

Nkx3-1 and *Fech* Genes Might be Switch Genes Involved in Pituitary Non-Functioning Adenoma Invasiveness

Nasibeh Khayer

Iran University of Medical Sciences

Maryam Jalessi

Iran University of Medical Sciences

Amin Jahanbakhshi

Iran University of Medical Sciences

Alireza Tabib khoeei

Iran University of Medical Sciences

Mehdi Mirzaie (✉ mirzaie@modares.ac.ir)

Tarbiat Modares University

Research Article

Keywords: Non-functioning pituitary adenomas, Invasiveness, fast liquid association, Random forest classification, Gene set enrichment analysis.

Posted Date: May 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-536826/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on October 22nd, 2021. See the published version at <https://doi.org/10.1038/s41598-021-00431-2>.

Abstract

Non-functioning pituitary adenomas (NFPAs) are typical pituitary macroadenomas in adults that are associated with increased mortality and morbidity. Although pituitary adenomas are commonly considered slow-growing benign brain tumors, numerous of them possess an invasive nature. Such tumors destroy sella turcica and invade the adjacent tissues such as the cavernous sinus and sphenoid sinus. The most critical obstacle for complete surgical removal in these cases is the high risk of damaging adjacent vital structures. Therefore, the development of novel therapeutic strategies for either early diagnosis through biomarkers or medical therapies to reduce the recurrence rate of NFPAs is imperative. Identification of gene interactions has paved the way for decoding complex molecular mechanisms, including disease-related pathways, and identifying the most momentous genes involved in a specific disease. Currently, our knowledge of the invasion of the pituitary adenoma at the molecular level is not sufficient. The current study aimed to identify critical biomarkers and biological pathways associated with invasiveness in the NFPAs using a three-way interaction model for the first time. In the current study, the Liquid association method was applied to capture the statistically significant triplets that are involved in NFPAs invasiveness. Subsequently, Random Forest analysis was applied to select the most important switch genes. Finally, gene set enrichment (GSE) and gene regulatory network (GRN) analyses were applied to trace the biological relevance of the statistically significant triplets. The results of this study suggest that “mRNA processing” and “spindle organization” biological processes are important in NFPAs invasiveness. Specifically, our results suggest *Nkx3-1* and *Fech* as two switch genes in NFPAs invasiveness that may be a potential biomarkers or target genes, in this pathology.

Introduction

Pituitary adenomas (PAs) are the second most common primary brain tumors with substantial mortality rates [1, 2]. PAs are categorized into non-functioning and functioning types based on clinical and biochemical features. Non-functioning pituitary adenomas (NFPAs) are the most common type of PAs in adults. In contrast with the functioning pituitary adenomas (FPAs), which release additional levels of endocrine hormones, NFPAs are not hormonally active [3]. The absence of any clinical and biochemical signs of hormone-excess leads to the late detection of NFPAs.

Furthermore, PAs are commonly considered slow-growing benign brain tumors, but a large number of them exhibit a local invasive behavior that is unpredictable with the aid of current tumor biomarkers [4]. The invasive PAs destroy sella turcica and invade the adjacent tissues such as the cavernous and sphenoid sinus. The most critical obstacle for total surgical removal is the high risk of involvement of adjacent vital nervous or vascular structures. On the other hand, despite technological improvements in surgical approaches and radiotherapy, the recurrence risk of invasive NFPAs remains high [5]. Therefore, the development of novel therapeutic strategies for early diagnosis as well as decreasing the recurrence rate of NFPAs is imperative. Hence, a comprehensive biological insight into the NFPAs invasiveness procedure is a primary step to achieve the above purpose.

High throughput gene expression data (i.e., the transcriptome) provide genome-scale snapshots of gene expression, rich sources of information for inferring gene relationships [6, 7]. Identification of gene interactions has paved the way for decoding complex molecular mechanisms including disease-related pathways, and identifying the most momentous genes involved in a specific disease [8]. With the purpose of developing diagnostic and therapeutic strategies, several biomarkers and pathways have been reported associated with invasiveness in NFPA through gene expression data analysis. Some of the most momentous potential biomarkers related to the aggressive nature of NFPA are pituitary tumor transforming gene 1 (PTTG1) [9], Ezrin (EZR) [10], Ectoderm-Neural Cortex 1 (ENC1) [11], WNT Inhibitory Factor 1 (WIF1) [12], E-cadherin (CDH1) and Neural cell adhesion molecule (NCAM) [13]. Moreover, previous studies identified a perturbation in some signaling pathways that can make NFPA prone to invasiveness. The main reported pathways include the “WNT signaling pathway” [12], “local suppression of the immune response pathway”, “TGF- β signaling” [14], “PI3K-Akt signaling pathway” and “chemokine signaling pathway” [15]. However, notwithstanding that the molecular markers and pathways associated with NFPA invasiveness are extensively studied, much remains unknown.

Depending on applied mathematical and statistical methods, various gene expression patterns can be traced from the same biological dataset [16]. It should be noted that the above studies were done based on two-way gene interaction approach. In the current study, we aimed to trace three-way gene interaction pattern in the NFPA microarray gene expression dataset. We used the Liquid Association method [17]. The three-way gene interaction pattern draws the dynamic nature of the co-expression relation of two genes by proposing a third gene known as a switch gene [16]. Such pattern deciphers the sophisticated molecular relations at a higher level than the conventional two-way gene interaction pattern, including co-expression [18] and differentially co-expression [19] patterns. Therefore, it can lead to a more comprehensive and explicit biological insight in to the cause of cellular changes. The successful identification of the switch genes in diseases can be consequential because they can be regarded as potential drug targets. In the meantime, switching genes can be helpful in decoding biological complexities [20, 21].

To the best of our knowledge, the three-way gene interaction model is not investigated for NFPA's gene expression data. The main challenge to implementing a three-way interaction model is presumably a large number of possible interactions for more than two genes at the genome-scale that result to a high computational load.

The current study aimed to identify critical biomarkers and biological pathways associated with invasiveness in the NFPA, using three-way interaction model. We hope that the results of this study provide efficient therapeutic targets and diagnostic or prognostic biomarkers.

Materials And Methods

Gene Expression Profiling Dataset

The selected dataset includes gene expression of 22 invasive and 18 non-invasive NFPA, which is available at the Array Express database [87] under accession number E-TABM-899 [88]. Additionally, it was generated using the A-AGIL-11- Agilent Human Whole-genome microarray platform. The background correction on the raw microarray dataset was carried out using the Normex method [89]. Furthermore, the expression profiles were normalized within- and between- arrays using loess [90] and quantile normalization[91] methods, respectively. It should be noted that above-mentioned methods were implemented in the Limma R package [92].

Moreover, the duplicate probes were removed using the genefilter package [93]. Accordingly, the highest interquartile range (IQR) across probes corresponds to each gene is retained. Furthermore, unchanged genes were removed from the microarray dataset because they do not provide valuable information to decipher gene expression relationships. For this purpose, the empirical Bayes method [94] was used to detect differentially expressed genes (DEGs). Additionally, the Benjamini-Hochberg method [95] was used to control the false discovery rate. By considering p -value < 0.01 as the threshold, DEGs include 2321 genes that were selected for further consideration.

Liquid association triplets

Three-way interactions between all genes involved in the dataset were calculated using the fastMLA function in the fastMLA R package [96]. This package uses a modified liquid association algorithm for determining changes in coexpression relations of a gene pair, X_1 and X_2 , based on the expression level of a third gene (X_3).

Indeed, the fast modified liquid association algorithm computes an MLA score for each gene triplet to assess the magnitude of the liquid association. More specifically, $MLA(X_1, X_2 | X_3)$ can be estimated as:

$$\hat{MLA} = \frac{\sum_i^M \hat{\rho}_i \overline{X_{3i}}}{M}$$

where M is the number of bins over X_3 , $\hat{\rho}_i$ is the Pearson's correlation coefficient of X_1 and X_2 in samples of the i th bin, and $\overline{X_{3i}}$ is the mean of expression values of X_3 in the i th bin.

