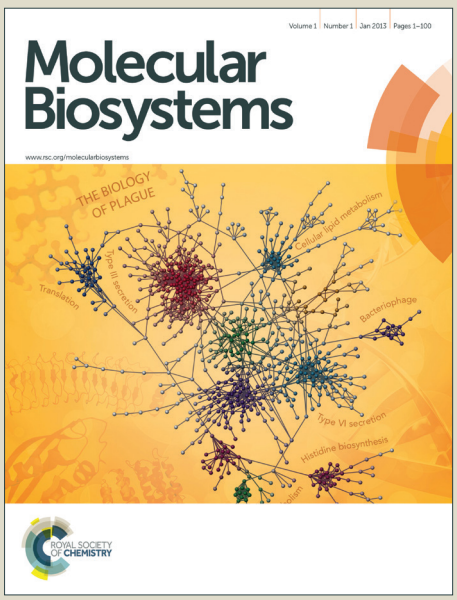
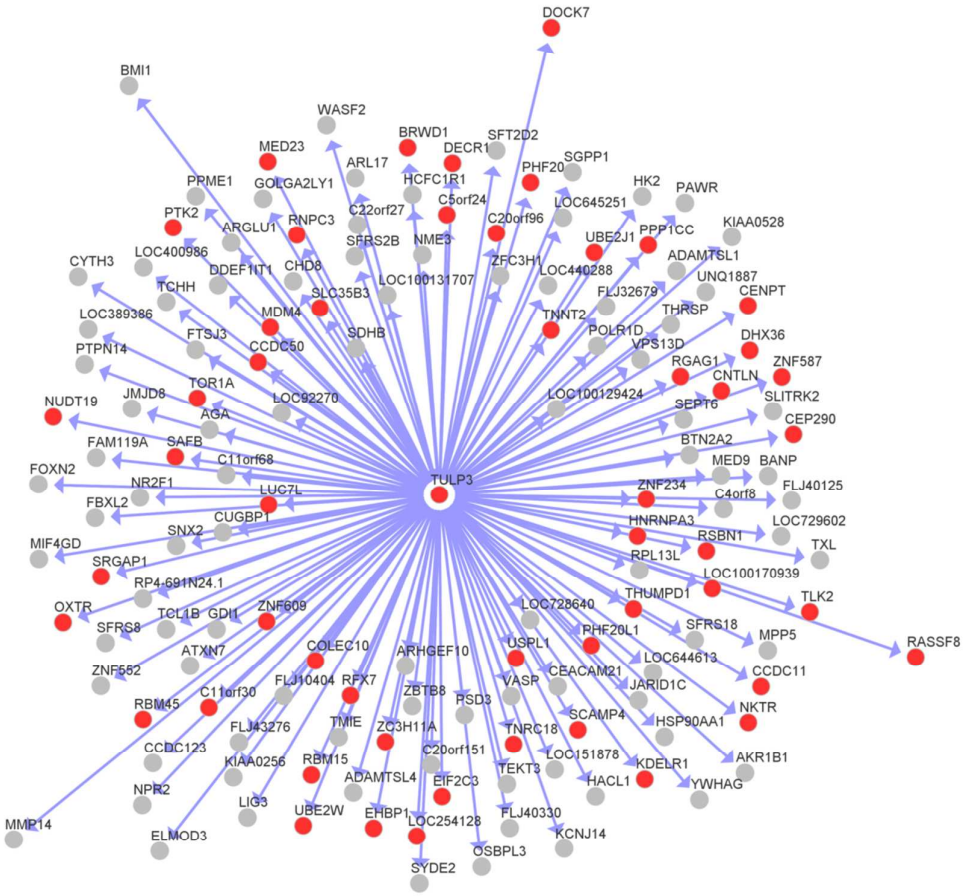
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems



Computational analyses identified TULP3 as Master Regulator of transcription in PDAC expression data and moreover its regulated-genes, assigning TULP3-prognostic value.

Computational Analyses Reveal a Prognostic Impact of TULP3 as a Transcriptional Master Regulator in Pancreatic Ductal Adenocarcinoma

I. T. S. Sartor[a], F. Zeidán-Chuliá[a], R. D. Albanus[a], R. J. S. Dalmolin[a] and J. C. F. Moreira[a].

[a]Centro de Estudos em Estresse Oxidativo. Departamento de Bioquímica, Instituto de Ciências Básicas da Saúde (ICBS), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

Corresponding author: Miss. Ivaine Taís Sauthier Sartor, Centro de Estudos em Estresse Oxidativo, Departamento de Bioquímica, Instituto de Ciências Básicas da Saúde (ICBS), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil; Ramiro Barcelos 2600, Porto Alegre, RS, Brazil. CEP: 90035-003; Phone +55(51)3308-5577, Fax: +55(51)3309-5535; ivaine.sauthier@yahoo.com.br

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is world-wide recognized as an aggressive disease with poor prognosis in patients with ou without resection. Further knowledge about the biological mechanisms of PDAC are necessary to enable the identification of novel molecular markers and therapeutic targets for early diagnosis and improved treatment. Transcription factors are the final effectors of signaling pathways and regulate a number of cellular functions. Changes in its expression may contribute to cellular transformation and tumor progression. Thus, the aim of the present study was to identify the Master Regulators (MRs) of transcription potentially involved in PDAC disease. To achieve this goal, we utilized microarray data to correlate MR genes with tumor phenotype. Analyses were performed with RTN, Limma, and Survival packages at the R environment. We identified Tubby-like protein 3 (TULP3) as MR of transcription in PDAC samples. Prognostic value of TULP3 was accessed in three independent cohort analyses. Our data demonstrated that pancreatic cancer patients exhibiting high transcriptional levels of TULP3 showed a poor overall survival. High expression levels of TULP3 may play an essential role in pancreatic cancer progression possibly leading to poor clinical outcome. Our results highlight the potential use of TULP3 as a clinical prognostic biomarker for pancreatic adenocarcinoma.

Keywords: PDAC, *regulon*, cancer, prognosis, transcription factor, master regulator.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) cancer is recognized as the 4th cause of death in United States due to its aggressive nature. In 2013, it was estimated that 45.220 north-Americans would be diagnosed with pancreatic cancer and 38.460 would die due to the disease[1]. In the UK, in 2010, 8.455 people were diagnosed with PDAC and 7.921 deaths were reported, and it is currently in the 5th place of cancer-related deaths[2]. In Brazil, PDAC is responsible for about 2% of diagnosis and 4% of the total deaths among all types of cancer. In 2010, 7.740 deaths were reported by PDAC[3].

