scientific reports

Check for updates

OPEN

Nkx3-1 and *Fech* genes might be switch genes involved in pituitary non-functioning adenoma invasiveness

Nasibeh Khayer [ID][1], Maryam Jalessi[1,2 ✉], Amin Jahanbakhshi[1,3], Alireza Tabib khooei[3] & Mehdi Mirzaie[4 ✉]

Non-functioning pituitary adenomas (NFPAs) are typical pituitary macroadenomas in adults 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. In these cases, the most critical obstacle for complete surgical removal 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 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 NFAPs invasiveness. Specifically, our results suggest *Nkx3-1 and Fech* as two switch genes in NFAPs invasiveness that may be potential biomarkers or target genes in this pathology.

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.

[1]Skull Base Research Center, The Five Senses Health Institute, Iran University of Medical Sciences, Tehran, Iran. [2]ENT and Head & Neck Research Center and Department, Hazrat Rasoul Hospital, Iran University of Medical Sciences, Tehran, Iran. [3]Neurology Department, Hazrat Rasoul Hospital, Iran University of Medical Sciences, Tehran, Iran. [4]Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran. ✉email: jalessi.m@iums.ac.ir; mirzaie@modares.ac.ir

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 NFPAs through gene expression data analysis. Some of the most momentous potential biomarkers related to the aggressive nature of NFPAs 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 NFPAs 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 the two-way gene interaction approach. In the current study, we aimed to trace the 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 a pattern deciphers the sophisticated molecular relations at a higher level than the conventional two-way gene interaction pattern, including co-expression[18,19] and differentially co-expression[20] patterns. Therefore, it can lead to a more comprehensive and explicit biological insight into 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[21,22].

The three-way gene interaction model is not investigated for NFPAs' gene expression data to the best of our knowledge. 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 in a high computational load.

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

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 ({X_1, X_2}) in the dataset. The top 200,000 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 Fig. S1. 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-value) for the first 200,000 triplets are shown in the Fig. S2.

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 the Table S3.

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[23,24]. 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[25]. Therefore, ROC curves were used to analyze the sensitivity and specificity of the Random Forest model. As demonstrated in the Fig. S4, 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-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 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 the switch 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 Table S5.

By tracing triplets in 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.
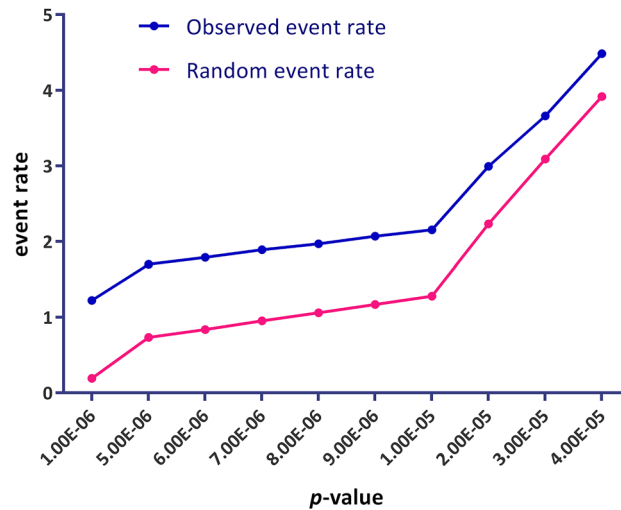
Figure 1. A survey of accuracy of fastLA analysis. In the wide range of the significant fastLA p-values, the observed event rate of X_3 position (switch) genes was compared with the random event rate. The random event rate is equal to the ratio of the number of statistically significant triplets (in a particular p-value) to the number of total examination genes; the observed event rate is equal to the ratio of statistically significant triplets to the number of unique X3. *Random event rate* $= \frac{num.\ of\ significant\ triplets}{num.\ of\ totals\ examined\ genes}$; *Observed event rate* $= \frac{num.\ of\ significant\ triplets}{num.\ of\ unique\ X_3}$. As shown, the observed event rate of switch genes is significantly different from random, confirming the accuracy of fastLA analysis.

As another attempt to analyze the functional relevance of three-way interactions, we reconstructed a GRN based on ARACNE. The regulatory relationship of significant triplets obtained from the 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 are available in the Table S6.

Conclusively, 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 the 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 the 46th triplet is observed in a nontrivial way.

To determine the possible associations of the identified switch genes with tumor grade, as the most important clinic-pathological feature, the mean expression level of two identified switch genes (i.e., *Nkx3-1* and *Fech* genes) were surveyed in the different grades of invasive and non-invasive pituitary adenomas. As shown in Fig. 6A, the samples correspond to each tumor grade were clearly separated based on the switch genes' expression level. Besides, to verify the accuracy of such a result, it was compared with ten randomly selected genes (Fig. 6B). The results show that the gene expression level of identified switch genes is correlated with tumor grade, but it is also significantly different from random. Such a feature is an advantage for any gene to be considered as a potential biomarker.

Figure 6B indicates two exemplary plots of mean expression levels of random genes in the various grades of NFPAs. Furthermore, additional information about all of the 15 randomly selected genes are presented in Fig. S7.

It should be noted that since the identified switch genes comprise approximately 0.1 percent of total surveyed genes (i.e., two from 2321 genes), we considered 15 random genes from the dataset, including 0.1 percent of all known genes in the dataset (ten from 15,215). The random genes were generated in the R environment.

Furthermore, the association of two identified switch genes with tumor volume, age at surgery, recurrence and gender were examined.

The results show that there might be an association between gene expression levels of both switch genes and tumors with volume greater than 20 cm³. Nevertheless, the number of samples corresponding to the aforementioned tumor is inadequate for such a conclusion (only three samples belong to the tumors greater than 20 cm³). Moreover, statistical analysis on the variance indicated significant changes in the expression levels of the *Fech* gene in the different groups of the recurrence feature (p-value = 0.016).