It should be not that before running fastMLA, performing two preprocessing steps are required: (i) to reduce the number of potential outliers in the data, the marginal distribution of each variable should be normal. Therefore, a normal quantile transformation was performed based on Li's approach [97]; (ii) each variable should be standardized to have mean 0 and variance 1 [98]. The first preprocessing was performed using an in-house implementation, while the second one by using the CTT package[99].

False discovery rate (FDR) was estimated using the Benjamini-Hochberg correction method, and liquid association triplets with $FDR < 0.001$ were chosen as statistically significant triplets. Subsequently, all triplets with the non-random observed rate in X_3 position genes were retained for further study.

Random Forests Clustering

Random Forest (RF) is a powerful ensemble algorithm based on machine learning. Such an algorithm generates a collection of decision trees that are learned independently by bootstrap sampling. Each tree recursively divides observations into more homogeneous subsets. Finally, the outcome is obtained by combining a collection of accurately chosen classification trees.

A random forests classifier was built using the randomForest R package [22]. The “number of decision trees” and “mtry” parameters were set to “10,000 trees” and “square root of the total number of features”, respectively [100]. Finally, the gene importance measure is computed by averaging the increase in the error rate over all the trees.

Pathway and Functional Enrichment Analysis

Functional enrichment analysis is utilized to ascertain biologically-relevant triplets and determine the central pathways and biological processes involved in PA. Functional enrichment analysis is a statistical method to classify genes (proteins) over-presented in a particular dataset using predefined annotations [101]. For all of the genes involved in all statistically significant triplets, functional enrichment analysis was performed based on the biological process using the gene ontology (GO) database. Furthermore, the same analyses were performed to find enriched pathways in the KEGG database [102]. For the analyses, as mentioned earlier, we used the ClueGO tool [103] (with a Kappa threshold of 0.4) within the Cytoscape v.3.3.0 environment [104]. The right-sided hypergeometric test and the Benjamini-Hochberg correction method [105] were used for validation of enrichment analysis. Subsequently, comparing the enriched GO terms and KEGG pathways was performed to recognize the different biological processes between the invasive and non-invasive samples.

Gene Regulatory Network Construction

A gene regulatory network (GRN) models complex regulatory mechanisms that control the gene expression levels of mRNA, which, in turn, govern the function of the cell. A GRN consists of nodes (genes) and edges (regulatory relations) that can help to predict changes in gene expression under different conditions [106]. Here, we used ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) [107] for constructing the GRN. ARACNE is a reverse engineering approach for the construction of cellular networks from gene expression data. This algorithm captures directed regulatory interactions between each transcriptional regulator and its potential targets based on mutual information. ARACNE runs in the geWorkbench_2.6.0 framework for all of the genes involved in the statistically significant triplets by considering $p\text{-value} < 0.05$.

Results

Determining statistically significant three-way interaction

Using the fastLA package, liquid association analysis was performed for every combination of a candidate switching gene (X_3) and every possible pair of genes in the dataset $\{X_1, X_2\}$. The top 200000 triplets with the highest significance levels based on p -value were defined as outputs of this analysis. A p -value histogram of these three-way interactions is available in S1 Fig. To survey the validity of fastLA analysis, the observed event rate of X_3 position (switch) genes was compared with random event rate in the wide range of the significant fastLA p -values. The plots of such comparison are presented in Fig. 1. Furthermore, changes in FDR using the Benjamini-Hochberg method versus $-\log(p\text{-value})$ for the first 200000 triplets are shown in the S2 Fig.

For the rest of our analysis, the set of all three-way interactions were chosen by considering $FDR < 0.001$ and, in addition, non-random observed rate in X_3 position genes, consisting of 124 triple combinations. The list of all statistically significant triplets is presented in S3 Table.

Gene selection using random forest

Several measures of variable importance are obtained using the random forest algorithm. The most reliable measure is Mean Decrease Accuracy (MDA), which is based on the decrease of classification accuracy when the expression values of a particular gene are randomly permuted [22, 23]. We reported 25 top importance genes selected based on MDA in Fig. 2. Furthermore, the area under the receiver operating characteristic (ROC) curve (AUC) is widely used as an assessment indicator to evaluate the performance of supervised classification models [24]. Therefore, ROC curves were used to analyze the sensitivity and specificity of the Random Forest model. As demonstrated in the S4 Fig, the AUC, sensitivity, and specificity of the classifier are 0.70, 67, and 82, respectively.

As continued, all 124 statistically significant triplets whose X_3 position gene belongs to 25 top importance genes as well as the observed event rate of X_3 position are far from random were selected to detect biologically relevant triplets.

Identification of biologically-relevant triplets

We used GSEA in order to find biologically-relevant triplets. Such analysis was performed using $p\text{-value} < 0.05$ and $FDR < 0.1$ for all of the involved genes in 124 statistically significant triplets (including 199 individual genes). Since the terms in lower levels of gene ontology are more general, ones in levels lower than level 6 are not reported. As reported in Fig. 3, the enriched terms based on "biological process" as follows: "spindle organization", "steroid hormone mediated signaling pathway" and "mRNA processing". Based on the proposed definition of three-way interactions of switching mechanism model, it is expected that in biologically-relevant triplets, X_1 and X_2 are in the same biological process. The complete list of enriched terms is available in the S5 Table.

By tracing triplets in the enriched terms, three triplets in which X_1 and X_2 are involved in the same biological process were determined. Such triplets including *Nkx3-1*, {*Ckap5*, *Dlg1*} triplet, *Znf347*, {*Safb*, *Dnaja1*} triplet and *Fech*, {*Safb*, *Cdk9*} triplet that are involved in "spindle organization", "steroid hormone mediated signaling pathway" and "mRNA processing", respectively.

As another attempt for analyzing the functional relevance of three-way interactions, we reconstructed a GRN based on ARACNE. The regulatory relationship of significant triplets obtained from liquid association method was traced in this network and the results are shown as a sub-network in Fig. 4. The details of construction GRN and detection of significant triplets in this network is available in S6 Table.

Taken together, the biological relevancy of two statistically significant triplets was confirmed using both GSEA and GRN, including 22th and 46th triplets. The scatter plots of these triplets in three different ranges of associated X_3 expression levels are shown in Fig. 5, which indicates a considerable change in the correlation of X_1 and X_2 as a result of a change in X_3 .

As it is observed, the regulatory relationship between X_3 from 22nd triplet (*Nkx3-1*) and two other genes in this triplet (*Ckap5* and *Dlg1*) can be seen with two intermediate genes. In addition, a regulatory interaction between X_3 (*Fech*) and X_2 (*Cdk9*) from 46th triplet is observed in a nontrivial way.

Discussion

Although pituitary adenomas, including NFPAs, are commonly considered slow-growing benign brain tumors, a large number of them exhibit a local invasive behavior. Notwithstanding that transcriptome changes associated with NFPA invasiveness have been extensively studied in the NFPAs, that is unpredictable with the aid of current tumor biomarkers [4]. Therefore, current study for the first time utilized the three-way interaction model to provide insights in to upon the biological pathways as well as critical genes associated with invasive nature in the NFPAs.

The validity of fastLA analysis was confirmed by comparing the observed event rate of X_3 position (switch) genes in a wide range of significant fastLA p -values and the random one. As presented in Fig. 1, the observed event rate for switch genes is far greater than random. Such a result means that certain genes occupy most X_3 positions in the statistically significant triplets.

The biological relevancy of two statistically significant triplets was confirmed using both GSEA and GRN (see Fig. 3 and Fig. 4). Such results suggest that these two triplets may play a central role in PA invasiveness. In the following, we discussed the relationships between involved genes in such triplets separately.

Relationship between involved genes in triplet *Fech*, {*Safb*, *Cdk9*}

In such triplet, *Fech* is the switch gene that controls the co-expression relationship between gene pair {*Safb*, *Cdk9*}. The protein encoded by the *Fech* gene is ferrochelatase, which is a crucial enzyme that

catalyzes the conversion of protoporphyrin IX (PpIX) to heme. A significant down-regulation of *Fech* expression was found in several malignancies [25–28], resulting in PpIX accumulation in such tumor cells. Indeed, accumulated PpIX in tumor cells leads to photodynamic therapy as effective adjuvant therapy for treating various cancers through visualizing the extent and margins of tumors, including PA [29].

