS_Miss_120', 'lnGen_SS_Miss_119', 'lnGen_SS_Miss_118']	0.5	0.000	['intronicSense', 'intronicSense', 'intronicSense']
172	chr5:72816312-72916733	+	['lnGen_OS_1128', 'lnGen_OS_1127']	0.2	0.008	['exonicAntisense', 'exonicAntisense']
227	chr8:26514022-26658178	+	['lnGen_SS_333', 'lnGen_SS_332']	0.3	0.000	['intronicSense', 'intronicSense']
234	chr18:26133852-26193355	+	['lnGen_OS_3090', 'lnGen_OS_3091']	0.6	0.000	['intronicAntisense', 'intronicAntisense']
253	chr2:111036776-111523376	-	['lnGen_OS_506', 'lnGen_OS_507']	0.3	0.000	['intronicSense', 'intronicSense']
293	chr2:215360440-215436172	-	['lnGen_OS_430', 'lnGen_OS_429', 'lnGen_OS_427', 'lnGen_OS_431', 'lnGen_OS_428']	0.5	0.000	['exonicAntisense', 'exonicAntisense', 'exonicAntisense', 'exonicAntisense', 'exonicAntisense']
299	chr12:4248765-4276184	-	['lnGen_OS_2226', 'lnGen_OS_2225']	0.5	0.000	['intronicAntisense', 'intronicAntisense']
310	chr6:157381133-157678146	+	['lnGen_SS_Miss_904', 'lnGen_SS_Miss_905']	0.3	0.000	['intronicSense', 'intronicSense']
342	chr12:49265156-49273306	-	['lnGen_OS_2353', 'lnGen_OS_2354']	0.8	0.000	['intronicSense', 'intronicSense']
352	chr6:128883141-129516569	+	['lnGen_SS_Miss_311', 'lnGen_SS_Miss_312']	0.4	0.000	['intronicSense', 'intronicSense']
370	chr6:166409364-166906451	-	['lnGen_SS_275', 'lnGen_SS_276']	0.3	0.000	['intronicSense', 'intronicSense']
392	chr22:30969245-30979395	+	['lnGen_OS_3489', 'lnGen_OS_3488']	0.4	0.000	['exonicAntisense', 'exonicAntisense']
402	chr7:139777051-140020325	+	['lnGen_SS_300', 'lnGen_SS_301', 'lnGen_SS_302']	0.4	0.000	['intronicSense', 'intronicSense', 'intronicSense']
427	chr9:137618963-137870016	+	['lnGen_SS_393', 'lnGen_SS_Miss_427']	0.3	0.000	['intronicSense', 'intronicSense']
436	chr11:1074875-1110511	+	['lnGen_SS_Miss_497', 'lnGen_SS_Miss_498', 'lnGen_SS_Miss_499']	0.8	0.000	['intronicSense', 'intronicSense', 'intronicSense']
440	chr9:125234853-125241330	-	['lnGen_OS_1838', 'lnGen_OS_1837']	0.6	0.000	['exonicAntisense', 'exonicAntisense']
444	chr18:13217498-13652755	+	['lnGen_OS_3087', 'lnGen_OS_3088']	0.5	0.000	['intronicAntisense', 'intronicAntisense']
493	chr19:14147743-14206187	-	['lnGen_SS_Miss_801', 'lnGen_SS_Miss_800']	0.6	0.000	['intronicSense', 'intronicSense']
533	chr6:118460772-118710075	-	['lnGen_OS_1279', 'lnGen_OS_1278']	0.2	0.034	['exonicAntisense', 'exonicAntisense']