Finally, we did not find a considerable association between the expression levels of none of the switch genes and other clinic-pathological features, including gender and age at surgery.

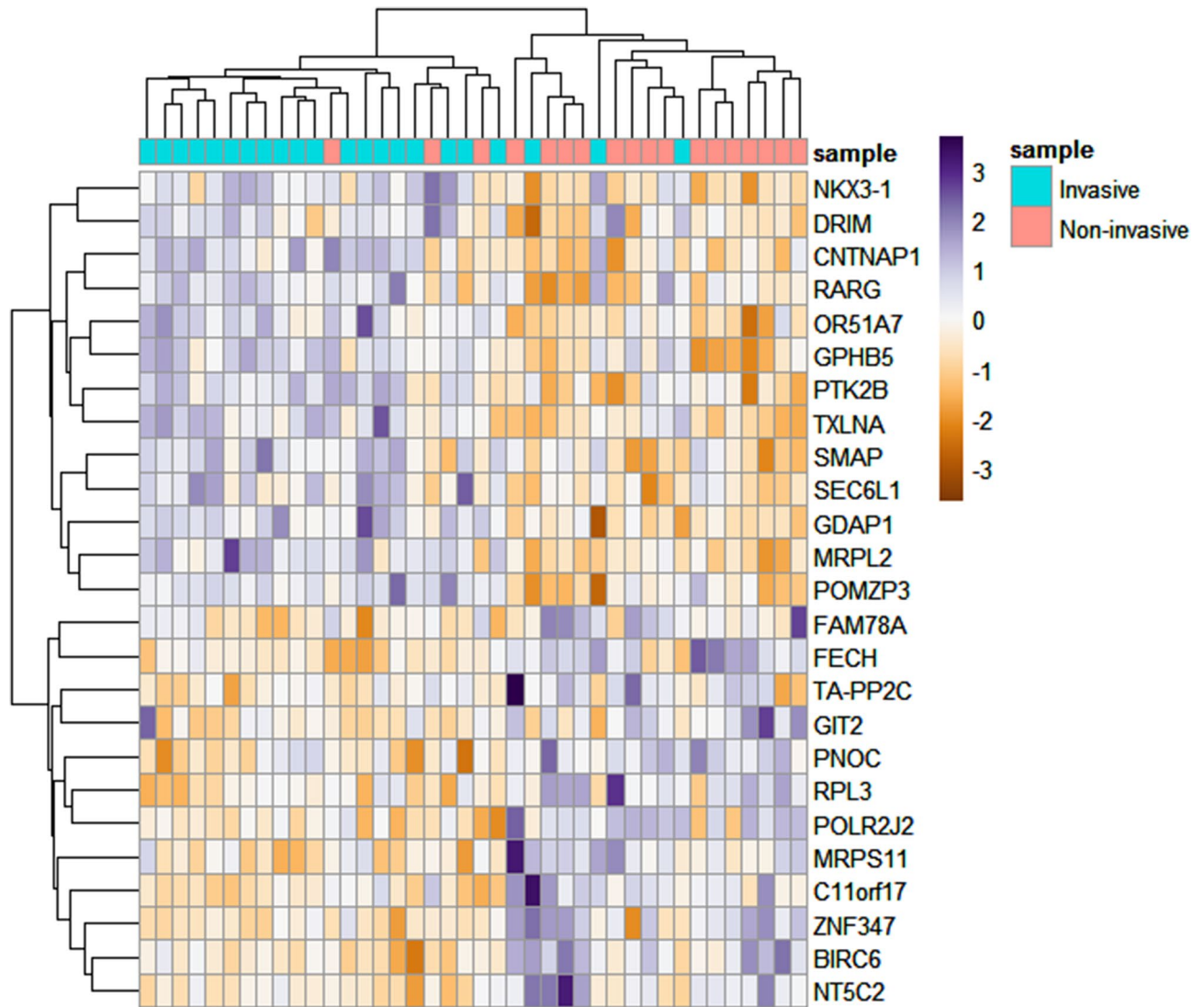The results of the above analyses were presented in the Fig. S8.

Figure 2. Random forest classification. This figure present 25 top importance genes selected based on Mean Decrease Accuracy measure.

