academicJournals

Vol. 16(25), pp. 1404-1427, 21 June, 2017
DOI: 10.5897/AJB2016.15860
Article Number: E56524C64872
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

DEXseq and Cuffdiff approaches weighing differential spliced genes exons modulation in estrogen receptor β (Erβ) breast cancer cells

Noel Dougba Dago[1,2]*, Nafan Diarrassouba[1], Inza Jesus Fofana[1], Jean-Luc Aboya Moroh[1], Hermann-Désiré Lallié[1], Didier Martial Yao Saraka[1], Oléfongo Dagnogo[3], Souleymane Silué[1], Armel Herve Nwabo Kamdje[4], Lamine Baba-Moussa[5] and Alessandro Weisz[2]

[1]Unité de Formation et de Recherche (UFR) Sciences Biologiques, Unité Pédagogique et de Recherche de Génétique, Université Péléforo Gon Coulibaly BP1328 Korhogo, Côte d'Ivoire.
[2]Laboratory of Molecular Medicine and Genomics, Department of Medicine and Surgery, University of Salerno, Via S. Allende, 1, Baronissi, SA 84081, Italy.
[3]Unité de Formation et de Recherche (UFR) Biosciences, Laboratoire de Pharmacodynamie Biochimie, Université Felix Houphoet-Boigny BP V 34 Abidjan 01, Côte d'Ivoire.
[4]Department of Biomedical Sciences, Faculty of Science, University of Ngaoundere P. O. Box 454 Ngaoundere, Cameroon.
[5]Laboratoire de Biologie et de Typage Moléculaire en Microbiologie, Faculté des Sciences et Techniques, Université d'Abomey-Calavi, Cotonou, Benin.

Received 28 December, 2016; Accepted 9 June, 2017

While the alternative transcription and splicing mechanisms have long been known for some genes like oncogenes, their prevalence in almost all multi-exon genes has been recently realized with the increasing application of high-throughput experimental methods, named Next Generation RNA Sequencing (NGS). Henceforth, understanding the regulation of these processes in comparisons between cell types and cancer requests, sensitive and specific bioinformatics as well as bio-statistic approaches, that is, Cufflinks/Cuffdiff, DEXseq and RESEM, detecting gene transcript/isoforms and exons abundance is necessary. Isoforms and exons expression analysis by NGS is complicated by several sources of measurement variability causing numerous statistical defies. Here, with the purpose of minimizing this statistical challenge, we integrated both Cufflinks/Cuffdiff and DEXseq bioinformatics approaches assessing whole alternative splicing (AS) events, focusing on alternative transcripts regulation and their exons modulation respectively, by processing our previous prepared Estrogen Receptor β ($Erβ^+$ and $Erβ^-$) breast cancer (BC) cells, stimulated by estradiol (E_2). Results showed that Estradiol (E_2) induced $Erβ^+$ BC ($Erβ^+E_2$), exhibited dissimilar reply as opposed to the other's analyzed BC cell lines in terms of intragenic, exons and junction reads count ratio. Relationship analysis between expressed genes and transcript isoforms, suggested a substantial role of alternative promoters in AS event occurrence in $Erβ^+$ BC as opposed to $Erβ^-$ BC. Indeed, merging Cufflinks/Cuffdiff and DEXseq approaches, 79 multi-exon genes were detected as statistically differentially modulated (spliced) in $Erβ^+$ hormone induced BC cell line, and around 38% of these spliced genes claimed to be induced by alternative promoters. The present survey discriminated between several cancer specific alternative splicing genes like *LIFR* a BC metastasis suppressor, *PBX1* a pioneer factor defining aggressive $Erβ^-$ BC and *PHLPP2* a tumor suppressor, as exhibiting significant exon modulation in early AS occurrence in hormone responsive $Erβ^+$ BC exclusively. Although, our findings supported dissimilar reply comparing both Cuflinks/Cuffdiff and DEXseq approaches called AS events, it is noteworthy to underline their relative agreement, evaluating spliced genes functional annotation as well as their complementarity performing whole AS survey. This study therefore proposed the integration between Cufflinks/Cuffdiff and DEXseq tools as a reasonable complementary methodology assessing full AS pattern in hormone responsive Erβ BC cells.

Key words: Cufflinks/Cuffdiff, DEXseq, RNA-Seq, alternative splicing (AS), exons, transcript isoforms, estrogen receptor β (ERβ), breast cancer (BC) cells.

INTRODUCTION

Alternative splicing is a central cellular process that produces different mRNA transcript isoforms from a single gene. The qualitative and quantitative identification of such transcript isoforms is more complex as well as essential for understanding the different roles of alternatively spliced genes occurrence in a cell. However, detection of disease-specific transcript isoforms is an important task because aberrant splicing is known to be responsible for various diseases (Kim et al., 2008) and associated with different cancer types (Christofk et al., 2008; Venables et al., 2009). Several studies provided an intriguing insight into the mechanism of cancer specific alternative transcription and alternative splicing, which have long been implicated in the development of cancer. Cancer-specific isoforms have enhanced proliferative, invasive, and migratory abilities and provide a survival advantage to the tumor cells, suggesting that there is specific manipulation of the alternative event regulation in cancer that is beyond the general lack of fidelity of the splicing or the alternative transcription regulatory machinery. It is conceivable that the balanced expression of isoforms, rather than just activation or inhibition of those genes, may hold the key to impeding tumor growth and accordingly it is important to target the disease associated genes at the isoform level rather than at the gene level. The well-known application of exon arrays (Moller-Levet et al., 2009) and the advent of massive parallel sequencing named Next Generation RNA Sequencing (NGS-RNA-Seq) are allowing whole cancer genomes and transcriptomes to be sequenced with extraordinary speed and accuracy, providing insight into the bewildering complexity of isoform-specific expression in cancer genomes (Cancer Genome Atlas, 2012). Detecting alternative isoform regulation is inherently difficult in RNA-Seq, as sequencer reads are often one or more orders of magnitude shorter than the transcripts themselves. While there are several utilities that attempt to de-convolute read data into isoform abundances, the accuracy and robustness of these methods is difficult to establish (Chandramohan et al., 2013; Zhang et al., 2014). In addition, transcript isoforms expression estimates seem to vary considerably between different tools, and generally depend on the quality and completeness of the transcript assembly (Kanitz et al., 2015; Rehrauer et al., 2013). Existing well-established methods detect alternative splicing process mainly by considering sequencing reads that map uniquely to single isoforms or by assembling transcripts and estimating the most likely isoform abundance levels according to the given sequencing reads (Jeffrey and Zhong, 2011). Short read assemblers have been developed for genome and transcriptome assembly, that is, Velvet (Zerbino and Birney, 2008), Scripture (Guttman et al., 2010), Cufflinks (Trapnell et al., 2010) and others (Jeffrey and Zhong, 2011). In literature, the term alternative splicing has been used to describe both alternative transcription and splicing events. However, the variation in the pattern of intron removal, exon joining, and the addition of a poly-A tail on a single pre-mRNA result in alternatively splice mature mRNAs. These various alternative events have been identified in different cells and tissues by application of RNA sequencing (RNA-Seq) next generation sequencing (NGS) based methods in genome wide studies (Sultan et al., 2008; Wang et al., 2008; Pal et al., 2011). Furthermore, it is estimated that there are 263,772 exons in the human genome and approximately, 22% of these exons participate in alternative splicing phenomena (Pal et al., 2012), suggesting that performance assessment of whole alternative splicing occurrence in a genomic survey needed strong bioinformatics/bio-statistical tools characterizing statistical exon modulation. However, a systematic assessment of transcriptome assemblies is difficult because appropriate quality metrics have not been established yet and require a well-defined gold standard that is difficult to find (Jeffrey and Zhong, 2011). Recently, Anders et al. (2012) published a bioinformatics tool called DEXseq, which process exon differential survey in RNA-Seq genomic and/or transcriptomic experimentation. Rather than using an assembly approach and comparing abundance levels of predicted transcripts, DEXseq avoids the assembly step and calculates probabilities values (p-values) for every annotated exon (statistical estimation of exons abundance). Then, the emergence of NGS provides an exciting new technology to analyse alternative splicing on a large scale including differential expression analysis of exons of multi-exon genes. Hence, we believe that combining transcript isoforms differential expression survey with gene and/or transcript isoform exons modulation, could increase researcher alternative splicing phenomena comprehension processing two or more analysed cells and/or tissues conditions. Here we investigated the relation-ship between significantly differentially spliced genes (DSGs) based on alternative transcript isoforms measurement of our previous developed Erβ^{+} breast cancer (BC) cell line model and their corresponding exons abundance modulation merging both Cufflikns/Cuffdiff (Trapnell et al., 2013) and DEXseq (Anders et al., 2012) bioinformatics/bio-statistical approaches respectively, emphasizing alternative splicing manifestation in developed Erβ^{+} and Erβ^{-} breast cancer cell line models (Grober et al., 2011; Paris et al., 2012;

*Corresponding author. E-mail: dgnoel7@gmail.com Tel: (+39) 3381426596 or (+225) 48397811 or (+225) 44595981.

Nassa et al., 2011; Tarallo et al., 2011).

MATERIALS AND METHODS

Erβ$^+$ breast cancer cell model preparation and gene and transcript isoforms differential expression analysis

MCF7 and 5B12 breast cancer (BC) cell line model have been prepared to mimic Erβ/Erβ; (Erβ$^+$) and Erβ/Erα; (Erβ$^-$) breast cancer (BC). The preparation of Erβ$^+$ and Erβ$^-$ BC cell lines have been entirely described in Nassa et al. (2011), Grober et al. (2011) and Dago et al. (2015). Differential genes, transcript isoforms and exons expression analysis, assessing the effect of early stimulation of estradiol (E_2) hormone on alternative splicing occurrence in breast tumor cell models, have been performed by Cufflinks/Cuffdiff (Trapnell et al., 2013) and DEXseq (Anders et al., 2012) bioinformatics/bio-statistic packages respectively. Next generation RNA sequencing (NGS RNA-Seq) data used for the present analysis have been deposited in the Gene Expression Omnibus genomics data public repository (http://www. ncbi.nlm.nih.gov/geo/) with Accession Number GSE64590.

RNA sequencing (RNA-Seq) data generation and gene and/or transcript isoforms expression measurement

1 µg micrograms of high-quality total mRNA was used as starting material for the Illumina mRNASeq library preparation kit and was prepared to manufacturer's directions (Illumina). Libraries were sequenced on the Illumina Genome Analyzer II as 101 base pair paired-end reads. Tophat v.2.0.8 (Kim et al., 2013) was used to align all reads including junction-spanning reads back to the human genome (*Homo sapiens* Ensembl GRCh37). The reads alignment quality and distribution were estimated using SAMtools. Cufflinks v2.1.1 (Trapnell et al., 2013) was used to identify differential spliced genes, isoform transcript and gene expression changes between analyzed experimental groups and/or conditions. We defined statistical significance in expression q-value and/or adjusted p-value for multiple testing ≤ 0.05, and Fragments per Kilobase of exon per Million reads mapped (FPKM)>0.5, since FPKM is a measure of expression used in high-throughput sequencing data that is normalized for both transcript length and total number of reads sequenced.

DEXseq measuring exons differential modulation usage by RNA-sequence

DEXseq is a package for the statistical programming language R (R Development Core Team, 2009) available as open source software via the Bioconductor project (Gentleman et al., 2004). For the preparation steps, namely the flattening of the transcriptome annotation to counting bins and the counting of the reads overlapping each counting bin, two Python scripts are provided, which are built on the HTSeq framework (Anders et al., 2015). The first script takes a GTF file with gene models and transforms it into a GFF file listing counting bins, the second takes such a GFF file and an alignment file in the SAM format and produces a list of counts. The R package is used to read these counts, estimate the size factors and dispersions, fit the dispersion-mean relation and test for differential exon usage. Then, exons with significant change and/or modulation at a false discovery rate lower or equal to 0.05 (FDR≤0.05) have been selected as involve in alternative splice events. All bioinformatics and biostatistics analyses and comparisons were implemented and performed using in-house scripts written in Unix and R.

Functional annotation gene ontology analysis by DAVID

We performed Database for Annotation, Visualization and Integrated Discovery (DAVID http://david.abcc.ncifcrf.gov/) analysis (Glynn et al., 2003) focusing exclusively on the set of differentially spliced gene transcripts and/or isoforms, discriminated by both Cufflinks/Cuffdiff and DEXseq approaches. Then, we performed a gene ontology (GO) survey (FDR ≤0.05 with at least 2 fold enrichment) by processing significantly differentially spliced genes, evaluating alternative transcript and exon change and/or modulation in multi-exon genes.

RESULTS

Reads sequences from both Erβ+ and Erβ- human BC cell line models high-throughput RNA sequencing (RNA-Seq) statistical analysis

mRNA sequencing experiment basing on illumina genome analyzer II (GAII) processing both Erβ+ and Erβ- hormone induced breast cancer (BC) cell lines, yielded approximately 94911602-99876796 million pair-end read (101 bp) sequences (Table 1). From these reads, low quality sequences, were eliminated, resulting in around 65-71 million reads corresponding to 68-70.2% of total reads for each processed replicate sample (Table 1). In total around 65-71 million reads were aligned to the *Homo sapiens* Ensembl GRCh37 reference genome (Table 1). The number of reads per genes and transcript isoforms were further normalized to Fragment per Kilobases of exon per Million mapped reads (FPKM). Then, in order to include a maximum number of genes and transcript isoforms, reducing as possible statistical type I error in calling differentially modulated and spliced genes as well as exon modulation events, we adopted 0.5 FPKM, as genes and/or transcript isoforms expression level threshold in both processed Erβ$^+$ and Erβ$^-$ BC cell exemplars (Figure 1). Student test analysis, based on both loaded junctions and found junction parameters from transcriptome and/or genome reconstruction through Cufflinks package, suggested a reasonable difference (p-value <0.01) between Erβ$^+$ BC cell line under estradiol stimulus (Erβ$^+E_2$) and the other's analyzed BC cell conditions (Table 1). In other words, the present result supported that Erβ$^+$ and Erβ$^-$ BC cell lines in normal conditions (Erβ$^+$noE_2, Erβ$^-$noE_2) as well as Erβ$^-$ BC cell line induced by E_2 (Erβ$^-E_2$), exhibited the same behaviors as opposed to Erβ$^+E_2$ breast cancer cell reacting to E_2 induction (Table 1).