549	chr12:93377883-93408146	+	['InGen_SS_506', 'InGen_SS_505']	0.6	0.000	['intronicSense', 'intronicSense']
566	chr1:3069168-3438621	+	['InGen_OS_166', 'InGen_OS_167']	1.0	0.000	['intronicAntisense', 'intronicAntisense']
573	chr5:150401637-150412929	-	['InGen_OS_1070', 'InGen_OS_1071']	0.2	0.043	['exonicAntisense', 'exonicAntisense']
594	chr17:58325450-58415766	+	['InGen_SS_Miss_723', 'InGen_SS_Miss_724']	0.4	0.000	['intronicSense', 'intronicSense']
601	chr6:21664772-22368328	+	['InGen_SS_Miss_297', 'InGen_SS_Miss_296']	0.3	0.000	['intronicSense', 'intronicSense']
622	chr7:55019021-55211628	+	['InGen_SS_279', 'InGen_SS_281', 'InGen_SS_280']	0.6	0.000	['intronicSense', 'intronicSense', 'intronicSense']
636	chr15:72199029-72231822	-	['InGen_OS_2666', 'InGen_OS_2665']	0.5	0.000	['exonicAntisense', 'exonicAntisense']
655	chr2:109129348-109504632	+	['InGen_SS_Miss_148', 'InGen_SS_Miss_147']	0.5	0.000	['intronicSense', 'intronicSense']
659	chr6:139271362-139667284	+	['OutGen_Ind_622', 'InGen_OS_1286']	0.4	0.000	['intronicSense', 'intronicSense']
664	chr18:14477955-14499278	-	['InGen_SS_Miss_761', 'InGen_SS_Miss_762']	0.6	0.000	['intronicSense', 'intronicSense']
672	chr22:49773283-49827512	-	['InGen_OS_3461', 'InGen_OS_3460']	0.2	0.005	['intronicAntisense', 'intronicAntisense']
699	chr19:35319261-35347355	+	['InGen_SS_Miss_793', 'InGen_SS_Miss_794']	0.8	0.000	['intronicSense', 'intronicSense']
706	chr15:96110040-96327361	-	['InGen_SS_Miss_646', 'InGen_SS_Miss_645']	0.5	0.000	['intronicSense', 'intronicSense']
710	chr1:11479166-11537584	+	['InGen_SS_Miss_6', 'InGen_SS_Miss_5']	0.7	0.000	['intronicSense', 'intronicSense']
720	chr1:110150486-110202202	+	['InGen_SS_Miss_24', 'InGen_SS_Miss_25']	0.4	0.000	['intronicSense', 'intronicSense']
722	chr7:43951910-44019151	-	['InGen_SS_Miss_364', 'InGen_SS_Miss_366', 'InGen_SS_Miss_367', 'InGen_SS_Miss_368']	0.3	0.000	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
742	chr16:69976297-70065948	-	['InGen_SS_Miss_922', 'InGen_SS_Miss_923', 'InGen_SS_Miss_921']	0.4	0.000	['intronicSense', 'intronicSense', 'intronicSense']
759	chr12:132822187-132956304	-	['InGen_SS_528', 'InGen_SS_529']	0.5	0.000	['intronicSense', 'intronicSense']
770	chr6:35573585-35728583	-	['InGen_SS_Miss_324', 'InGen_SS_Miss_325', 'InGen_SS_Miss_326', 'InGen_SS_261', 'InGen_SS_Miss_328', 'InGen_SS_Miss_323']	0.8	0.000	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
812	chr1:20719732-20732837	-	['InGen_SS_Miss_68', 'InGen_SS_Miss_67']	0.5	0.000	['intronicSense', 'intronicSense']
813	chr2:132671799-133568463	-	['InGen_SS_98', 'InGen_SS_97']	0.3	0.000	['intronicSense', 'intronicSense']
854	chrX:154389955-154409168	+	['InGen_SS_813', 'InGen_SS_814']	0.7	0.000	['intronicSense', 'intronicSense']
856	chr17:9021542-9244000	+	['InGen_SS_Miss_717', 'InGen_SS_Miss_718']	0.3	0.000	['intronicSense', 'intronicSense']
859	chr13:112690329-112887168	+	['InGen_SS_Miss_575', 'InGen_SS_Miss_574', 'InGen_SS_Miss_576']	0.5	0.000	['intronicSense', 'intronicSense', 'intronicSense']
865	chr12:111443485-111599676	-	['InGen_OS_2286', 'InGen_OS_2283']	0.2	0.008	['exonicAntisense', 'exonicAntisense']
877	chr10:119207589-119459742	+	['InGen_SS_423', 'InGen_SS_422']	0.4	0.000	['intronicSense', 'intronicSense']
880	chr7:69598296-70793495	+	['InGen_SS_283', 'InGen_SS_282']	0.2	0.007	['intronicSense', 'intronicSense']
912	chr7:66682164-66811464	+	['InGen_OS_1538', 'InGen_OS_1537']	0.4	0.000	['exonicAntisense', 'exonicAntisense']
921	chr7:135092363-135170795	-	['InGen_OS_1481', 'InGen_OS_1480']	0.5	0.000	['exonicAntisense', 'exonicAntisense']
933	chr2:131492813-131521573	+	['InGen_SS_Miss_155', 'InGen_SS_Miss_154']	0.5	0.000	['intronicSense', 'intronicSense']
963	chr4:118850688-119061247	+	['InGen_OS_957', 'InGen_OS_956']	0.6	0.000	['exonicAntisense', 'exonicAntisense']
990	chr4:15469865-15601557	+	['InGen_SS_Miss_236', 'InGen_SS_Miss_235']	0.3	0.000	['intronicSense', 'intronicSense']
992	chr9:14521-73865	-	['InGen_SS_Miss_434', 'InGen_SS_Miss_435', 'InGen_SS_Miss_431', 'InGen_SS_Miss_433']	0.4	0.000	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
1001	chr5:76083172-76353939	+	['InGen_OS_1137', 'InGen_OS_1131']	0.2	0.022	['intronicAntisense', 'intronicAntisense']

