GENOME-WIDE ANALYSIS OF THE NEUROD FAMILY OF BHLH TRANSCRIPTION FACTORS IN MAMMALS

Shouhartha Choudhury

Department of Biotechnology, Assam University, Silchar-788011, Assam, India

Abstract

The NeuroD (Beta2) family of helix-loop-helix transcription factor promotes neurogenesis. In NeuroD family, particular gene express in the specific parts of the neurons, beta-pancreatic cells, and enteroendocrine cells. The NeuroD genes induced in the central nervous system and development of the organs. Therefore, the investigation of the NeuroD family is mandatory to explore the functional regulation of particular genes of interest. The query gene (NeuroD1) found in glial cells and functional neurons in the brain. The NeuroD1 induced in the nervous system and regulate insulin by binding to a critical E-box motif on promoter and mutation result in Type 2 Diabetes. The query gene found in chromosome 2 of the human. I accumulate mammalian genome i.e. Homo sapiens and Mus musculus for comparative and functional analysis. My finding data provide documentation of the NeuroD family and their specific domain,motifs, phylogeny and the chromosome location and expression of the query gene. The genome-wide analysis of the NeuroD family of a transcription factor is an essential component for the identification of the specific neurogenic transcription factor in the genome. In this study, bioinformatics and computational techniques to identify the NeuroD family of bHLH transcription factor in the genome.

Keyword: NEUROD1; NeuroD Family; bHLH Transcription Factor; Neurogenesis; and Type-2 Diabetes

1.INTRODUCTION

The transcription factor is regulatory protein bind to the specific DNA sequence **[1, 2].** The functions of the transcription factor regulate turn on or off genes in a cell. Transcription factors are groups of proteins read and interpret genetic "blueprint" in DNA. Particularly transcription factor bind with DNA and initiate a program of increase or decrease gene transcription. By turning gene transcription on or off in a cell, transcription factors play major roles in the development and disease response. Transcription factors coordinate cell division, cell growth, and cell death of eukaryotes. Almost twenty-six thousand proteins in the human genome contain DNA-binding domains most are presumed function as transcription factors **[3-5]**. During embryo development, many novel transcription factors emerged and contributing complex ontogenesis and adaption provide an intriguing case to investigate, how transcription factors contribute to a major response in animal development **[6].** The identification of numerous validated transcription factor genes across the vast scientific literature concerning studies on "*Mus musculus*" model widely used and propose major responsibility to build high-confidence transcription factor data. In this study, conducted a compressive genome-wide survey of the NeuroD family of the bHLH transcription factor in *Homo sapiens* and *Mus musculus,* and reveal the heterogeneity of the neurogenic transcription factor and wiring preference of specific transcription factor. These results suggested a probable mechanism of the contribution of transcription factors in eukaryote organisms.

The bHLH (basic helix-loop-helix) is the largest transcription factors contain protein structural motif is characterized by two alpha-helices connected with a loop, bHLH domain dimeric each helix containing amino-acid that bind to the DNA. The bHLH TFs may be homo or heterodimerize with specific functions are conserved and characterizes largest transcription factors in eukaryotes. The bHLH transcription factor contains certain amino acid and two amphipathic alpha-helices

separated from a linker region of the length. The peptide sequence has specific motifs they ability to bind DNA sequence contains bHLH domain **[7-10]**.

In this study, reviewed the NeuroD family of bHLH transcription factors are responsible for neurogenesis in multicellular organisms. The NeuroD family is a basic helix-loop-helix transcription factor express in specific part in the neuron, beta-pancreatic cells, and enteroendocrine cells. Particularly, the neurogenic transcription factor involves the differentiation of the central nervous system and the development of the organisms. Inconsistent region of the adult central nervous system has disparate amounts of NeuroD transcription factor present. One of them, the mutation of NeuroD1 associated with a monogenic form of diabetes of the young with respect **[11].** NeuroD1 is found in glial cells into functional neurons and regulate the expression of insulin. Mutation of insulin results in Types 2 Diabetes. However, NeuroD1 expresses at embryo and persist in the adult central nervous system and potentially activate similar target genes composed of multimerized E-box **[12, 13]**.