About 50% of the total number of pancreatic cancer cases is diagnosed in emergency assistance, where patients exhibit diverse and general symptoms such as non-specific abdominal pain, jaundice, or both. Courvoisier's signal, which is characterized by the presence of palpable gall bladder accompanied with jaundice without abdominal pain, is an indicative of PDAC; nonetheless, it occurs in less than 25% of patients[4]. The majority of tumors is located in the pancreas head and usually cause jaundice due to blockage of bile duct[5].

PDAC can be defined as a solid and infiltrative neoplasia, arising from precursors named Pancreatic Intraepithelial Neoplasia (PanIN). The histological progression of low-grade intraepithelial neoplasia (PanIN-1) to intermediate-grade (PanIN-2) and high-grade (PanIN-3) and, subsequently, to invasive carcinoma is associated with accumulation of different genetic alterations[6]. For instance, over-expression of ErbB2 and Kras mutations are present in PanIN-1, whereas inactivation of p16 gene is observed in PanIN-2 and inactivation of p53, DPC4, and BRCA2 occur in PanIN-3 lesions[6].

Early diagnosis of PDAC still remains a challengefor clinicians, since the symptoms are generally non-specific (e.g., abdominal pains, weight loss and nausea)[4]. When diagnosed, approximately 80% of patients present unresectable disease[7]. In patients who did not undergo resection, the survival varies between 3 and 5% and those who had the organ resection, from 10 to 20%, all within 5 years[8], leading a poor prognosis to PDAC disease.

The choice for PDAC treatment depends on the disease stage. For example, primary tumors limited to the pancreas (stage I) and locally invasive tumors (stage II) are usually resectable, while locally advanced (stage III) and metastatic (stage IV) cannot be removed surgically[7]. Conventional therapies such as radio and chemotherapy have palliative effects for more advanced stages, rendering the surgery the only treatment with chances of cure[9].

Transcription factors are the final effectors of signaling pathways and are involved in regulating cellular functions such as proliferation, differentiation, and apoptosis[11]. Therefore, changes in transcription factors expression may impact cell biology possibly leading to tumor progression. In the present study, we aimed to identify the transcriptional Master Regulators involved in pancreatic adenocarcinoma. Our analyses highlighted Tubby-like protein 3 (TULP3) as a transcription factor with a potential contribution to the PDAC phenotype.

Materials and methods

Data acquisition. PDAC microarray data were obtained from Gene Expression Omnibus database – GEOdatabase[12], under accession number GSE21501, and was originally contributed by Stratford et al. 2010[13]. Only human PDAC samples with clinical data were used in the present study. The genome-wide human transcription factors were obtained from the Animal Transcription Factor DataBase[14]. These data were essential to develop the transcriptional network.

Transcriptional Network. The transcriptional network was constructed at the R environment[15] with RTN package[16], which applies Mutual Information (MI) measures for a pair of random variables expression data, generating a degree of statistical dependency between these variables. These values were transformed in MI estimated values. Pearson correlation was used as Gaussian estimator[17]. Statistical relations, applied in expression data and transcription factors, were the basis for the transcriptional network development.

ARACNe algorithm was used to eliminate the majority of indirect interactions inferred, since it explores the MI estimated values of a gene triplet (triangle of two TFs and a gene target) and removes the smallest one[17].Therefore, the PDAC transcriptional network was comprised by the greatest MI estimated values of TF-target pair. This method infers candidate interactions of a TF and target genes from transcriptional network. A TF and genes that are directly regulated by it is referred as *regulon*[17].

Master Regulator Analysis. GSE16515 microarray data, obtained at GEOdatabase[12], which present normal and tumor samples, were used as a phenotype data to determine

the statistical significance of the overlap between each *regulon* and the phenotype data. Limma R-package[18] was used to estimate the differentially expressed genes between normal and tumor samples (i.e. hits). Master Regulator Analysis (MRA) identified the differentially expressed genes in each *regulon* by using Fischer's exact test, pointing a TF as Master Regulator (MR) in the PDAC phenotype array[19,20] (p-value<0.001).

Filters. GSE10780 normal breast tissue expression data, publicly available at GEOdatabase[12], was used as negative control network of non-related tissue. Since our goal was to maintain the genetic features of pancreas tissue, then characteristics shared by tissues of different embryonic derivation were discarded. Thus, common data between transcriptional network and negative control network were removed. The MetaPCNA[21], which is a proliferation-based signature presenting genes that are involved in cancer progression, was used to discard the proliferative MRs; whereas it is known that these genes are related to tumorigenesis.

Survival Analysis. GSE21501, GSE28735 and MEXP2780 microarrays data, publicly available in GEOdatabase[12] and ArrayExpress[22] were used for survival analysis, taking as parameter the expression of MR genes. Analysis was performed by using the R-Survival package[23]. To compare the prognostic value of different gene expressions of MRs, an optimal cut-off was used to dichotomize the cohort, which is defined as the point with the most significant split[24]. Kaplan-Meier method was used to estimate survival curves for the patients and LogRank test was used to compare the survival. Hazard ratio (HR), which describe the measurements of how often the event occurs in

one group compared to another over time, were calculatedwith 95% confidence intervals.

Statistical Analysis. Statistical analysis was performed by using R-Student's *t* test in GSE15471 normal and tumor expression data, obtained at GEOdatabase[12]. In order to select an optimal microarray probe to represent a gene we used the JetSet score on Affymetrix gene expression microarrays, assigning the 221964_at probe for TULP3 gene[25].

Kaplan-Meier Plotter web tool. Prognostic value of TULP3 expression was analyzed in three different types of cancer (breast, ovarian and lung cancer), using the Kaplan-Meier Plotter web tool, which performs a meta-analysis based *in silico* biomarker assessment. Overall survival and auto select best cut-off were used as parameters for analysis. All clinical and gene expression data were accessed from GEO[12], EGA[26] and TCGA[27] public repositories. The dichotomized patient cohorts were then compared by LogRank test and the HR with 95% confidence intervals were calculated[28,29].