1003	chr7:4682309-4771443	+	['InGen_SS_Miss_337', 'InGen_SS_Miss_336']	0.4	0.000	['intronicSense', 'intronicSense']
1012	chr19:45079288-45305283	+	['InGen_OS_3250', 'InGen_OS_3251']	0.4	0.000	['intronicAntisense', 'intronicAntisense']
1013	chr5:139846779-140043299	-	['InGen_SS_Miss_287', 'InGen_SS_Miss_286']	0.3	0.000	['intronicSense', 'intronicSense']
1023	chr17:50426158-50474845	+	['InGen_SS_645', 'InGen_SS_644']	0.3	0.000	['intronicSense', 'intronicSense']
1031	chr14:105703995-105708665	-	['InGen_OS_2568', 'InGen_OS_2554', 'InGen_OS_2555']	0.5	0.000	['exonicAntisense', 'exonicAntisense', 'exonicAntisense']
1040	chr1:208022242-208244320	-	['InGen_SS_Miss_101', 'InGen_SS_Miss_102']	0.4	0.000	['intronicSense', 'intronicSense']
1045	chr3:4979116-4985323	+	['InGen_OS_678', 'InGen_OS_679']	0.3	0.000	['exonicAntisense', 'exonicAntisense']
1048	chr14:91965991-92040896	-	['InGen_SS_572', 'InGen_SS_571']	0.4	0.000	['intronicSense', 'intronicSense']
1069	chr9:39983793-40106611	-	['OutGen_Ind_412', 'OutGen_Ind_411', 'OutGen_Ind_403', 'OutGen_Ind_407']	0.2	0.021	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
1109	chr22:17787649-18024559	-	['InGen_SS_Miss_841', 'InGen_SS_Miss_843', 'InGen_SS_Miss_842', 'InGen_SS_Miss_844']	0.3	0.002	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
1182	chr2:158968671-159232659	+	['InGen_SS_Miss_159', 'InGen_SS_Miss_156']	0.4	0.000	['intronicSense', 'intronicSense']
1189	chr2:132147591-132257969	-	['InGen_OS_402', 'InGen_OS_401', 'InGen_OS_400']	0.9	0.000	['intronicAntisense', 'intronicAntisense', 'intronicAntisense']
1199	chr13:101053776-101416492	-	['InGen_SS_549', 'InGen_SS_550']	0.2	0.007	['intronicSense', 'intronicSense']
1230	chr19:12938578-12944489	+	['InGen_OS_3210', 'InGen_OS_3211']	0.6	0.000	['exonicAntisense', 'exonicAntisense']
1232	chr16:27313668-27364778	+	['InGen_SS_613', 'InGen_SS_614']	0.5	0.000	['intronicSense', 'intronicSense']
1249	chr16:89912119-89920977	+	['InGen_OS_2874', 'InGen_OS_2873']	0.6	0.000	['exonicAntisense', 'exonicAntisense']
1282	chr7:27181510-27185223	-	['InGen_SS_312', 'InGen_SS_311']	0.5	0.000	['intronicSense', 'intronicSense']
1304	chrM:1671-3229	+	['InGen_OS_3657', 'InGen_OS_3658']	0.8	0.000	['exonicAntisense', 'exonicAntisense']
1312	chr10:49942033-49974850	+	['InGen_SS_Miss_456', 'InGen_SS_Miss_483']	0.5	0.000	['intronicSense', 'intronicSense']
1317	chr7:11370357-11832198	-	['InGen_SS_Miss_360', 'InGen_SS_Miss_361']	0.3	0.000	['intronicSense', 'intronicSense']
1321	chr7:92246234-92401384	+	['InGen_SS_Miss_348', 'InGen_SS_Miss_347']	0.7	0.000	['intronicSense', 'intronicSense']
1372	chr3:114314501-115147271	-	['InGen_SS_Miss_220', 'InGen_SS_Miss_218', 'InGen_SS_Miss_219']	0.8	0.000	['intronicSense', 'intronicSense', 'intronicSense']
1402	chr1:174159410-174995308	+	['InGen_OS_271', 'InGen_OS_272']	0.2	0.015	['intronicAntisense', 'intronicAntisense']
1407	chr17:59707192-59842255	+	['InGen_OS_3030', 'InGen_OS_3031']	0.2	0.004	['exonicAntisense', 'exonicAntisense']
1438	chr1:2050470-2185395	+	['InGen_SS_2', 'InGen_SS_Miss_916', 'InGen_SS_Miss_915']	0.5	0.000	['intronicSense', 'intronicSense', 'intronicSense']
1448	chr11:78652831-79440948	-	['InGen_SS_Miss_528', 'InGen_SS_Miss_527']	0.4	0.000	['intronicSense', 'intronicSense']
1489	chr13:78596129-78599619	+	['InGen_OS_2468', 'InGen_OS_2469']	0.6	0.000	['exonicAntisense', 'exonicAntisense']
1498	chr19:53519527-53580269	+	['InGen_SS_717', 'InGen_SS_716']	0.6	0.000	['intronicSense', 'intronicSense']
1511	chr2:26562582-26579532	+	['InGen_SS_Miss_115', 'InGen_SS_Miss_116']	0.6	0.000	['intronicSense', 'intronicSense']
1546	chr14:105300563-105398147	+	['InGen_SS_Miss_597', 'InGen_SS_Miss_596', 'InGen_SS_Miss_598']	0.4	0.000	['intronicSense', 'intronicSense', 'intronicSense']
1548	chr11:77066941-77126155	+	['InGen_SS_458', 'InGen_SS_459']	0.2	0.040	['intronicSense', 'intronicSense']
1567	chr9:130053426-130140169	+	['InGen_SS_Miss_422', 'InGen_SS_Miss_421']	0.3	0.000	['intronicSense', 'intronicSense']
1600	chr1:159854316-159862657	-	['InGen_SS_Miss_96', 'InGen_SS_Miss_95']	0.6	0.000	['intronicSense', 'intronicSense']
1615	chr1:151399534-151401944	+	['InGen_OS_251', 'InGen_OS_250']	0.3	0.000	['exonicAntisense', 'exonicAntisense']