Since each analyzed condition have been processed in three technical replicates, Erβ$^+E_2$_Rep1 indicates the replicate 1 of Erβ$^+$ BC cell line condition under estradiol (E_2) treatment, while Erβ$^+$noE_2_Rep1 indicates replicate 1 of the same cell line with any E_2 treatment The same nomenclature has been adopted for Erβ$^-E_2$ and Erβ$^+$noE_2 analyzed samples conditions.

Table 1. Summary of RNA-Seq reads sequences for each analyzed experimental condition ($Erβ^+E_2$, $Erβ^+noE_2$, $Erβ^-E_2$ and $Erβ^-noE_2$ conditions).

Samples	Sequenced fragment	Aligned fragment	Percent aligned (%)	Loaded junctions	Found junctions
$Erβ^+E_2$_Rep1	97634516	68420165	70.21	197356	165750
$Erβ^+E_2$_Rep2	98586566	68906448	70.02	198164	166736
$Erβ^+E_2$_Rep3	96977352	71363401	70.10	197142	165686
$Erβ^+noE_2$_Rep1	95754654	65171906	68.15	229834	190798
$Erβ^+noE_2$_Rep2	96703300	65626636	68.00	244701	198966
$Erβ^+noE_2$_Rep3	94911602	64472829	68.08	243329	198181
$Erβ^-E_2$_Rep1	97994626	69120928	70.67	243336	198938
$Erβ^-E_2$_Rep2	99083778	69790775	70.56	243590	199088
$Erβ^-E_2$_Rep3	97277796	68582921	70.62	243082	198391
$Erβ^-noE_2$_Rep1	98893200	67641680	68.44	242723	197682
$Erβ^-noE_2$_Rep2	99876796	68208308	68.41	243016	197921
$Erβ^-noE_2$_Rep3	98117332	67079584	68.48	241857	196947

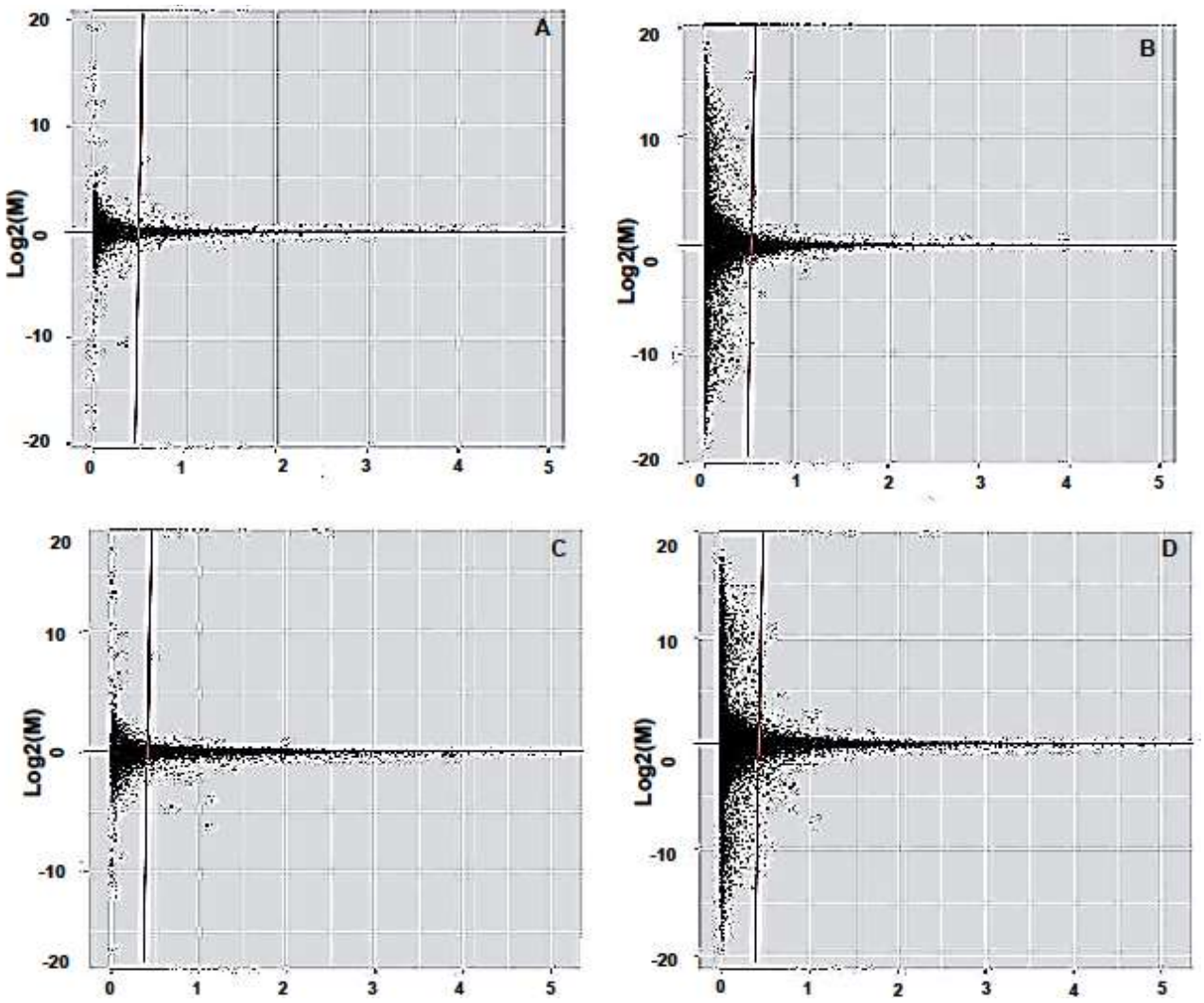

Figure 1. Panels A, B, C and D (MAplot graphics) indicate log 2 mensuration (normalized expression value) referred to genes and/or gene transcript isoforms expression of processed hormone responsive $Erβ^-$ and $Erβ^+$ breast cancer (BC) cell line models respectively.

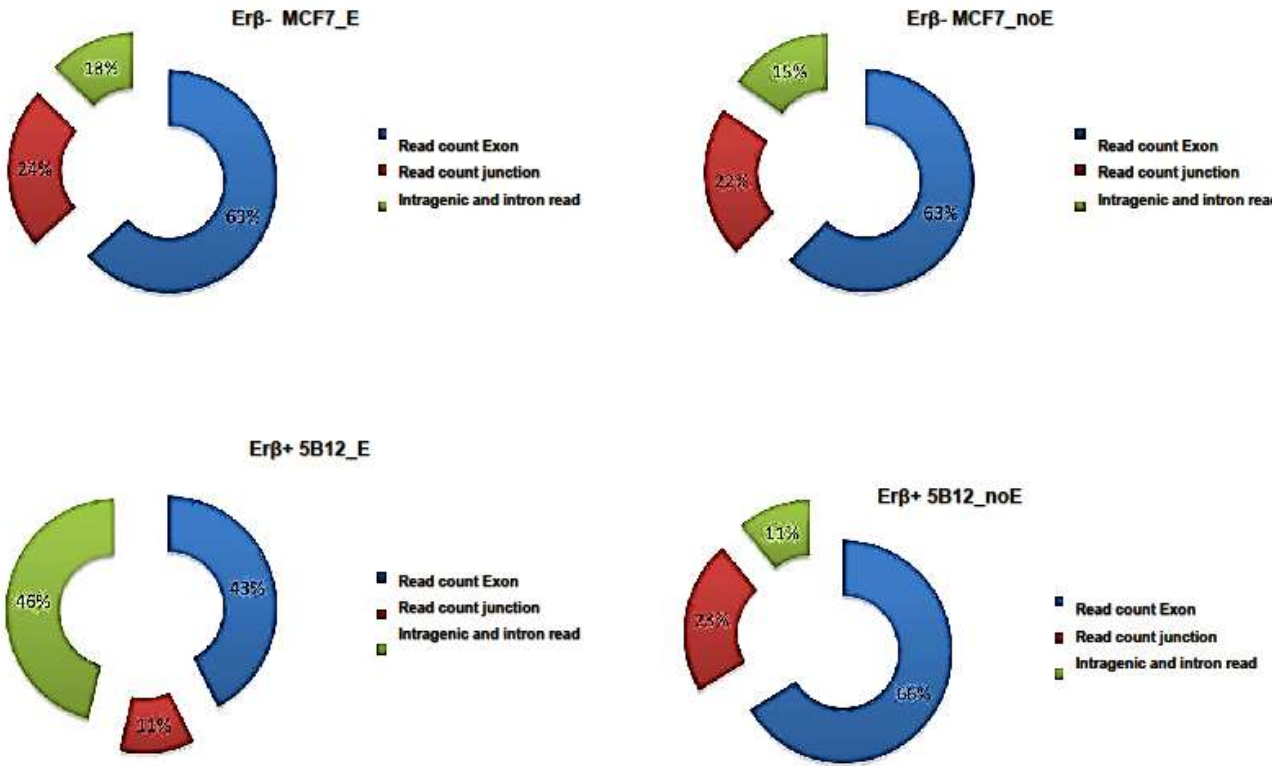

Figure 2. Summary of RNA-Seq reads sequences re-distribution on *Homo sapiens* Ensembl GRCh37 genome assessing alternative splicing occurrence in hormone responsive BC cells by Tophat/Cufflinks and DEXseq/HTseq tools.

Genomic re-distribution of RNA sequencing (RNA-Seq) reads sequences evaluating breast cancer cell line reply to estradiol (E_2) hormone stimulus

Next, we focused on genomic re-distribution of reads sequences, merging read count analysis results from HTSeq/DEXseq with those from Tophat2/Cufflinks software's for each analyzed hormone induced Erβ BC cell line conditions ($Erβ^+E_2$, $Erβ^+noE_2$, $Erβ^-E_2$ and $Erβ^-noE_2$) appraising early alternative splicing events monitoring differential spliced genes and exons change in breast tumor. The present survey highlighted a strong difference in read distributions between $Erβ^+E_2$ BC cells and the other's analyzed conditions ($Erβ^+noE_2$, $Erβ^-E_2$ and $Erβ^-noE_2$) (Figure 2). These results seem to be in agreement with above reported loaded and/or found junction analysis (Table 1), supporting subtly an agreement between HTSeq/DEXseq and Tophat2/ Cufflinks approaches, in genome reconstruction process and/or in genomic reads re-distribution analysis. Furthermore, the present survey exhibited a constancy rate, measuring exon read count as well as intragenic and intron reads values comparing $Erβ^+noE_2$, and both $Erβ^-E_2$ and $Erβ^-noE_2$ breast cancer cells (Figure 2), hypothesizing a similitude between the former's assessing alternative splicing pathway in the present analyzed BC cells, reinforcing the link between estrogen receptor beta (Erβ) and early transcription and mRNA splicing events in hormone responsive BC cells. Considering as a whole, the present results supported a dynamic molecular reply of $Erβ^+$ BC cells as opposed to $Erβ^-$ BC in terms of alternative splicing events (Figure 2).

Assessment of expressed genes and their transcript isoforms proportion comparing hormone responsive $Erβ^+$ and $Erβ^-$ breast cancer cell lines

We evaluated change in differential expressed gene and transcript isoforms including exclusively genes and transcript isoforms that exhibit a FPKM expression value ≥0.5 and significant statistical differential change at an adjusted p-value ≤0.05 by Cufflinks/Cuddiff approach in processed hormone responsive BC cell lines. Then, basing on previous analysis (Figure 1), we processed in total 6714 and 52499 genes and transcript/isoforms respectively (Figure 3). As expected, change in gene and transcript isoforms strongly contrast between both analyzed estrogen hormone responsive $Erβ^+$ and $Erβ^-$ BC cell lines (Figure 3A and B) since, estimated Pearson correlations values comparing log 2 fold change assessing measurement between expressed genes and transcript isoforms resulted to lower 0.5 (R^2 =0.23 and R^2=0.31 for $Erβ^+$ and $Erβ^-$ BC cell respectively), advising