Results and discussion

Transcription factors identified as Master Regulators of PDAC by computational analyses

A total universe of 19.751 genes and 5.476 hits were analyzed and it were identified 15 transcription factors as significant MRs (p-value<0.001), essentials for

PDAC signature (**Table 1** and **Fig. 1**). These TFs regulate genes that possibly have an essential role in promoting and sustaining the transformed cells. For such a reason, we decided to study how these MRs were expressed in PDAC samples and how its expressions could be related to survival. Therefore, survival analyses were done for the 15 MR genes, in GSE21501, GSE28735 and MEXP2780 arrays and, Tubby-like protein 3 (TULP3) was the only MR which presented the same survival curve profile for all arrays. Those MRs which did not present the same profile curve or were not verified in all arrays, were then excluded (Supplementary Figures).

TULP3 is a master regulator with prognostic value in PDAC

Analysis of TULP3 expression in subjects with PDAC revealed that lower transcription levels of TULP3 is associated with better prognosis. GSE21501 cohort was dichotomized for TULP3 gene and patients with low- (n=75) vs. high-expression (n=27) were different (*p*-value=0.0007) and exhibited 2.4 greater chances of survival (**Fig. 2A**). For the GSE28735 cohort, patients with low- (n=24) vs. high-expression of TULP3 (n=18) showed a significance of (*p*-value=0.00474) and 2.98 more chances of survival (**Fig. 2B**). Finally, the MEXP2780 cohort was divided in low- (n=20) vs. high-expression of TULP3 (n=10) were different (with *p*-value=0.00496) and exhibited 3.4 greater chances of survival for patients with low expression of TULP3 (**Fig. 2C**).

Next step was to verify the levels of TULP3 (mRNA) in normal and PDAC individual. To achieve this goal, we used the GSE15471 array with both normal and

PDAC samples. The mean value for log_2-TULP3 expression in normal patients was 6.54 and 6.86 for PDAC patients, with difference between these two groups (*p*-value=2.096e-07). This provides further evidence of alterations in TULP3 expression in those individuals with PDAC (**Fig. 3**).

In order to confirm that aberrant expression of TULP3 may serve as a pancreatic cancer biomarker, we analyzed its expression levels in others types of cancer by using the Kaplan-Meier Plotter web tool. Breast and ovarian cancer presented no statistical difference in TULP3 expression comparisons (**Fig. 4A** and **4B**). Nonetheless, we found significant difference in lung cancer samples (*p*-value=0.033) (**Fig. 4C**). However, at overall survival of 60 months, which is the same time used for PDAC dataset, no statistical significance in TULP3 expression was detected on neither breast, ovarian, nor lung cancers, respectively (*p*-value=0.099, *p*-value=0.18 and *p*-value=0.079) (**Fig. 3D-F**). All these results taken together, may suggest TULP3 as a potentially specific biomarker for pancreatic cancer prognosis.

Developing a *regulon* model to identify the potential TULP3-associated target genes

Very little is known about the TULP3-associated target genes. Therefore, we here provide the possible TULP3-regulated genes in PDAC (**Fig. 5**). Among the overexpressed genes associated to TULP3 *regulon* which contribute to tumor phenotype, we found Dedicator of Cytokinesis Protein 7 (DOCK7) with *p*-value=0.0002

and Ras-Association Domain Family protein 8 (RASSF8) with *p*-value=0.006 (**Supplementary Table**).

Discussion

Tulp3 is a member of the mammalian tubby-like proteins (Tulp's), which present a carboxy-terminal tubby domain, include Tub and Tulp1 to Tulp4 proteins. TULP3 gene has an essential role during mammalian development, since mutations in TULP3 exhibit embryonic lethality with defects in neural tube closure[30-32]. Tubby domain is positively charged at carboxy-terminal region and possesses a nuclear localization conserved sequence, which enables its role as transcription factor. This region of Tulp3 proteins binds to the plasma membrane, specifically to phosphatidylinositol-4,5-bisphosphate (PIP_2), a phospholipid highly enriched in the membrane[30]. Tulp3 can interact with IFT-A or be dislodged from the membrane, enabling nuclear translocation[33]. A conserved domain in amino-terminus enables Tulp3 to bind to the intraflagellar transport A-complex (IFT-A), which is a microtubule-based transport essential for ciliogenesis[34]. Primary cilia are a microtubule-based organelle ubiquitously expressed in epithelial cells, including the pancreatic tissue[35]. These sensory compartments receive extracellular signals and transduce the information leading to transcriptional regulation of downstream genes[33,34].

When Tulp3-(IFT-A complex) binding is impaired, Tulp3 cannot exert its function in primary cilia, and it is then translocated into the nucleus. Its absence in

primary cilia stimulates the Hedgehog (Hh) signaling since Tulp3 negatively regulates this pathway[34,35].

PDAC arises when genetic alterations occur at the level of signal transduction proteins, which participate in normal pancreas embryonic development (e.g., Hh pathway)[30,34-36]. Hh signaling proteins are commonly undetectable in normal ductal epithelia, although its expression is found in PanIN lesions and invasive PDAC[36]. Besides Tulp3 induces proliferation through Hh signaling, the role of some TULP3 regulated genes, which effectively contribute to PDAC phenotype, are describe bellow:

(i) DOCK7, a member of Dock180-related superfamily of Guanine nucleotide Exchange Factors (GEFs), acts generating active GTPases like Rac1, Cdc42, and RhoA, which are responsible for actin cytoskeleton regulation. Moreover, it has been reported that ErbB2 receptors bind and activate Dock7, promoting activation of Rho GTPases and thus inducing migration of Schwann cells[37]. Cancer patients with overexpression of ErbB2 tend to have a poor clinical outcome[38]. In addition, the overexpression of ErbB2 is observed in early steps of PanIN lesions progression until the infiltrating pancreatic adenocarcinoma[6]. This may suggest a potential role of ErbB2 in Dock7 activation and promotion of invasiveness.

(ii) RASSF8, which contains the Ras-Association (RA) domain at N-terminal region, is involved in Ras signaling[39]. As for GTPases, the Ras family proteins are activated by the GEFs proteins. When activated, Ras proteins induce proliferative signals which promote tumor initiation and progression through the stimulation of transcription factors (e.g., Transforming Growth Factor-α, TGFα)[40].

(iii) Among the TULP3-regulated genes, we also found MMP14 (Matrix Metalloproteinase-14 also known as Membrane-Type 1 Matrix Metalloproteinase - MT1-MMP) and BMI1 (B-cell-specific Moloney murine leukemia virus insertion site 1). Despite the non-statistical significant between PDAC vs. normal individuals in our study, overexpression of MMP14[41,42] and BMI1[43,44] are observed in human pancreatic adenocarcinomas, possibly playing a role on enhancing proliferation and invasiveness.