2.MATERIALS AND METHODS

2.1. Primary sequence and database

The primary sequence retrieved from different specialization database (UniProt, EMBL, KEGG, GenBank, and NCBI), and performed web base application SMART for identification of a specific domain in the query sequence. Pfam searched for retrieving protein family information. PROSITE performed for the identification of domain, family, and functional sites and associated pattern and profile. PROCHECK perform for the stereochemical property of primary peptide sequence. The genome sequences were downloaded from genomic data in different specialized databases (NCBI and Ensemble).

2.2. Standalone tools and GO annotation

HMMER executed using multiple sequence alignments of the specific domain as a profile search. HMMER is a statistical algorithm, making multiple sequence alignment (MSA) of the specific domain as a profile search, an implement methods using probabilistic models called the profile is hidden Markov model. Standalone BLAST performed for homologs gene in selected organisms. The BLAST2GO performed for the gene ontology annotation of the particular gene. BLAST2GO is bioinformatics and computational tool for high-throughput gene ontology annotation of the novel sequence data. The functional information retrieves via Gene Ontology (GO) annotation a controlled vocabulary of the functional attribute.

2.3. Domain, motif, and phylogeny

Multiple sequence alignment (MSA) methods to calculate the best match of the homologs sequences and line them up so identities, similarities, and differences can be seen. MSA of highest hits sequence analysis carried out by web-based tool MultAlin for identification of conserved domain. The identification of the molecular evolutionary relationship between *Homo sapiens* and *Mus musculus*, MEGA7 performed for constructing a phylogenetic tree using *Neighbor-Joining Methods*. The MEME suite performed for the sequence motifs is a computational web-based tool for discovery and analysis of specific motifs

2.4. Gene expression and chromosome location

The gene expression analysis carried out using the GENEVESTIGATOR tool is a high-performance search engine of gene expression in different biological contexts.GENEVESTIGATOR use to identify and characterize validate novel target and biomarkers. Chromosome location retrieves using gene card is a database of the human genes provide genomic information of all known and predicted human genes. This database is currently available for biomedical information such as gene, encoded protein, and relevant disease.

3.RESULTS

3.1. Identification of the primary and secondary structure

The primary structure demonstrated the composition of the nucleotide and peptide. The total of 1071 nucleotides and 356 peptides (**Table: 1)** with 53 peptides bind to the specific DNA sequence called bHLH domain. The bHLH domain consists of 53 peptides

characterized by two alpha-helix connected by a loop; the adaptability of the loop allows dimerization of folding against another helix **(Fig. 1)**.

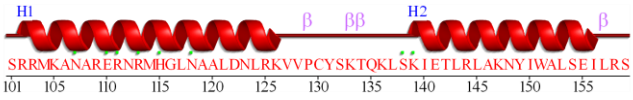

Fig.1

Table 1:Primary sequence

>NeuroD1
Atgaccaaatcgtacagcgagagtgggctgatgggcgagcctcagccc
caaggtcctccaagctggacagacgagtgtctcagttctcaggacgagg
agcacgaggcagacaagaaggaggacgacctcgaagccatgaacgca
gaggaggactcactgaggaacgggggagaggaggaggacgaagat
gaggacctggaagaggaggaagaagaggaagaggaggatgacgatc
aaaagcccaagagacgcggccccaaaaagaagaagatgactaaggctc
gcctggagcgttttaaattgagacgcatgaaggctaacgcccgggagcg
gaaccgcatgcacggactgaacgcggcgctagacaacctgcgcaaggt
ggtgccttgctattctaagacgcagaagctgtccaaaatcgagactctgcg
cttggccaagaactacatctgggctctgtcggagatcctgcgctcaggca
aaagcccagacctggtctccttcgttcagacgctttgcaagggcttatccca
acccaccaccaacctggttgcgggctgcctgcaactcaatcctcggacttttt
ctgcctgagcagaaccaggacatgccccccacctgccgacggccagcg
cttccttccctgtacacccctactcctaccagtcgcctgggctgcccagtccg
ccttacggtaccatggacagctcccatgtcttccacgttaagcctccgccgc
acgcctacagcgcagcgctggagcccttctttgaaagccctctgactgatt
gcaccagcccttcctttgatggacccctcagcccgccgctcagcatcaatg
gcaacttctctttcaaacacgaaccgtccgccgagtttgagaaaaattatgc
ctttaccatgcactatcctgcagcgacactggcaggggcccaaagccacg
gatcaatcttctcaggcaccgctgcccctcgctgcgagatccccatagaca
atattatgtccttcgatagccattcacatcatgagcgagtcatgagtgccca
gctcaatgccatatttcatgattag