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, which is unpredictable with the aid of current tumor biomarkers[4]. Therefore, the current study, for the first time, utilized the three-way interaction model to provide insights into 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. It is expected that the number of genes that occupy the X_3 position be significantly lower than random because, as a biological concept, a limited number of genes controls most biological processes. As presented in Fig. 1, the observed event rate for switch genes is significantly different from 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 Figs. 3 and 4). Furthermore, the gene expression level of identified switch genes is significantly correlated with tumor grade (see Fig. 6). 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[26–29], result-
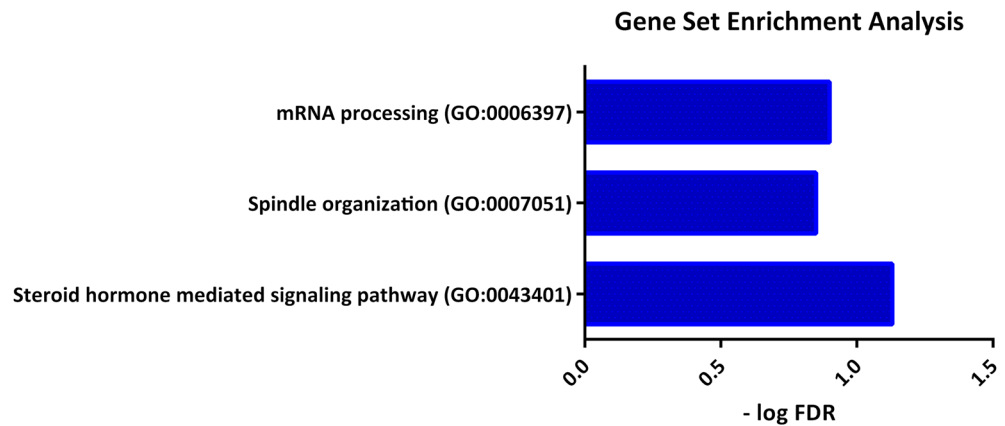
Figure 3. Biological process enrichment analysis. Enriched terms based on biological process for all genes involved in the statistically significant triplets.
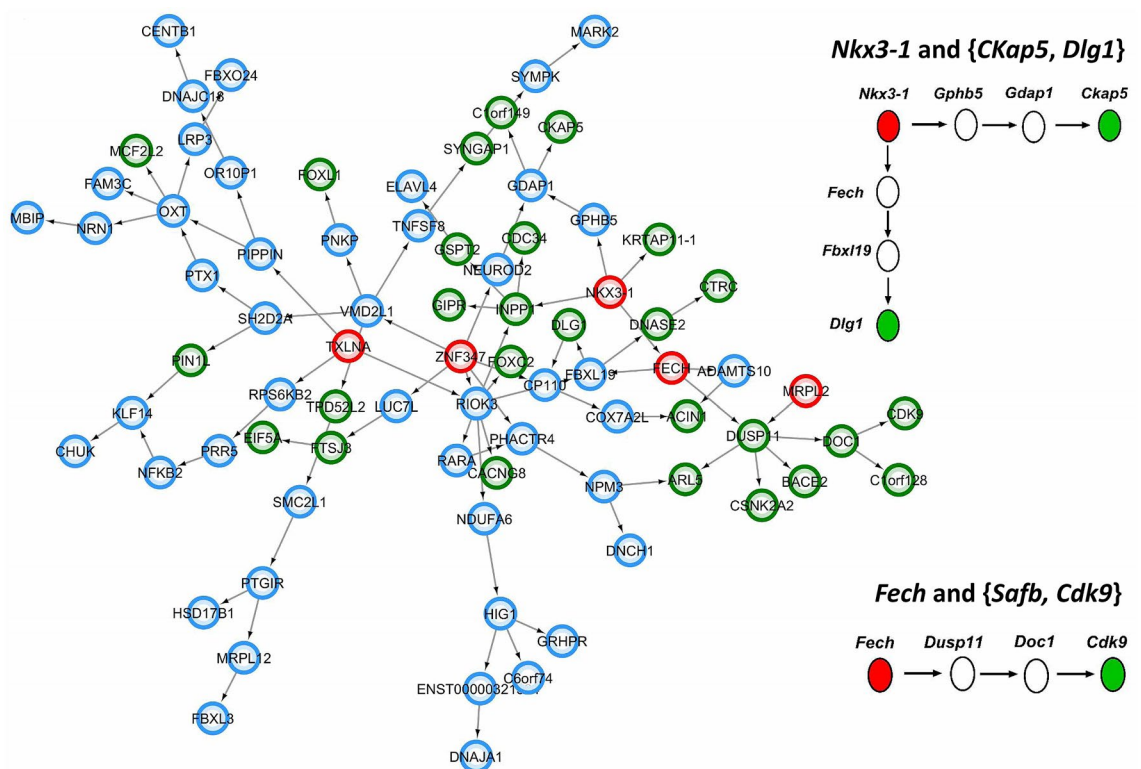

Figure 4. The position of biologically relevant triplets in Gene Regulatory Network (GRN). The biological relevance of thirty statistically significant triplets was confirmed GRN analysis. A subnetwork of GRN that includes the regulatory relations of such triplets is shown here. Red nodes represent the X_3 position gene in each triplet, green nodes represent the X_1 and X_2 position genes, and other genes are presented by blue nodes.

ing 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[30].

