BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(14)

Analysis of gene expression on ngn3 gene signaling pathway in endocrine pancreatic cancer

Prashantha Nagaraja[1]*, Kavya Parashivamurthy[2], Nandini Sidnal[2], Siddappa Mali[2], Dakshyani Nagaraja[1] & Sivarami Reddy[3]

[1]Department of Bioinformatics, Scientific Bio-Minds, Bangalore-560092; [2]Department of Biotechnology, KLE Dr.M.S. Sheshagiri College of Engineering and Technology, Belgaum-590008; [3]Department of Biotechnology, Scientific Bio-Minds, Bangalore-560092; Prashantha Nagaraj – Email: prashanthacn@gmail.com; Phone: +91-98441 58444; *Corresponding author

Received July 17, 2013; Revised July 20, 2013; Accepted July 20, 2013; Published August 07, 2013

Abstract:
In order to define the undifferentiated transcriptional factors present in neurogenesis of pancreatic β-islet cells, we studied the effect of Pdx1 in embryonic stem cell derived endocrine lineage. There are undifferentiated transcriptional progenitors Pdx1+/Ptf1a+/Cpa1+ tracking the growth of acini, ducts, α and β-islet cells. The upregulated transcriptional factors Pdx1 and ngn3 specify consequences of cell cycle regulation in early gut endocrine cells. The undifferentiated transcriptional factors basic helix loop helix (bHLH) protein regulate Ptf1a+/Cpa1+ in acini, ducts and it also regulate ngn3 to decrease expression of insulin and other pancreas secretions. The Pdx1+ and other unknown gene mutations show abnormal growth of neurogenesis in endocrine lineages. Using microarray based gene expression analysis to determine undifferential gene ontology in tissue specific gene regulation and disease progression that common in both metabolic and biological signaling pathways. The data expression profiles of ngn3 of wild- type pancreatic islet and islet derived tumor stem cells provide information on endocrine specific ngn3 genes. Therefore, 3755 genes were significantly regulated by Ngn3 induced pancreatic islet cell development. Moreover 317 upregulated and 175 downregulated, 757 genes deemed as undifferential expressions in endocrine cell. Furthermore to predict signaling pathways that associates with diabetes is highlighted.

Key words: Pancreatic regeneration, neurogenesis, ngn3, bHLH, Notch Pathway, Embryonic stem cells, islet cells, pancreatic beta cells, diabetes, microarray.

Background:
Pancreatic and duodenal homeobox 1 (Pdx1 also known insulin promoter factor1) gene is a major role in pancreas development **[1].** Eventually Pdx1+ is expressed in epithelial and mesenchymal cells and helps for development of pancreas buds **[2, 3].** The transcriptional factors Pdx1, SRY, SOX9 **[4]** and WNT7 **[5]** regulate morphological assessment of embryogenesis in human pancreas. The initiation of embryogenesis shows gene conversion is taking place in chromosomes, but it did not affect to flanking genes (Pyr1 and Col4) located at either side of Pdx-1+. During mitosis the heterozygous conversion of two Pdx-1 and six Pdx-1+ alleles show mutations show undifferentiated genes expressions leads to apoptosis **[6].** The heterozygous mutation of pdx1 show many pancreatic associated diseases including adenocarcinoma and diabetes mellitus **[7].**

The development of pancreatic cell neurogenesis is a major part of organogenesis to differentiate both exocrine and endocrine tissue to develop acini, ducts, α and β-islet cells **[8].** Both tissues are present in early endoderm epithelium through cell-cell signaling with mesenchyme that directs a cascade of transcriptional regulatory events **[9].** During early stage development of epithelial cells is undergo control of pancreas-specific transcriptional complex ptf1 which binds to nuclear transcriptional factors bHLH **[10]** it inhibits Pdx-1+ to form undifferential growth factors and morphogenesis **[11].** Out of these factors, the bHLH factor neurogenin3 (ngn3) plays a central role in initiating the differentiation of the pancreatic cells **[12].** Until recently, may intracellular signals that regulate pancreatic development and function **[13],** these include transforming growth factor-β (TGF- β), Notch, Hedgehog,

fibroblast growth factor (FGF), and epidermal growth factor (EGF) pathways, and investigation of pancreas developmental biology demands familiarity with these signaling pathways **[14-18].** In early sets of gene expression and differentiation shows Jack-STAT and notch signaling pathways play a crucial role in neuronal cell development. The loss of function on premature endocrine cell differentiation conversely, forced expression of the Notch intracellular domain (NICD) blocks on endocrine cell development **[19].** The differential expression of notch1 in pdx1+ epithelial cells suggests that Notch signaling could inhibit a ngn-neuroD cascade in neurogenesis and thus differentiate endocrine cells **[20].**

In this study, we examined the influence of ngn3 on early β-cell development. The results demonstrate that differential gene expression on Pdx1 and ngn3 on β-cell proliferation in embryonic cell development and regulate signaling pathway in early post transcriptional modifications. Our findings provide evidence of ngn3 is necessary for activation of pancreatic acini, ductal and beta cells to develop normal pancreas. Furthermore, we classified the undifferential genes based on the gene regulatory network and expression patterns in different tissues and these genes shows potential drug targets in pancreatic cancer.