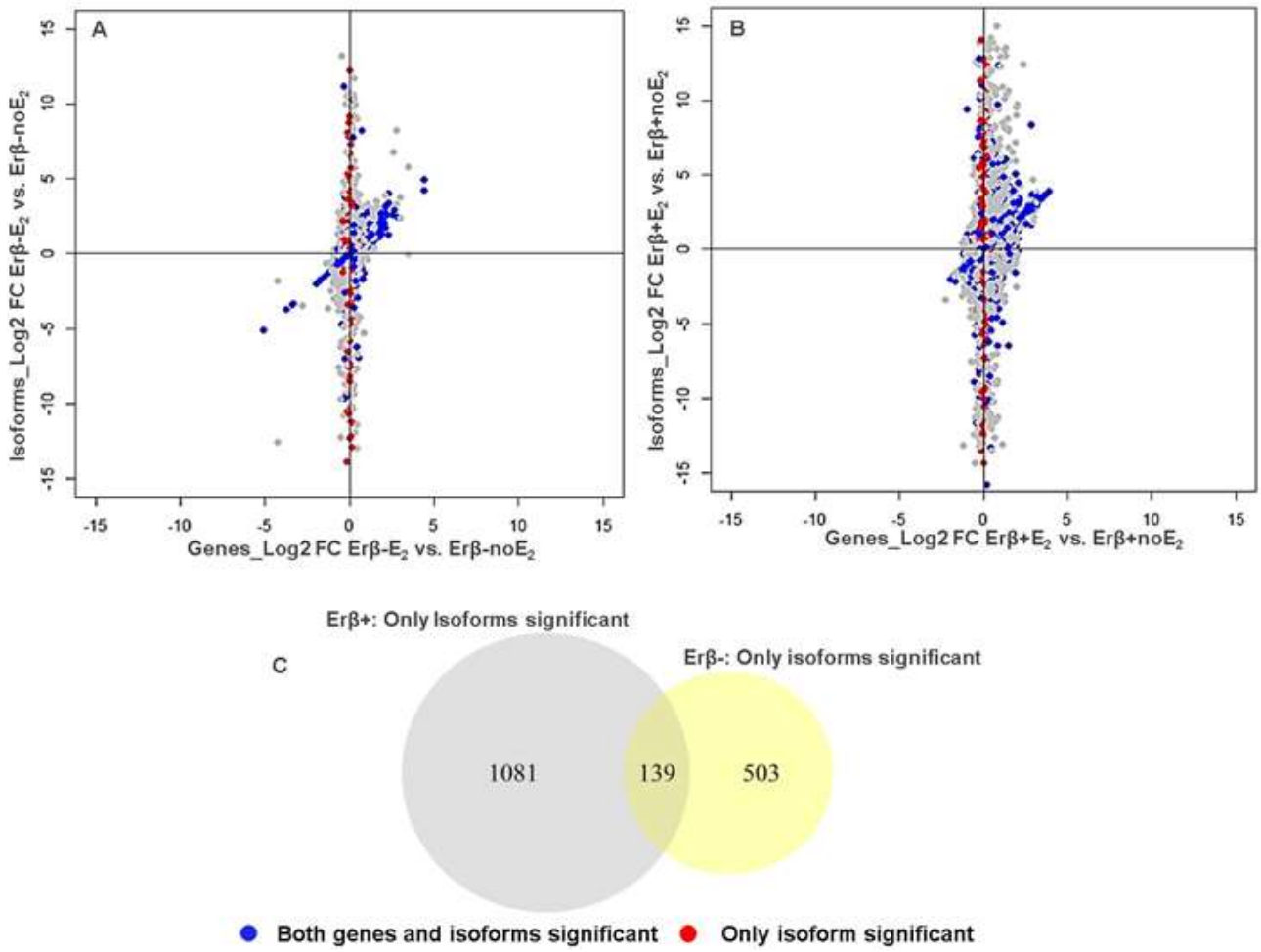

Figure 3. Log_2 fold change scatterplot comparing genes and transcript/isoforms expression in both hormone induced; $Erβ^-$ (A) and $Erβ^+$ or (B) BC cells. The x- coordinate is the gene expression fold change relative to $Erβ^-$ (A) or $Erβ^+$ (B) BC cells and the y-coordinate is the transcript isoform fold change relative to the same BC cells respectively. Blue dots represent the situation where both isoforms and genes, with an expression value FPKM ≥ 0.5, have been selected as statistically significantly differentially expressed at a p-adjusted value≤0.05. Red dots represent situation where only isoform expression is significantly altered. (C) Venn diagrams showing overlapping/divergence between hormone induced Erβ- and Erβ+ BC models analyzing cases where only transcript isoforms expression is significantly altered.

weak agreement between change in genes and their respective transcript isoforms, especially in $Erβ^+$ BC cells (Figure 3). Hence, we focused on the cases where only transcript isoforms expression were significantly altered, alerting alternative promoter usage in alternative transcript isoforms modulation as well as in alternative splicing event occurrence in estrogen responsive BC cell line. Indeed, we showed that situations for which transcript isoform expression was significantly alerted as opposed to their corresponding genes resulted 2 fold more in hormone induced $Erβ^+$ BC cell (Figure 3C). Taking together, the present results suspected a considerable involvement of alternative promoter usage in early alternative splicing occurrence in hormone responsive $Erβ^+$ BC cell lines as opposed to $Erβ^-$ BC cell lines (Figure 3).

Identification of differentially spliced genes (DSGs) in hormone responsive $Erβ^+$ breast cancer cell lines by Cufflinks/Cuffdiff approach

To identify the differences in splice ratios between $Erβ^+E_2$ and $Erβ^+noE_2$ BC cell line ($Erβ^+E_2$ vs. $Erβ^+noE_2$), we employed Cufflinks/Cuffdiff v2.1.1 package, which calculates the changes in the relative splice abundances by quantifying the square root of the Jensen Shannon divergence on all the primary transcripts that produce two or more isoforms. It is essential to note that the distributions of genes, and primary transcripts, and isoform expression level (FPKM) are comparable between the samples that are taken for the differential splicing test. Then, 213 genes randomly distributed on human chromosomes have been detected as significantly

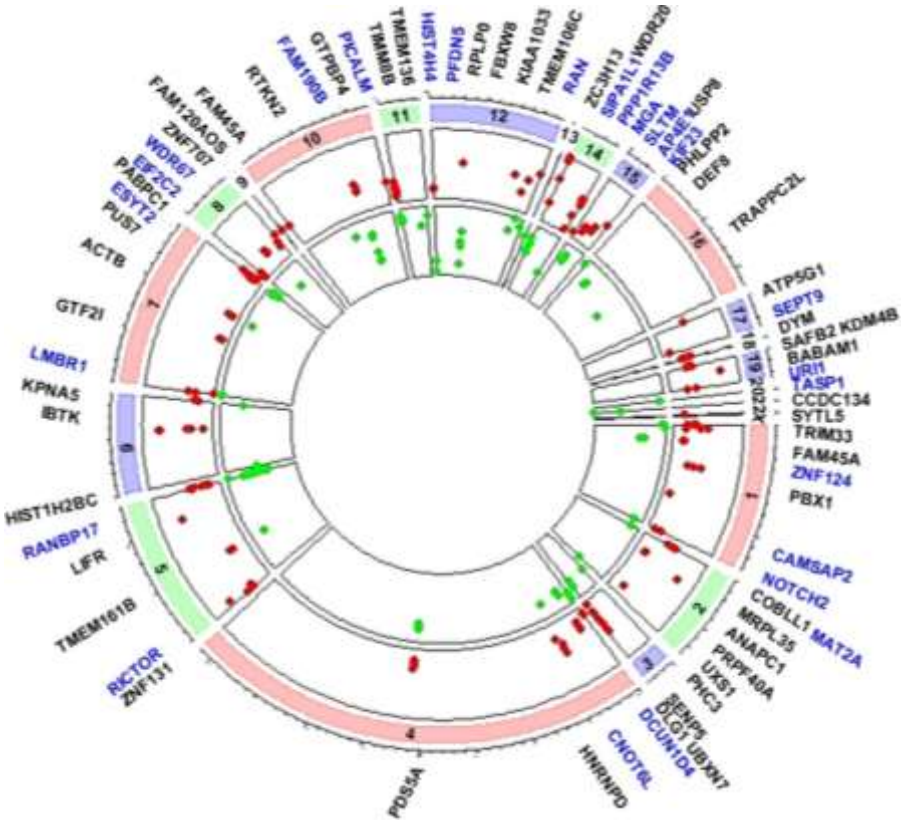

Figure 4. Exons modulation of differentially spliced genes discriminated by both Cufflinks/Cuffdiff and DEXseq in hormone responsive Erβ$^+$ BC cells. Red dots indicate up regulated exons while green dots designate down regulated exons in Erβ$^+$E$_2$ vs. Erβ$^+$noE$_2$ contrast assessing early AS pattern in hormone responsive BC cells. Genes in blue represent spliced multi-exonic genes under alternative promoter usage.

differentially spliced (DSGs) at a false discovery rate ≤0.05 (FDR ≤0.05) monitoring early AS occurrence in hormone responsive Erβ$^+$ BC cell lines (Figure 4 and Supplementary Table 1). Furthermore, basing on this result as well as on previous one (data non shown), we showed that Erβ$^+$ BC cell lines exhibited 3 fold more differentially spliced genes with respect to Erβ- BC cells replying to estradiol hormone stimulus (Supplementary Tables 1 and 2) confirming previous results and observations (Figures 2 and 3). Therefore, we were interested in investigating DSGs exons change monitoring early AS pathway in hormone responsive Erβ+ BC, since exons modulation have been recognized as suitable process understating mechanism of transcript isoforms expression, proposing the former's as potential tumors biomarkers as well as therapeutic target.

Assessment of differential spliced multi-exonic genes (DSGs), exons modulation by DEXseq in Erβ breast cancer cell lines

Here, we analyzed through the DEXseq approach, differential exons expression change, referring to previous detected differential spliced multi-exonic genes (Cufflinks/Cuffdiff) in hormone induced Erβ$^+$ BC cell line. This analysis looks for difference across conditions (in Erβ$^+$E$_2$ vs. in Erβ$^+$noE$_2$) between quantities that are directly observable from shotgun data (read count data), such as the relative usage of each exon. Then, performing the test for differential exons used considering exon with at least 10 reads count in at least one analyzed condition and controlling false discovery rate (FDR) with the Benjamini Hochberg method, at 5% threshold (statistical stringency), 3682 and 122 exons from 1438 and 75 multi-exonic genes, claimed to be significantly differentially modulated in hormone responsive Erβ$^+$ and Erβ$^-$ BC cell lines respectively (Supplementary Materials 1 and 2). Also, this result suggested that exons change ratio in analyzed differentially spliced multi-exonic genes was 1.6 fold more higher in hormone responsive Erβ$^+$BC cell lines when compared to Erβ$^-$ BC cell line, supporting a potential high number of significantly alternatively modulated transcript isoforms involvement characterizing processed hormone responsive Erβ$^+$BC cell line, highlighting the links between Erβ and early AS

Table 2. Assessment of spliced exon position (exon gene ID) in differentially spliced multi-exonic genes called by DEXseq approach.

Gene	DSGs Exon change Exons Gene ID E000-E010	DSGs Exons Change Exon Gene ID E011-E20	DSGs Exons Change Exon Gene ID E021-E30	DSGs Exons Change Exon Gene ID E031-E40	DSGs Exons Change Exon Gene ID E041-E50	DSGs Exons Change Exon Gene ID > E050
(Erβ/Erβ): Erβ$^+$ BC Cell Line	28.04%	22.4%	16.61%	10.77%	6.06%	16.12%
(Erβ/Erα): Erβ$^-$ BC Cell Line	25.62%	18.18%	19%	13.22%	8.26%	15.72%

occurrence in E_2 hormone induced breast cancer. However, significantly modulated exons proportion survey by processing differentially spliced multi-exonic genes revealed that exons change in the analyzed hormone responsive BC cell lines, meanly regard the first 10 gene transcripts exons (more than 25%), advising and/or suspecting weakly solicitation of ending transcripts exons yielding alternative transcript isoforms in hormone responsive BC cells (Table 2).

DAVID analysis assessing functional annotation of differentially spliced genes discriminated by both Cufflinks/Cuffdiff and DEXseq tools

68 genes out the 73 processed differentially spliced genes (DSGs) discriminated by Cufflinks/ Cuffdiff approach at 5% false discovery rate in hormone responsive Erβ$^+$ BC (Supplementary Table 2) were converted in a new list for functional annotation analysis by DAVID package showing at Benjamini correction referring to a p-adjusted value at 5%, that (i) 75% of differentially alternatively spliced genes called by Cufflinks/ Cuffdiff were genes that code for at least two isoforms due to pre-mRNA splicing event and that (ii) 74% of these genes were recognized as alternative splice variant. Furthermore, 42.64% of alternative spliced genes categorized by Cufflinks/ Cuffdiff approach have been discriminated to be localized in the nucleus (Figure 5A). In the same tendency and focusing on alternative spliced exons events called by DEXseq methodology at 5% false discovery rate threshold with 2 fold exon change (high alternative splicing evidence), we showed that 55.35% of analyzed alternative spliced genes, coded for at least two isoforms due to pre-mRNA splicing event and that 54.91% of their transcript isoform variants were recognized as alternative splice variant. We also showed that 37.84% of all analyzed alternative spliced genes called by DEXseq approach have been discriminated to be localized in the nucleus (Figure 5B). These results, delicately suggested a relative agreement between the two considered and/or analyzed bioinformatics tools (Cufflinks/ Cuffdiff and DEXseq) assessing the performance of alternative splicing pathway in breast cancer cell line, since mRNA maturation in eukaryotic cells happened in cell's nucleus. Finally, even if the present results suggested DEXseq approach as discriminating highest number of potential alternative splicing event in term of functional annotation as opposed to Cufflinks/Cuffdiff (Figure 5), it is noteworthy to underline their similitude evaluating (i) splice and isoform variant, (ii) alternative spliced genes localization in the nucleus and (iii) nucleolus in the present functional annotation survey (Figure 5).