Identification of transcriptional MRs associated with prognosis of pancreatic cancer patients may shed light on the biological mechanisms involved in pancreatic adenocarcinoma and contribute to the identification of novel molecular targets. Our results are consistent with another study reporting the deregulation of TULP3 transcriptional levels in both PDAC samples and PanIN lesions[45]. Thus, we highlight the possible role of TULP3 on tumor progression and maintenance of pancreatic cancer phenotype.

Conclusions

Our results indicated that high TULP3 expression may play a critical role in pancreatic cancer progression since it is significantly correlated with a poor clinical outcome.To date, TULP3 low-high expression levels have not been associated with prognostic value in other types of cancer such as breast, ovarian and lung. Moreover, we believe that TULP3 expression could be explored in the future as a prognostic biomarker for PDAC patients.Nevertheless, further studies will be necessary for TULP3

biomarker validation as well as its definition as therapeutic target in pancreaticadenocarcinoma.

Additional Information

Supplementary data for this article include Supplementary Figures and Table.

Conflict of interest statement

The authors declare no conflict of interests.

Acknowledgements

We thank the Brazilian research funding agencies FAPERGS (PqG 1008860, PqG 1008857, ARD11/1893-7, PRONEX1000274), CAPES (PROCAD 066/2007), CNPq (558289/2008-8 and 302330/2009-7) and PROPESQ-UFRGS for financial support. We thank Mauro A. A. Castro for reviewing the manuscript.

References

1. National Cancer Institute. SEER Cancer Statistics Factsheets: Pancreas Cancer. http://seer.cancer.gov/statfacts/html/pancreas.html. Accessed September 2013.

2. Pancreatic Cancer Action. *UK Pancreatic Cancer Statistics 2012/13*. http://pancreaticcanceraction.org/wp-content/uploads/2013/05/PCA-Pancreatic-Cancer-Statistics-A.pdf. Accessed September 2013.
3. INCA. Câncer. Tipos de câncer. *Pâncreas*. http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pancreas. Accessed September 2013.
4. G. Bond-Smith, N Banga, T.M. Hammond, J.C. Imber. *BMJ*. 2012, 344, e2476.
5. M. Hidalgo. *N Engl J Med*. 2010, 362, 1605-1617.
6. R.H. Hruban, M. Goggins, J. Parsons and S.E. Kern. *Clin Cancer Res*. 2000, 6, 2969-2972.
7. K.Y. Bilimoria, D.J. Bentrem, C.Y. Ko, J. Ritchey, A.K. Stewart, D.P. Winchester and M.S. Talamonti. *Cancer*. 2007, 110(4), 738-744.
8. A.F. Hezel, A.C. Kimmelman, B.Z. Stange, N. Bardeesy and R.A. DePinho. *Genes Dev.* 2006, 20, 1218-1249.
9. Y. Yokoyama, Y. Nimura and M. Nagino. *Surg Today*. 2009, 39(6), 466-475.
10. D.S. Lachtman. *Eukaryotic Transcription Factors*. London, UK; Elsevier; 2008.
11. M.E. Fernandez-Zapico, P.S. Bramati, S. Zakaria, J.A. Kaczynski and R. Urrutia. *Pancreatology*. 2003, 3(4), 276-283.
12. R. Edgar, M. Domrachev and A.E. Lash. *Nucleic Acids Res*. 2002, 30(1), 207-210.
13. J.K. Stratford, D.J. Bentrem, J.M. Anderson, C. Fan, K.A. Volmar, J.S. Marron, E.D. Routh, L.S. Caskey, J.C. Samuel, C.J. Der, L.B. Thorne, B.F. Calvo, H.J.

Kim, M.S. Talamonti, C.A. Iacobuzio-Donahue, M.A. Hollingsworth, C.M. Perou and J.J. Yeh. *PLoS Medicine*. 2010, 7(7), e1000307.

14. H.M. Zhang, H. Chen, W. Liu, J. Gong, H. Wang and A.Y. Guo. *Nucleic Acids Res*. 2012, 40, D144-149.
15. R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing. Viena, Austria. ISBN 3-900051-07-0. http://www.r-project.org/.
16. M.N.C. Fletcher, M.A.A. Castro, X. Wang, I. Santiago, M. O'Reilly, S. Chin, O.M. Rueda, C. Caldas, B.A.J. Ponder, F. Markowetz and K.B. Meyer. *Nat Commun*. 2013, 4, 2464.
17. A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R.D. Favera and A. Califano. *BMC Bioinformatics*. 2006, 7(Suppl 1), S7.
18. G.K. Smyth, in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* , eds. R. Gentleman , V. Carey, S. Dudoit, R. Irizarry and W. Huber. New York: Springer; 2005, pp. 397-420.
19. W.K. Lim, E. Lyashenko and A. Califano A. *Pac Symp Biocomput*. 2009, 504-515.
20. M.S. Carro, W.K. Lim, M.J. Alvarez, R.J. Bollo, X. Zhao, E.Y. Snyder, E.P. Sulman, S.L. Anne, F. Doetsch, H. Colman, A. Lasorella,, K. Aldape, A. Califano and A. Iavarone. *Nature*. 2010, 463, 318-325.
21. D. Venet, J.E. Dumont and V. Detours. *PLoS Comput Biol*. 2011, 7(10), e1002240.

22. G. Rustici, N. Kolesnikov, M. Brandizi, T. Burdett, M. Dylag, I. Emam, A. Farne, E. Hastings, J. Ison, M. Keays, N. Kurbatova, J. Malone, R. Mani, A. Mupo, R.P. Pereira, E. Pilicheva, J. Rung, A. Sharma, Y.A. Tang, T. Ternent, A. Tikhonov, D. Welter, E. Williams, A. Brazma, H. Parkinson and U. Sarkans. *Nucleic Acids Res*. 2013, 41, D987–D990.

23. T.M. Therneau and P.M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. New York: Springer; 2000. ISBN 0-387-98784-3.

24. J. Budczies, F. Klauschen, B.V. Sinn, B. Györffy, W.D. Schmitt, S. Darb-Esfahani and C. Denkert. *PLoS One*. 2012, 7(12), e51862.