Methodology:

In order to study signal transduction and analysis of various disease mechanisms, we need (i) gene expression measurements (ii) definitions of signaling pathways and (iii) drug target identification. The pancreatic cancer gene expression data analysis helps to provide signaling transductions which contains information about relation of neuronal development in pancreatic cancer and target gene product.

Microarray Raw data searching

We evaluated all published case control studies and diseased datasets were selected using GEO (GDS2276) contains 16 datasets of embryonic stem cells were processed for microarray analysis using Affymetrix Mouse Genome 430 2.0 GeneChip. The samples include parental stem cells (1) undifferentiated with 0 days (2) differentiated for 3 days as embryonic stem cells without ngn3. (3) 10 days as embryoid bodies. The remaining samples included tetracycline inducible Ngn3 ESC line selected from the parental stem cell line (4) after differentiation and addition of deoxycycline for 3 days without embryoid body formation (Ngn3 ES3 ON), (5 and 6) after differentiation as embryoid bodies for 3 days with (Ngn3 EB3 ON) or without (Ngn3 EB3 OFF) deoxycycline, and (7 and 8) after differentiation as embryoid bodies for 10 days with (Ngn3 EB10 ON) or without (Ngn3 EB10 OFF) deoxycycline. The ngn3 probe sets were selected from GenBank, dbEST, and RefSeq. The sequences of clusters were developed from the UniGene database and then refined by analysis and comparison with a number of other publicly available databases.

Identification of ngn3 gene expression data

BioConductor packages used to calculate the processed data. GCRMA (Gene-Chip Robust Multiarray Average) was used for signal normalization. Ainv15 ES0 data were defined as the baseline to generate differential expression values for all hybridizations. A patient sample pair was excluded from further analysis since one of the samples did not meet the quality controls. The microarray data was subsequently normalized using the Robust Microarray Analysis (RMA) algorithm. A multiclass analysis of variance was performed to generate p values for every probe set. To evaluate transcriptional factors among treatment groups, differentially expressed genes with expression of at least 2 and not greater than -2 and p values of not greater than 0.001 were selected from each group and combined for hierarchical clustering using a correlation similarity metric, an average clustering method, and by clustering both rows and columns.

Since the expression data is approximately log normally distributed, we used the log-transformed data as produced by the RMA algorithm for all subsequent statistical tests. For visualization purposes, we centered the log-transformed expression data by subtracting the average probeset log-expression values. Probe sets with relatively low expression (average expression values below 100 Affymetrix units) or with nearly constant expression values (standard deviation below 50) were excluded from further consideration. Of the 45101 probe sets on the mouse430, 18000 were thus retained. An unpaired t-test was used to determine the probesets (genes) that are differentially expressed between the normal and the tumor tissue samples. The first 400 probesets with the lowest t-test p-values (corresponding to a p-value cutoff of 9x10-12) were selected for further analysis. We also used a more stringent fold-change constraint that excluded the probesets with log-fold change < 1 (roughly corresponding to a fold change < 2), where the log-fold change of gene g between classes N ('normal') and T ('tumor') is defined as:

Log-fc (*g*, N-T) =log2 *g* (T) – log2 *g* (N),

Log2 *g*(C) = ∑si∑clog2g (Si)/|C|

The average log-expression value of gene g in the samples Si of class C.

Also, probe set lists of Ngn3-induced differentially expressed genes were produced for each differentiation context. The intersecting probe sets representing "context-independent" differentially expressed genes were systematically evaluated for significant enrichment of canonical signaling pathways.

Functional enrichment analyses

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov) and Concept Gen (http://conceptgen.ncibi.org) were used to identify over-represented biological functions and pathways among the differentially expressed genes.

Network analysis

A gene co-citation network of the differentially expressed genes was generated by using a sentence level co-citation filter. This network allows visualization of the differentially expressed genes and their potential associations with each other identified in the literature. The topology of the network was further analyzed using the Fast Greedy community structure identification algorithm, implemented in the Cytoscape plug-in GLay (http://brainarray.mbni.med.umich.edu/ sugang/glay) to identify coherent sub networks. Identified sub networks were subjected to functional enrichment analyses by DAVID to reveal over-represented biological functions within each subnetwork.

Results & Discussion:

The genes over expressed in pancreatic cancer whole tumor tissue contain a NGN3-beta signature

We performed an expression analysis of 16 pancreatic parental stem cells of 1) differentiated (2).undifferentiated (3). Genetically modied Ngn3 induced stem cells of series of time intervals (0, 3, and 10 days). Using microarray data analysis of RNA profiling on genesets was annotated using Affymetrix Mouse430 2.0 whole-genome chip. We generated datasets on epithelial cell lines that is capable of Dox-inducible; uniform expression of Ngn3 at different time points (0, 3days, 10days) after culture with the site-specific integration of Ngn3 mRNA into the HPRT1 locus of the Ainv15 ESC line was confirmed using a genome specific embryoid stem cells. Single-copy integration of murine Ngn3 datasets was calculated using R and BioConductor.