(a) Nucleotide

>NeuroD1
MTKSYSESGLMGEPQPQGPPSWTDECLSSQDEEHEADKKE
DDLEAMNAEEDSLRNGGEEEDEDEDLEEEEEEEEDDDQKP
KRRGPKKKKMTKARLERFKLRRMKANARERNRMHGLNAA
LDNLRKVVPCYSKTQKLSKIETLRLAKNYIWALSEILRSGKSP
DLVSFVQTLCKGLSQPTTNLVAGCLQLNPRTFLPEQNQDM
PPHLPTASASFPVHPYSYQSPGLPSPPYGTMDSSHVFHVKP
PPHAYSAALEPFFESPLTDCTSPSFDGPLSPPLSINGNFSFKH
EPSAEFEKNYAFTMHYPAATLAGAQSHGSIFSGTAAPRCEIP
IDNIMSFDSHSHHERVMSAQLNAIFHD

(b) Peptide

3.2. Genome-wide identification and gene ontology annotation

Genome-wide identification of both organisms performed using standalone tools. The HMMER results obtain a total of 82, 75 bHLH domains in *Homo sapiens*and *Mus musculus* respectively. The standalone BLAST results represent a total 45, 43 homologs in *Homo sapiens* and *Mus musculus* respectively. The gene ontology annotation demonstrated the sequence accuracy of the basic helix-loop-helix transcription factor in both organisms (**Table: 3**).The principle of transcription factor data analysis suggested multiple hits of 8 NEUROD and 75, 69 specific bHLH domain in *Homo sapiens* and *Mus musculus* respectively (**Table 2**).

Table 2:Summary of the bHLH transcription factors

Gene	*Homo sapiens*	*Mus musculus*
NEUROD1	1	1
NEUROD2	1	1
NEUROD4	1	1
NEUROD6	1	1
NEUROG1	1	1
NEUROG2	1	1
NEUROG3	1	2
ATOH1	1	1
ATOH7	1	1
ATOH8	1	1
BHLHA15	2	1
BHLHE23	2	1
BHLHE22	1	1
FER3	1	1
OLIG1	1	1
OLIG2	2	1
OLIG3	1	1
SCX	1	1

PTF1A	1	1
TAL1	2	3
TAL2	1	1
TWIST1	2	1
TWIST2	2	2
NHLH1	1	1
NHLH2	2	3
TCF15	1	1
TCF21	2	2
TCF23	1	1
TCF24	1	1
ASCL1	1	1
ASCL2	1	2
ASCL3	1	1
ASCL4	1	1
ASCL5	2	1
HAND1	2	2
HAND2	1	1
LYL1	1	1
MSC	1	1
MESP1	1	1
MESP2	1	1
FIGLA	1	1
TFAP4	1	1
MYOD1	1	1
MYF5	1	1
MYF6	1	1
ARNT	6	4
MYOG	1	1
ID1	2	2
ID2	3	0
ID3	2	1
ID4	1	1
MYCN	1	3
BHLHA9	1	1
MSGN1	1	1
USF3	0	1
KIAA2018	1	0
Total	**74**	**69**

3.3. Identification of the domain, motifs, and phylogeny:

The NEUROD family of transcription factors selected from both organisms for multiple sequence alignment (MSA). The high consensus sequence indicated conserved extended basic helix-loop-helix domain (**Fig. 2**) and their specific motifs (**Fig. 3**). The phylogenetic tree demonstrated the molecular evolutionary relationship between *Homo sapiens* and *Mus musculus*. Particular clade defines multifunctional bHLH domain involved in both organisms genome (**Fig. 4**).