1679	chr1:245749342-246507312	-	['InGen_SS_54', 'InGen_SS_55', 'InGen_SS_53']	0.6	0.000	['intronicSense', 'intronicSense', 'intronicSense']
1688	chr10:61901300-62096944	+	['InGen_SS_Miss_458', 'InGen_SS_Miss_459', 'InGen_SS_Miss_461', 'InGen_SS_Miss_460']	0.5	0.000	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
1746	chr15:60488284-61229319	-	['InGen_SS_Miss_635', 'InGen_SS_Miss_636', 'InGen_SS_Miss_637', 'InGen_SS_Miss_634']	0.3	0.001	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
1751	chr1:149054027-149103561	-	['InGen_SS_Miss_31', 'InGen_SS_Miss_30', 'InGen_SS_Miss_33', 'InGen_SS_Miss_32']	0.4	0.000	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
1788	chr1:14404-29570	-	['InGen_SS_Miss_59', 'InGen_SS_Miss_60']	0.2	0.009	['intronicSense', 'intronicSense']
1798	chr12:121712752-121783001	+	['InGen_SS_Miss_548', 'InGen_SS_Miss_549', 'InGen_SS_Miss_547']	0.4	0.000	['intronicSense', 'intronicSense', 'intronicSense']
1836	chr3:151085697-151437072	+	['InGen_OS_762', 'InGen_OS_763']	0.3	0.000	['exonicAntisense', 'exonicAntisense']
1840	chr12:57128493-57213351	+	['InGen_SS_504', 'InGen_SS_503']	0.3	0.000	['intronicSense', 'intronicSense']
1841	chr1:205712819-205750276	-	['InGen_OS_127', 'InGen_OS_128']	0.4	0.000	['exonicAntisense', 'exonicAntisense']
1852	chr11:65422774-65445540	+	['InGen_OS_2184', 'InGen_OS_2183']	0.6	0.000	['exonicAntisense', 'exonicAntisense']
1869	chr2:69457997-69674349	-	['InGen_SS_91', 'InGen_SS_90']	0.2	0.005	['intronicSense', 'intronicSense']
1871	chr4:88007668-88077777	+	['InGen_OS_944', 'InGen_OS_945']	0.3	0.000	['exonicAntisense', 'exonicAntisense']
1872	chr2:148875250-149026759	+	['InGen_SS_70', 'InGen_SS_69']	0.7	0.000	['intronicSense', 'intronicSense']
1874	chr16:88715338-88785220	-	['InGen_SS_Miss_711', 'InGen_SS_Miss_710']	0.4	0.000	['intronicSense', 'intronicSense']
1875	chr12:132489551-132585188	+	['InGen_SS_511', 'InGen_SS_512']	0.5	0.000	['intronicSense', 'intronicSense']
1896	chr5:61332273-61546170	+	['InGen_SS_Miss_267', 'InGen_SS_Miss_269', 'InGen_SS_Miss_268']	0.3	0.032	['intronicSense', 'intronicSense', 'intronicSense']
1941	chr2:95836919-95991831	-	['InGen_SS_Miss_179', 'InGen_SS_Miss_178', 'InGen_SS_Miss_180', 'InGen_SS_Miss_181']	0.5	0.000	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
1953	chr6:89642499-89819723	-	['InGen_OS_1272', 'InGen_OS_1273', 'InGen_OS_1271']	0.4	0.000	['intronicAntisense', 'intronicAntisense', 'intronicAntisense']
1968	chr8:27596917-27615031	-	['InGen_OS_1614', 'InGen_OS_1615']	0.4	0.000	['exonicAntisense', 'exonicAntisense']
1974	chr11:64244480-64246941	-	['InGen_OS_2112', 'InGen_OS_2113']	0.5	0.000	['exonicAntisense', 'exonicAntisense']
1978	chr1:41361922-41383590	+	['InGen_SS_Miss_15', 'InGen_SS_Miss_14', 'InGen_SS_Miss_13']	0.2	0.005	['intronicSense', 'intronicSense', 'intronicSense']
1991	chr7:72969696-73005922	-	['InGen_SS_Miss_374', 'InGen_SS_Miss_376']	0.6	0.000	['intronicSense', 'intronicSense']
2015	chr11:67583595-67586656	+	['InGen_OS_2187', 'InGen_OS_2186']	0.2	0.009	['exonicAntisense', 'exonicAntisense']
2026	chr6:34417454-34426125	-	['InGen_OS_1245', 'InGen_OS_1246']	0.3	0.000	['exonicAntisense', 'exonicAntisense']
2061	chr6:87152833-87264196	+	['InGen_SS_Miss_304', 'InGen_SS_Miss_305']	0.8	0.000	['intronicSense', 'intronicSense']
2086	chr16:8788823-8849331	+	['InGen_SS_608', 'InGen_SS_607']	0.7	0.000	['intronicSense', 'intronicSense']
2091	chr6:33299694-33314387	-	['InGen_OS_1240', 'InGen_OS_1241']	0.5	0.000	['exonicAntisense', 'exonicAntisense']
2130	chr22:19756703-19783593	+	['InGen_SS_Miss_826', 'InGen_SS_Miss_825']	0.4	0.000	['intronicSense', 'intronicSense']
2132	chr12:112160188-112382439	-	['InGen_SS_Miss_562', 'InGen_SS_Miss_563']	0.3	0.000	['intronicSense', 'intronicSense']
2136	chr5:140401814-140539856	+	['InGen_OS_1183', 'InGen_OS_1184']	0.4	0.000	['exonicAntisense', 'exonicAntisense']
2147	chrX:101390824-101396154	+	['InGen_OS_3623', 'InGen_OS_3622']	0.3	0.000	['exonicAntisense', 'exonicAntisense']
2169	chr6:31665236-31670343	+	['InGen_OS_1335', 'InGen_OS_1334', 'InGen_OS_1336']	0.3	0.001	['exonicAntisense', 'exonicAntisense', 'exonicAntisense']
2181	chr22:50217689-50244992	-	['InGen_SS_799', 'InGen_SS_800']	0.4	0.000	['intronicSense', 'intronicSense']
2195	chr8:139727725-140458579	-	['InGen_SS_362', 'InGen_SS_363', 'InGen_SS_364', 'InGen_SS_365']	0.3	0.001	['intronicSense', 'intronicSense', 'intronicSense',