On the other hand, ferrochelatase is involved in endothelial cell growth and choroidal neovascularization[31]. Pusha and coworker[32] 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[33], thyroid tumors[34], and the development of multiple hepatomas[35].
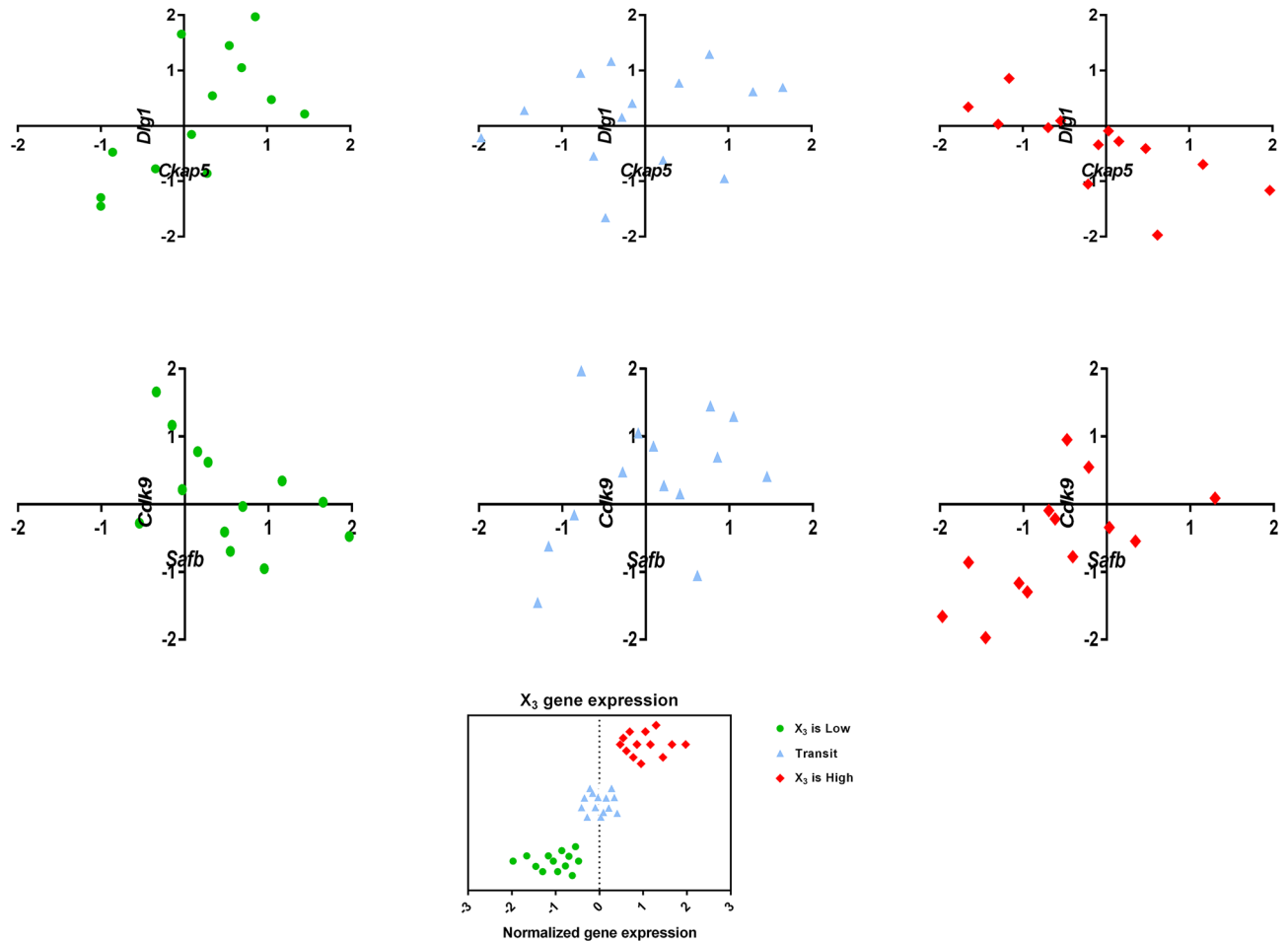
Figure 5. Scatter plot of two biologically relevant triplets. In each case, there is a considerable change in the correlation of X_1 and X_2 as a result of change in X_3 expression level.

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[36].

The results of GSEA (see Fig. 3) show that such triplet is involved in the "mRNA processing" biological process. In the following, we discuss 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[37]. However, besides the genetic changes, abnormalities in the mRNA processing can also trigger cancer formation and motive tumor progression[38]. 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[39]. Approximately forty years after recognizing the RNA processing, it is clear that post-transcriptional mechanisms are disrupted in cancer biology[40,41]. In other words, mRNA processing is frequently altered in the tumors. These alterations lead to the formation of numerous cancer-specific mRNAs translated to misfunction 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[42,43].

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[44]. 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[45]. Furthermore, ubiquitin-specific peptidase 39 (USP39) serves critical roles in mRNA processing[46] and additionally is involved in tumorigenesis of multiple solid malignancies[47,48], including human renal cell
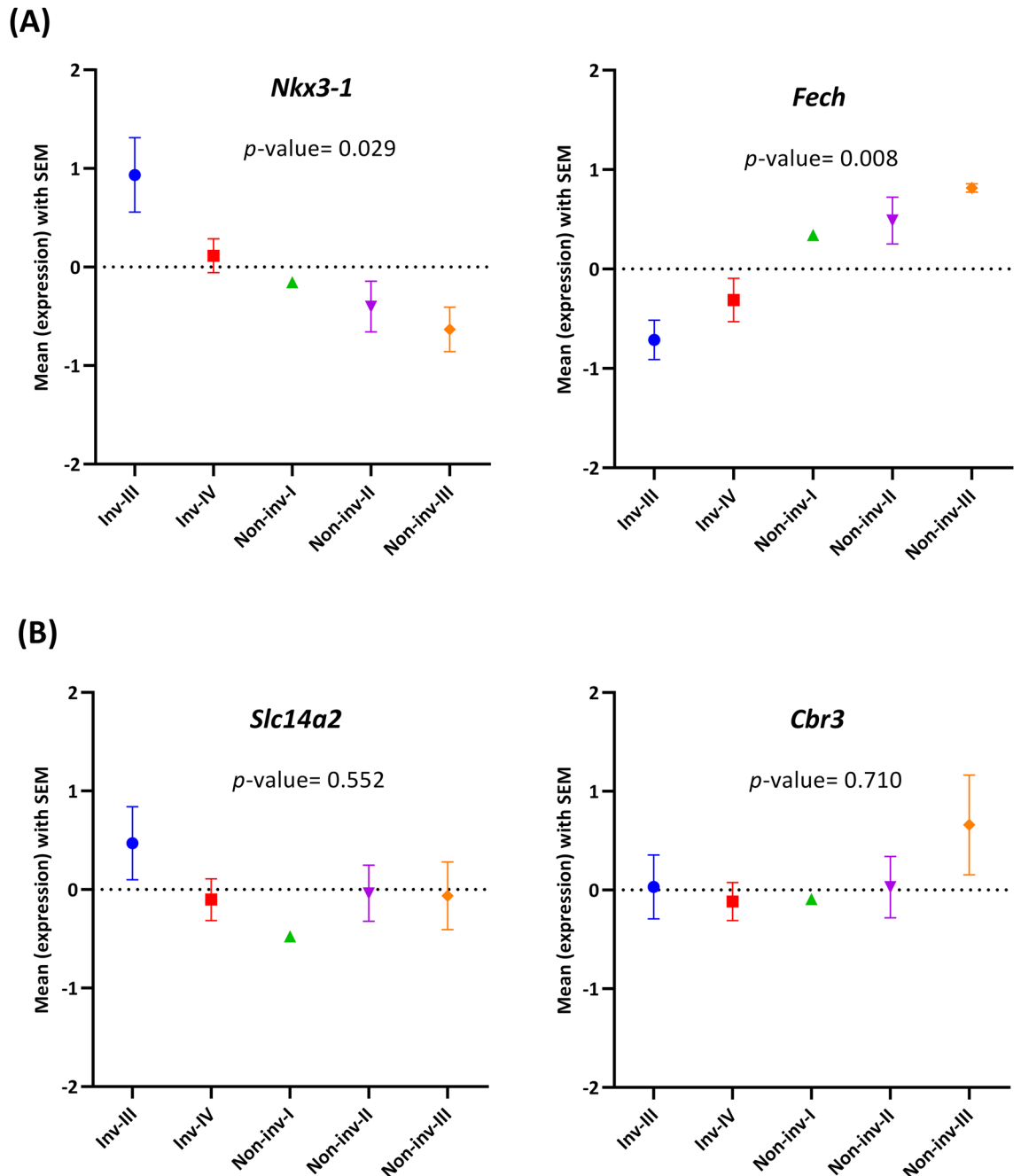
Figure 6. Mean expression level plot in the different grades of invasive and non-invasive pituitary adenomas for (**A**) two identified switch genes (i.e., Nkx3-1 and Fech genes); and (**B**) two exemplary random genes (i.e., Cl14a1 and Cbr3 genes).