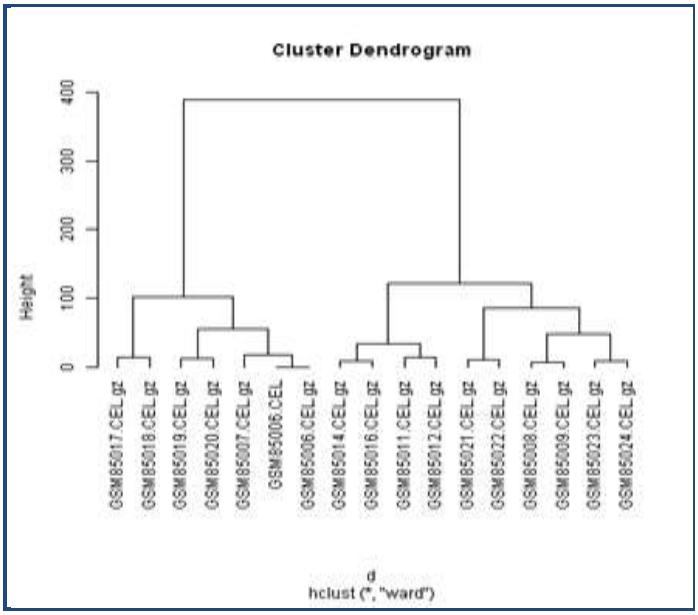

Figure 1: Hierarchical clustering of Grouped data based on gene expression.

The histological analysis confirmed that, the Ngn3 ESC cell lines was capable of forming tissues specific mRNA from Ngn3 cells. The mRNA samples were pooled, amplified and labeled before duplicate hybridizations were performed on spotted (pre-synthesized oligonucleotide) microarrays. The sample sets contains 45101 probe sets and it was genes in 16 datasets, all the probe sets were represented on GeneChip Mouse 430®.

The ngn3 mRNA hybridized data contains two outliers and one mislabeled sample is identified during the quality assessment process. The data sets along with the appropriate controls were clustered based on genome-wide gene expression in each treatment group. The mRNA probes were hybridized with simultaneous measurement of the relative expression levels of thousands of individual genes. For the RNA probes with parental embryoid stem cells and ngn3 induced embryoid stem cell processed shows more degradation is present, the 95%. For the degraded samples, the 95% confidence intervals exclude 1 for all but normal parental cell with high amount of degradation of Affymetrix Mouse430 2.0 when log2 GC% was used for adjusting probe affinity, suggesting more RNA quality, again as expected. Also note that the point estimates of the 3′/5′ ratios for all nine higher-quality GeneChip are much higher than the point estimates of their corresponding GeneChip with good quality, as expected.

Ngn3 regulation of gene expression in pancreatic cell line

Analysis of ngn3 gene expression results is identifying the most relevant pancreas developmental context led to microarray analysis of gene expression after induction of Ngn3 from EB7 to EB10 (Ngn3 EB10 ON). To evaluate Ngn3 expression in homogenous and pluripotent cellular context, using gene expression of induction of Ngn3 in undifferentiated ESCs either in conjunction with EB formation for 3 days (Ngn3 EB3 ON). The changes of gene expression were described based on representation changes between parental undifferentiated embryonic stem cells (n=2), differentiated embryoid bodies (n=2) embryoid bodies of parental stem cell (n=4), embryoid bodies of 3 days and 10 days differentiation (n=8) groups **Table 1 (see the supplementary material).**

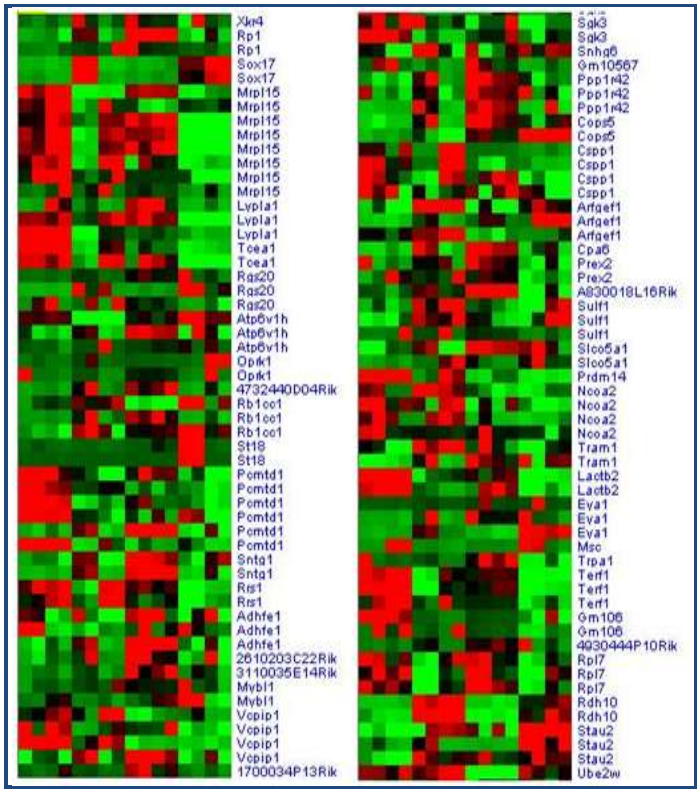

Figure 2: up regulated genes that led to the identification of similarities within the transcriptome of Ngn3-induced ESCs and uninduced EBs, indicating that Ngn3 can initiate a gene marker expression profile that is similar to EB formation profiles.