On the other hand, ferrochelatase is involved in endothelial cell growth and choroidal neovascularization [30]. Pusha and coworker [31] found that inhibition of *Fech* reduces retinal neovascularization and endothelial cell proliferation in the oxygen-induced retinopathy (OIR) mouse model. Furthermore, they suggested griseofulvin as a *Fech*-inhibiting drug that could be repurposed to treat retinal neovascularization by blocking pathological tuft formation and revascularized areas of vaso-obliteration. Additionally, the inhibitory effect of griseofulvin is reported in skin carcinogenesis [32], thyroid tumors [33], and the development of multiple hepatomas [34].

Although any direct effect of *Fech* gene expression on PA invasiveness is not reported until now, the positive association between gene expression of *Fech* gene and epidermal growth factor receptor (EGFR) as a critical gene involving in PA progression has been reported [35].

The results of GSEA (see Fig. 3) show that such triplet is involved in “mRNA processing” biological process. In the following, we discuss about the importance of “mRNA processing” in PA invasiveness.

mRNA processing and PA invasiveness

Intrinsically, cancer evolves through successive genetic alterations that are advantageous to tumor cells. DNA sequence perturbations, as well as epigenomic disruption, are two significant cancer-related alterations [36]. However, besides the genetic changes, abnormalities in the mRNA processing can also trigger cancer formation and motive tumor progression[37]. Indeed, the mRNA processing known as a post-transcriptional mechanism is a crucial biological process during which pre-mRNA undergoes a series of chemical modifications to form the mature mRNA. Subsequently, mature mRNA can be transported to the cytoplasm and translated into the corresponding protein. Such biological processes comprise three critical steps: removing introns by splicing, cleavage the 3'end of mRNA, and polyadenylation [38]. Approximately forty years after recognizing the RNA processing, it is clear that post-transcriptional mechanisms are disrupted in cancer biology [39, 40]. In other words, mRNA processing is frequently altered in the tumors. These alterations lead to the formation of numerous cancer-specific mRNAs translated to malfunction proteins and/or proteins with changed expression levels. Such changed proteins can result in the activation of oncogenes or the inactivation of tumor-suppressor genes [41, 42].

Moreover, abnormality in mRNA processing can be associated with cancer therapeutic resistance. Pre-mRNA processing factor 4 (PRPF4) is known as a novel therapeutic target for breast cancer treatment. The PRPF4 gene was overexpressed in various breast cancer cell lines. The PRPF4 gene was overexpressed in various breast cancer cell lines. Furthermore, Knockdown of the PRPF4 gene reduced migration and breast cancer invasion via suppressing the p38 MAPK phosphorylation pathway [43]. On

the other hand, heterogeneous ribonucleoproteins (hnRNPs) that participate in different steps of pre-mRNA processing are involved in human malignancies and metastasis. Many reports also suggested several hnRNAs as promising therapeutic targets in numerous metastatic cancer types [44]. Furthermore, ubiquitin-specific peptidase 39 (USP39) serves critical roles in mRNA processing [45] and additionally, is involved in tumorigenesis of multiple solid malignancies [46, 47], including human renal cell carcinomas (RCC) [48]. XU and colleagues [48] show that silencing of USP39 by siRNA induced cell apoptosis and decreased invasive capacity of RCC cells. Hence, they suggested USP39 as an oncogenic factor that can play a pivotal role in human RCC treatment. Moreover, pre-mRNA processing factor (PRPF) 4B kinase [49], pre-mRNA processing factor 19 (PRP19) [50, 51], and pre-mRNA processing factor 31 (PRP31) [52] are the other oncogenic factors that are involved in mRNA processing pathway. Furthermore, the central role of the factors mentioned above is reported in previous studies in invasiveness and metastatic of numerous malignancies, including prostate cancer, melanoma, hepatocellular carcinoma, and invasive ovarian cancer.

To the best of our knowledge, there is no direct report on the role of the “mRNA processing” pathway in PA so far. Nevertheless, according the above studies, there is considerable evidence to support that such biological process may be associated with the invasiveness of PA.

Relationship between involved genes in triplet *Nkx3-1*, {*Ckap5*, *Dlg1*}

The other significant triplet is *Nkx3-1* as the switch gene that controls the co-expression relationship between gene pair {*Ckap5*, *Dlg1*}. The switch gene (*Nkx3-1*) is a homeodomain transcription factor with tumor suppressor function [53]. Homeobox genes comprise a large family of developmental regulators that are essential for cell differentiation and are often aberrantly expressed in cancer [54]. Furthermore, the *Nkx3-1* gene is a marker for diagnosing metastatic tumors [55, 56]; besides, loss of *Nkx3-1* expression occurs in the early tumorigenesis, suggesting such gene plays a role in malignant initiation [55]. Surprisingly, such evidence is consistent with the concept of disease-related-switch genes.

In specific, dysregulation of *Nkx3-1* is known as a biomarker for prostate cancer progression [56–58]. Hereupon, anti-NKX3-1 antibodies are used as a method for diagnosing metastatic prostatic adenocarcinomas. Nevertheless, previous studies reported that loss of *Nkx3.1* expression correlates with several other malignancies, including breast cancer [56] and salivary duct carcinoma [59].

To the best of our knowledge, no direct link was reported between the *Nkx3-1* gene and the PA, although there is an indirect association. The *Fgf-2* gene, that plays a central role in the angiogenesis of invasive PA [60–62], is an upstream regulator of NKX genes [63, 64]. Furthermore, the importance of *Fgf-2* was reported in human prostate cancer progression [62]. Therefore, it can be inferred that *Fgf-2* might control the angiogenesis procedure by regulating the gene expression level of *Nkx3-1*.

The other aim of the current study was to comprehensively characterize which biological processes may be involved in the invasiveness of PA.

As shown in Fig. 3, above triplet is involved in “spindle organization” biological process. We discussed about such biological process in the PA invasiveness. See below.

Spindle organization and PA invasiveness

Another enriched biological process is “spindle organization”, which assists the arrangement, assembly, and disassembly of spindle components. The spindle, which belongs to cytoskeletal components, is composed of an array of microtubules and associated molecules that forms between opposite poles of a eukaryotic cell during DNA segregation. Accordingly, the spindle plays a pivotal role in separating duplicated chromosomes apart. Hereupon, the correctness of spindle organization and its associated molecules during cell division is crucial for cell fate determination, tissue organization, and cell development. On the other hand, deregulation of cytoskeletal components is associated with several oncogenic phenotypes, including increased migration and invasion of cancer cells [65–67].

Nucleolar and spindle-associated protein 1 (NUSAP1), a microtubule-binding protein, is selectively expressed in proliferating cells. Moreover, it plays a critical role in spindle microtubule organization [68]. The expression levels of NUSAP1 are increased in the G2 to mitosis transition and then immediately decreased after cell division [69]. Previous studies have reported that dysregulation of NUSAP1 is associated with invasion, proliferation, and migration in several malignancies [70–77], including pituitary adenomas [78]. Additionally, Lee and colleagues [78] showed that the *NUSAP1* gene upregulated in 95% of patients with pituitary adenomas using the qRT-PCR technique. On the other hand, a pyrrolopyrimidine-based microtubule-depolymerizing agent (PP-13) reduces the metastatic dissemination of invasive cancer cells. PP-13, through binding to the colchicine site of β -tubulin, disturbs microtubules organization; and consequently induces spindle multipolarity, mitotic cell cycle blockade, and apoptosis [79]. Moreover, Gilson and colleagues [80] illustrated that low concentration PP-13 (130 nmol.L⁻¹) treatment significantly decreased the metastatic invasiveness of human cancer cells. Furthermore, they suggested that PP-13 might be a potential alternative to standard chemotherapy in drug-refractory tumors.