25. Q. Li, N.J. Birkbak, B. Györffy, Z. Szallasi and A.C. Eklund. *BMC Bioinformatics*. 2011, 12, 474.

26. I. Lappalainen, J. Almeida-King, V. Kumanduri, P. Marin-Garcia and P. Flicek. *Nuclei Acids Res*. ISSN 1362-4962. https://www.ebi.ac.uk/ega/.

27. The Cancer Genome Atlas Research Network. *Nature*. 2008, 455, 1061-1068.

28. B. Györffy, A. Lanczky, A.C. Eklund, C. Denkert, J. Budczies, Q. Li and Z. Szallasi. *Breast Cancer Res Treat*. 2010, 123(3), 725-731.

29. B. Györffy, A. Lanczky and Z. Szallasi. *Endocr Relat Cancer*. 2012, 19(2), 197-208.

30. T.J. Boggon, W.S. Shan, S. Santagata, S.C. Myers and L. Shapiro. *Science*. 1999, 286, 2119.

31. A. Ikeda, P.M. Nishina and J.K. Naggert. *J Cell Sci*. 2002, 115(Pt 1), 9-14.

32. S. Mukhopadhyay and P.K. Jackson. *Genome Biol*. 2011, 12(6), 225.

33. S. Santagata, T.J. Boggon, C.L. Baird, C.A. Gomez, J. Zhao, W.S. Shan, D.G. Myszka and L. Shapiro. *Science*. 2001, 292(5524), 2041-2050.

34. S. Mukhopadhyay, X. Wen, B. Chih, C.D. Nelson, W.S. Lane, S.J. Scales and P.K. Jackson. *Genes Dev*. 2010, 24(19), 2180-2193.

35. E.S. Seeley, C. Carrière, T. Goetze, D.S. Longnecker and M. Korc. *Cancer Res*. 2009, 69(2), 422-430.

36. S.P. Thayer, M.P. di Magliano, P.W. Heiser, C.M. Nielsen, D.J. Roberts, G.Y. Lauwers, Y.P. Qi, S. Gysin, C.F. Castillo, V. Yajnik, B. Antoniu, M. McMahon, A.L. Warshaw and M. Hebrok. *Nature*. 2003, 424(6960), 851-856.

37. J. Yamauchi, Y. Miyamoto, J.R. Chan and A. Tanoue. *J Cell Biol*. 2008, 181(2), 351-365.

38. T. Holbro, G. Civenni and N.E. Hynes. *Exp Cell Res*. 2003, 284(1), 99-110.

39. A.M. Richter, G.P. Pfeifer and R.H. Dammanna. *Biochim Biophys Acta*. 2009, 1796(2), 114-128.

40. Y. Pylayeva-Gupta, E. Grabocka and D. Bar-Sagi. *Nat Rev Cancer*. 2011, 11(11), 761-774.

41. S. Curran and G.I. Murray. *Eur J Cancer*. 2000, 36(13 Spec No), 1621-1630.

42. N.E. Sounni and A. Noel. Biochimie. 2005, 87(3-4), 329-342.

43. T. Konuma, H. Oguro and A. Iwana. *Dev Growth Differ*. 2010, 52(6), 505-516.

44. Z. Wu, Q. Wang, L. Wang, G. Li, H. Liu, F. Fan, Z. Li, Y. Li and Y. Tu. *J Neurol Sci*. 2013, 355(1-2), 192-196.

45. W. Song, K. Tao, H. Li, C. Jin, Z. Song, J. Li, H. Shi, X. Li, Z. Dang and K. Dou. *Cancer Sci*. 2010, 101(7), 1754-1760.

46. E. Proctor, E. Waghray, C.J. Lee, D.G. Heidt, M. Yalamanchili, C. Li, F. Bednar and D.M. Simeone. *PLoS One*. 2013, 8(2), e55820.

47. T. Harada, C. Chelala, T. Crnogorac-Jurcevic and N.R. Lemoine. *Pancreatology*. 2009, 9(1-2), 13-24.

MR	Probes	*Regulon* size	Observed hits	*p*-value	Adjusted *p*-value
ZNF407	A_23_P380954	108	67	2.09e-14	2.04e-11
MYSM1	A_23_P348992	60	42	1.84e-12	5.99e-10
ZNF148	A_23_P139408	122	63	6.51e-09	1.27e-06
ZZZ3	A_23_P11507	128	65	9.44e-09	1.53e-06
ZFP91	A_24_P56052	255	106	6.15e-07	7.49e-05
ZNF41	A_23_P45234	63	35	8.83e-07	9.57e-05
HSF4	A_23_P3592	195	82	5.53e-06	4.90e-04
ARID4B	A_23_P201951	126	57	8.39e-06	6.82e-04
PRDM10	A_32_P228699	102	48	9.80e-06	7.35e-04
SMARCE1	A_23_P333063	231	93	1.31e-05	7.98e-04
ZBTB25	A_23_P48628	66	34	1.23e-05	7.98e-04
ZNF3	A_23_P219084	71	36	1.18e-05	7.98e-04
ZNF280D	A_23_P14708	131	58	1.61e-05	8.71e-04
TULP3	A_23_P116980	200	82	1.74e-05	8.94e-04
NFE2L2	A_23_P5761	67	34	1.89e-05	9.21e-04

Table 1. Significant Master Regulators of transcription in PDAC expression data (*p*-value<0.001). Each MR was assigned to the corresponding probes of Agilent gene expression microarray (GSE21501), to the total number of genes which composed the *regulon*, to the differentially expressed genes observed in each *regulon*, to the *p*-value and the adjusted *p*-value.