In probesets contains 45101 genes, using normalization techniques after hybridization shows ~32691 genes will express in all datasets, using clustering technique shows in murine stem cells shows less expression in during embryogenesis and induced ngn3 in stem cells shows more expression. Among them, 3755 genes were identified as significantly regulated by GCRMA of differentially expressed probes sets in induced ngn3 to genetically modified embryonic stem cells and 1282 genes were identified in GCRMA Normalized signal ratio of probe sets of regulated ESC on Tissue /cell lines data. Out of these 317 genes were significantly upregulated 175 genes

differentially down regulated in ngn3 induced ESCs **Table 2 (see supplementary material).**

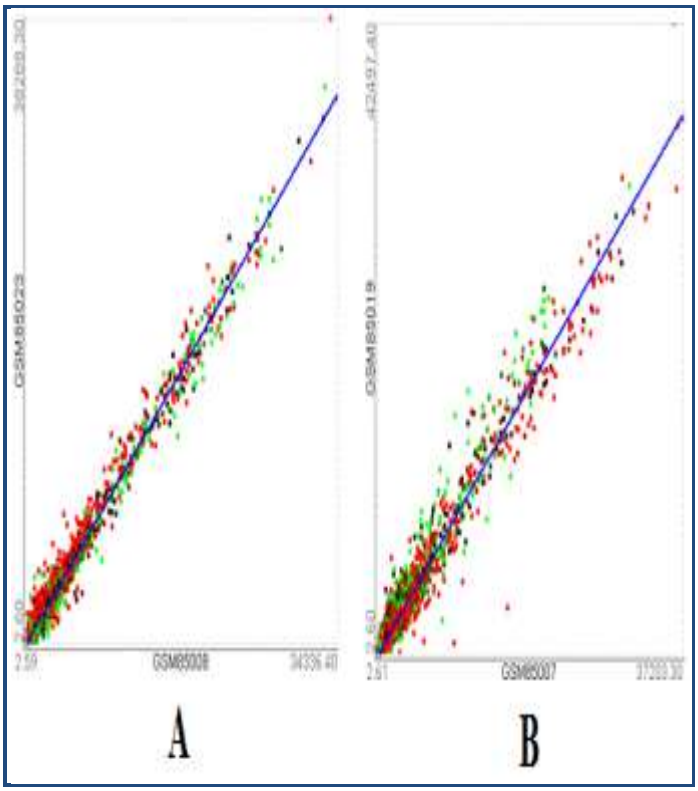

Figure 3: Scatter plot: Differential Gene expression with lower expression to higher expression. (II) GSE2276 dataset expression based on the Heat map. The rank list of all expressed datasets were predicted according to p values ($p<0.001$). The list of upregulated genes were predicted in **Table 1 (see supplementary material)**

In probe sets contains 45101 genes, using normalization techniques after hybridization shows ~32691 genes will express in all datasets, using clustering technique shows in murine stem cells shows less expression in during embryogenesis and induced ngn3 in stem cells shows more expression. Among them, 3755 genes were identified as significantly regulated by GCRMA of differentially expressed probes sets in induced ngn3 to genetically modified embryonic stem cells and 1282 genes were identified in GCRMA Normalized signal ratio of probe sets of regulated ESC on Tissue /cell lines data. Out of these 317 genes were significantly upregulated 175 genes differentially down regulated in ngn3 induced ESCs. **(Figure 1)**

Based on functional analysis of upregulated and down regulated genes includes *Pdx1cre, LSL-Kras^{G12D} PRRX1A* and *PRRX1B, SOX9 and CK19* are transient requirement during endocrine cell differentiation. The most relevant pancreas oriented ngn3 expression with endoderm genes Soc17 and FoxA2 and pancreatic endoderm and endocrine development Pdx1 and ngn3. Based on gene cluster, the clusters of 0days is unique in down regulation and 3 and 10 days datasets of functional characters were presented in **(Figure 2 & 3).**

The pancreas-specific gene Ptf1a, Sox17, FoxA2, Pdx1, and Ngn3 mRNA increased in early EB formation suggests that ngn3 induced ESC line helps for endoderm development **(Figure 4).**