The “spindle organization” is defined as a child term for the “cell cycle” process according to the Gene Ontology Databank categories [81]. Several studies confirmed the significant role of the “cell cycle” in PA invasion and migration. See below.

Zhang and colleagues [82] compared differentially expressed microRNAs (DEMs) in the invasive and non-invasive PA. They report that DEMs were significantly associated with the “cell proliferation” and “cell cycle” pathway. On the other hand, Zheng and colleagues [83] showed that MiR-106b is upregulated in the invasive PA patients compare to non-invasive ones, associated with migration and invasion of pituitary adenoma cells. Moreover, they illustrated the inhibition of miR-106b remarkably suppressed proliferation and migration through the arrest of cell cycles. Some other biological molecules that can affect migration and invasion of PA through disturbing the “cell cycle” process include S100 calcium-binding protein A9[84], cyclin B1[85], Lactate dehydrogenase A[86].

Taken together, above evidence confirms the significant role of “spindle organization” in invasiveness and migration of tumor cells.

Conclusion

The existence of a considerable number of disease-related high-throughput "omics" datasets has provided studies about disease-related pathways and genes. In the current study, for the first time, we used the three-way interaction model to identify critical biomarkers and biological pathways associated with invasiveness in the NFPAs. The main advantage of such approach compared to the pairwise co-expression approach is that the three-way interaction model can cope with the dynamic nature of co-expression relations by introducing a third gene known as the switch gene. Therefore, the three-way interaction model can lead to a more comprehensive and precise understanding of the cause of cellular changes. The switch genes can be considered potential drug targets; therefore, the successful identification of them in a disease can be momentous. More specifically, in the present study, we identified two triplets associated with the invasive nature of NFPAs; consequently, we suggested their corresponding switch gene (i.e., *Fech* and *Nkx3-1* genes) as drug targets for invasive NFPAs. Moreover, we introduced two biological processes, “mRNA processing” and “spindle organization”, which might play a central role in the NFPAs invasiveness.

Declarations

Acknowledgements

We would like to thank the Institute for Research in Fundamental Sciences (IPM) for providing us with research facilities.

Authors contributions

All authors contributed to the study conception and design. M.J. and M.M. conceived and directed the study. M.M. and N.Kh. designed the data analysis framework. N.Kh. performed the data analyses and evaluated the results. A.J. and A.T. discussed the results. N.Kh. wrote the first draft of manuscript. All authors read and approved the final manuscript.

Additional information

Funding: This research was supported by a grant from the Five Senses Institute, Iran university of Medical Sciences.

Competing interests Statement: The authors declare no conflicts interests.

Availability of data and material: The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

References

1. Ostrom, Q. T. *et al.* CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015. *Neuro-oncology*. **20** (suppl_4), 1–86 (2018).
2. Fernandez, A., Karavitaki, N. & Wass, J. A. Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). *Clinical endocrinology*. **72** (3), 377–382 (2010).
3. Ezzat, S. *et al.* The prevalence of pituitary adenomas: a systematic review. *Cancer: Interdisciplinary International Journal of the American Cancer Society*. **101** (3), 613–619 (2004).
4. Galland, F. *et al.* Differential gene expression profiles of invasive and non-invasive non-functioning pituitary adenomas based on microarray analysis. *Endocrine-related cancer*. **17** (2), 361–371 (2010).
5. Androulakis, I. I., Kaltsas, G. & C.E.A.K. and *Aggressive pituitary tumors*. *Neuroendocrinology*. **101**, 87–104 (2015).
6. Lander, E. S. Array of hope. *Nature genetics*. **21** (1), 3–4 (1999).
7. Alm, E. & Arkin, A. P. *Biological networks*. Current opinion in structural biology, 2003. **13**(2): p.193–202.
8. Emmert-Streib, F. & Glazko, G. V. Pathway analysis of expression data: deciphering functional building blocks of complex diseases. *PLoS Comput Biol*. **7** (5), 1002053 (2011).
9. Wang, H. *et al.* Diagnosis of Invasive Nonfunctional Pituitary Adenomas by Serum Extracellular Vesicles. *Anal Chem*. **91** (15), 9580–9589 (2019).
10. Chen, Y. *et al.* A Novel Invasive-Related Biomarker in Three Subtypes of Nonfunctioning Pituitary Adenomas. *World Neurosurg*. **100**, 514–521 (2017).
11. Feng, J. *et al.* Identification of a subtype-specific ENC1 gene related to invasiveness in human pituitary null cell adenoma and oncocytomas. *J Neurooncol*. **119** (2), 307–315 (2014).
12. Song, W. *et al.* Aberrant expression of the sFRP and WIF1 genes in invasive non-functioning pituitary adenomas. *Mol Cell Endocrinol*. **474**, 168–175 (2018).
13. Ongaratti, B. R. *et al.* Gene and protein expression of E-cadherin and NCAM markers in non-functioning pituitary adenomas. *Ann Diagn Pathol*. **38**, 59–61 (2019).
14. Kim, Y. H. & Kim, J. H. Transcriptome Analysis Identifies an Attenuated Local Immune Response in Invasive Nonfunctioning Pituitary Adenomas. *Endocrinol Metab (Seoul)*. **34** (3), 314–322 (2019).
15. Joshi, H., Vastrad, B. & Vastrad, C. Identification of Important Invasion-Related Genes in Non-functional Pituitary Adenomas. *J Mol Neurosci*. **68** (4), 565–589 (2019).
16. Khayer, N. *et al.* Three-way interaction model with switching mechanism as an effective strategy for tracing functionally-related genes. *Expert review of proteomics*. **16** (2), 161–169 (2019).
17. Ho, Y. Y. *et al.* Modeling liquid association. *Biometrics*. **67** (1), 133–141 (2011).

18. Heyer, L. J., Kruglyak, S. & Yooseph, S. Exploring expression data: identification and analysis of coexpressed genes. *Genome research*. **9** (11), 1106–1115 (1999).
19. Watson, M. CoXpress: differential co-expression in gene expression data. *BMC bioinformatics*. **7** (1), 1–12 (2006).
20. Khayer, N. *et al.* Three-way interaction model to trace the mechanisms involved in Alzheimer's disease transgenic mice. *PloS one*. **12** (9), 0184697 (2017).
21. Khayer, N. *et al.* Rps27a might act as a controller of microglia activation in triggering neurodegenerative diseases. *Plos one*. **15** (9), 0239219 (2020).
22. Breiman, L. Random forests. *Machine learning*. **45** (1), 5–32 (2001).
23. Bureau, A. *et al.* Mapping complex traits using Random Forests. in *BMC genetics* (BioMed Central, 2003).
24. Janitza, S., Strobl, C. & Boulesteix, A. L. An AUC-based permutation variable importance measure for random forests. *BMC bioinformatics*. **14** (1), 1–11 (2013).
25. Kondo, M. *et al.* Heme-biosynthetic enzyme activities and porphyrin accumulation in normal liver and hepatoma cell lines of rat. *Cell biology and toxicology*. **9** (1), 95–105 (1993).
26. El-Sharabasy, M. *et al.* Porphyrin metabolism in some malignant diseases. *British journal of cancer*. **65** (3), 409–412 (1992).
27. Smith, A. *Mechanisms of Toxicity of Photoactivated Artificial Porphyrins Role of Porphyrin-Protein Interactions* 514p. 309–322(Annals of the New York Academy of Sciences, 1987). 1
28. Van Hillegersberg, R. *et al.* Selective accumulation of endogenously produced porphyrins in a liver metastasis model in rats. *Gastroenterology*. **103** (2), 647–651 (1992).
29. Nemes, A. *et al.* 5-ALA fluorescence in native pituitary adenoma cell lines: resection control and basis for photodynamic therapy (PDT). *PLoS One*. **11** (9), 0161364 (2016).
30. Basavarajappa, H. D. *et al.* Ferrochelatase is a therapeutic target for ocular neovascularization. *EMBO molecular medicine*. **9** (6), 786–801 (2017).
31. Babu, S. P. S. P., White, D. & Corson, T. W. *Ferrochelatase regulates retinal neovascularization*. bioRxiv, 2020.
32. Vesselinovitch, S. & Mihailovich, N. The inhibitory effect of griseofulvin on the “promotion” of skin carcinogenesis. *Cancer research*. **28** (12), 2463–2465 (1968).
33. Rustia, M. & Shubik, P. Thyroid tumours in rats and hepatomas in mice after griseofulvin treatment. *British Journal of Cancer*. **38** (2), 237–249 (1978).
34. Knasmüller, S. *et al.* Toxic effects of griseofulvin: disease models, mechanisms, and risk assessment. *Critical reviews in toxicology*. **27** (5), 495–537 (1997).
35. Gu, W. *et al.* Sex Difference in the Expression and Gene Network of Epidermal Growth Factor Receptor in Pituitary Gland in Mice. *Mol Biol*. **6** (179), 2 (2017).
36. Vogelstein, B. *et al.* Cancer genome landscapes. *science*. **339** (6127), 1546–1558 (2013).