carcinomas (RCC)[49]. Xu et al.[49] 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[50], pre-mRNA processing factor 19 (PRP19)[51,52], and pre-mRNA processing factor 31 (PRP31)[53] 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 to the above studies, there is considerable evidence to support that such a 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 the gene pair {*Ckap5*, *Dlg1*}. The switch gene (*Nkx3-1*) is a homeodomain transcription factor with tumor suppressor function[54]. Homeobox genes comprise a large family of developmental regulators that are essential for cell differentiation and are often aberrantly expressed in cancer[55]. Furthermore, the *Nkx3-1* gene is a marker for diagnosing metastatic tumors[56,57]; besides, loss of *Nkx3-1*expression occurs in the early tumorigenesis, suggesting such gene plays a role in malignant initiation[56]. 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[57–59]. 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[57] and salivary duct carcinoma[60].

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, which plays a central role in the angiogenesis of invasive PA[61–63], is an upstream regulator of NKX genes[64,65]. Furthermore, the importance of *Fgf-2* was reported in human prostate cancer progression[63]. 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, the above triplet is involved in the "spindle organization" biological process. We discussed such a biological process in 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[66–68].

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[69]. The expression levels of NUSAP1 are increased in the G2 to mitosis transition and then immediately decreased after cell division[70]. Previous studies have reported that dysregulation of NUSAP1 is associated with invasion, proliferation, and migration in several malignancies[71–78], including pituitary adenomas[79]. Additionally, Lee et al.[79] 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[80]. Moreover, Gilson et al.[81] 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[82]. Several studies confirmed the significant role of the "cell cycle" in PA invasion and migration. See below.

Zhang et al.[83] 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" pathways. On the other hand, Zheng et al.[84] 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[85], cyclin B1[86], Lactate dehydrogenase A[87].

Altogether, the above evidence confirms the significant role of "spindle organization" in the invasiveness and migration of tumor cells.

A comparison of results with the initial study. Galland et al.[4] performed genome-wide expression analysis using microarray technology to determine possible biomarkers in NFPAs invasiveness. Although their study only focused on differentially expressed genes and the correlation among such genes was not studied, parts of their results are comparable with our study. See below.

With the purpose of identifying a specific gene expression profile in the invasive NFPAs, Galland et al. traced DEGs between invasive and non-invasive tumors. Furthermore, they verified the gene expression levels of the top 44 DEGs using qRT-PCR. Consequently, the overexpression of four genes, namely *Igfbp5*, *Myo5a*, *Flt3* and *Nfe2l1*, was confirmed.

A major part of DEGs, which were reported in the Galland et al. study, have coverage our results. More precisely, among them, 36 genes, including *Myo5a*, *Flt3* and *Nfe2l1*, were found in common across our results. Nevertheless, two identified switch genes (i.e., the *Fech* and *Nkx3-1* genes) were not found among the top 44 DEGs. There might be two possible explanations for such observation: (i) although DEGs play a critical role in switch mechanisms, there is not necessarily a direct relation between DEG significance and switch gene importance.

In other words, a gene may be significantly expressed between two conditions (e.g., invasive and non-invasive samples), but it does not act as a considerable switch gene; (ii) a small proportion of DEGs were verified using the qRT-PCR technique in Galland et al. study. It is possible that if more genes had been examined, the two switch genes found in our study were also included.

Moreover, Galland et al. reported that the DEGs have known molecular functions such as "cell cycling and cell death", "cellular assembly", "morphology and motility" and "gene expression regulation". Surprisingly, "cell cycling" and "cellular assembly" functions support the "spindle organization" biological process that was indicated in our GSEA results.

Conclusion and further work

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 an 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.

Although our study provides new insight into the invasive nature of NFPAs using computational methods, more efforts should be performed to validate such findings. A reasonable way to in-silico validation of our results is verifying them in other NFPAs' datasets. However, some crucial prerequisites should be considered for selecting a reliable dataset. The most critical prerequisite is an adequate sample size. The LA algorithm is based on correlation coefficient; on the other hand, the samples are divided into at least three bins during the LA analysis procedure. Since the statistical significance of the correlation coefficient is related to the sample size, such a parameter should be considered in choosing a decent dataset. Another significant prerequisite is the association between the samples of datasets following two approaches: (i) features, which in turn is related to the design similarity of corresponding studies; (ii) gene expression profile, which in turn, can be affected by variation across platforms used to generate data[88].