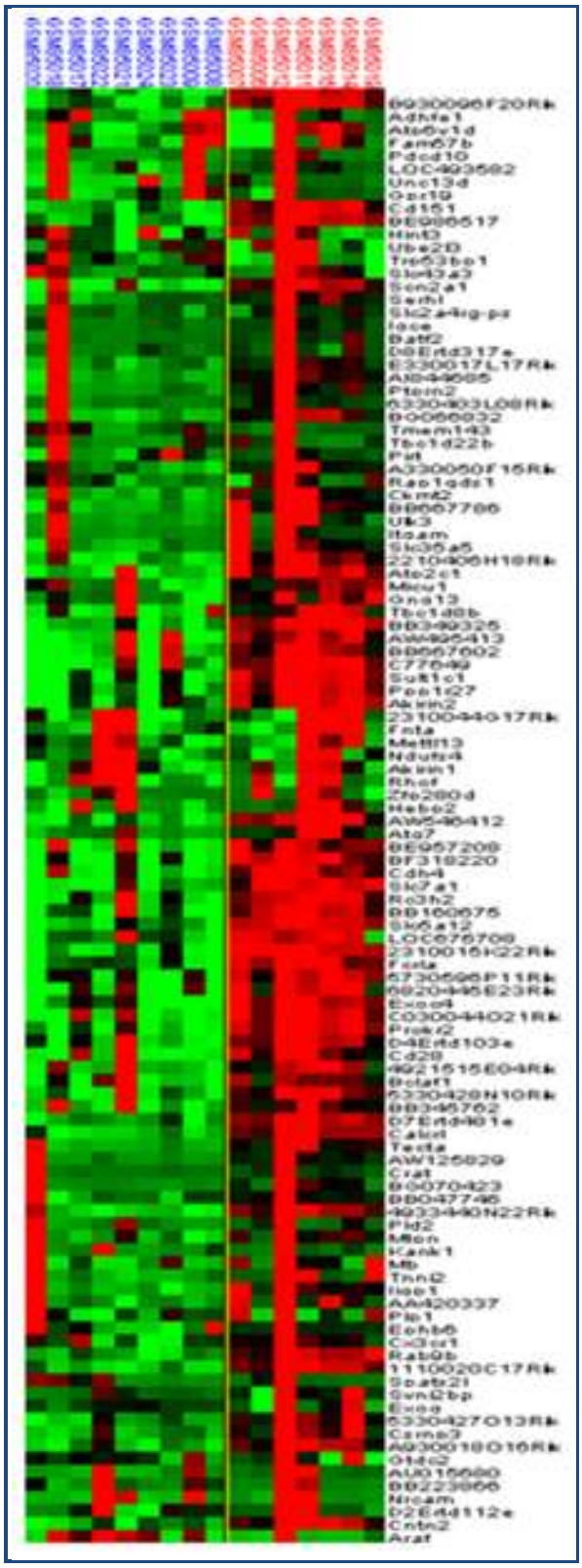

Figure 4: Upregulated and downregulated genes that led to the identification of similarities within the transcriptome of Ngn3-induced ESCs and uninduced EBs, indicating that Ngn3 can initiate a gene marker expression profile that is similar to EB formation profiles. The list of genes was predicted in **Table 2.**

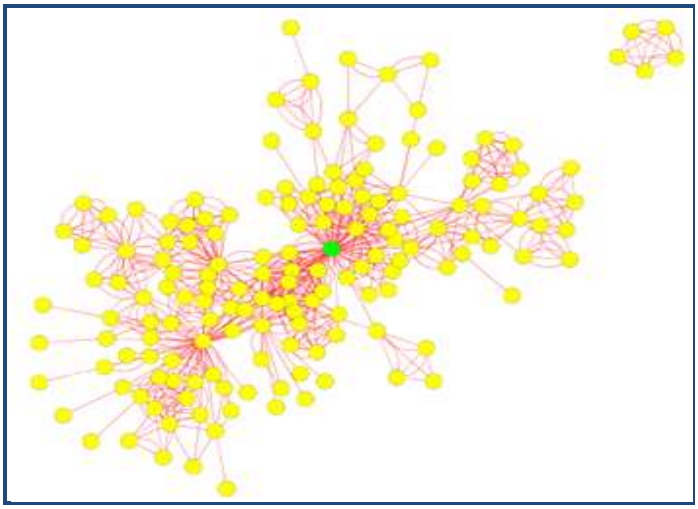

Figure: 5: Gene networks present in ngn3 induced pancreatic cancer development. The green spot represents ngn3 induction in regulation of pancreas development.

Functional enrichment

Functional enrichment analyses of the 1216 differentially expressed genes were performed to identify over-represented biological functions using Gene Ontology terms and pathways **Table 3 (see supplementary material).**

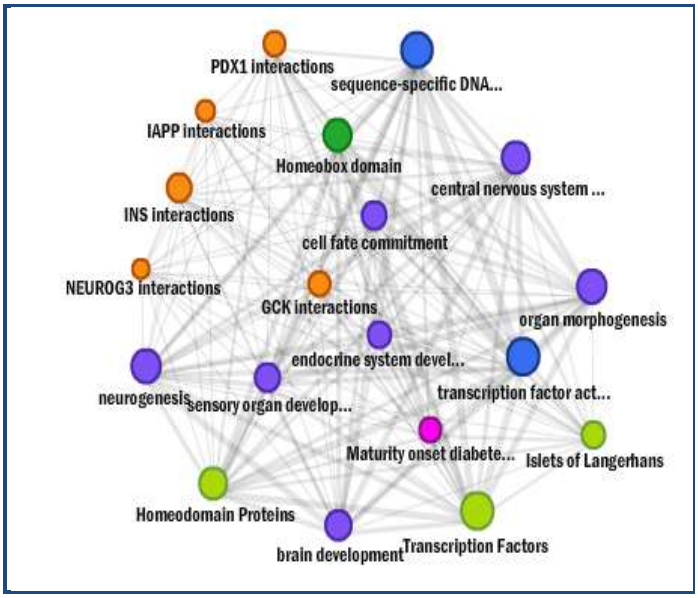

Figure 6: Functional networks involved in many notch signaling pathway and other associated networks.*Network*