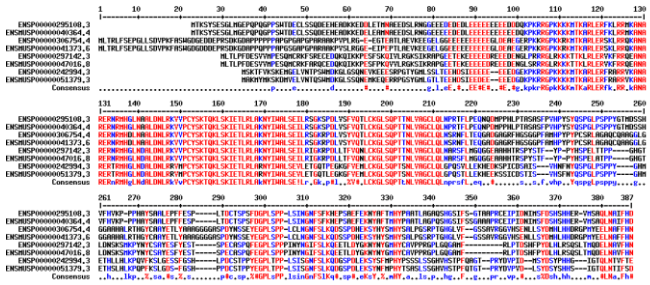

Fig. 2

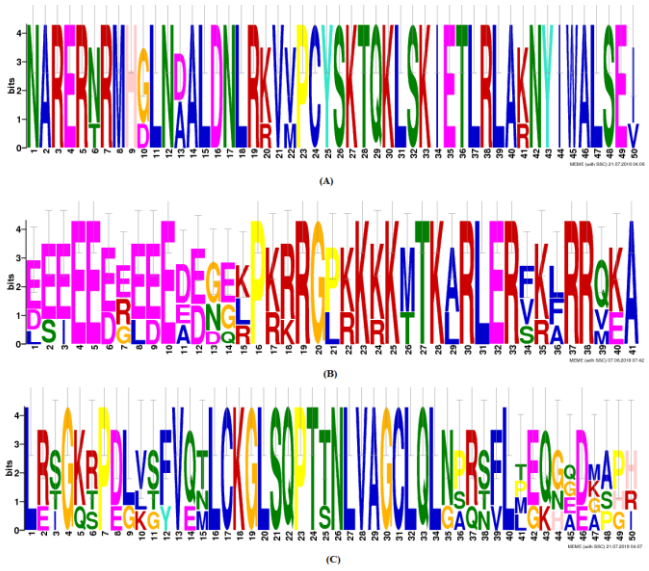

Fig. 3

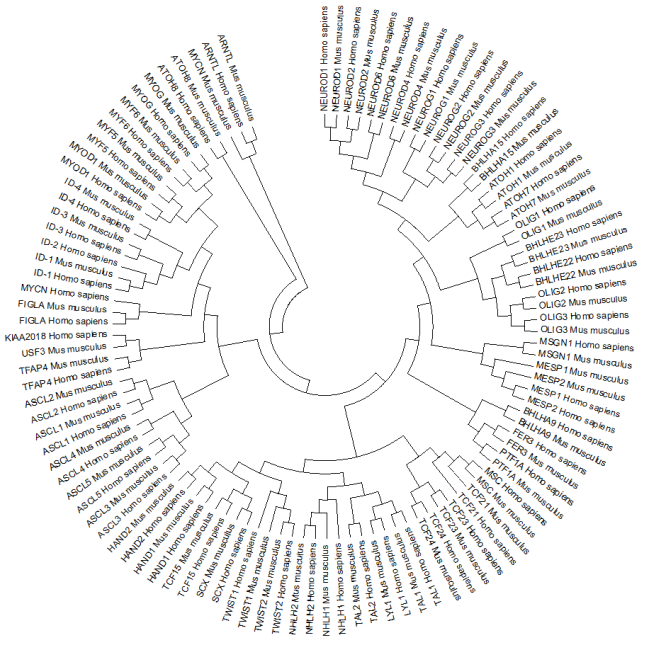

Fig. 4

3.4. Gene expression and chromosome location:

The gene expression analysis demonstrated that the query gene expresses in the tissue, nervous system, peripheral nervous system (PNS), and sural nerve of the human (**Fig. 5**). The chromosome localization studies confirm that the NeuroD1 located in band 2q31.3 of the human (**Fig. 6**).