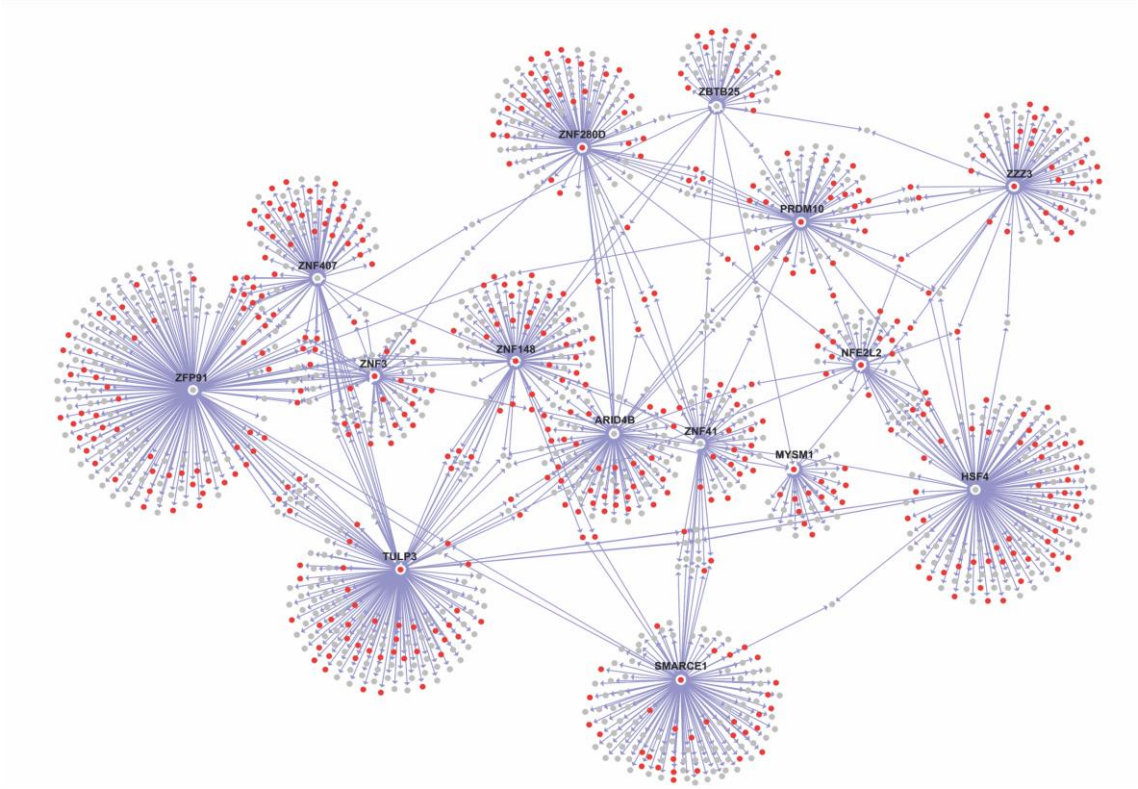
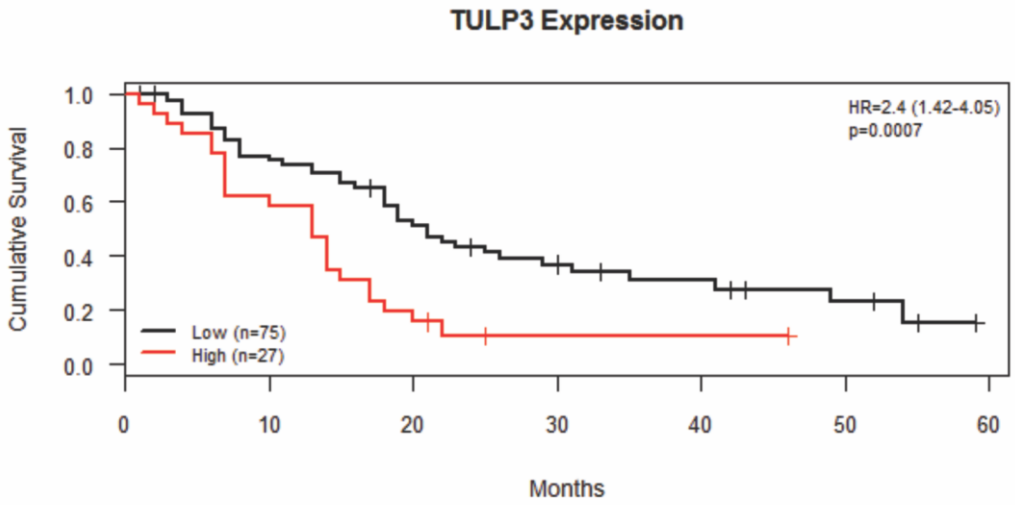
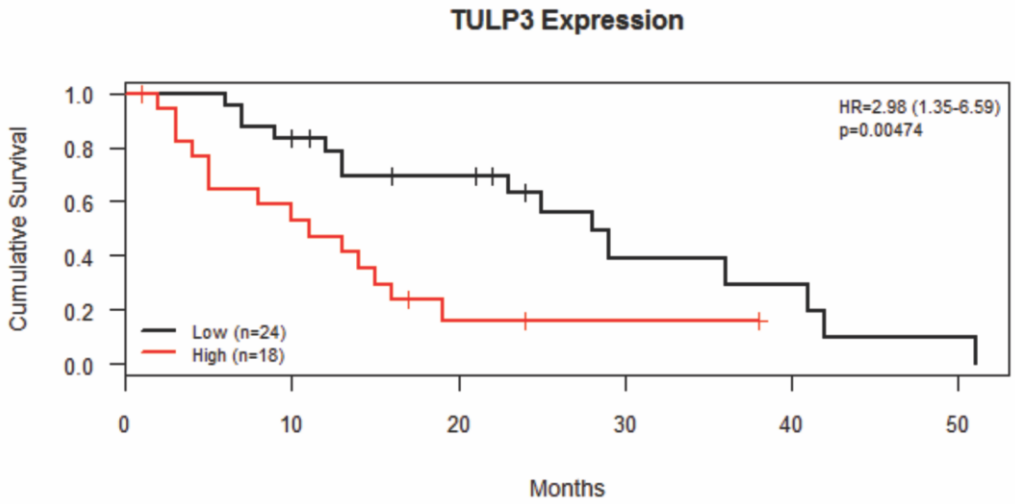


Fig. 1. Master Regulators network. The 15 significant MRs of transcription and the respectively regulated genes (p-value<0.001). Hits are showed in red.