						'intronicSense']
2197	chr1:146064699-146144804	-	['InGen_SS_Miss_35', 'InGen_SS_Miss_34']	0.3	0.000	['intronicSense', 'intronicSense']
2205	chr12:6533927-6538374	+	['InGen_OS_2323', 'InGen_OS_2322']	0.7	0.000	['exonicAntisense', 'exonicAntisense']
2208	chr1:93561786-93847150	-	['InGen_SS_44', 'InGen_SS_45']	0.3	0.000	['intronicSense', 'intronicSense']
2230	chr1:3489920-3611495	-	['InGen_SS_Miss_64', 'InGen_SS_Miss_65', 'InGen_SS_Miss_63']	0.3	0.000	['intronicSense', 'intronicSense', 'intronicSense']
2253	chr5:179110851-179345430	-	['InGen_SS_Miss_290', 'InGen_SS_Miss_291']	0.2	0.004	['intronicSense', 'intronicSense']
2287	chr21:33229901-33265675	+	['InGen_SS_Miss_821', 'InGen_SS_759']	0.5	0.000	['intronicSense', 'intronicSense']
2297	chr1:145992435-145996600	-	['InGen_OS_245', 'InGen_OS_246']	0.5	0.000	['exonicAntisense', 'exonicAntisense']
2303	chr17:72290091-72640472	-	['InGen_SS_Miss_739', 'InGen_SS_Miss_738', 'InGen_SS_Miss_741', 'InGen_SS_Miss_740']	0.5	0.000	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
2313	chr2:131625957-131628147	+	['OutGen_Conc_Miss_27', 'OutGen_Conc_Miss_26']	0.7	0.000	['intronicSense', 'intronicSense']
2321	chr2:199269500-199471266	-	['InGen_SS_Miss_192', 'InGen_SS_Miss_190']	0.3	0.000	['intronicSense', 'intronicSense']
2348	chr5:14704800-14871785	-	['InGen_SS_210', 'InGen_SS_211', 'InGen_SS_209']	0.4	0.000	['intronicSense', 'intronicSense', 'intronicSense']
2349	chr20:32443059-32585074	-	['InGen_SS_753', 'InGen_SS_754']	0.3	0.000	['intronicSense', 'intronicSense']
2356	chr12:91140484-91183123	-	['InGen_OS_2267', 'InGen_OS_2266']	0.2	0.002	['exonicAntisense', 'exonicAntisense']
2364	chr10:17318383-17454330	-	['InGen_OS_1925', 'InGen_OS_1924']	0.8	0.000	['intronicAntisense', 'intronicAntisense']
2378	chr11:1012821-1036706	-	['InGen_OS_2081', 'InGen_OS_2080']	0.2	0.027	['exonicAntisense', 'exonicAntisense']
2383	chr17:81703371-81720539	+	['InGen_OS_3045', 'InGen_OS_3044']	0.2	0.019	['intronicAntisense', 'intronicAntisense']
2394	chr14:94376747-94390693	-	['InGen_OS_2539', 'InGen_OS_2541', 'InGen_OS_2540']	0.8	0.000	['exonicAntisense', 'exonicAntisense', 'exonicAntisense']
2416	chr3:195959748-195990318	-	['InGen_SS_Miss_232', 'InGen_SS_Miss_231']	0.2	0.007	['intronicSense', 'intronicSense']
2433	chr19:13795460-13832230	+	['InGen_SS_Miss_783', 'InGen_SS_Miss_785']	0.3	0.001	['intronicSense', 'intronicSense']
2497	chr1:228208130-228378874	+	['InGen_SS_Miss_54', 'InGen_SS_Miss_53', 'InGen_SS_Miss_52']	0.5	0.000	['intronicSense', 'intronicSense', 'intronicSense']
2517	chrX:73946555-74293574	-	['InGen_SS_Miss_857', 'InGen_SS_Miss_858', 'InGen_SS_Miss_868', 'InGen_SS_Miss_866', 'InGen_SS_Miss_865', 'InGen_SS_Miss_864', 'InGen_SS_Miss_863', 'InGen_SS_Miss_862', 'InGen_SS_Miss_861', 'InGen_SS_Miss_860', 'InGen_SS_Miss_869', 'InGen_SS_Miss_870', 'InGen_SS_Miss_871', 'InGen_SS_Miss_872', 'InGen_SS_Miss_873', 'InGen_SS_Miss_874', 'InGen_SS_Miss_875', 'InGen_SS_Miss_876', 'InGen_SS_Miss_859']	0.9	0.000	['intronicSense', 'intronicSense']
2520	chr9:99585786-99819889	-	['InGen_SS_Miss_449', 'InGen_SS_Miss_448']	0.5	0.000	['intronicSense', 'intronicSense']
2525	chr9:16409503-16870843	-	['InGen_SS_Miss_440', 'InGen_SS_Miss_441']	0.4	0.000	['intronicSense', 'intronicSense']
2528	chr20:1325405-1378734	+	['InGen_SS_Miss_809', 'InGen_SS_Miss_810']	0.2	0.002	['intronicSense', 'intronicSense']
2539	chr9:137306896-137423262	-	['InGen_SS_Miss_912', 'InGen_SS_Miss_911', 'InGen_SS_Miss_910', 'InGen_SS_Miss_909']	0.3	0.007	['intronicSense', 'intronicSense', 'intronicSense', 'intronicSense']
2542	chr2:110007675-110010783	+	['OutGen_Disc_Miss_71', 'OutGen_Disc_Miss_333']	0.3	0.000	['intronicSense', 'intronicSense']
2557	chr8:11842524-11869448	-	['InGen_OS_1607', 'InGen_OS_1606']	0.6	0.000	['exonicAntisense', 'exonicAntisense']
2560	chr22:39312882-39320389	-	['InGen_OS_3441', 'InGen_OS_3440']	0.6	0.000	['exonicAntisense', 'exonicAntisense']