Integration between Cufflinks/Cuffdiff and DEXseq data assessing alternative splicing event in Erβ breast cancer cell line

More than 34.74% of DSGs (multi-exonic genes) discriminated by Cufflinks/Cuffdiff approach in the hormone responsive Erβ$^+$ BC cell, exhibited at least one significant exons change as DEXseq exon differential analysis (Figure 4). Indeed, *LIFR* gene, knows as a breast cancer metastasis suppressor, *MATA2* gene involves in apoptosis process in human hepatoma, *NBPF10* gene associated with several types of cancer, *PBX1* gene that results engaged in the progression of breast cancer and *PHLPP2* gene that inhibits cancer cell proliferation acting as tumor suppressor, detected as significantly differentially spliced by Cufflinks/Cuffdiff approach, displayed a significant exons change in the present analyzed hormone responsive BC cell line (Figure 4), proposing both Cufflinks/Cuffdiff and DEXseq data merging process as a valid methodology monitoring AS pathway in breast cancer. As DEXseq provides exons splicing visualization, we reported in Figure 6 an example of exons modulation evaluating alternative splicing happening in two fully differentially spliced genes *MATA2* and *PBX1* in E_2 hormone induced BC exclusively (Erβ$^+$ BC). Moreover, several generic tumor biomarkers such as *SEPT9*, known as a candidate for the ovarian tumor suppressor, *PPP1R13B* and *NOTCH2* genes affecting cells differentiation implementation, proliferation and apoptotic programs and *NBPF10* gene which results associated with several type of cancer, were recognized as exhibiting significant exons change in the hormone responsive Erβ$^+$ BC cell line (Figure 4). We also showed that 36.11% of detected spliced genes merging both Cuffdiff and

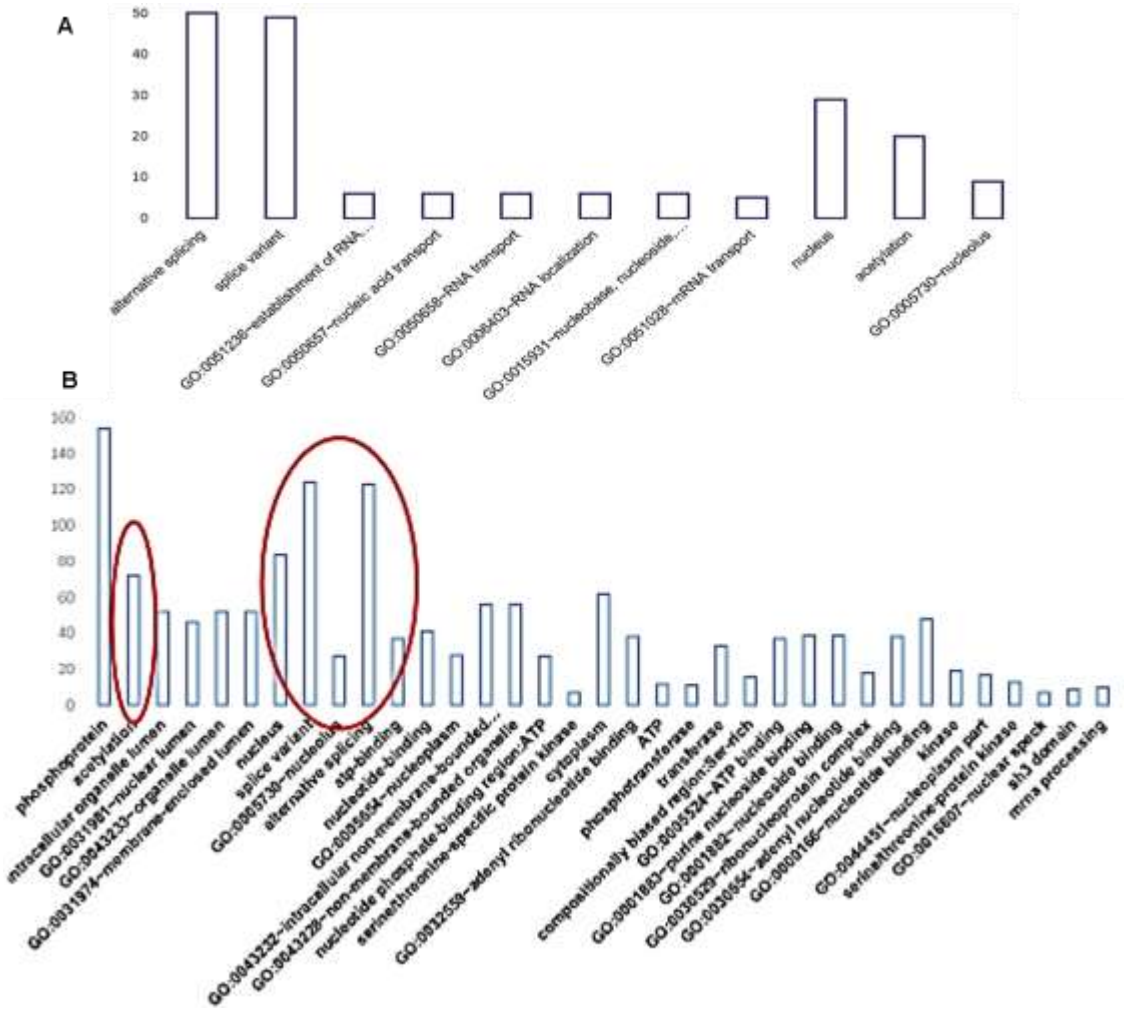

Figure 5. Functional annotation survey by DAVID tool by processing differentially spliced genes (DSGs) candidates comparing both Cufflinks/Cuffdiff (A) and DEXseq (B) bioinformatics approaches.

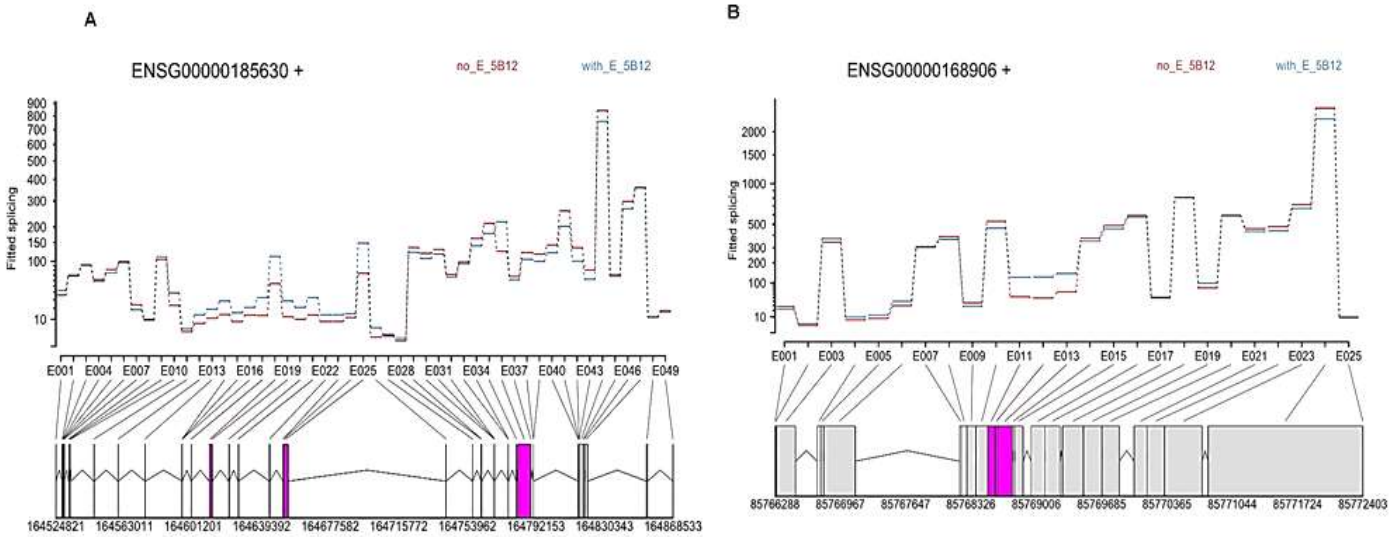

Figure 6. DEXseq representation of significant exon change of two differentially spliced genes PBX1 (A) and MATA2 (B) discriminated by Cufflinks/Cuffdiff and DEXseq. Shown in pink is the differential expressed exons (exon reads count ≥10 at a FDR ≤ 0.05) involved in the potential alternative splicing events.

DEXseq approaches were regulated by significant alternative promoter use (Figure 4). Interestingly, well-noted biomarker genes like *MATA2*, *SEPT9*, *NOTCH2* and *PPP1R13B* claimed to be under alternative promoter usage (p-value≤0.05) (Figure 4 and Supplementary Table 3), assessing alternative splicing pathway in Erβ^{+} BC cells. So, the present results suspecting the involvement of some remarkable cancer biomarkers controlling AS path in hormone responsive BC cell lines, proposed Cufflinks/Cuffdiff and DEXseq approaches data integration as reasonable and complementary scheme weighing whole AS pathways in breast cancer cells.

DISCUSSION

Alternative splicing (AS) is a means of expressing several or many different transcripts from the same genomic DNA and results from the inclusion of a subset of the available exons for a particular protein. By excluding one or more exons, certain protein domains may be lost from the encoded protein, which can result in protein function damage or gain. Several types of alternative splicing have been described resulting in exon skipping; alternative 5′ or 3′ splice sites; mutually exclusive exons and much more rarely and intron retention, approving the complexity of AS study and the needed of accurate bioinformatics or bio-statistics systems evaluating this phenomena in eukaryotic cells. However, many alternative splicing events have been noted in human development, especially in the brain and the testes (Grabowski and Black, 2001; Venables, 2002) as well as in cancer, including the use of alternative individual splice sites, alternative exons, and alternative introns. Therefore, whole alternative splicing survey must include capable sensitive and specific bioinformatics tools, able to measure accurately alternative transcript isoforms as well as multi-exonic genes transcripts exons abundance, since recent Next Generation RNA Sequencing (NGS RNA-Seq) analysis provides innovative platform exploring in detail cell transcriptome and/or genome. However, a recurrent challenge in RNA-Seq experimentation regards application of adequate bioinformatics tools analyzing huge quantity and complex yielded data. Also, RNA-Seq coupled with well-established bioinformatics approaches; that is, Cufflinks/Cuffdiff (Trapnell et al., 2013) and DEXseq (Anders et al., 2012), allowed a suitable whole genome and transcriptome reconstruction measuring gene transcript isoforms as well as exons expression level, performing differential expression analysis between cells types. In such analysis, annotated regions of the considered genome can be expressed (that is, exons), describing how the pre-mRNAs are spliced into transcripts. While there are several utilities that attempt to de-convolute read data into isoform abundances, the accuracy and robustness of these methods is difficult to establish (Chandramohan et al., 2013; Zhang et al., 2014). Isoform expression estimates seem to vary considerably between different tools, and generally depend on the quality and completeness of the transcript assembly (Kanitz et al., 2015; Rehrauer et al., 2013). Also, Cufflinks/Cuffdiff methodology is not more informative regarding exons modulation measuring alternative splice transcript isoforms pattern. Simon Anders et al. (2012) demonstrate the versatility of DEXseq package by applying it to several data sets facilitating the study of regulation and function of alternative exon usage on a genome-wide scale. Guided by these observations, we proposed merging between both Cufflinks/Cuffdiff and DEXseq investigating meticulously alternative transcript isoforms exons abundance in our previous studied hormone responsive Erβ+ breast cancer cells, since Erβ significantly affects estrogen-induced early transcription and mRNA splicing in hormone-responsive BC cells (Dago et al., 2015). Then, genomic re-distribution of RNA-Seq reads sequences from hormone responsive Erβ BC cells by HTSeq/DEXseq and TopHat/Cufflinks, suggested strong difference (p-value <0.05) between estrogen induced Erβ^{+} BC cells (Erβ$^{+}E_2$) and the other's analyzed breast cancer cells conditions (Erβ^{+} with any E_2 stimulus; Erβ^{+}noE_2, Erβ^{-} induced by E_2; Erβ$^{-}E_2$ and Erβ^{-} without any E_2 stimulus; Erβ^{-}noE_2), since exhibiting dissimilar attitude and/or behaviors considering loaded and fund reads junction as well as exon intragenic and intron reads distribution (Figure 2 and Table 1), subtly evoking agreement between both Cufflinks and DEXseq approaches in genomic and transcriptomic reconstruction analysis. In addition, this analysis suggested high variability between Erβ$^{+}E_2$ and Erβ^{+}noE_2 conditions suspecting a high number of differentially expressed genes and/or transcript isoforms in the latter's as opposed to hormone induced Erβ^{-} BC cells. However, comparative analysis assessing the relationship between change in expressed genes and their respective transcript isoforms, showed a strong evidence of alternative promoter used in alternative splicing pattern processing hormone responsive Erβ^{+} BC cell line, as the ratio between expressed transcript isoforms as oppose to expressed genes, was 2 fold more higher in the latter's (Erβ^{+} BC) when compared to Erβ^{-}BC cell line (Figure 3). Taking together, these results reinforced alternative splicing evidence in hormone responsive Erβ^{+}BC as opposite to Erβ^{-} BC cell lines alerting a significant participation of alternative promoter regulating AS events occurrence in breast cancer process. Several studies have shown that the occurrence of alternative transcriptional termination and splicing is higher in genes with alternative promoters, and the choice of alternative promoter and transcription termination can influence the alternative splicing pattern of the pre-mRNA (Winter et al., 2007; Albulescu et al., 2012). Nevertheless, it is estimated that participate to the alternative splicing phenomena proposing genes transcript isoforms exons modulation as a suitable approach understanding