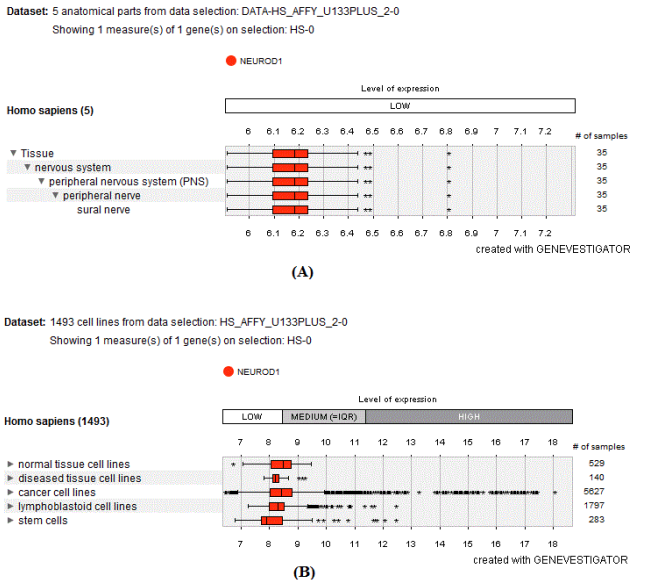

Fig. 5

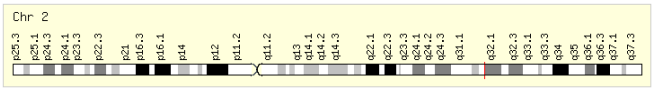

Fig. 6

My finding data suggested that the NeuroD family of transcription factors present in mammals. I built high confidence NeuroD family of transcription factor data, such as the *conserved domain, motifs, phylogeny, chromosome location*, and *gene expression*. Therefore, the NeuroD family of basic helix-loop-helix transcription factor study is essential for a better understanding of the neurogenic transcription factor in a particular organism.

4.DISCUSSION

The bHLH transcription factors govern cell types during development. The NeuroD family of basic helix-loop-helix transcription factor express transiently in a subset of neurons and central and peripheral nervous system at the time of terminal differentiation. The ectopic expression of the NeuroD in embryo causes the conversion of epithelial cells and neurons. The NeuroD1 is identical with hamster beta 2 genes is a regulator of insulin gene transcription. A predicted 357 amino acids polypeptide shares 97% identity of the bHLH region. The NeuroD homologs are widely expressed during development of mammalian brain and pancreas.

NeuroD1 has abnormal pancreatic islet ontogenesis and overt diabetes due to an inadequate expression of the insulin gene. The deletion of NeuroD failed to develop the granule cell layer in the dentate gyrus, one of the principals of the hippocampal formation. The different markers were found in the cell population of the dentate gyrus appeared normal. However, the dramatic defects in the proliferation of precursor cells reach dentate and significantly differentiate of granule cell. This process and assessment led malformation in dentate granule cell and excess cell death. The limbic seizure is associated with seizure activity in hippocampus and cortex **[14-17]**. During neurogenesis positive and negative regulations of bHLH domain are essential for the development of the organism. The identification of NEUROD1 deficient binding to the polypeptide for target promoter in pancreatic islet leads to the development of Type 2 Diabetes of the human. Two mutations in NeuroD1 are associated with the development of Type 2 Diabetes in the heterozygous state. The first missense mutation at Arginine111 in the DNA abolishes E-box binding activity. The second mutation rise polypeptide lacking carboxy-terminal transactivation domain region associated with the co-activators CBP and p300. NeuroD1 was more severe and suggestive with arginine111 to leucine mutation was typical of Type 2 Diabetes. Besides, the NeuroD2 initially expresses at an embryo and persist its expression in the adult nervous system and appear to mediate neuron. NeuroD2 is neuron-specific transcription factor can induce neural differentiation in undifferentiated cells and involved in neurogenesis and neuroblastoma cell line. NeuroD2 alter brain organization and affects brain size slightly smaller and rounder hippocampus and absence of corpus callosum. NeuroD2 segregates somatosensory cortex and postsynaptic barrel organization and reduces total excitatory synaptic currents layer due to the reduced contribution of AMPA receptors compared with NMDA receptors **[18, 19]**. The human NeuroD4 shares 88.5% amino acid identity with 100% identity of the bHLH region initially express throughout developing nervous system and gradually restricted in the neural retina. A functional analysis identified a proximal region of Math3 promoter for neuron-specific expression and upstream region for retinal expression. NeuroD4 actively participate in specific motor neuron subtype in the embryonic spinal cord and stem cells **[20-22]**. Particularly, the NeuroD6 induced in neuronal differentiation during brain development. The sequence analysis defines NeuroD6 is called MATH2. The deduced 337 amino acids contain N-terminal region in glutamic acid followed by a bHLH domain. NeuroD6 shares 98% identity with Math2 ortholog and 100% identity of bHLH domain share 95% identity with NeuroD1 and NeuroD2 **[23-27]**. Therefore, my finding data suggested that the NeuroD family is associated with neurogenesis in mammals. The genome-wide study of the NeuroD family of a bHLH transcription factor is an essential component for a better understanding of the neurogenesis in a multicellular organism.

5.CONCLUSION

My finding data demonstrated that the NeuroD family of transcription factors associated with neurogenesis in mammals. In this report, I documented several bHLH transcription factors in *Homo-sapiens* and *Mus musculus*. In contrast, neurogenic transcription factor regulates neural development and the special regulation of NeuroD1 mutation in Type-2 Diabetes of the young with respect. Therefore, genome-wide study of the species-specific transcription factors is essential for clinical research and development.