Supplementary Table 1. Encode Dataset description

RNA	Classification	Tumor_Normal	Tissue
A172	cell	tumor	brain
A172_rep	cell	tumor	brain
A375	cell	tumor	skin
A375_rep	cell	tumor	skin
BipNeuFromIPSC	cell		neuron
BipNeuFromIPSC_rep2	cell		neuron
CD34	cell		blood
Caki2	cell	tumor	kidney
Caki2_rep	cell	tumor	kidney
CardioMyocytesFromhESC	cell		cardiomyocite
CardioMyocytesFromhESC_rep2	cell		cardiomyocite
Daoy	cell	tumor	brain
Daoy_rep	cell	tumor	brain
G401	cell	tumor	kidney
G401_rep	cell	tumor	kidney
GM12878	cell		lymphoblast
GM12878_rep	cell		lymphoblast
H4	cell	tumor	brain
H4_rep	cell	tumor	brain
H7-hESC	ESC		ESC
H7-hESC_rep2	ESC		ESC
HSMMdiffInMyotubes	cell		myoblast
HSMMdiffInMyotubes_rep2	cell		myoblast
HT1080	cell	tumor	connective_tissue
HT1080_rep	cell	tumor	connective_tissue
HT_29	cell	tumor	colon
HT_29_rep	cell	tumor	colon
HVMFfemale	cell		villous_placental_tissue
HVMFmale	cell		villous_placental_tissue
HepFromH9	cell		
HepFromH9_rep2	cell		
K562CytosolLongTotLecuyer	fraction		
K562CytosolLongTotLecuyer_rep	fraction		
K562InsolLongTotLecuyer	fraction		
K562InsolLongTotLecuyer_rep2	fraction		
K562LongTotalGraveleyB	fraction		
K562LongTotalGraveleyB_rep2	fraction		
K562LongTotalGraveley	fraction		
K562LongTotalGraveley_rep2	fraction		
K562LongTotalLecuyer	fraction		
K562LongTotalLecuyer_rep2	fraction		
K562MembraneLongTotalLecuyer	fraction		
K562MembraneLongTotalLecuyer_rep2	fraction		
K562chromatin_rep2	fraction		