alternative transcript isoform expression in human genome. Hence, detailed alternative splicing analysis by NGS RNA-Seq need innovative bioinformatics scheme capable to integrate accurately transcript isoforms and exon expression analysis. Based on this, we combined differentially spliced genes and transcript isoforms exons modulation assessing early AS occurrence in estradiol hormone induced Erβ+ BC cells, emphasizing strong involvement of transcript isoforms exons change in alternative splicing mechanism in breast tumor. Then, integrating Cufflinks/ Cuffdiff and DEXseq approaches, our findings proposed alternative promoter's usage as well as significant exon change as key molecular events favoring AS pattern in hormone responsive breast cancer cells (Figure 4). Also, it is noteworthy to underline the weak involvement of ended exons of transcript isoforms evaluating alternative splicing occurrence in the present hormone induced Erβ BC cells (Table 2). Furthermore, around 34.74% of significantly differentially spliced genes by Cufflinks/ Cuffdiff have shown significant exons change in DEXseq analysis testing for differential usage of exon regions as a proxy for alternative isoform regulation as well as providing a powerful suite of visualization tools (Figure 6) (Love et al., 2014). Interestingly, the present analysis identified several remarkable cancer biomarkers, like *MDM2* and *NF1* genes, known as cancer specific alternative splicing genes (Venables, 2004), *PBX1* gene, revealed as a novel pioneer factor defining aggressive $Er\beta^-$ breast tumors, as it guides Erα genomic activity to unique genomic regions promoting a transcriptional program favorable to breast cancer progression (Magnani et al., 2011), *LIFR* gene, knows as a breast cancer metastasis suppressor (Chen et al., 2012), *MATA2* gene, involved in human colon cancer progression (Chen et al., 2007) and *PHLPP2* tumor suppressor (Liu et al., 2011) and *AIB1/NCOA3* hormone signaling in breast cancer as exhibiting significant exon change in hormone responsive Erβ BC cells, demonstrating the key role of exon modulation in early estrogen induced breast cancer alternative splicing pattern (Figure 4, Supplementary Materials 1 and 2). Furthermore, 36.11% of detected differentially spliced genes by merging Cuffdiff and DEXseq approaches were recognized as including alternative promoter, and some well-noted cancer biomarker (*MATA2*, *SEPT9*, *NOTCH2* and *PPP1R13B*) claimed to be under alternative promoter usage ($p\text{-value} \leq 0.05$) (Figure 4 and Supplementary Table 3), monitoring early alternative splicing pathway in our processed hormone responsive $Er\beta^+$ BC cells (Figure 4). So, the present results by evoking the involvement of some remarkable cancer biomarkers, controlling AS pattern in the present hormone responsive BC cell lines, proposed Cufflinks/Cuffdiff and DEXseq genomic data integration, as a reasonable complementary scheme assessing whole AS occurrence in hormone responsive breast cancer cell. Moreover, concordance between Cufflinks/ Cuffdiff and DEXseq has been supported in part by DAVID functional annotation analysis since 42.64 and 37.84% of alternative spliced genes categorized by previous mentioned approaches respectively, have been discriminated to be localized in cell's nucleus, suggesting a strong contribution of the latter's (alternative sliced genes) regulating AS pattern occurrence in hormone responsive BC cell line (Figure 5). Also, even if the present study admitted conflicting results between Cufflinks/Cuffdiff and DEXseq approaches in alternative splicing analysis (Figure 5), we showed that an adequate integration between these bioinformatics/biostatistics tools can help to wholly investigate alternative splicing occurrence in breast cancer disease reducing false discovery event rate. Furthermore, accuracy in statistical analysis processing alternative transcript isoforms regulation in genomic and transcriptomic studies, has been supported by recent work introducing a new method that builds on the statistical techniques used by the well-established DEXseq package to detect differential usage of both exonic regions and splice junctions helping differential usage of novel splice junctions without the need for an additional isoform assembly step (Stephen and James, 2016).

Conclusion

RNA-Seq providing more than simple measurements of gene and/or transcript-level expression, can be used to study more complex regulatory phenomena at the isoform level, even when the isoforms in question are unannotated. Numerous tools have been developed to detect alternative isoform regulation exhibiting in some cases conflicting results. In contrast to this tendency, the present study proposed integration between well-established Cufflinks/Cuffdiff and DEXseq approaches as reasonable system assessing alternative splicing events in hormone responsive Erβ breast cancer cells. Finally, our findings exhibited significant exon modulation of multi-exonic gene transcripts regulated by alternative promoters, as recurrently solicited in early AS pattern in estrogen induced BC cells.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENT

Thanks to all members of the Laboratory of Molecular Medicine and Genomics of the Department of Medicine and Surgery, University of Salerno (SA) Italy, for providing information regarding RNA sequencing data for the present study.

REFERENCES

Albulescu LO, Sabet N, Gudipati M, Stepankiw N, Bergman ZJ, Huffaker TC, Pleiss JA (2012). A quantitative high-throughput reverse genetic screen reveals novel connections between Pre-mRNA splicing and 5′ and 3′ end transcript determinants. PLoS Genet. 8(3):e1002530.

Anders S, Reyes A, Huber W (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. 22(10):2008-2017.

Anders S, Pyl PT, Huber W (2015). HTSeq A Python framework to work with high-throughput sequencing data. Bioinformatics 31(2):166-169.

Cancer Genome Atlas (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330-337.

Chandramohan R, Wu PY, Phan H, Wang MD (2013). Benchmarking RNA-Seq quantification tools. Conf. Proc. IEEE Eng. Med. Biol. Soc. pp. 647-650.

Chen D, Sun Y, Wei Y, Zhang P, Rezaeian AH, Teruya-Feldstein J, Gupta S, Liang H, Lin HK, Hung MC, Ma L (2012). LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. Nat. Med. 18(10):1511-1517.

Chen H, Xia M, Lin M, Yang H, Kuhlenkamp J, Li T, Sodir NM, Chen YH, Josef-Lenz H, Laird PW, Clarke S, Mato JM, Lu SC (2007). Role of methionine adenosyltransferase 2A and S-adenosylmethionine in mitogen-induced growth of human colon cancer cells. Gastroenterology 133(1):207-218.

Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature 452:230-233.

Dago DN, Claudio S, Antonio R, DomenicoM, Giorgio G, Giovanni N, Maria R, Francesca R, Roberta T, Alessandro W (2015). Estrogen receptor beta impacts hormone-induced alternative mRNA splicing in breast cancer cells. BMC Genomics 16(367):1-13.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004). Bioconductor: open software development for computational biology and bioinformatics Genome Biol. 5(10):1-39.

Glynn DJ, Brad TS, Douglas AH, Jun Y, Wei G, Clifford HL and Richard AL (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4:1-11.

Grabowski PJ, Black DL (2001). Alternative RNA splicing in the nervous system. Prog. Neurobiol. 65:289-308.

Grober OM, Mutarelli M, Giurato G, Ravo M, Cicatiello L, De Filippo MR, Ferraro L, Nassa G, Papa MF, Paris O, Tarallo R (2011). Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation. BMC Genomics 12(36):1-16.

Jeffrey AM, Zhong W (2011). Next-generation transcriptome assembly. Nat. Rev. Genet. 12:671-682.

Kanitz A, Gypas F, Gruber AJ, Gruber AR, Martin G and Zavolan M (2015). Comparative assessment of methods for the computational inference of transcript isoform abundance from RNAseq data. Genome Biol. 16:1-26.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14(4):1-13.

Kim E, Goren A, Ast G (2008). Insights into the connection between cancer and alternative splicing. Trends Genet. 24:7-10.

Liu J, Stevens PD, Li X, Schmidt MD, Gao T (2011). PHLPP-mediated dephosphorylation of S6K1 inhibits protein translation and cell growth. Mol. Cell. Biol. 31(24):4917-4927.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550.

Magnani L, Elizabeth BB, Xiaoyang Z, Mathieu L (2011). PBX1 Genomic Pioneer Function Drives ERα Signaling Underlying Progression in Breast Cancer PLoS Genet. 7(11):e1002368.

Moller-Levet CS, Betts GN, Harris AL, Homer JJ, West CM, Miller CJ (2009). Exon array analysis of head and neck cancers identifies a hypoxia related splice variant of LAMA3 associated with a poor prognosis. PLoS Comput. Biol. 5(11):e1000571.

Nassa G, Tarallo R, Ambrosino C, Bamundo A, Ferraro L, Paris O, et al. (2011). A large set of estrogen receptor beta-interacting proteins identified by tandem affinity purification in hormone-responsive human breast cancer cell nuclei. Proteomics 11(1):159-165.

Pal S, Gupta R, Kim H, Wickramasinghe P, Baubet V, Showe LC, et al. (2011). Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. Genome Res. 21(8):1260-72.

Pal S, Ravi G, Ramana VD (2012). Alternative transcription and alternative splicing in cancer. Pharmacol. Ther. 136:283-294.

Paris O, Ferraro L, Grober OM, Ravo M, De Filippo MR, Giurato G, Nassa G, Tarallo R, Cantarella C, Rizzo F, Di Benedetto A (2012). Direct regulation of microRNA biogenesis and expression by estrogen receptor beta in hormone-responsive breast cancer. Oncogene 31(38):4196-206.

R Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL http://www.R-project.org.

Rehrauer H, Opitz L, Tan G, Sieverling L, Schlapbach R (2013). Blind spots of quantitative RNA-seq: the limits for assessing abundance, differential expression, and isoform switching. BMC Bioinformatics 14:370.

Stephen WH, James CM (2016). Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucleic Acids Res. 44(15):1-15.

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956-960.

Tarallo R, Bamundo A, Nassa G, Nola E, Paris O, Ambrosino C, Facchiano A, Baumann M, Nyman TA, Weisz A (2011). Identification of proteins associated with ligand-activated estrogen receptor alpha in human breast cancer cell nuclei by tandem affinity purification and nano LC-MS/MS. Proteomics 11(1):172-179.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31(1):46-53.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28(5):511-515.

Venables JP (2002). Alternative splicing in the testes. Curr. Opin. Genet. Dev.12:615-619.

Venables PJ (2004). Aberrant and Alternative Splicing in Cancer. Cancer Res. 64:7647-7654.

Venables,JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, Lucier JF, Thibault P, Rancourt C, Tremblay K, Prinos P, Chabot B, Elela SA (2009). Cancer-associated regulation of alternative splicing. Nat. Struct. Mol. Biol.16:670-676.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008). Alternative isoform regulation in human tissue transcriptomes. Nature 456:470-476.

Winter J, Kunath M, Roepcke S, Krause S, Schneider R, Schweiger S (2007). Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz syndrome gene MID1. BMC Mol. Biol. 8(105):1-12.

Zerbino DR, Birney E (2008).Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res.18(5):821-829.

Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, Robinson GJ, Lundberg AE, Bartlett PF, Wray NR, Zhao QY (2014). A comparative study of techniques for differential expression analysis on RNA-Seq data. PLoS One 9(8):e103207.

Supplementary Table 1. Erβ+ BC cell line spliced genes at q_value ≤0.05.

Gene_id	Gene	Locus	sqrt(JS)	p_value	q_value	Significant
ENSG00000234608	C12orf47	12:112277570-112334343	0.215053	0.00255	0.0318605	yes
ENSG00000234741	GAS5	1:173831289-173866494	0.138086	0.00005	0.00104714	yes
ENSG00000068745	IP6K2	3:48725435-48777786	0.27223	0.00075	0.0116144	yes
ENSG00000119285	HEATR1	1:236681299-236767804	0.0748117	0.00005	0.00104714	yes
ENSG00000169689	STRA13	17:79976578-79980794	0.0627464	0.0001	0.00199909	yes
ENSG00000105750	ZNF85	19:21106058-21133503	0.388189	0.00075	0.0116144	yes
ENSG00000075711	DLG1	3:196769430-197030618	0.384547	0.00005	0.00104714	yes
ENSG00000136051	KIAA1033	12:105501101-105562912	0.0564964	0.0043	0.0446024	yes
ENSG00000091039	OSBPL8	12:76745576-76953589	0.0515044	0.002	0.0261786	yes
ENSG00000224831	RP11-651P23.4.1	3:149478891-149942977	0.39333	0.00005	0.00104714	yes
ENSG00000007392	LUC7L	16:238967-279462	0.133702	0.0013	0.018325	yes
ENSG00000092978	GPATCH2	1:217600333-217804424	0.0213977	0.00365	0.0393449	yes
ENSG00000196712	NF1	17:29421944-29709134	0.210908	0.00005	0.00104714	yes
ENSG00000112414	GPR126	6:142622990-142767403	0.498713	0.00005	0.00104714	yes
ENSG00000138592	USP8	15:50716576-50838905	0.193283	0.00005	0.00104714	yes
ENSG00000154743	TSEN2	3:12525930-12705725	0.337813	0.00005	0.00104714	yes
ENSG00000122008	POLK	5:74664310-74896969	0.370688	0.00005	0.00104714	yes
ENSG00000185722	ANKFY1	17:4066664-4167274	0.166863	0.00005	0.00104714	yes
ENSG00000073350	LLGL2	17:73521782-73571289	0.10957	0.00005	0.00104714	yes
ENSG00000100325	ASCC2	22:30184596-30234271	0.578665	0.00005	0.00104714	yes
ENSG00000147274	RBMX	X:135923089-135962923	0.143233	0.00025	0.004398	yes
ENSG00000218739	AC007390.5.1	2:37394962-37551951	0.167284	0.0027	0.0333556	yes
ENSG00000249846	RP11-77P16.4.1	3:129800673-129838359	0.559348	0.00295	0.0346901	yes
ENSG00000185305	ARL15	5:53179774-53606412	0.311492	0.00295	0.0346901	yes
ENSG00000172748	ZNF596	8:182136-197342	0.258979	0.00005	0.00104714	yes
ENSG00000188266	AGPHD1	15:78799905-78829714	0.616911	0.00005	0.00104714	yes
ENSG00000005700	IBTK	6:82879699-82957471	0.15325	0.00005	0.00104714	yes
ENSG00000137776	SLTM	15:59063390-59389618	0.329689	0.00005	0.00104714	yes
ENSG00000182307	C8orf33	8:146277763-146281416	0.0615689	0.0001	0.00199909	yes
ENSG00000090263	MRPS33	7:140702195-140715028	0.0773085	0.00285	0.0346251	yes
ENSG00000175467	SART1	11:65729159-65747299	0.111518	0.00025	0.004398	yes
ENSG00000153107	ANAPC1	2:112523847-112642267	0.153132	0.00005	0.00104714	yes
ENSG00000088808	PPP1R13B	14:104200088-104313927	0.176983	0.0025	0.0315948	yes
ENSG00000123908	EIF2C2	8:141541263-141645718	0.145366	0.0012	0.0174755	yes
ENSG00000074621	SLC24A1	15:65903703-66184329	0.475611	0.0009	0.013649	yes
ENSG00000163867	ZMYM6	1:35439956-35497569	0.273954	0.00005	0.00104714	yes
ENSG00000197548	ATG7	3:11313994-11770134	0.134414	0.0003	0.00515391	yes
ENSG00000221983	UBA52	19:18682613-18688269	0.0527833	0.0003	0.00515391	yes
ENSG00000143367	TUFT1	1:151512780-151556059	0.191093	0.00005	0.00104714	yes
ENSG00000116191	RALGPS2	1:178694299-179067158	0.0718482	0.0041	0.0433457	yes
ENSG00000167792	NDUFV1	11:67374322-67380006	0.23601	0.00095	0.0143086	yes
ENSG00000198160	MIER1	1:67390577-67454302	0.235012	0.00005	0.00104714	yes
ENSG00000108799	EZH1	17:40852293-40897071	0.208434	0.00005	0.00104714	yes
ENSG00000118939	UCHL3	13:76123618-76434004	0.117807	0.00005	0.00104714	yes
ENSG00000119979	FAM45A	10:120863597-120897496	0.205479	0.00215	0.0278109	yes
ENSG00000162063	CCNF	16:2479394-2509980	0.260364	0.00005	0.00104714	yes
ENSG00000254986	DPP3	11:66234215-66313709	0.458483	0.00005	0.00104714	yes
ENSG00000109184	DCUN1D4	4:52709165-52783003	2.14375	0.00005	0.00104714	yes
ENSG00000119772	DNMT3A	2:25455844-25565459	0.0649625	0.00175	0.0237546	yes
ENSG00000077809	GTF2I	7:74071993-74306729	0.103656	0.00325	0.03665	yes
ENSG00000213762	ZNF134	19:58125618-58134721	0.515367	0.00005	0.00104714	yes

Supplementary Table 1. Contd.