In order to perform in-silico validation, three well-known omics databases, including ArrayExpress[89], Gene Expression Omnibus (GEO)[90] and The Cancer Genome Atlas (TCGA)[91], were explored to find genomics and/or transcriptomics datasets associated with PA invasiveness. Unfortunately, the NFPAs dataset is scarce in the publicly available database. Consequently, no dataset was found that meets all of the above requirements. The scarcity of omics NFPAs dataset indicates the urgent need for further efforts on such data, which in turn pave the way for a better understanding of this disease.

In the next step, the relationship between the *Fech* gene and the {*Cdk9*,*Safb*} gene pair as well as the *Nkx3-1* gene and the {*Ckap5*,*Dlg1*} gene pair need to be experimentally validated.

Materials and methods

Gene expression profiling dataset. The selected dataset includes gene expression of 22 invasive and 18 non-invasive NFPA (with no hormone secretion), which is available at the Array Express database[89] under accession number E-TABM-899[4]. Additionally, it was generated using the A-AGIL-11- Agilent Human Whole-genome microarray platform. The samples belong to grade I to III of non-invasive NFPAs and grade III and IV of invasive NFPAs. The related clinical data of invasive and non-invasive PA samples were reported in Table S9. Additionally, detailed information about tumor size and grade selection criteria is found in Galland and co-worker's study[4].

The background correction on the raw microarray dataset was carried out using the Normex method[92]. Furthermore, the expression profiles were normalized within- and between—arrays using loess[93] and quantile normalization[94] methods, respectively. It should be noted that the above-mentioned methods were implemented in the Limma R package[95].

Moreover, the duplicate probes were removed using the genefilter package[96]. Accordingly, the highest inter-quartile 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[97] was used to detect differentially expressed genes (DEGs). Additionally, the Benjamini–Hochberg method[98] was used to control the false discovery rate. After removing duplicated probes as well as probes corresponding to unknown genes, the dataset included 15,215 genes. Furthermore, by considering p-value < 0.01 as the threshold, DEGs include 2321 genes that were selected for further consideration.

Moreover, to assess the association of identified switch genes with clinic-pathological features, a statistical analysis on the variance was performed using Kruskal–Wallis ANOVA[99]. A p-value of less than 0.05 was considered statistically significant. Data are presented as means ± SEM.

Liquid association triplets. Three-way interactions between all genes involved in the dataset were calculated using the fastMLA function in the fastMLA R package[100]. 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:

$$\widehat{MLA} = \frac{\sum_{i}^{M} \widehat{\rho_i}\overline{X_{3i}}}{M}$$

where M is the number of bins over X_3, $\widehat{\rho_i}$ is the Pearson's correlation coefficient of X_1 and X_2 in samples of the ith bin, and $\overline{X_{3i}}$ is the mean of expression values of X_3 in the ith 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[101]; (ii) each variable should be standardized to have mean 0 and variance 1[17]. The first preprocessing was performed using an in-house implementation, while the second one by using the CTT package[102].

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[23]. The "number of decision trees" and "mtry" parameters were set to "10,000 trees" and "square root of the total number of features", respectively[103]. 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[104,105]. 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[106]. For the analyses, as mentioned earlier, we used the ClueGO tool[107] (with a Kappa threshold of 0.4) within the Cytoscape v.3.3.0 environment[108]. The right-sided hypergeometric test and the Benjamini–Hochberg correction method[109] were used to validate 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[110]. Here, we used ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks)[111] to construct the GRN. ARACNE is a reverse engineering approach for the construction of cellular networks from gene expression data. This algorithm capture directed regulatory interactions between each transcriptional regulator and its potential targets based on mutual information. ARACNE run in the geWorkbench_2.6.0 framework for all of the genes involved in the statistically significant triplets by considering p-value < 0.05.