37. Desterro, J., Bak-Gordon, P. & Carmo-Fonseca, M. Targeting mRNA processing as an anticancer strategy. *Nature Reviews Drug Discovery*. **19** (2), 112–129 (2020).
38. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*. **136** (4), 688–700 (2009).
39. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences* 1977. **74**(8): p. 3171–3175.
40. Chow, L. T. *et al.* An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*. **12** (1), 1–8 (1977).
41. Kahles, A. *et al.* Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer cell*. **34** (2), 211–224 (2018). e6
42. Mayr, C. & Bartel, D. P. Widespread shortening of 3' UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*. **138** (4), 673–684 (2009).
43. Park, S. *et al.* PRPF4 is a novel therapeutic target for the treatment of breast cancer by influencing growth, migration, invasion, and apoptosis of breast cancer cells via p38 MAPK signaling pathway. *Molecular and cellular probes*. **47**, 101440 (2019).
44. Han, N., Li, W. & Zhang, M. The function of the RNA-binding protein hnRNP in cancer metastasis. *Journal of cancer research and therapeutics*. **9** (7), 129 (2013).
45. Makarova, O. V., Makarov, E. M. & Lührmann, R. The 65 and 110 kDa SR-related proteins of the U4/U6·U5 tri-snRNP are essential for the assembly of mature spliceosomes. *The EMBO journal* 2001. **20**(10): p. 2553–2563.
46. Yuan, X. *et al.* USP39 promotes the growth of human hepatocellular carcinoma in vitro and in vivo. *Oncology reports*. **34** (2), 823–832 (2015).
47. Wang, H. *et al.* Lentivirus-mediated inhibition of USP39 suppresses the growth of breast cancer cells in vitro. *Oncology reports*. **30** (6), 2871–2877 (2013).
48. Xu, Y. *et al.* Knockdown of ubiquitin-specific peptidase 39 inhibits the malignant progression of human renal cell carcinoma. *Molecular medicine reports*. **17** (3), 4729–4735 (2018).
49. Islam, S. U. *et al.* PRPF overexpression induces drug resistance through actin cytoskeleton rearrangement and epithelial-mesenchymal transition. *Oncotarget*. **8** (34), 56659 (2017).
50. Yin, J. *et al.* Prp19 facilitates invasion of hepatocellular carcinoma via p38 mitogen-activated protein kinase/twist1 pathway. *Oncotarget*. **7** (16), 21939 (2016).
51. Cai, Y. *et al.* Prp19 is an independent prognostic marker and promotes neuroblastoma metastasis by regulating the Hippo-YAP signaling pathway. *Frontiers in oncology*. **10**, 1872 (2020).
52. Peedicayil, A. *et al.* Risk of ovarian cancer and inherited variants in relapse-associated genes. *PloS one*. **5** (1), 8884 (2010).
53. Bhatia-Gaur, R. *et al.* Roles for Nkx3. 1 in prostate development and cancer. *Genes & development*. **13** (8), 966–977 (1999).

54. Abate-Shen, C. Deregulated homeobox gene expression in cancer: cause or consequence? *Nature Reviews Cancer*. **2** (10), 777–785 (2002).
55. Gurel, B. *et al.* NKX3. 1 as a marker of prostatic origin in metastatic tumors. *The American journal of surgical pathology*. **34** (8), 1097 (2010).
56. Gelmann, E. P., Bowen, C. & Bubendorf, L. Expression of NKX3. 1 in normal and malignant tissues. *The Prostate*. **55** (2), 111–117 (2003).
57. Bowen, C. *et al.* Loss of NKX3. 1 expression in human prostate cancers correlates with tumor progression^{1, 2}. *Cancer research*. **60** (21), 6111–6115 (2000).
58. He, W. W. *et al.* A novel human prostate-specific, androgen-regulated homeobox gene (NKX3. 1) that maps to 8p21, a region frequently deleted in prostate cancer. *Genomics*. **43** (1), 69–77 (1997).
59. Takada, N. *et al.* Immunohistochemical Reactivity of Prostate-Specific Markers for Salivary Duct Carcinoma. *Pathobiology*. **87** (1), 30–36 (2020).
60. McCabe, C. *et al.* Expression of pituitary tumour transforming gene (PTTG) and fibroblast growth factor-2 (FGF-2) in human pituitary adenomas: relationships to clinical tumour behaviour. *Clinical endocrinology*. **58** (2), 141–150 (2003).
61. Tanase, C. *et al.* Angiogenic markers: molecular targets for personalized medicine in pituitary adenoma. *Personalized Medicine*. **10** (6), 539–548 (2013).
62. Huss, W. J. *et al.* Differential expression of specific FGF ligand and receptor isoforms during angiogenesis associated with prostate cancer progression. *The Prostate*. **54** (1), 8–16 (2003).
63. Pradhan, A. *et al.* FGF signaling enforces cardiac chamber identity in the developing ventricle. *Development*. **144** (7), 1328–1338 (2017).
64. Keren-Politansky, A., Keren, A. & Bengal, E. Neural ectoderm-secreted FGF initiates the expression of Nkx2. 5 in cardiac progenitors via a p38 MAPK/CREB pathway. *Developmental biology*. **335** (2), 374–384 (2009).
65. Hall, A. The cytoskeleton and cancer. *Cancer Metastasis Rev*. **28** (1–2), 5–14 (2009).
66. Yilmaz, M. & Christofori, G. EMT, the cytoskeleton, and cancer cell invasion. *Cancer Metastasis Rev*. **28** (1–2), 15–33 (2009).
67. Fielding, A. B. & Dedhar, S. The mitotic functions of integrin-linked kinase. *Cancer Metastasis Rev*. **28** (1–2), 99–111 (2009).
68. Iyer, J. *et al.* What's Nu(SAP) in mitosis and cancer? *Cell Signal*. **23** (6), 991–998 (2011).
69. Raemaekers, T. *et al.* NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization. *J Cell Biol*. **162** (6), 1017–1029 (2003).
70. Qian, Z. *et al.* Prognostic value of NUSAP1 in progression and expansion of glioblastoma multiforme. *J Neurooncol*. **140** (2), 199–208 (2018).
71. Zhu, T. *et al.* Nucleolar and spindle-associated protein 1 is a tumor grade correlated prognosis marker for glioma patients. *CNS Neurosci Ther*. **24** (3), 178–186 (2018).