K562nuclearLecuyer	fraction		
K562nuclearLecuyer_rep2	fraction		
K562nucleolus	fraction		
K562nucleolus_rep2	fraction		
K562nucleoplasmic	fraction		
K562nucleoplasmic_rep2	fraction		
Karpas422	cell	tumor	blood
Karpas422_rep	cell	tumor	blood
LHCN-M24dDiff	cell		Skeletal myoblasts
LHCN-M24dDiff_rep2	cell		Skeletal myoblasts
LHCN_M2	cell		Skeletal myoblasts
LHCN_M2_rep	cell		Skeletal myoblasts
M059J	cell	tumor	brain
M059J_rep	cell	tumor	brain
MG63	cell	tumor	bone
MG63_rep	cell	tumor	bone
NCI_H460	cell	tumor	lung
NCI_H460_rep	cell	tumor	lung
NeuronFromH9	cell		neuron
NeuronFromH9_rep2	cell		neuron
OCI_LY7	cell	tumor	blood
OCI_LY7_rep	cell	tumor	blood
PC_3	cell	tumor	prostate
PC_3_rep	cell	tumor	prostate
RPMI_7951	cell	tumor	skin
RPMI_7951_rep	cell	tumor	skin
SJCRH30	cell	tumor	muscle
SJCRH30_rep	cell	tumor	muscle
SJSA1	cell	tumor	bone
SJSA1_rep	cell	tumor	bone
SK_MEL5	cell	tumor	skin
SK_MEL5_rep	cell	tumor	skin
SK_N_DZ_DMSO	cell	tumor	brain
SK_N_DZ_DMSO_rep	cell	tumor	brain
SmooMusFromH9	cell		
SmooMusFromH9_rep2	cell		
hMNC-PB	cell		
hMSC-AT	cell		
hMSC-AT_rep2	cell		
hMSC-BMfemale	cell		
hMSC-BMmale	cell		
hMSC-UCfemale	cell		
hMSC-UCmale	cell		
iPSC_GM23338	iPSC		
iPSC_GM23338_rep	iPSC		
adrenal_gland_fem_51	tissue	normal	adrenal_gland
adrenal_gland_fem_53	tissue	normal	adrenal_gland
ascending_aorta_fem_53	tissue	normal	aorta
bodyPancreas_fem_51	tissue	normal	pancreas
breastEpithelium_ma_37	tissue	normal	breast
cerebellum_fem_19w	tissue	normal	brain
cerebellum_fem_37w	tissue	normal	brain
diencephalon_fem_20w	tissue	normal	brain
diencephalon_ma_22w	tissue	normal	brain

esophagusMuscularisMucosa_fem_51	tissue	normal	muscle
esophagusMuscularisMucosa_fem_53	tissue	normal	muscle
esophagusMuscularisMucosa_ma_37	tissue	normal	muscle
esophagusMuscularisMucosa_ma_54	tissue	normal	muscle
esophagusSquamousEpithelium_ma_37	tissue	normal	gastric
frontalCortex_fem_20w	tissue	normal	brain
frontalCortex_ma_22w	tissue	normal	brain
gastrocnemius_medialis_ma_37_rep	tissue	normal	muscle
gastrocnemius_medialis_ma_37	tissue	normal	muscle
gastrocnemius_medialis_fem_51	tissue	normal	muscle
gastrocnemius_medialis_fem_53	tissue	normal	muscle
gastroesophagealSphincter_fem_51	tissue	normal	gastric
gastroesophagealSphincter_fem_53	tissue	normal	gastric
gastroesophagealSphincter_ma_37	tissue	normal	gastric
gastroesophagealSphincter_ma_54	tissue	normal	gastric
heart_fem_fetal_19w	tissue	normal	heart
heart_fem_fetal_28w	tissue	normal	heart
heart_ma_34	tissue	normal	heart
heartLeftVentricle_fem_51	tissue	normal	heart
heartLeftVentricle_fem_53	tissue	normal	heart
liver_fem_6	tissue	normal	liver
liver_ma_32	tissue	normal	liver
lowLegSkin_fem_51	tissue	normal	skin
lowLegSkin_fem_53	tissue	normal	skin
metanephros_fem_fetal_20w	tissue	normal	fetal
metanephros_fem_fetal_24w	tissue	normal	fetal
omentalFatPad_fem_53	tissue	normal	omental
omentalFatPad_ma_37	tissue	normal	omental
omentalFatPad_ma_54	tissue	normal	omental
ovary_fem_53	tissue	normal	ovary
parietalLobe_fem_fetal_24w	tissue	normal	brain
parietalLobe_ma_fetal_22w	tissue	normal	brain
peyerPatch_fem_51	tissue	normal	peyer_patch
peyerPatch_fem_53	tissue	normal	peyer_patch
peyerPatch_ma_37	tissue	normal	peyer_patch
peyerPatch_ma_54	tissue	normal	peyer_patch
prostate_37	tissue	normal	prostate
prostate_54	tissue	normal	prostate
rightAtriumAuricular_fem_53	tissue	normal	heart
sigmoidColon_fem_51	tissue	normal	colon
sigmoidColon_ma_37	tissue	normal	colon
skeletalMuscle_fem_19w	tissue	normal	muscle
skeletalMuscle_ma_22w	tissue	normal	muscle
skin_fem_fetal_22w	tissue	normal	fetal
skin_fem_fetal_24w	tissue	normal	fetal
spinalCord_fem_24w	tissue	normal	bone_marrow
spinalCord_ma_22w	tissue	normal	bone_marrow
spleen_fem_53	tissue	normal	spleen
spleen_ma_54	tissue	normal	spleen
stomach_fem_51	tissue	normal	stomach