Data availability
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 Oncol.* **20**(suppl_4), iv1–iv86 (2018).
2. Fernandez, A., Karavitaki, N. & Wass, J. A. Prevalence of pituitary adenomas: A community-based, cross-sectional study in Banbury (Oxfordshire, UK). *Clin. Endocrinol.* **72**(3), 377–382 (2010).
3. Ezzat, S. et al. The prevalence of pituitary adenomas: A systematic review. *Cancer* **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. *Endocr. Relat. Cancer* **17**(2), 361–371 (2010).
5. Androulakis, C. E. A. K. & Kaltsas, G. Aggressive pituitary tumors. *Neuroendocrinology* **101**, 87–104 (2015).
6. Lander, E. S. Array of hope. *Nat. Genet.* **21**(1), 3–4 (1999).
7. Alm, E. & Arkin, A. P. Biological networks. *Curr. Opin. Struct. Biol.* **13**(2), 193–202 (2003).
8. Emmert-Streib, F. & Glazko, G. V. Pathway analysis of expression data: Deciphering functional building blocks of complex diseases. *PLoS Comput. Biol.* **7**(5), e1002053 (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 Rev. Proteom.* **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 Res.* **9**(11), 1106–1115 (1999).
19. Alavi Majd, H. *et al.* Two-way gene interaction from microarray data based on correlation methods. *Iran Red Crescent Med. J.* **18**(6), e24373 (2016).
20. Watson, M. CoXpress: Differential co-expression in gene expression data. *BMC Bioinform.* **7**(1), 1–12 (2006).
21. Khayer, N. *et al.* Three-way interaction model to trace the mechanisms involved in Alzheimer's disease transgenic mice. *PLoS ONE* **12**(9), e0184697 (2017).
22. Khayer, N. *et al.* Rps27a might act as a controller of microglia activation in triggering neurodegenerative diseases. *PLoS ONE* **15**(9), e0239219 (2020).
23. Breiman, L. Random forests. *Mach. Learn.* **45**(1), 5–32 (2001).
24. Bureau, A. *et al.* Mapping complex traits using Random Forests. *BMC Genet.* **4**, S64 (2003).
25. Janitza, S., Strobl, C. & Boulesteix, A.-L. An AUC-based permutation variable importance measure for random forests. *BMC Bioinform.* **14**(1), 1–11 (2013).
26. Kondo, M. *et al.* Heme-biosynthetic enzyme activities and porphyrin accumulation in normal liver and hepatoma cell lines of rat. *Cell Biol. Toxicol.* **9**(1), 95–105 (1993).
27. El-Sharabasy, M. *et al.* Porphyrin metabolism in some malignant diseases. *Br. J. Cancer* **65**(3), 409–412 (1992).
28. Smith, A. Mechanisms of toxicity of photoactivated artificial porphyrins role of porphyrin-protein interactions. *Ann. N. Y. Acad. Sci.* **514**(1), 309–322 (1987).
29. Van Hillegersberg, R. *et al.* Selective accumulation of endogenously produced porphyrins in a liver metastasis model in rats. *Gastroenterology* **103**(2), 647–651 (1992).
30. 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), e0161364 (2016).
31. Basavarajappa, H. D. *et al.* Ferrochelatase is a therapeutic target for ocular neovascularization. *EMBO Mol. Med.* **9**(6), 786–801 (2017).
32. Pran Babu, S. P. S., White, D. & Corson, T. W. Ferrochelatase regulates retinal neovascularization. *FASEB J.* **34**(9), 12419–12435 (2020).
33. Vesselinovitch, S. & Mihailovich, N. The inhibitory effect of griseofulvin on the "promotion" of skin carcinogenesis. *Cancer Res.* **28**(12), 2463–2465 (1968).
34. Rustia, M. & Shubik, P. Thyroid tumours in rats and hepatomas in mice after griseofulvin treatment. *Br. J. Cancer* **38**(2), 237–249 (1978).
35. Knasmüller, S. *et al.* Toxic effects of griseofulvin: Disease models, mechanisms, and risk assessment. *Crit. Rev. Toxicol.* **27**(5), 495–537 (1997).
36. 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).
37. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**(6127), 1546–1558 (2013).
38. Desterro, J., Bak-Gordon, P. & Carmo-Fonseca, M. Targeting mRNA processing as an anticancer strategy. *Nat. Rev. Drug Discov.* **19**(2), 112–129 (2020).
39. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back totranscription and ahead to translation. *Cell* **136**(4), 688–700 (2009).
40. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* **74**(8), 3171–3175 (1977).
41. Chow, L. T. *et al.* An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**(1), 1–8 (1977).
42. Kahles, A. *et al.* Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**(2), 211–224 (2018).
43. 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).
44. 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. *Mol. Cell. Probes* **47**, 101440 (2019).
45. Han, N., Li, W. & Zhang, M. The function of the RNA-binding protein hnRNP in cancer metastasis. *J. Cancer Res. Ther.* **9**(7), 129 (2013).
46. 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. *EMBO J.* **20**(10), 2553–2563 (2001).
47. Yuan, X. *et al.* USP39 promotes the growth of human hepatocellular carcinoma in vitro and in vivo. *Oncol. Rep.* **34**(2), 823–832 (2015).
48. Wang, H. *et al.* Lentivirus-mediated inhibition of USP39 suppresses the growth of breast cancer cells in vitro. *Oncol. Rep.* **30**(6), 2871–2877 (2013).
49. Xu, Y. *et al.* Knockdown of ubiquitin-specific peptidase 39 inhibits the malignant progression of human renal cell carcinoma. *Mol. Med. Rep.* **17**(3), 4729–4735 (2018).
50. Islam, S. U. *et al.* PRPF overexpression induces drug resistance through actin cytoskeleton rearrangement and epithelial-mesenchymal transition. *Oncotarget* **8**(34), 56659 (2017).
51. Yin, J. *et al.* Prp19 facilitates invasion of hepatocellular carcinoma via p38 mitogen-activated protein kinase/twist1 pathway. *Oncotarget* **7**(16), 21939 (2016).
52. Cai, Y. *et al.* Prp19 is an independent prognostic marker and promotes neuroblastoma metastasis by regulating the Hippo-YAP signaling pathway. *Front. Oncol.* **10**, 1872 (2020).

53. Peedicayil, A. *et al.* Risk of ovarian cancer and inherited variants in relapse-associated genes. *PLoS ONE* **5**(1), e8884 (2010).

54. Bhatia-Gaur, R. *et al.* Roles for Nkx3. 1 in prostate development and cancer. *Genes Dev.* **13**(8), 966–977 (1999).

55. Abate-Shen, C. Deregulated homeobox gene expression in cancer: Cause or consequence?. *Nat. Rev. Cancer* **2**(10), 777–785 (2002).

56. Gurel, B. *et al.* NKX3. 1 as a marker of prostatic origin in metastatic tumors. *Am. J. Surg. Pathol.* **34**(8), 1097 (2010).

57. Gelmann, E. P., Bowen, C. & Bubendorf, L. Expression of NKX3. 1 in normal and malignant tissues. *Prostate* **55**(2), 111–117 (2003).

58. Bowen, C. *et al.* Loss of NKX3. 1 expression in human prostate cancers correlates with tumor progression1, 2. *Cancer Res.* **60**(21), 6111–6115 (2000).

59. 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).

60. Takada, N. *et al.* Immunohistochemical reactivity of prostate-specific markers for salivary duct carcinoma. *Pathobiology* **87**(1), 30–36 (2020).

61. 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. *Clin. Endocrinol.* **58**(2), 141–150 (2003).