Analysis of ngn3 induced signaling pathways

Based on literature and microarray techniques, we analyzed the genes present in gene networks of the differentially expressed genes was created by Cytoscape based on sentence level co-citations of differentially expressed genes to examine their potential relationships. This network is composed of subnetworks centered on the nine most connected genes **Table: 4 (see supplementary material).** The transcriptional factors Casq1,NeuroG3,Trp53bp1,PrkCi,4933426M11Rik, 33426M11Rik, Taf1, Znhit1, and A230108P19Rik. The complete network was further analyzed by Fast Greedy algorithm, implemented in the Cytoscape plug-in GLay, to cluster the genes into subgroups based on their network structure. 41 clusters with a minimum of 317 genes were identified **(Figure 5& 6)**

Conclusion:

We analyzed differential gene expression patterns present in ngn3 induced ESCs on neurogenesis and post translational modifications in different signaling cascades. Total of 3755 genes is functional gene expression, with these 317 genes shows positive up regulation and 175 genes negative down regulation in embryogenesis. 176 genes involved in notch signaling pathway helps to differentiate tissues in both exocrine and endocrine, our gene classification results shows that, the functionally enriched gene functions may express acini, ductal and β-cells. In functional network analysis shows 19 functional genes is used as regulator in notch signaling pathway and it is potentially used for drug targets.

References:

[1] D'Amour KA *et al. Nat Biotechnol.* 2006 **24:** 1392 [PMID: 17053790]
[2] Jonsson J *et al. Nature.* 1994 **371:** 606 [PMID: 7935793]
[3] Offield MF *et al. Development.* 1996 **122:** 983 [PMID: 8631275]
[4] Hanley NA *et al. Mech Dev.* 2000 **91:** 403 [PMID: 10704874]
[5] Fougerousse F *et al. Human Mol Genet.* 2000 **9:** 165 [PMID: 10607827]
[6] Fajans SS *et al. N Engl J Med.* 2001 **345:** 971 [PMID: 11575290]
[7] Liew CG *et al. PLoS ONE.* 2008 **3:** e1783 PMC2262135
[8] Jacquemin P *et al. Mol Cell Biol.* 2000 **20:** 4445 [PMID: 10825208]
[9] Kemp DM *et al. Rev Endocr Metab Dis.* 2003 **4:** 5 [PMID: 12618555]
[10] Cockell M *et al. Mol Cell Biol.* 1989 **9:** 2464 [PMID: 2788241]
[11] Teitelman G, *Tumor Biol.* 1993 **14**: 167 [PMID: 8210949]
[12] Arias AE & Bendayan M, *Lab Investig*. 1993 **69:** 518 [PMID: 8246444]
[13] Gradwohl G *et al. Proc Natl Acad Sci USA.* 2000 **97:** 1607 [PMID: 10677506]
[14] Artavanis-Tsakonas S *et al. Science.* 1999 **284:** 770 [PMID: 10221902]
[15] Bailey EC *et al. Ernst Schering Res Found Workshop.* 2000 *29***:** 211 [PMID: 10943312]
[16] Hackel PO *et al. Curr Opin Cell Biol.* 1999 **11:** 184 [PMID: 10209149]
[17] McMahon AP, *Cell.* 2000 **100:** 185 [PMID: 10660040]
[18] Massague J & Chen YG, *Genes & Dev.* 2000 **14:** 627 [PMID: 10733523]
[19] Edlund H, *Nat Rev Genet.* 2002 **3:** 524 [PMID: 12094230]
[20] Guz Y *et al. Development.* 1995 **121:** 11 [PMID: 7867492]

Edited by P Kangueane
Citation: **Nagaraja *et al.*** Bioinformation 9(14): 739-747 (2013)

BIOINFORMATION

open access

Supplementary material:

Table 1: List of functional genes present in embryogenesis in pancreas development.

Concept Name	Gene List Size	Overlap	P-Value	Q-Value
Maturity onset diabetes of the young	25	6	3.73196E-11	1.04E-08
Homeobox domain	206	4	2.42079E-05	2.54E-03
endocrine system development	34	4	2.98246E-07	4.70E-05
Transcription Factors	716	6	3.08718E-06	3.99E-04
Islets of Langerhans	44	4	3.82314E-07	5.92E-05
cell fate commitment	71	4	2.77867E-06	3.63E-04
IAPP interactions	9	3	3.61197E-06	4.61E-04
sequence-specific DNA binding	508	5	1.6611E-05	1.82E-03
GCK interactions	23	3	2.97318E-05	3.06E-03
neurogenesis	261	4	0.000164139	1.40E-02
organ morphogenesis	359	4	0.000432185	3.29E-02
sensory organ development	82	3	0.000569111	4.20E-02
Homeodomain Proteins	233	6	4.54058E-09	9.54E-07
transcription factor activity	928	5	0.000198552	1.66E-02
brain development	136	4	2.03887E-05	2.18E-03
INS interactions	66	3	0.000265868	2.14E-02
PDX1 interactions	20	3	2.20198E-05	2.34E-03
central nervous system development	232	4	0.000108894	9.71E-03

Table 2: Results of identifying significant ($p < 0.05$) signaling and metabolic pathways or network processes using systematic gene ontology analysis of 757 context-independent, neurogenin 3–regulated genes.