ENSG00000196911	KPNA5	6:117002349-117063029	0.320921	0.00005	0.00104714	yes
ENSG00000164180	TMEM161B	5:87485449-87794514	0.389591	0.00005	0.00104714	yes
ENSG00000049283	EPN3	17:48609903-48633213	0.369319	0.00005	0.00104714	yes
ENSG00000141446	ESCO1	18:19109263-19180845	0.0673011	0.00205	0.0266743	yes
ENSG00000181264	TMEM136	11:120195837-120204391	0.291405	0.00005	0.00104714	yes
ENSG00000070756	PABPC1	8:101698043-101735037	0.0677226	0.00075	0.0116144	yes
ENSG00000008294	SPAG9	17:49039534-49198226	0.163075	0.00005	0.00104714	yes
ENSG00000040199	PHLPP2	16:71657610-71758604	0.244715	0.0029	0.0346901	yes
ENSG00000132313	MRPL35	2:86426579-86440917	0.197046	0.0042	0.0437716	yes
ENSG00000164327	RICTOR	5:38845959-39074510	0.55906	0.00005	0.00104714	yes
ENSG00000105576	TNPO2	19:12810007-12834810	0.226942	0.0007	0.0113184	yes
ENSG00000089123	TASP1	20:13202417-13619587	0.222231	0.00085	0.0129802	yes
ENSG00000010404	IDS	X:148558520-148632055	0.0921044	0.0002	0.003665	yes
ENSG00000067066	SP100	2:231280656-231444721	0.189363	0.0003	0.00515391	yes
ENSG00000047230	CTPS2	X:16606125-16731059	0.123897	0.00275	0.0335958	yes
ENSG00000073921	PICALM	11:85668726-85780924	0.108588	0.00105	0.015601	yes
ENSG00000083896	YTHDC1	4:69176104-69215807	0.025357	0.00305	0.0354865	yes
ENSG00000134291	TMEM106C	12:48357351-48362661	0.0742842	0.00005	0.00104714	yes
ENSG00000107771	FAM190B	10:86088341-86278273	0.0696925	0.00005	0.00104714	yes
ENSG00000107937	GTPBP4	10:1034337-1095110	0.107753	0.00005	0.00104714	yes
ENSG00000139697	SBNO1	12:123773655-123834988	0.150288	0.00005	0.00104714	yes
ENSG00000196290	NIF3L1	2:201754049-201768655	0.136813	0.00005	0.00104714	yes
ENSG00000101333	PLCB4	20:9049409-9461889	0.165373	0.00005	0.00104714	yes
ENSG00000157657	ZNF618	9:116638561-116818871	0.117957	0.00405	0.0430239	yes
ENSG00000123349	PFDN5	12:53689074-53700961	0.16216	0.00185	0.0245069	yes
ENSG00000103365	GGA2	16:23474862-23533316	0.107243	0.00005	0.00104714	yes
ENSG00000167863	ATP5H	17:73032144-73043074	0.0369063	0.00255	0.0318605	yes
ENSG00000164171	ITGA2	5:51971026-52390609	0.212906	0.0019	0.0250186	yes
ENSG00000156787	WDR67	8:124014799-124164393	0.160779	0.0042	0.0437716	yes
ENSG00000101596	SMCHD1	18:2655885-2805015	0.0751137	0.00005	0.00104714	yes
ENSG00000117143	UAP1	1:162531320-162569627	0.117689	0.0037	0.0396893	yes
ENSG00000196504	PRPF40A	2:153508106-153617688	0.104612	0.00005	0.00104714	yes
ENSG00000100147	CCDC134	22:42196682-42222303	0.328042	0.00005	0.00104714	yes
ENSG00000109184	DCUN1D4	4:52709165-52783003	0.539569	0.00005	0.00104714	yes
ENSG00000173145	NOC3L	10:95753745-96122716	0.0601538	0.00085	0.0129802	yes
ENSG00000167515	TRAPPC2L	16:88880141-88933068	0.0375659	0.00295	0.0346901	yes

Supplementary Table 2. Erβ- BC Cell line Spliced genes at q_value ≤0.05.

Gene_id	Gene	Locus	sqrt(JS)	Test_stat	p_value	q_value	Significant
ENSG00000234741	GAS5	1:173831289-173866494	0.136996	0	0.00105	0.0398523	yes
	HEATR1	1:236681299-236767804	0.121115	0	5,00E-05	0.00305488	yes
ENSG00000224831	RP11-651P23.4.1	3:149478891-149942977	0.418664	0	5,00E-05	0.00305488	yes
ENSG00000131263	RLIM	X:73805051-73834452	0.065689	0	0.00115	0.04175	yes
ENSG00000104738	MCM4	8:48872744-48890720	0.336343	0	5,00E-05	0.00305488	yes
ENSG00000213782	DDX47	12:12878850-12982915	0.484054	0	5,00E-05	0.00305488	yes
ENSG00000178660	ARMC10P1	3:94225609-94226464	0.289791	0	5,00E-05	0.00305488	yes
ENSG00000068878	PSME4	2:54091203-54307601	0.12927	0	0.00045	0.0201295	yes
ENSG00000063601	MTMR1	X:149861434-149933576	0.100801	0	5,00E-05	0.00305488	yes
ENSG00000155085	AKD1	6:109809108-110012420	0.239652	0	5,00E-05	0.00305488	yes
ENSG00000160948	VPS28	8:145648999-145653931	0.598512	0	5,00E-05	0.00305488	yes
ENSG00000138069	RAB1A	2:65283499-65357240	0.0738017	0	0.00015	0.00835	yes
ENSG00000058804	TMEM48	1:54231132-54304533	0.115543	0	5,00E-05	0.00305488	yes
ENSG00000131269	ABCB7	X:74273108-74376567	0.170756	0	5,00E-05	0.00305488	yes
ENSG00000185480	C12orf48	12:102513955-102591623	0.360215	0	0.00065	0.0266926	yes
ENSG00000145734	BDP1	5:70751441-70863649	0.0933384	0	5,00E-05	0.00305488	yes
ENSG00000115275	MOGS	2:74688183-74692537	0.117025	0	5,00E-05	0.00305488	yes
ENSG00000223658	AC011242.6.1	2:43864411-43995126	0.251771	0	0.0003	0.0147353	yes
ENSG00000156787	WDR67	8:124014799-124164393	0.160299	0	0.0001	0.00582558	yes
ENSG00000151503	NCAPD3	11:133938819-134117686	0.775431	0	0.00045	0.0201295	yes
ENSG00000196975	ANXA4	2:69686413-70053596	0.0578661	0	0.0002	0.0102245	yes

Supplementary Table 3. Erβ+ BC cell line Differential Promoter using at q-value ≤0.05.

gene_id	gene	locus	sqrt(JS)	test_stat	p_value	q_value	significant
ENSG00000234741	GAS5	1:173831289-173866494	0.136996	0	0.00105	0.0398523	yes
	HEATR1	1:236681299-236767804	0.121115	0	5,00E-05	0.00305488	yes
ENSG00000224831	RP11-651P23.4.1	3:149478891-149942977	0.418664	0	5,00E-05	0.00305488	yes
ENSG00000131263	RLIM	X:73805051-73834452	0.065689	0	0.00115	0.04175	yes
ENSG00000104738	MCM4	8:48872744-48890720	0.336343	0	5,00E-05	0.00305488	yes
ENSG00000213782	DDX47	12:12878850-12982915	0.484054	0	5,00E-05	0.00305488	yes
ENSG00000178660	ARMC10P1	3:94225609-94226464	0.289791	0	5,00E-05	0.00305488	yes
ENSG00000068878	PSME4	2:54091203-54307601	0.12927	0	0.00045	0.0201295	yes
ENSG00000063601	MTMR1	X:149861434-149933576	0.100801	0	5,00E-05	0.00305488	yes
ENSG00000155085	AKD1	6:109809108-110012420	0.239652	0	5,00E-05	0.00305488	yes
ENSG00000160948	VPS28	8:145648999-145653931	0.598512	0	5,00E-05	0.00305488	yes
ENSG00000138069	RAB1A	2:65283499-65357240	0.0738017	0	0.00015	0.00835	yes
ENSG00000058804	TMEM48	1:54231132-54304533	0.115543	0	5,00E-05	0.00305488	yes
ENSG00000131269	ABCB7	X:74273108-74376567	0.170756	0	5,00E-05	0.00305488	yes
ENSG00000185480	C12orf48	12:102513955-102591623	0.360215	0	0.00065	0.0266926	yes
ENSG00000145734	BDP1	5:70751441-70863649	0.0933384	0	5,00E-05	0.00305488	yes
ENSG00000115275	MOGS	2:74688183-74692537	0.117025	0	5,00E-05	0.00305488	yes
ENSG00000223658	AC011242.6.1	2:43864411-43995126	0.251771	0	0.0003	0.0147353	yes
ENSG00000156787	WDR67	8:124014799-124164393	0.160299	0	0.0001	0.00582558	yes
ENSG00000151503	NCAPD3	11:133938819-134117686	0.775431	0	0.00045	0.0201295	yes
ENSG00000196975	ANXA4	2:69686413-70053596	0.0578661	0	0.0002	0.0102245	yes

Supplementary material 1. (Erβ-) with selected differentially spliced genes (significant exon change) at an adjusted p-value≤0.05.

geneID	exonID	Dispersion	p-value	padjust	log2fold(E.MCF7TO/noE.MCF7TO)
ENSG00000005022	E007	0.0147055470136266	0	0	1.23698E+14
ENSG00000062716	E010	0.000760338066554652	0	0	0.188890253253861
ENSG00000062716	E015	0.000295163964057167	0	0	-0.104675732380832
ENSG00000062716	E020	0.000234579474480542	0	0	-0.0943458090571052
ENSG00000063177	E027	0.00891411550729496	0	0	0.0553905614482828
ENSG00000071082	E007	0.0144210199887183	0	0	0.0700737410615521
ENSG00000075624	E019	0.00455444228811321	0	0	-0.0644923967077791
ENSG00000109475	E002	0.00907538924849258	0	0	-0.0779595909508594
ENSG00000109475	E003	0.00546314773064575	0	0	-0.0656868304620877
ENSG00000111907	E024	0.00501112637065375	0	0	0.670055812031743
ENSG00000128609	E023	0.0311252561789971	0	0	1.65431E+14
ENSG00000128609	E024	0.0342686024037692	0	0	1.64574E+12
ENSG00000128641	E036	0.00603682819389217	0	0	-0.772440536766849
ENSG00000134333	E051	0.0447131631998469	0	0	0.438335394659404
ENSG00000140988+ENSG00000207405+ENSG00000255513+ENSG00000206811	E006	0.0348438876275743	0	0	0.0659382766423838
ENSG00000161016	E011	0.0106634397668682	0	0	-1.46983E+14
ENSG00000170889	E015	0.0109430196428278	0	0	0.112189002755497
ENSG00000206941+ENSG00000149273	E045	0.00157025629193141	0	0	-0.678273482281045
ENSG00000206941+ENSG00000149273	E048	0.000274692972329399	0	0	0.103829799147221
ENSG00000259001+ENSG00000252678	E001	0.000336625879494382	0	0	-0.0977170086797675
ENSG00000259001+ENSG00000252678	E002	0.000402833507196914	0	0	0.0422333243280596

Supplementary material 2. (Erβ+) with selected differentially spliced genes (significant exon change) at an adjusted p-value≤0.05.

geneID	exonID	Dispersion	pvalue	padjust	log2fold (with_E_5B12/no_E_5B12)
ENSG00000001631+ENSG00000243107	E029	0.0148526122373315	0	0	0.865040077151733
ENSG00000005700	E017	0.015898293844675	0	0	0.985529091793383
ENSG00000015153	E025	0.0150988457158889	0	0	0.953724663614144
ENSG00000031003	E022	0.00479163865635899	0	0	0.453751262738086
ENSG00000031003	E023	0.00336889926466394	0	0	0.554888688343613
ENSG00000033170	E016	0.0128992399400473	0	0	0.887002581926375
ENSG00000062716	E007	0.00171864314294444	0	0	0.386073943547944
ENSG00000062716	E008	0.00311704081967452	0	0	0.389584872758711
ENSG00000062716	E009	0.00190088447167943	0	0	0.289544969931232
ENSG00000062716	E015	0.00286459949724193	0	0	-0.30851002640611
ENSG00000062716	E016	0.00425652256325295	0	0	-0.305026234124033
ENSG00000062716	E017	0.00356481803370272	0	0	-0.335980847377032
ENSG00000062716	E018	0.00304181606709994	0	0	-0.3485959073919
ENSG00000062716	E019	0.00324253922385614	0	0	-0.362150147316055
ENSG00000062716	E020	0.00367726498344995	0	0	-0.366566409915974
ENSG00000063177	E018	0.0180659071070234	0	0	0.843553863217536
ENSG00000064419	E024	0.00509427483521333	0	0	0.501746375626447
ENSG00000064419	E025	0.00387433347133388	0	0	0.478221693303528

Supplementary material 2. Contd.