A



B



C

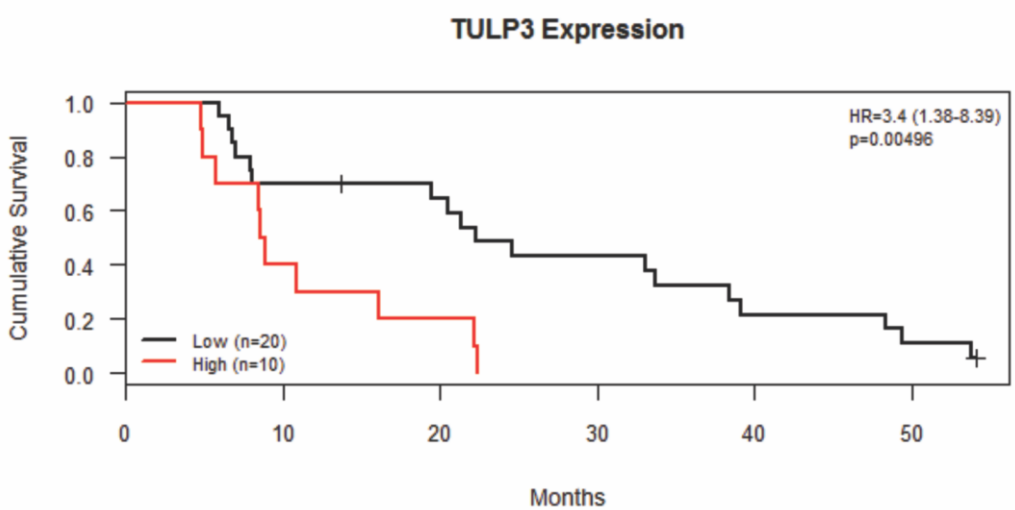


Fig. 2. Kaplan-Meier curves in three independent arrays comparing TULP3 expression. (**A**) in the GSE21501 data we observed 2.4 greater chances of survival for patients who had low TULP3 expression levels and a difference (*p*-value=0.0007) between low-high TULP3 expression, identified by the LogRank test. (**B**) in the GSE28735 microarray, patients with low expression levels had 2.98 more chances of survival than patients who had high expression levels, with a difference between low-high TULP3 transcriptional levels of *p*-value=0.00474. (**C**) in the MEXP2780 samples data, patients who had low TULP3 expression levels ehxibited 3.4 greater chances of survival comparing to those who had highest expression levels and a difference between two expression groups of *p*-value=0.00496.

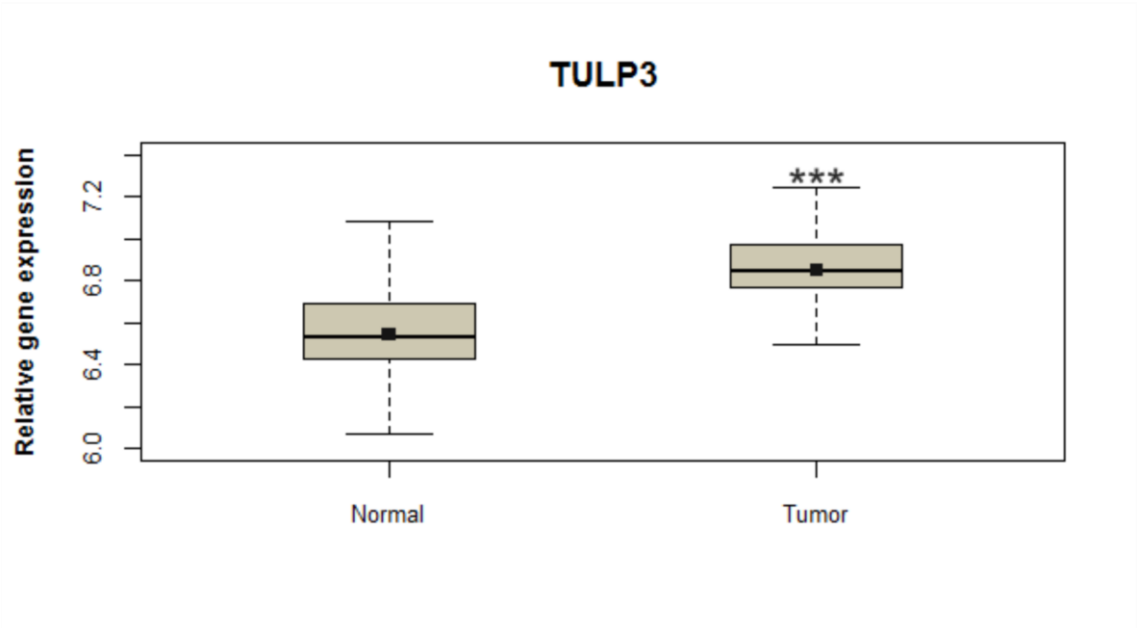


Fig. 3. Boxplot of relative TULP3 mRNA levels in normal and PDAC patients in the GSE15471 array. T-test pointed a significant difference between normal and tumor expression values. The mean value of $\log_2$-TULP3-expression in

normal patients was 6.54 whereas in PDAC patients the mean value log_2-TULP3-expression was 6.86. Mean: square black, Median: solid black line, whiskers: maximum and minimum values, bottom and top of each boxes: lower and upper quartiles, respectively. ****p*-value=2.096e-07.