Gene Functions	P-Vales	% of Gene expression
Phosphoprotein	1.70E-15	43.4986
Acetylation	3.10E-13	19.45744
cytoplasm	8.24E-10	22.45089
Nucleus	5.57E-08	26.09916
Apoptosis	2.13E-05	3.648269
cytoskeleton	3.30E-05	5.332086
lipoprotein	4.42E-05	5.332086
DNA-binding	8.55E-05	10.38354
metal-binding	8.65E-05	17.86717
transferase	1.83E-04	10.1029
zinc	2.46E-04	13.00281
Actin-binding	2.98E-04	2.525725
zinc-finger	3.21E-04	8.88681
heparin-binding	5.39E-04	1.028999
coated pit	9.89E-04	0.748363
serine/threonine-specific protein kinase	0.001289	0.841908
nucleotide-binding	0.001512	11.03835
cell division	0.001909	2.525725
activator	0.002286	4.022451
neurogenesis	0.002378	1.683817
cell cycle	0.002975	3.741815
repressor	0.003161	3.367633
amino-acid biosynthesis	0.004824	0.561272
pyridoxal phosphate	0.005203	0.935454
kinetochore	0.005814	0.935454
kinase	0.005816	5.238541
duplication	0.006144	1.403181
transcription regulation	0.007115	10.1029
Proto-oncogene	0.007316	1.122544
GPI-anchor	0.008165	1.403181
ATP-binding	0.010105	8.512629
selenocysteine	0.010501	0.561272
triple helix	0.010821	0.374181
alternative splicing	0.014096	25.91207

magnesium	0.014491	3.180543
sh3 domain	0.017303	1.870907
DNA replication	0.017911	1.028999
LIM domain	0.018194	0.935454
muscle protein	0.018758	0.748363
phosphotransferase	0.018781	1.309635
selenium	0.019594	0.561272
angiogenesis	0.019659	0.841908
lipid transport	0.020745	0.748363
DNA damage	0.02084	1.777362
hydroxylation	0.023398	0.841908
annexin	0.025055	0.374181
extracellular matrix	0.025844	1.870907
DNA repair	0.02894	1.590271
acetylated amino end	0.030019	0.467727
calcium/phospholipids-binding	0.031268	0.374181
methylation	0.035828	1.870907
Rotamase	0.036699	0.561272
Aminotransferase	0.040008	0.467727
Mitochondrion	0.042119	5.238541
endoplasmic reticulum	0.042451	4.583723
mitochondrion inner membrane	0.044907	1.590271
coiled coil	0.046058	10.47708
cell adhesion	0.046428	2.806361
serine/Threonine-protein kinase	0.051478	2.806361
tumor suppressor	0.052199	0.935454
heparan sulfate	0.054209	0.374181
prenylation	0.056133	1.309635
SH2 domain	0.056573	1.028999
Isomerase	0.059846	1.122544
steroid metabolism	0.060826	0.561272
translocation	0.065026	0.748363
dimer	0.067722	0.280636
lipid degradation	0.06798	0.841908
Endocytosis	0.069319	0.935454
differentiation	0.071056	3.180543
ATP	0.072527	1.309635
Transcription	0.076147	10.47708
cholesterol metabolism	0.079891	0.467727
hydroxyproline	0.083979	0.280636
developmental protein	0.084063	5.144995
blocked amino end	0.08793	0.467727
collagen	0.089974	0.841908
nucleotidyltransferase	0.099791	0.748363

Table 3: Results of identifying significant (p<0 .05) signaling and metabolic pathways or network processes using systematic gene ontology analysis of 919 context-independent, neurogenin down regulated genes.

Term	Count	% of Identity	P.Value
PDZ/DHR/GLGF	18	1.683816651	0.001882629
Calponin-like actin-binding	11	1.028999065	0.002665983
Glutathione S-transferase, C-terminal-like	7	0.654817587	0.004037648
Glutathione S-transferase/chloride channel, C-terminal	7	0.654817587	0.007660217
Protein kinase, ATP binding site	38	3.554724041	0.007761197
Amyloidogenic glycoprotein	3	0.280636109	0.007895454
Amyloidogenic glycoprotein, extracellular domain conserved site	3	0.280636109	0.007895454
Amyloidogenic glycoprotein, copper-binding	3	0.280636109	0.007895454
Amyloidogenic glycoprotein, intracellular domain, conserved site	3	0.280636109	0.007895454
Amyloidogenic glycoprotein, heparin-binding	3	0.280636109	0.007895454
Amyloidogenic glycoprotein, extracellular	3	0.280636109	0.007895454
Beta-amyloid precursor protein C-terminal	3	0.280636109	0.007895454
EMI domain	5	0.467726848	0.010202164