62. Tanase, C. *et al.* Angiogenic markers: Molecular targets for personalized medicine in pituitary adenoma. *Pers. Med.* **10**(6), 539–548 (2013).

63. Huss, W. J. *et al.* Differential expression of specific FGF ligand and receptor isoforms during angiogenesis associated with prostate cancer progression. *Prostate* **54**(1), 8–16 (2003).

64. Pradhan, A. *et al.* FGF signaling enforces cardiac chamber identity in the developing ventricle. *Development* **144**(7), 1328–1338 (2017).

65. 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. *Dev. Biol.* **335**(2), 374–384 (2009).

66. Hall, A. The cytoskeleton and cancer. *Cancer Metastasis Rev.* **28**(1–2), 5–14 (2009).

67. Yilmaz, M. & Christofori, G. EMT, the cytoskeleton, and cancer cell invasion. *Cancer Metastasis Rev.* **28**(1–2), 15–33 (2009).

68. Fielding, A. B. & Dedhar, S. The mitotic functions of integrin-linked kinase. *Cancer Metastasis Rev.* **28**(1–2), 99–111 (2009).

69. Iyer, J. *et al.* What's Nu(SAP) in mitosis and cancer?. *Cell Signal* **23**(6), 991–998 (2011).

70. Raemaekers, T. *et al.* NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization. *J. Cell Biol.* **162**(6), 1017–1029 (2003).

71. Qian, Z. *et al.* Prognostic value of NUSAP1 in progression and expansion of glioblastoma multiforme. *J. Neurooncol.* **140**(2), 199–208 (2018).

72. 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).

73. 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).

74. 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).

75. Gordon, C. A. *et al.* NUSAP1 promotes invasion and metastasis of prostate cancer. *Oncotarget* **8**(18), 29935–29950 (2017).

76. Yu, Z. *et al.* NUSAP1 promotes lung cancer progression by activating AKT/mTOR signaling pathway. *Zhonghua Zhong Liu Za Zhi* **42**(7), 551–555 (2020).

77. 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).

78. 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. *Med. Sci. Monit.* **24**, 8553 (2018).

79. 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).

80. Gilson, P. *et al.* Identification of pyrrolopyrimidine derivative PP-13 as a novel microtubule-destabilizing agent with promising anticancer properties. *Sci. Rep.* **7**(1), 1–14 (2017).

81. 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).

82. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**(1), 25–29 (2000).

83. Zhang, C. *et al.* Analysis of whole genome-wide microRNA transcriptome profiling in invasive pituitary adenomas and non-invasive pituitary adenomas. *Chin. Neurosurg. J.* **5**(1), 1–7 (2019).

84. Zheng, Z. *et al.* Effect of miR-106b on invasiveness of pituitary adenoma via PTEN-PI3K/AKT. *Med. Sci. Monit.* **23**, 1277 (2017).

85. 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* **12**, 23114 (2020).

86. Li, B. *et al.* Regulating the CCNB1 gene can affect cell proliferation and apoptosis in pituitary adenomas and activate epithelial-to-mesenchymal transition. *Oncol. Lett.* **18**(5), 4651–4658 (2019).

87. An, J. *et al.* Lactate dehydrogenase A promotes the invasion and proliferation of pituitary adenoma. *Sci. Rep.* **7**(1), 1–12 (2017).

88. Rohart, F. *et al.* MINT: A multivariate integrative method to identify reproducible molecular signatures across independent experiments and platforms. *BMC Bioinform.* **18**(1), 128 (2017).

89. Brazma, A. *et al.* ArrayExpress: A public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**(1), 68–71 (2003).

90. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**(1), 207–210 (2002).

91. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **19**(1A), A68 (2015).

92. Ritchie, M. E. *et al.* A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23**(20), 2700–2707 (2007).

93. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**(4), 265–273 (2003).

94. 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).

95. Smyth, G. K. Limma: Linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (Springer, 2005).

96. Gentleman, R. *et al.* Genefilter: Methods for Filtering Genes From HIGH-Throughput Experiments. R package version (2015).

97. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**(1), 1–25 (2004).

98. Benjamini, Y. *et al.* Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* **125**(1–2), 279–284 (2001).

99. McKight, P. E. & Najab, J. Kruskal-wallis test. In *The Corsini Encyclopedia of Psychology* 1 (Wiley, 2010).

100. Gunderson, T. & Ho, Y. Y. An efficient algorithm to explore liquid association on a genome-wide scale. *BMC Bioinform.* **15**(1), 371 (2014).
101. Li, K. C. Genome-wide coexpression dynamics: Theory and application. *Proc. Natl. Acad. Sci. U S A* **99**(26), 16875–16880 (2002).
102. Willse, J. T. & Willse, M. J. T. *Package 'CTT'*. (2014).
103. Wright, M. N. & Ziegler, A. *ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R.* arXiv preprint, 2015.
104. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**(43), 15545–15550 (2005).
105. Zarnegarnia, Y. *et al.* Application of fuzzy clustering in analysis of included proteins in esophagus, stomach and colon cancers based on similarity of Gene Ontology annotation. *Koomesh* **12**(1), 14–21 (2010).
106. Kanehisa, M. *et al.* KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**(1), 109–114 (2012).
107. 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).
108. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**(11), 2498–2504 (2003).
109. Thissen, D., Steinberg, L. & Kuang, D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat.* **27**(1), 77–83 (2002).
110. Remo, A. *et al.* Systems biology analysis reveals NFAT5 as a novel biomarker and master regulator of inflammatory breast cancer. *J. Transl. Med.* **13**(1), 138 (2015).
111. Margolin, A. A. *et al.* ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7**, S7 (2006).

Author 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.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00431-2.

Correspondence and requests for materials should be addressed to M.J. or M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.