ENSG00000065559	E019	0.0047346883521754	0	0	-0.3892548049488
ENSG00000065833	E015	0.00638213123629173	0	0	0.643325750042247
ENSG00000065833	E016	0.00713957796110783	0	0	0.759016398261631
ENSG00000067225	E042	0.0203550550372131	0	0	110,983,190,688,681
ENSG00000068784	E013	0.00640862683473831	0	0	0.588913719428765
ENSG00000068784	E015	0.00490848909695605	0	0	0.530878094223965
ENSG00000068784	E016	0.0057852745979348	0	0	0.519287317592044
ENSG00000069020	E008	0.00799151004978879	0	0	0.621478519504317
ENSG00000070018	E037	0.00805879332343016	0	0	0.595272496170778
ENSG00000073921	E049	0.00239473416240031	0	0	0.309446443510305
ENSG00000073921	E051	0.00272295771365416	0	0	0.343864111345239
ENSG00000075415+ENSG00000212443	E022	0.00152116185896419	0	0	-0.268723862804229
ENSG00000075415+ENSG00000212443	E027	0.00179945036335906	0	0	-0.288335397378916
ENSG00000075415+ENSG00000212443	E028	0.00202095372077982	0	0	-0.296109552567584
ENSG00000075415+ENSG00000212443	E029	0.00143577303941475	0	0	-0.24120068415607
ENSG00000075415+ENSG00000212443	E030	0.00105444398579539	0	0	13,134,421,512,385
ENSG00000075415+ENSG00000212443	E031	0.00267605182220506	0	0	151,362,757,630,403
ENSG00000075415+ENSG00000212443	E033	0.00130871368914569	0	0	-0.267558868234547
ENSG00000075415+ENSG00000212443	E035	0.00143944275597132	0	0	-0.296074491191454
ENSG00000075415+ENSG00000212443	E036	0.00129603734922975	0	0	-0.298857446926918
ENSG00000077454+ENSG00000205307	E023	0.0129472931725166	0	0	0.937501657360712
ENSG00000080815	E025	0.00480298722660499	0	0	0.706879670381831
ENSG00000080815	E027	0.00688613769902666	0	0	0.79462813741011
ENSG00000080815	E028	0.00408933751263672	0	0	0.749617930831614
ENSG00000080815	E029	0.00382638192974004	0	0	0.633359607590383
ENSG00000082996	E047	0.00246856015033472	0	0	-0.302374927302371
ENSG00000083544	E010	0.00552035550206321	0	0	0.533214394515569
ENSG00000083544	E011	0.00432208822840506	0	0	0.444470863022972
ENSG00000084676	E008	0.0113129515504324	0	0	0.742318956342851
ENSG00000087206	E020	0.00610244376068148	0	0	0.511025033513407
ENSG00000088808	E052	0.00719296593685567	0	0	0.668990644176052
ENSG00000088808	E054	0.00875987294228689	0	0	0.741674100350788
ENSG00000089280	E039	0.00877485746148689	0	0	0.950136473057835
ENSG00000096746	E021	0.00812591195113053	0	0	117,421,325,765,102
ENSG00000099901	E031	0.00305195841705114	0	0	0.983367839219162
ENSG00000100941	E006	0.0206335783689133	0	0	11,693,174,711,559
ENSG00000101236	E002	0.0109607296293204	0	0	-0.531431758519199
ENSG00000101745	E026	0.00294489428487539	0	0	-0.334440688443602
ENSG00000104738	E015	0.0282472557665178	0	0	140,102,862,544,272
ENSG00000105176	E005	0.00561896839100511	0	0	0.52574623890558
ENSG00000105176	E006	0.00286607974584592	0	0	0.367031436796265
ENSG00000106462	E033	0.0068757494087806	0	0	0.549818906633497
ENSG00000107077+ENSG00000225489	E026	0.00784186024918819	0	0	0.672843604902932
ENSG00000107077+ENSG00000225489	E027	0.00862568037441722	0	0	0.742512595954742
ENSG00000109184	E048	0.00224309599285509	0	0	-0.241803208462778
ENSG00000109381	E032	0.00371487388040335	0	0	0.583175186270463
ENSG00000109670	E016	0.00288603265525797	0	0	0.300404990880352
ENSG00000110422+ENSG00000223134	E004	0.00598991292389415	0	0	0.528736156353527
ENSG00000110422+ENSG00000223134	E005	0.00287797953828343	0	0	0.412427216460058
ENSG00000111057	E018	0.00683382279778425	0	0	0.760593435804187
ENSG00000111371	E021	0.00172571436230456	0	0	0.285422353013813
ENSG00000111371	E022	0.00155471180260163	0	0	0.304166388693279

Supplementary material 2. Contd.

ENSG00000111371	E023	0.00121258732135146	0	0	0.382947940494121
ENSG00000111371	E024	0.00177287313329524	0	0	0.429832701792157
ENSG00000111371	E025	0.00673870834043841	0	0	0.702039552277118
ENSG00000111907	E024	0.00976249551033376	0	0	0.95357227979476
ENSG00000112851	E012	0.00232381998253929	0	0	0.490348157953446
ENSG00000112851	E013	0.00347586013562747	0	0	0.456944892834843
ENSG00000112893	E002	0.00317823266743197	0	0	0.492430327273209
ENSG00000112893	E003	0.00334748098552705	0	0	0.478273486420261
ENSG00000113643	E009	0.00225700164346493	0	0	0.396964193857818
ENSG00000113643	E010	0.00164000710748988	0	0	0.375127949708799
ENSG00000113643	E011	0.00187734282512215	0	0	0.370650749366847
ENSG00000113643	E013	0.00197520780529728	0	0	0.336827188191702
ENSG00000113643	E014	0.00227764236649451	0	0	0.320264470637763
ENSG00000113643	E029	0.00302403413694174	0	0	-0.351054375360191
ENSG00000113643	E032	0.00298329231540746	0	0	-0.350005775040493
ENSG00000113643	E033	0.00349229022771841	0	0	-0.391157974689573
ENSG00000114062	E002	0.00230106068441794	0	0	-0.327897303276237
ENSG00000114062	E016	0.00626506375865901	0	0	0.696718717460068
ENSG00000114062	E019	0.00727796516475796	0	0	0.670851499558417
ENSG00000114062	E020	0.00666757498149153	0	0	0.636686595223821
ENSG00000114062	E021	0.00294456300796946	0	0	0.619166465060826
ENSG00000114062	E022	0.0047453885810777	0	0	0.714667115837895
ENSG00000115053	E011	0.0174307505260267	0	0	0.956372961702347
ENSG00000115053	E013	0.0156844461759044	0	0	0.937540618360008
ENSG00000115109	E028	0.00236605544736564	0	0	-0.250438862134361
ENSG00000115109	E031	0.00250881042409631	0	0	0.817459630556919
ENSG00000115109	E033	0.0025187534100837	0	0	0.809713281704265
ENSG00000115109	E034	0.00252918259297746	0	0	0.767372756661953
ENSG00000115109	E035	0.00359136144768794	0	0	0.838983065087876
ENSG00000115109	E037	0.00379696035904649	0	0	0.750357993394999
ENSG00000115109	E038	0.00426666549720381	0	0	0.726117824877599
ENSG00000115109	E039	0.00494002930958599	0	0	0.703335604309351
ENSG00000115109	E040	0.00757486945342753	0	0	0.706793028856247
ENSG00000115109	E041	0.00560834337136765	0	0	0.653626614057737
ENSG00000115109	E044	0.00328057860142173	0	0	-0.431820103842305
ENSG00000115947	E027	0.00393841091047314	0	0	0.394557249798872
ENSG00000115947	E029	0.014359638168039	0	0	104,574,791,083,796
ENSG00000117868	E036	0.00332960442370569	0	0	0.409338273052346
ENSG00000120071	E013	0.00888880095647917	0	0	0.650660762856621
ENSG00000120071	E014	0.00359246151293338	0	0	-0.318044300187656
ENSG00000120438+ENSG00000206910+ENSG00000207392	E010	0.0143232369671226	0	0	155,002,194,713,202
ENSG00000120438+ENSG00000206910+ENSG00000207392	E011	0.0129360786438524	0	0	149,977,401,705,643
ENSG00000121741	E033	0.016180250688066	0	0	116,552,030,219,602
ENSG00000121989	E008	0.0136890307268968	0	0	0.836706580293666
ENSG00000122566	E004	0.00386124696591086	0	0	0.436883965780834
ENSG00000123066	E048	0.00161924012474238	0	0	0.757927671681865
ENSG00000128585	E014	0.00588291761095006	0	0	0.524555117399378
ENSG00000133316+ENSG00000222328	E054	0.00328592820223505	0	0	193,294,955,185,962
ENSG00000134108	E018	0.0249809117446936	0	0	146,428,235,596,882
ENSG00000134222	E014	0.0281004655331887	0	0	105,114,629,849,682

Supplementary material 2. Contd.

ENSG00000134419+ENSG00000170540+ENSG00000260342	E006	0.0015086881878047	0	0	0.210328225098956
ENSG00000134419+ENSG00000170540+ENSG00000260342	E009	0.00125801049076872	0	0	0.183361081044624
ENSG00000134419+ENSG00000170540+ENSG00000260342	E012	0.00102332113069753	0	0	0.176785804925073
ENSG00000134419+ENSG00000170540+ENSG00000260342	E013	0.00105111512651973	0	0	0.181709375341326
ENSG00000134419+ENSG00000170540+ENSG00000260342	E014	0.00100104210755371	0	0	0.191534482214219
ENSG00000134419+ENSG00000170540+ENSG00000260342	E024	0.00254824865578304	0	0	-0.226258218108307
ENSG00000134419+ENSG00000170540+ENSG00000260342	E025	0.00192833614241941	0	0	-0.274120390856121
ENSG00000134419+ENSG00000170540+ENSG00000260342	E026	0.0029083726592252	0	0	-0.272566284242751
ENSG00000134419+ENSG00000170540+ENSG00000260342	E028	0.00195603894854	0	0	-0.257212555473744
ENSG00000134419+ENSG00000170540+ENSG00000260342	E031	0.00176796633179388	0	0	-0.239324489671118
ENSG00000134758	E004	0.00675923797830655	0	0	0.610963881687083
ENSG00000134909	E040	0.0200404299607068	0	0	106,806,777,849,084
ENSG00000134909	E041	0.0164959595399716	0	0	0.984878220836215
ENSG00000135821	E023	0.00652021775202899	0	0	105,818,815,620,248
ENSG00000135829	E033	0.00821932343757605	0	0	0.737277622503705
ENSG00000136021	E007	0.00361595781726588	0	0	0.557131981141342
ENSG00000136021	E008	0.00533979098241841	0	0	0.73494290008176
ENSG00000136021	E009	0.00304139629458152	0	0	0.615794672529598
ENSG00000136021	E012	0.00386544580126646	0	0	0.466288891855297
ENSG00000136492	E007	0.00188835582227356	0	0	0.391173324979618
ENSG00000136492	E008	0.00200962436209889	0	0	0.399591328093891
ENSG00000136699	E029	0.0190478639124207	0	0	101,999,477,752,199
ENSG00000137776	E056	0.00459299355839347	0	0	0.453702984870248
ENSG00000138346	E033	0.00462626175999363	0	0	0.468281872765351
ENSG00000138376	E001	0.00262350485773804	0	0	-0.337749437085777
ENSG00000139597+ENSG00000139617+ENSG00000244754	E052	0.00349480013417698	0	0	0.432517682531562
ENSG00000139597+ENSG00000139617+ENSG00000244754	E053	0.00359886441124523	0	0	0.494331300388985
ENSG00000139597+ENSG00000139617+ENSG00000244754	E055	0.00324763120904549	0	0	0.485813293929168
ENSG00000139597+ENSG00000139617+ENSG00000244754	E060	0.00444709803935796	0	0	0.441739703725799
ENSG00000140396	E035	0.00393256248044073	0	0	0.59557174165757
ENSG00000140396	E036	0.00583870341377709	0	0	0.711897360662693
ENSG00000140396	E037	0.00713046629089922	0	0	0.698407826058823
ENSG00000140988+ENSG00000207405+ENSG00000255513+ENSG00000206811	E003	0.0048916412936643	0	0	-0.415414360922689
ENSG00000140988+ENSG00000207405+ENSG00000255513+ENSG00000206811	E004	0.00307670268826956	0	0	-0.307003565666234
ENSG00000140988+ENSG00000207405+ENSG00000255513+ENSG00000206811	E010	0.00450419661335741	0	0	0.833760532630357
ENSG00000140988+ENSG00000207405+ENSG00000255513+ENSG00000206811	E019	0.0158543700016008	0	0	0.753205094081311

Supplementary material 2. Contd.