Glutathione S-transferase, N-terminal	6	0.561272217	0.011904405
Glutathione S-transferase, C-terminal	6	0.561272217	0.013883613
Zinc finger, LIM-type	10	0.935453695	0.014670166
Zinc finger, PHD-type, conserved site	11	1.028999065	0.015140516
Tropomyosin	3	0.280636109	0.015244498
Protein kinase, core	38	3.554724041	0.016110226
Protein kinase, C-terminal	7	0.654817587	0.016787282
Fibrillar collagen, C-terminal	4	0.374181478	0.017106048
Netrin domain	5	0.467726848	0.018344362
bZIP transcription factor, bZIP-1	6	0.561272217	0.01847949
Pyridoxal phosphate-dependent transferase, major region, subdomain 1	7	0.654817587	0.018821884
AGC-kinase, C-terminal	8	0.748362956	0.019683656
Zinc finger, PHD-type	11	1.028999065	0.020216917
SNF2-related	6	0.561272217	0.02111221
Protein kinase C-like, phorbol ester/diacylglycerol binding	9	0.841908326	0.021176935
Basic-leucine zipper (bZIP) transcription factor	8	0.748362956	0.021629709
Ras small GTPase, Rho type	5	0.467726848	0.021746035
Annexin repeat, conserved site	4	0.374181478	0.021936734
Annexin repeat	4	0.374181478	0.021936734
Annexin	4	0.374181478	0.021936734
Zinc finger, PHD-finger	10	0.935453695	0.025140605
Src homology-3 domain	19	1.777362021	0.028198906
SH2 motif	12	1.122544434	0.030462769
Netrin module, non-TIMP type	4	0.374181478	0.033586876
Tudor subgroup	4	0.374181478	0.033586876
Ubiquitin	7	0.654817587	0.034597556
Fibronectin, type III-like fold	17	1.590271282	0.036151053
Ras GTPase	13	1.216089804	0.039473447
Low-density lipoprotein receptor, YWTD repeat	4	0.374181478	0.040393907
Fibronectin, type III	17	1.590271282	0.042559704
Cellular retinaldehyde-binding/triple function, C-terminal	5	0.467726848	0.044266699
Ubiquitin supergroup	7	0.654817587	0.044958162
Zinc finger, RING-type	23	2.151543499	0.04500814
Six-bladed beta-propeller, TolB-like	6	0.561272217	0.046347968
Glutathione S-transferase, Mu class	3	0.280636109	0.048053225
HR1-like rho-binding repeat	3	0.280636109	0.048053225
Growth-arrest-specific protein 2	3	0.280636109	0.048053225
Ras	12	1.122544434	0.05181024
Serine/threonine protein kinase-related	28	2.619270346	0.055723923
Pleckstrin homology	21	1.96445276	0.055774381
Serine/threonine protein kinase, active site	27	2.525724977	0.059441529
Ephrin, conserved site	3	0.280636109	0.061894357
Ephrin	3	0.280636109	0.061894357
Paraneoplastic encephalomyelitis antigen	3	0.280636109	0.061894357
Tyrosine protein kinase, active site	10	0.935453695	0.081493379
Peptidyl-prolyl cis-trans isomerase, FKBP-type	4	0.374181478	0.083593957
Small GTP-binding protein	14	1.309635173	0.088378432
Low density lipoprotein-receptor, class A, cysteine-rich	6	0.561272217	0.090562934
K Homology, type 1, subgroup	5	0.467726848	0.091335251
Sterile alpha motif/pointed	3	0.280636109	0.092875643
Follistatin-like, N-terminal	3	0.280636109	0.092875643
Zinc finger, CXXC-type	3	0.280636109	0.092875643
Myb-type HTH DNA-binding domain	3	0.280636109	0.092875643

Table 4: Functional Enriched 317 genes 175 over-represented biological functions among the up and down regulated differentially expressed genes (DAVID P<50.05). The table lists a subset of over-represented biological functions regulated in the induced ngn3 in pancreatic cells.

Concept Name	Concept Type Name	Gene List Size	P-Value	Q-Value
Maturity onset diabetes of the young	KEGG Pathway	25	3.73196E-11	1.04E-08
Homeobox domain	pFAM	206	2.42079E-05	2.54E-03
endocrine system development	GO Biological Process	34	2.98246E-07	4.70E-05

Transcription Factors	MeSH	716	3.08718E-06	3.99E-04
Islets of Langerhans	MeSH	44	3.82314E-07	5.92E-05
cell fate commitment	GO Biological Process	71	2.77867E-06	3.63E-04
IAPP interactions	Protein Interaction (MiMI)	9	3.61197E-06	4.61E-04
sequence-specific DNA binding	GO Molecular Function	508	1.6611E-05	1.82E-03
GCK interactions	Protein Interaction (MiMI)	23	2.97318E-05	3.06E-03
neurogenesis	GO Biological Process	261	0.000164139	1.40E-02
organ morphogenesis	GO Biological Process	359	0.000432185	3.29E-02
sensory organ development	GO Biological Process	82	0.000569111	4.20E-02
Homeodomain Proteins	MeSH	233	4.54058E-09	9.54E-07
transcription factor activity	GO Molecular Function	928	0.000198552	1.66E-02
brain development	GO Biological Process	136	2.03887E-05	2.18E-03
INS interactions	Protein Interaction (MiMI)	66	0.000265868	2.14E-02
PDX1 interactions	Protein Interaction (MiMI)	20	2.20198E-05	2.34E-03
central nervous system development	GO Biological Process	232	0.000108894	9.71E-03