72. Wu, X. *et al.* Nucleolar and spindle associated protein 1 promotes the aggressiveness of astrocytoma by activating the Hedgehog signaling pathway. *J Exp Clin Cancer Res.* **36** (1), 127 (2017).
73. Gulzar, Z. G., McKenney, J. K. & Brooks, J. D. Increased expression of NuSAP in recurrent prostate cancer is mediated by E2F1. *Oncogene.* **32** (1), 70–77 (2013).
74. Gordon, C. A. *et al.* NUSAP1 promotes invasion and metastasis of prostate cancer. *Oncotarget.* **8** (18), 29935–29950 (2017).
75. Yu, Z. *et al.* [NUSAP1 promotes lung cancer progression by activating AKT/mTOR signaling pathway]. *Zhonghua Zhong Liu Za Zhi*, 2020. **42**(7): p. 551–555.
76. Xu, Z. *et al.* NUSAP1 knockdown inhibits cell growth and metastasis of non-small-cell lung cancer through regulating BTG2/PI3K/Akt signaling. *J Cell Physiol.* **235** (4), 3886–3893 (2020).
77. Zhang, X. *et al.* Nucleolar and spindle associated protein 1 (NUSAP1) inhibits cell proliferation and enhances susceptibility to epirubicin in invasive breast cancer cells by regulating cyclin D kinase (CDK1) and DLGAP5 expression. *Medical science monitor: international medical journal of experimental and clinical research.* **24**, 8553 (2018).
78. Lee, M. *et al.* Transcriptome analysis of MENX-associated rat pituitary adenomas identifies novel molecular mechanisms involved in the pathogenesis of human pituitary gonadotroph adenomas. *Acta Neuropathol.* **126** (1), 137–150 (2013).
79. Gilson, P. *et al.* Identification of pyrrolopyrimidine derivative PP-13 as a novel microtubule-destabilizing agent with promising anticancer properties. *Scientific reports.* **7** (1), 1–14 (2017).
80. Gilson, P. *et al.* The pyrrolopyrimidine colchicine-binding site agent PP-13 reduces the metastatic dissemination of invasive cancer cells in vitro and in vivo. *Biochem. Pharmacol.* **160**, 1–13 (2019).
81. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature genetics.* **25** (1), 25–29 (2000).
82. Zhang, C. *et al.* Analysis of whole genome-wide microRNA transcriptome profiling in invasive pituitary adenomas and non-invasive pituitary adenomas. *Chinese Neurosurgical Journal.* **5** (1), 1–7 (2019).
83. Zheng, Z. *et al.* Effect of miR-106b on invasiveness of pituitary adenoma via PTEN-PI3K/AKT. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research.* **23**, 1277 (2017).
84. Huang, N. *et al.* Intracellular and extracellular S100A9 trigger epithelial-mesenchymal transition and promote the invasive phenotype of pituitary adenoma through activation of AKT1. *Aging*, 2020. **12**.
85. Li, B. *et al.* Regulating the CCNB1 gene can affect cell proliferation and apoptosis in pituitary adenomas and activate epithelial-to-mesenchymal transition. *Oncology Letters.* **18** (5), 4651–4658 (2019).
86. An, J. *et al.* Lactate dehydrogenase A promotes the invasion and proliferation of pituitary adenoma. *Scientific reports.* **7** (1), 1–12 (2017).

87. Brazma, A. *et al.* ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic acids research*. **31** (1), 68–71 (2003).
88. Galland, F. *et al.* Differential gene expression profiles of invasive and non-invasive non-functioning pituitary adenomas based on microarray analysis. *Endocr Relat Cancer*. **17** (2), 361–371 (2010).
89. Ritchie, M. E. *et al.* A comparison of background correction methods for two-colour microarrays. *Bioinformatics*. **23** (20), 2700–2707 (2007).
90. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods*. **31** (4), 265–273 (2003).
91. Bolstad, B. M. *et al.* A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. **19** (2), 185–193 (2003).
92. Smyth, G. K. *Limma: linear models for microarray data*, in *Bioinformatics and computational biology solutions using R and Bioconductor*. 397–420(Springer, 2005).
93. Gentleman, R. *et al.* *Genefilter: methods for filtering genes from high-throughput experiments*. R package version, 2015. **1**(1).
94. Smyth, G. K. *Linear models and empirical bayes methods for assessing differential expression in microarray experiments*. Statistical applications in genetics and molecular biology, 2004. **3**(1).
95. Benjamini, Y. *et al.* Controlling the false discovery rate in behavior genetics research. *Behavioural brain research*. **125** (1–2), 279–284 (2001).
96. Gunderson, T. & Ho, Y. Y. An efficient algorithm to explore liquid association on a genome-wide scale. *BMC Bioinformatics*. **15** (1), 371 (2014).
97. Li, K. C. Genome-wide coexpression dynamics: theory and application. *Proc Natl Acad Sci U S A*. **99** (26), 16875–16880 (2002).
98. Ho, Y. Y. *et al.* Modeling liquid association. *Biometrics*. **67** (1), 133–141 (2011).
99. Willse, J. T. & Willse, M. J. T. Package ‘CTT’. 2014.
100. Wright, M. N. & Ziegler, A. *ranger: A fast implementation of random forests for high dimensional data in C++ and R*. arXiv preprint arXiv:1508.04409, 2015.
101. Subramanian, A. *et al.* *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545–15550.
102. Kanehisa, M. *et al.* KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. **40** (1), 109–114 (2012).
103. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. **25** (8), 1091–1093 (2009).
104. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. **13** (11), 2498–2504 (2003).
105. Thissen, D., Steinberg, L. & Kuang, D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of educational and behavioral statistics*. **27** (1), 77–83 (2002).

106. Remo, A. *et al.* Systems biology analysis reveals NFAT5 as a novel biomarker and master regulator of inflammatory breast cancer. *Journal of translational medicine.* **13** (1), 138 (2015).
107. Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics.* **7**, 7 (2006).