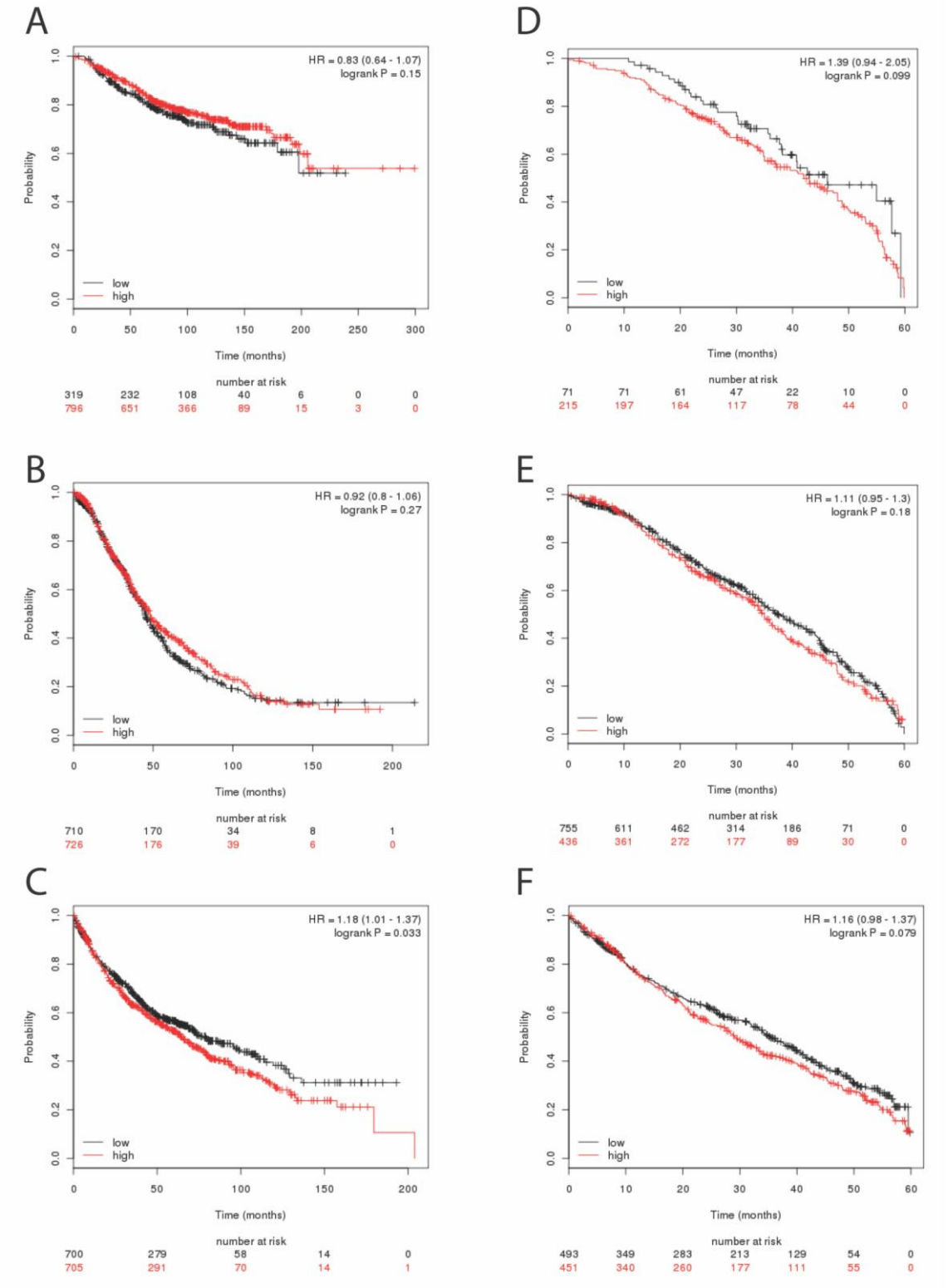


Fig. 4. Kaplan-Meier curves. (**A**) Cohort of 1115 breast cancer patients showed no statistical difference between low (n=319) vs. high expression of TULP3 (n=796) as well as the (**B**) Cohort of 1436 ovarian cancer patients wich were divided into low (n=710) and high TULP3 transcriptional levels (n=726). (**C**) Difference (*p*-value=0.033) between low (n=700) and high (n=705) TULP3 expressions of 1405 lung cancer patients cohort. **A**, **B** and **C** were plotted with maximum overall survival. **D**, **E** and **F** were plotted with 60 months of overall survival. No statistical significance was detected in (**D**) breast, (**E**) ovarian and (**F**) lung cancer.

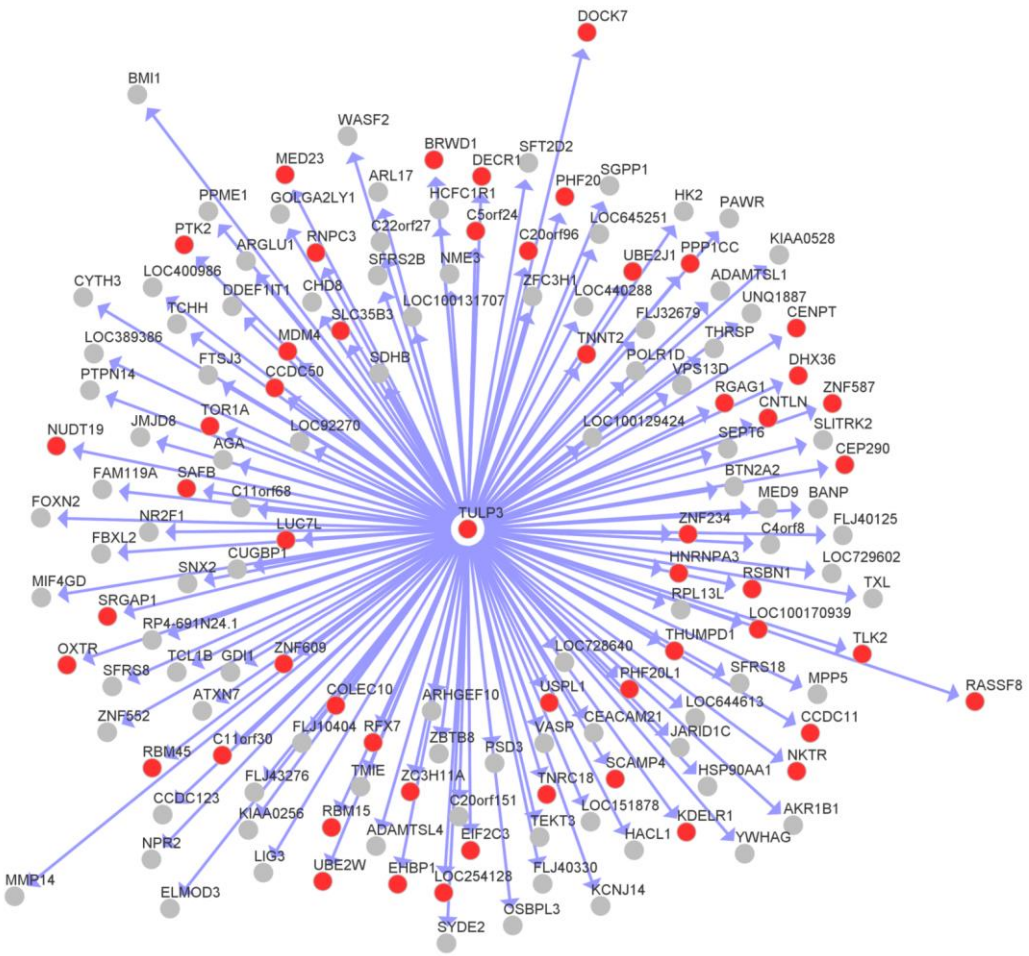


Fig. 5. TULP3 *regulon*. TULP3 transcription factor-regulated genes were inferred by ARACNe algorithm. Hits are showed in red; in gray are shown genes which presented no difference when comparing tumor to normal pancretic tissue.