ENSG00000140988+ENSG00000207405+ENSG00000255513+ENSG00000206811	E020	0.0049790530462157 2	0	0	100,042,608,404,356
ENSG00000143771	E019	0.00580520097517474	0	0	-0.5041393883146
ENSG00000143797	E026	0.00302999310622773	0	0	0.640097884989721
ENSG00000143797	E027	0.00296215379944861	0	0	0.801703371605897
ENSG00000143797	E028	0.00485746264540363	0	0	0.805104858767256
ENSG00000144036	E031	0.00546444790283555	0	0	0.502356874014504
ENSG00000144893	E006	0.00714940644321631	0	0	0.703862541135284
ENSG00000144935	E005	0.00600887095068696	0	0	0.581695853779746
ENSG00000144935	E006	0.00983878029905456	0	0	0.731898124492107
ENSG00000144935	E007	0.00452401038048158	0	0	0.520503916422583
ENSG00000145833+ENSG00000181904	E040	0.0168345677037895	0	0	103,200,845,784,413
ENSG00000146247	E041	0.00586044114182191	0	0	0.52106281646343
ENSG00000146247	E042	0.00621775722472215	0	0	0.527588929981903
ENSG00000146433	E006	0.00533184361147513	0	0	0.610389158762169
ENSG00000148334	E011	0.0170996959916619	0	0	0.98654097837898
ENSG00000148334	E012	0.00848502258958842	0	0	0.974722622401965
ENSG00000151292	E005	0.00238770654861252	0	0	0.447009837865454
ENSG00000151292	E006	0.00202357155119877	0	0	0.465030886133111
ENSG00000151292	E028	0.00513829761767353	0	0	-0.465483799698602
ENSG00000151466	E027	0.00671282815114215	0	0	0.547465798993094
ENSG00000153147	E023	0.002629307396414	0	0	0.341231102665827
ENSG00000153147	E024	0.00300773535154315	0	0	0.447890073547592
ENSG00000153147	E025	0.00256410264377357	0	0	0.376260255571111
ENSG00000155313	E004	0.00525180488605466	0	0	0.48555581709588
ENSG00000156011+ENSG00000244018	E005	0.00240412419426961	0	0	-0.281446493575435
ENSG00000156011+ENSG00000244018	E036	0.00689863041310596	0	0	0.597189559696127
ENSG00000156011+ENSG00000244018	E037	0.00547028551311603	0	0	0.617785477889397
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E016	0.00275340047434549	0	0	-0.38303906670846
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E030	0.00279960607623052	0	0	-0.408457196702123
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E031	0.00249095801198147	0	0	-0.431363680997929
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E032	0.00265598586358416	0	0	-0.370202825829498
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E033	0.00261195016889289	0	0	-0.414224393549785
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E034	0.0035776946911356	0	0	-0.422482033521442
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E037	0.00289849893607011	0	0	-0.406822286247417
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E038	0.00334877306929115	0	0	-0.41262176799185
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E044	0.00243606317247174	0	0	-0.467753878000779

Supplementary material 2. Contd.

ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E045	0.00345762767455533	0	0	-0.488182481794944
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E047	0.00472895529548749	0	0	0.435805022974449
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E049	0.00223173170649903	0	0	-0.498517262070707
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E052	0.00115189549505119	0	0	0.935724987229728
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E053	0.00137491633317313	0	0	0.691034785698287
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E054	0.0013975561364819	0	0	0.691073847995375
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E055	0.0014030438546344	0	0	0.68928044614001
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E056	0.00345856540979694	0	0	0.42783795705693
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E062	0.00245951902558292	0	0	-0.478019279500372
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E066	0.00327164795828963	0	0	-0.478799452946402
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E067	0.00350605632252888	0	0	-0.542164258795581
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E068	0.00515699694280688	0	0	-0.555088249985218
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E070	0.00203003877138065	0	0	-0.435397169463498
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E071	0.00380642449170613	0	0	-0.500010757652924
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E072	0.00191963090273959	0	0	-0.505857709281614
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E073	0.00202839488708531	0	0	-0.49491682781671
ENSG00000156976+ENSG00000200418+ENSG00000221420+ENSG00000238942+ENSG00000200320	E074	0.00179083331632279	0	0	-0.467558407351389
ENSG00000157107	E035	0.0064007240924872	0	0	0.741962665901774
ENSG00000157107	E036	0.00562177269318031	0	0	0.770913659764632
ENSG00000157107	E037	0.00798939232042488	0	0	0.653821831849602
ENSG00000161016	E027	0.00143609841806585	0	0	0.223354970922777
ENSG00000161016	E028	0.00279141700605149	0	0	0.338134525510153
ENSG00000161960+ENSG00000207152	E015	0.00721797104986602	0	0	0.865907954109202

Supplementary material 2. Contd.

ENSG00000163399	E022	0.0083465251426942	0	0	0.843606390042178
ENSG00000163527	E006	0.00184387440180405	0	0	0.241655481290931
ENSG00000163960	E001	0.00215200436341718	0	0	-0.22135756559334
ENSG00000163960	E011	0.00639705509140356	0	0	0.571678661886178
ENSG00000163960	E012	0.00700355643766015	0	0	0.728965166959754
ENSG00000163960	E013	0.00763245451948193	0	0	0.70157385489661
ENSG00000164898+ENSG00000146963	E009	0.00703578540729069	0	0	0.639251231077684
ENSG00000164898+ENSG00000146963	E011	0.00526343079222993	0	0	0.520757057855848
ENSG00000165322	E035	0.0016795661792196	0	0	0.461377389312872
ENSG00000165322	E036	0.00399604584864479	0	0	0.657302131724785
ENSG00000165322	E037	0.00802494090171092	0	0	0.653562893945478
ENSG00000165458	E030	0.0249500277094995	0	0	132,310,588,956,013
ENSG00000165458	E031	0.00997478438278549	0	0	127,435,235,988,798
ENSG00000166441+ENSG00000200983	E015	0.0127008087518793	0	0	116,897,921,519,154
ENSG00000166441+ENSG00000200983	E016	0.0127099666227081	0	0	10,988,627,912,785
ENSG00000166508	E026	0.0107712172226513	0	0	0.776312080658173
ENSG00000168234	E004	0.00799266268460179	0	0	0.613508174053404
ENSG00000168234	E017	0.00706536296863918	0	0	-0.54695122080361
ENSG00000168300	E015	0.0017959425060779	0	0	0.390997291771277
ENSG00000170571	E001	0.0039451242413448	0	0	-0.417074363235088
ENSG00000170571	E003	0.0024786415473555	0	0	-0.378387135132571
ENSG00000171132	E052	0.00739983119544396	0	0	-0.551582891048824
ENSG00000171456	E018	0.00777623331476681	0	0	102,567,100,973,926
ENSG00000174748	E033	0.00127123288715843	0	0	-0.175037895385076
ENSG00000174748	E034	0.00138393505120624	0	0	-0.27209252320527
ENSG00000174748	E035	0.0016730790310939	0	0	-0.252183364250906
ENSG00000175029	E029	0.0220718808482376	0	0	144,558,759,284,373
ENSG00000177600+ENSG00000199785	E014	0.00764616695379164	0	0	120,084,782,180,684
ENSG00000180573	E002	0.00221143715264896	0	0	0.2314128736999
ENSG00000180573	E004	0.00187775536125918	0	0	-0.213662493387306
ENSG00000182568+ENSG00000131374	E053	0.0154306133649054	0	0	0.916574865396555
ENSG00000184990	E016	0.0116844702678487	0	0	-0.715276347403957
ENSG00000185122	E011	0.00667644944274783	0	0	0.801609701353347
ENSG00000185122	E012	0.00869077668690626	0	0	0.75800522779588
ENSG00000188486	E002	0.002501804818815	0	0	-0.246722814514588
ENSG00000188486	E004	0.00465967186614012	0	0	0.660763644804424
ENSG00000188994	E014	0.00665864352520593	0	0	0.65137328793776
ENSG00000188994	E015	0.00664451917287129	0	0	0.609717313954432
ENSG00000188994	E028	0.00160039093305463	0	0	-0.174113784354817
ENSG00000196305	E018	0.0106171881395179	0	0	108,663,149,435,791
ENSG00000196323	E030	0.00264437984914019	0	0	0.32865159143385
ENSG00000196323	E033	0.00399908687553767	0	0	0.431836331628507
ENSG00000196507+ENSG00000172465	E007	0.00316649487083061	0	0	-0.334080891083039
ENSG00000196562	E047	0.00250212083003243	0	0	0.323648412117151
ENSG00000197323	E027	0.00145788738232438	0	0	0.333261343772383
ENSG00000197323	E028	0.00685203627436341	0	0	0.492107700640735
ENSG00000197323	E029	0.00139928204319553	0	0	0.450979355582536
ENSG00000197323	E030	0.00183679922227043	0	0	0.428880069858297
ENSG00000197409+ENSG00000196866	E001	0.00172784322418301	0	0	-0.225309873926191
ENSG00000197555	E008	0.0153539080947535	0	0	12,773,880,146,778
ENSG00000197555	E009	0.0138347962079179	0	0	118,283,138,392,894
ENSG00000197555	E010	0.00693835978409769	0	0	120,925,044,407,473

Supplementary material 2. Contd.

ENSG00000197555	E013	0.00795128120591918	0	0	123,557,179,344,223
ENSG00000198162	E004	0.00121154922061305	0	0	0.188235732881353
ENSG00000198162	E005	0.00153945011246118	0	0	0.215594596846282
ENSG00000198218	E011	0.00601535964135193	0	0	0.927244174929793
ENSG00000198218	E012	0.010144416450924	0	0	0.99351403204386
ENSG00000198363	E062	0.0043063981614386	0	0	0.553159372699458
ENSG00000198363	E063	0.00698653261112429	0	0	119,685,344,263,619
ENSG00000198363	E065	0.00349962656590338	0	0	0.661822814389495
ENSG00000198815	E019	0.00646598764600319	0	0	0.591750628117065
ENSG00000200259+ENSG00000201675+ENSG00000142541+ENSG00000202503+ENSG00000199631	E007	0.00779977227275279	0	0	0.718331946549828
ENSG00000200259+ENSG00000201675+ENSG00000142541+ENSG00000202503+ENSG00000199631	E008	0.00672936930243383	0	0	0.657555795340283
ENSG00000200463+ENSG00000179029	E002	0.0140137156483014	0	0	0.734536179627057
ENSG00000201129+ENSG00000143569	E048	0.0179917164916879	0	0	0.971192775925756
ENSG00000201808+ENSG00000242125+ENSG00000180198	E008	0.00226670779248858	0	0	22,836,062,220,647
ENSG00000201808+ENSG00000242125+ENSG00000180198	E011	0.00432483098196286	0	0	-405,813,828,106,499
ENSG00000201808+ENSG00000242125+ENSG00000180198	E048	0.00309453449892614	0	0	-465,753,535,559,348
ENSG00000204764	E054	0.00294568721692079	0	0	0.538583738222738
ENSG00000204764	E055	0.00348766153794291	0	0	0.593705920647754
ENSG00000204764	E056	0.00366240026945257	0	0	0.562404109509662
ENSG00000204842	E062	0.00368399705868222	0	0	0.515099379524403
ENSG00000204842	E064	0.00387052946002826	0	0	0.613313482164015
ENSG00000204842	E065	0.00804495883922123	0	0	0.669394526913633
ENSG00000207165+ENSG00000147403	E027	0.00169692053801728	0	0	0.617375222522592
ENSG00000212487+ENSG00000199437+ENSG00000233016	E004	0.00260564290326098	0	0	-0.59854798372185
ENSG00000212487+ENSG00000199437+ENSG00000233016	E005	0.00387684972471763	0	0	-0.607955590561775
ENSG00000212487+ENSG00000199437+ENSG00000233016	E007	0.0014830511297634	0	0	0.775476001818444
ENSG00000212487+ENSG00000199437+ENSG00000233016	E008	0.00622622892324662	0	0	0.55820527085017
ENSG00000212487+ENSG00000199437+ENSG00000233016	E009	0.00371124919009088	0	0	-0.594810946999488
ENSG00000215021+ENSG00000238795	E015	0.00507281302953734	0	0	0.415999483987769
ENSG00000215021+ENSG00000238795	E016	0.00462822138549259	0	0	0.446848247007522
ENSG00000215845+ENSG00000158769	E027	0.00236208580856115	0	0	0.272870922649241
ENSG00000234741+ENSG00000200729+ENSG00000200710+ENSG00000201692+ENSG00000202394+ENSG00000208313+ENSG00000200016+ENSG00000200954+ENSG00000206607	E048	0.00484913822193553	0	0	0.526348587933101
ENSG00000234741+ENSG00000200729+ENSG00000200710+ENSG00000201692+ENSG00000202394+ENSG00000208313+ENSG00000200016+ENSG00000200954+ENSG00000206607	E049	0.00365343042135109	0	0	0.524242044103493
ENSG00000241111+ENSG00000241572	E001	0.00389814531962009	0	0	0.353547187314918
ENSG00000246203+ENSG00000163374	E040	0.0061591245144764	0	0	0.441834318257926

Supplementary material 2. Contd.

ENSG00000251790+ENSG00000234912+ENSG00000129657	E012	0.00124939886827933	0	0	105,522,595,700,129
ENSG00000251790+ENSG00000234912+ENSG00000129657	E013	0.00461526473667066	0	0	0.48065205874616
ENSG00000251790+ENSG00000234912+ENSG00000129657	E039	0.012870208741594	0	0	-0.913531141877969
ENSG00000251790+ENSG00000234912+ENSG00000129657	E041	0.0169985343704731	0	0	-0.986462226319171
ENSG00000251790+ENSG00000234912+ENSG00000129657	E059	0.0111969522820811	0	0	-0.857889014514752
ENSG00000258508+ENSG00000198604	E047	0.00394845052589824	0	0	0.641303219622624
ENSG00000258941+ENSG00000150527+ENSG00000150526	E034	0.00786058374970822	0	0	0.829882498722205
ENSG00000258941+ENSG00000150527+ENSG00000150526	E035	0.0116095457657273	0	0	0.838478673247404
ENSG00000259001+ENSG00000252678	E001	0.00133201199818049	0	0	-0.132048289405817
ENSG00000259001+ENSG00000252678	E002	0.000823130909139826	0	0	0.0330027322644232
ENSG00000259932+ENSG00000259539+ENSG00000138606	E042	0.00779545365127278	0	0	0.513429082210731