stomach_fem_53	tissue	normal	stomach
stomach_ma_37	tissue	normal	stomach
stomach_ma_54	tissue	normal	stomach
subcutaneousAdipose_fem_51	tissue	normal	fat
subcutaneousAdipose_ma_54	tissue	normal	fat
suprapubicSkin_fem_51	tissue	normal	skin
suprapubicSkin_ma_37	tissue	normal	skin
temporalLobe_fem_fetal_20w	tissue	normal	brain
temporalLobe_fem_fetal_24w	tissue	normal	brain
testis_37	tissue	normal	testis
testis_54	tissue	normal	testis
thoracicAorta_ma_37_rep	tissue	normal	aorta
thoracicAorta_ma_37	tissue	normal	aorta
thyroid_fem_51	tissue	normal	thyroid
thyroid_fem_53	tissue	normal	thyroid
tibialNerve_ma_54	tissue	normal	nerve
tongue_fem_20w	tissue	normal	tongue
tongue_fem_24w	tissue	normal	tongue
transverse_colon_fem_51	tissue	normal	colon
transverse_colon_ma_37	tissue	normal	colon
upperLobeLeftLung_fem_51	tissue	normal	lung
upperLobeLeftLung_ma_54	tissue	normal	lung
uterus_51	tissue	normal	uterus
uterus_53	tissue	normal	uterus
vagina_53	tissue	normal	vagina

PhD publications:

1) SNPs and Somatic Mutation on Long Non-Coding RNA: New Frontier in the Cancer Studies?
Minotti L, Agnoletto C, Baldassari F, Corrà F and Volinia S
High Throughput. 2018 Nov 16;7(4). pii: E34. doi: 10.3390/ht7040034.

2) Heterogeneous expression of EPCAM in human circulating tumour cells from patient-derived xenografts
Agnoletto C, **Minotti L**, Soumare LB, Pasquali L, Galasso M, Corrà F, Baldassari F, Judde JB, Cairo S and Volinia S Biomark Res. 2018 Oct 30;6:31.. eCollection 2018 doi: 10.1186/s40364-018-0145-8

3) The network of non-coding RNAs in cancer drug resistance.
Corrà F, Agnoletto C, **Minotti L**, Baldassari F, Volinia S
Front Oncol. 2018 Aug 29;8:327. eCollection 2018. Review. doi: 10.3389/fonc.2018.00327.

4) Screen for MicroRNA and Drug Interactions in Breast Cancer Cell Lines Points to miR-126 as modulator of CDK4/6 and PIK3CA Inhibitors.
Baldassari F, Zerbinati C, Galasso M, Corrà F, **Minotti L**, Agnoletto C, Previati M, Croce CM, Volinia S. Front Genet. 2018 May 18;9:174. eCollection 2018. doi:10.3389/fgene.2018.00174.

5) Loss of miR-204 expression is a key event in melanoma.
Galasso M, Morrison C, **Minotti L**, Corrà F, Zerbinati C, Agnoletto C, Baldassari F, Fassan M, Bartolazzi A, Vecchione A, Nuovo GJ, Di Leva G, D'Atri S, Alvino E, Previati M, Nickoloff BJ, Croce CM, Volinia S. Mol Cancer. 2018 Mar 9;17(1):71. doi: 10.1186/s12943-018-0819-8.

6) A long non-coding RNA inside the type 2 tranglutaminase gene tightly correlates whit the expression of its transcriptional variants.
Minotti L, Baldassari F, Galasso M, Volinia S, Bergamini CM, Bianchi N. Amino Acids. 2018 Apr;50(3-4):421-438. Epub 2018 Jan 8. doi: 10.1007/s00726-017-2528-9.

7) Profiling of the Predicted Circular RNAs in Ductal In Situ and Invasive Breast Cancer: A Pilot Study.
Galasso M, Costantino G, Pasquali L, **Minotti L**, Baldassari F, Corrà F, Agnoletto C, Volinia S. Int J Genomics. 2016;2016:4503840. Epub 2016 Nov 14.