Figures

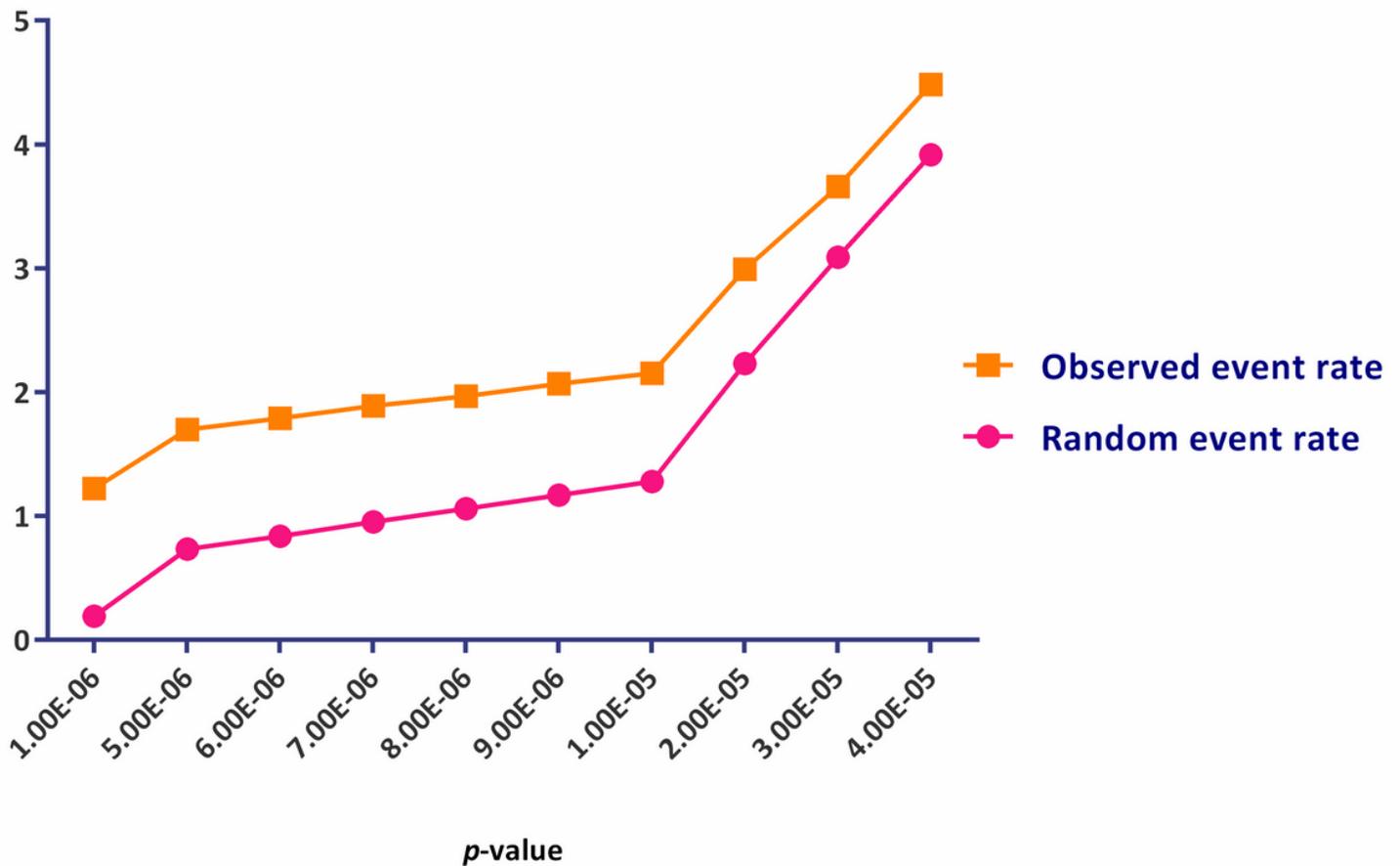


Figure 1

A survey of accuracy of fastLA analysis. In the wide range of the significant fastLA p-values, the observed event rate of X3 position (switch) genes was compared with the random event rate. As shown, the observed event rate of switch genes is far from random, confirming the accuracy of fastLA analysis.

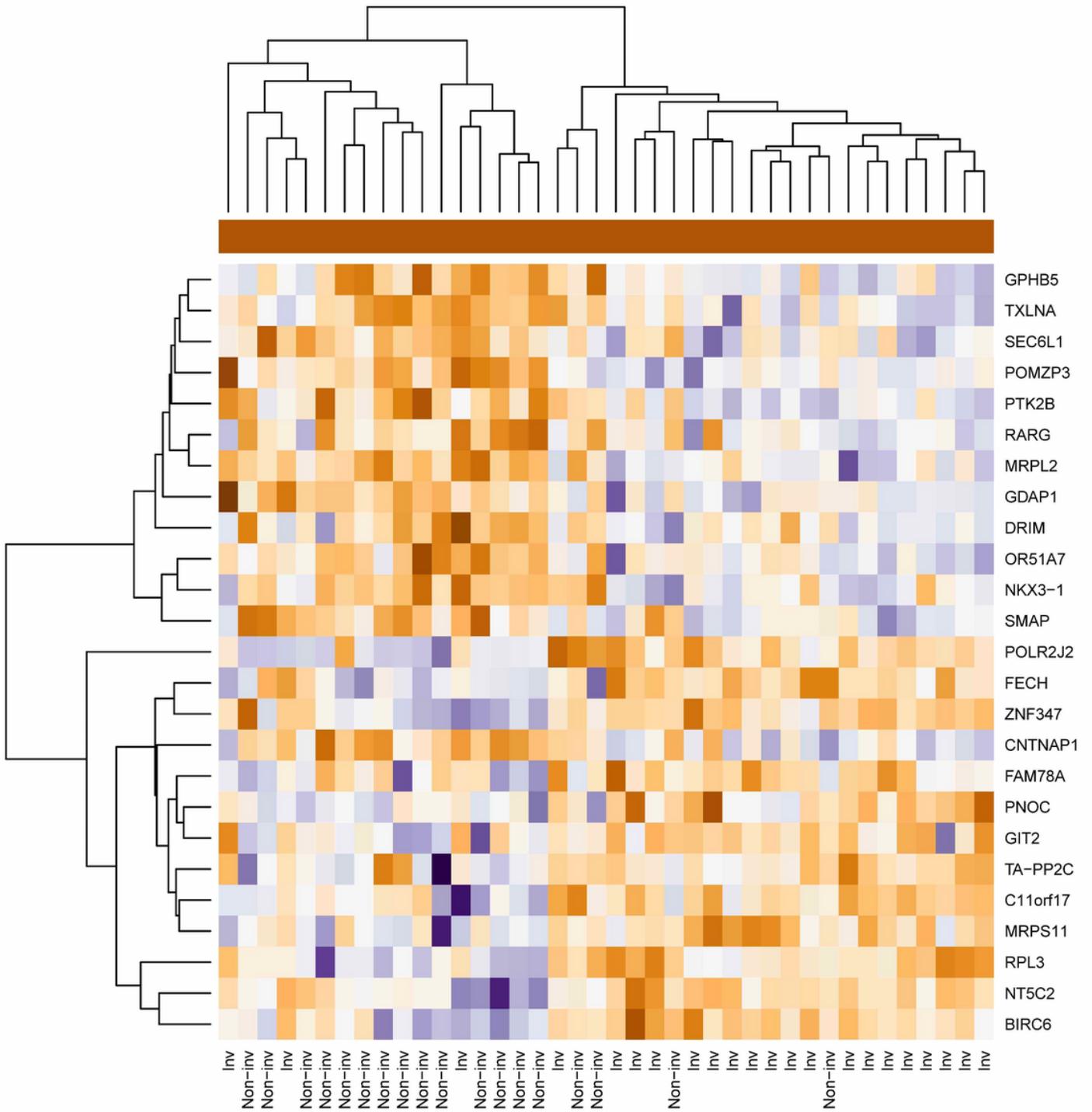


Figure 2

Random forest classification. This figure present 25 top importance genes selected based on Mean Decrease Accuracy measure.

Gene Set Enrichment Analysis

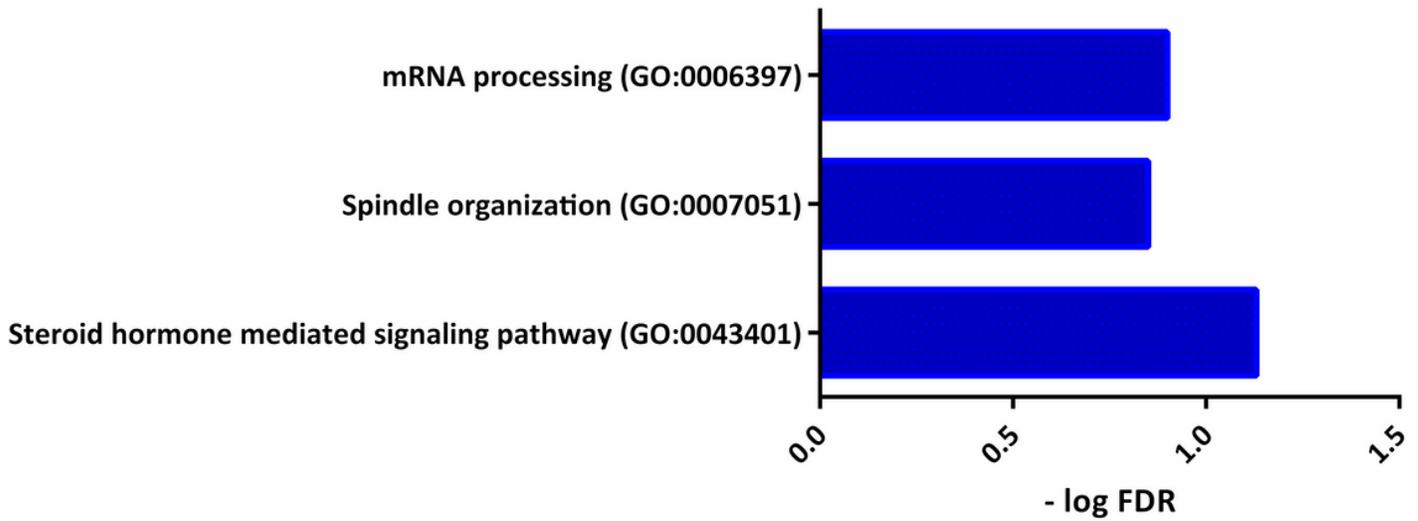


Figure 3

Biological process enrichment analysis. Enriched terms based on biological process for all genes involved in the statistically significant triplets.

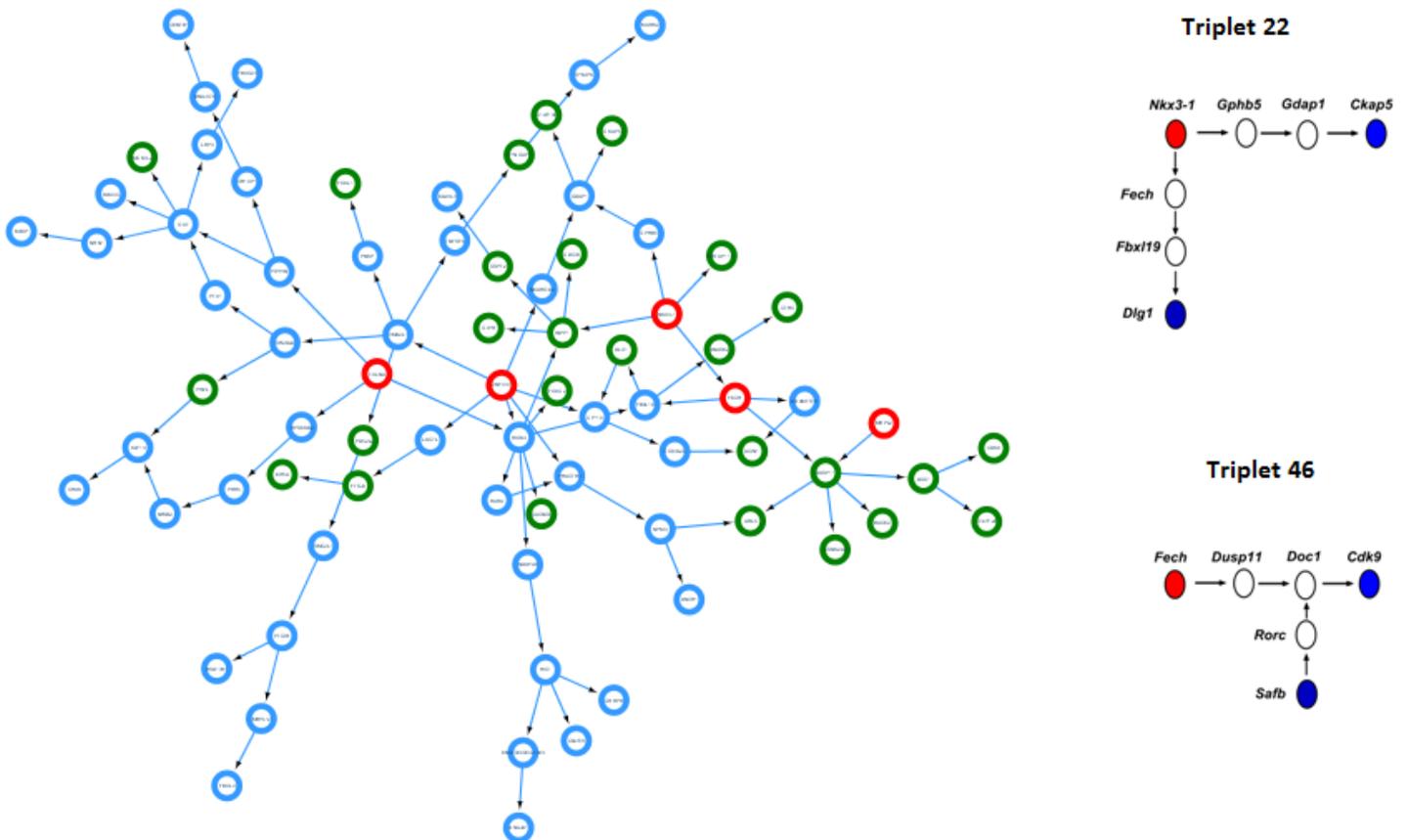


Figure 4

The position of biologically relevant triplets in Gene Regulatory Network (GRN). The biological relevance of thirty statistically significant triplets was confirmed GRN analysis. A subnetwork of GRN that includes the regulatory relations of such triplets is shown here. Red nodes represent the X3 position gene in each triplet, green nodes represent the X1 and X2 position genes, and other genes are presented by blue nodes.

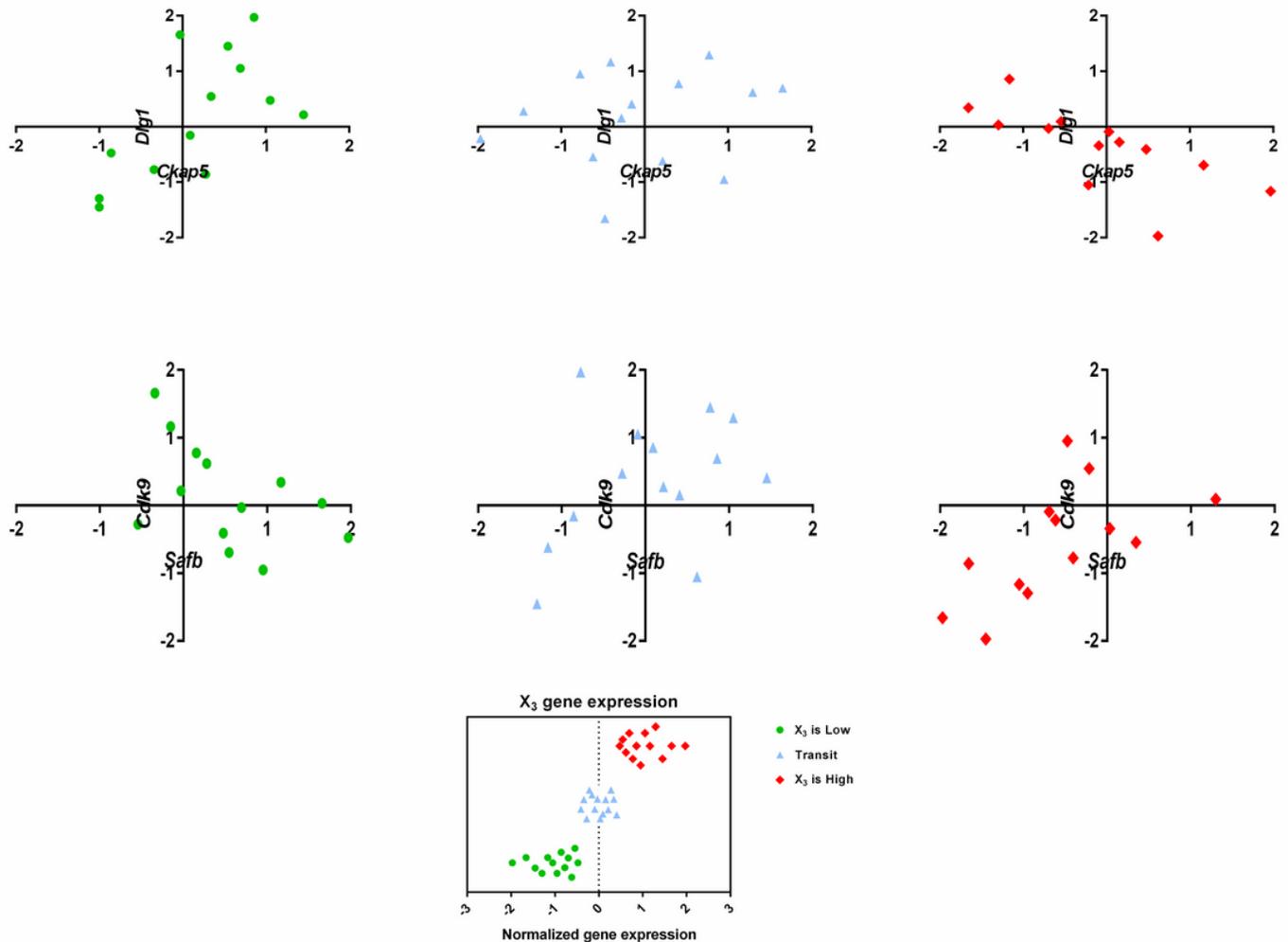


Figure 5

Scatter plot of two biologically relevant triplets. In each case, there is a considerable change in the correlation of X1 and X2 as a result of change in X3 expression level.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFiles.pdf](#)