6.COMPETING FINANCIAL INTEREST'S STATEMENT

The author did not avail of any financial assistance from any source in undertaking the present study. The author stated there is no dispute of interests concerning the publication of this paper.

7.ACKNOWLEDGMENT

The author grateful to Assam University, Silchar, Assam, India for providing the requisite lab facilities in carrying out this research work and also thankful to the editor and anonymous reviewer for valuable comments and suggestions.

REFERENCES

[1] Latchman, D.S., *Transcription factors: an overview*. The international journal of biochemistry & cell biology, 1997. **29**(12): p. 1305-1312. PMID: 9570129

[2] Karin, M., *Too many transcription factors: positive and negative interactions*. The New biologist, 1990. **2**(2): p. 126-131. PMID: 2128034

[3] Mitchell, P.J. and R. Tjian, *Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins*. Science, 1989. **245**(4916): p. 371-378. PMID: 2667136

[4] Ptashne, M. and A. Gann, *Transcriptional activation by recruitment*. Nature, 1997. **386**(6625): p. 569. PMID: 9121580

[5] Babu, M.M., et al., *Structure and evolution of transcriptional regulatory networks*. Current opinion in structural biology, 2004. **14**(3): p. 283-291. PMID: 15193307

[6] Brivanlou, A.H. and J.E. Darnell, *Signal transduction and the control of gene expression*. Science, 2002. **295**(5556): p. 813-818. PMID: 11823631

[7] Murre, C., et al., *Structure and function of helix-loop-helix proteins*. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 1994. **1218**(2): p. 129-135. PMID: 8018712

[8] Massari, M.E. and C. Murre, *Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms*. Molecular and cellular biology, 2000. **20**(2): p. 429-440. PMID: 10611221

[9] Amoutzias, G.D., et al., *Choose your partners: dimerization in eukaryotic transcription factors*. Trends in biochemical sciences, 2008. **33**(5): p. 220-229. PMID: 18406148

[10] Chaudhary, J. and M.K. Skinner, *Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells*. Molecular endocrinology, 1999. **13**(5): p. 774-786. PMID: 10319327

[11] Horikawa, Y., et al., *beta-cell transcription factors and diabetes: no evidence for diabetes-associated mutations in the gene encoding the basic helix-loop-helix transcription factor neurogenic differentiation 4 (NEUROD4) in Japanese patients with MODY*. Diabetes, 2000. **49**(11): p. 1955-1957. PMID: 11078465

[12] Poulin, G., B. Turgeon, and J. Drouin, *NeuroD1/beta2 contributes to cell-specific transcription of the proopiomelanocortin gene*. Molecular and cellular biology, 1997. **17**(11): p. 6673-6682. PMID: 9343431

[13] Guo, Z., et al., *In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model*. Cell stem cell, 2014. **14**(2): p. 188-202. PMID: 24360883

[14] Lee, J.E., et al., *Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein*. Science, 1995. **268**(5212): p. 836-844. PMID: 7754368

[15] Tamimi, R., et al., *TheNEURODGene Maps to Human Chromosome 2q32 and Mouse Chromosome 2*. Genomics, 1996. **34**(3): p. 418-421. PMID: 8786144

[16] Naya, F.J., C. Stellrecht, and M.-J. Tsai, *Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor*. Genes & development, 1995. **9**(8): p. 1009-1019. PMID: 7774807

[17] Liu, M., et al., *Loss of BETA2/NeuroD leads to malformation of the dentate gyrus and epilepsy*. Proceedings of the National Academy of Sciences, 2000. **97**(2): p. 865-870. PMID: 10639171

[18] McCormick, M.B., et al., *NeuroD2 and neuroD3: distinct expression patterns and transcriptional activation potentials within the neuroD gene family*. Molecular and cellular biology, 1996. **16**(10): p. 5792-5800. PMID: 8816493

[19] Ince-Dunn, G., et al., *Regulation of thalamocortical patterning and synaptic maturation by NeuroD2*. Neuron, 2006. **49**(5): p. 683-695. PMID: 16504944

[20] Tsuda, H., et al., *Structure and Promoter Analysis of Math3 Gene, a Mouse Homolog of Drosophila Proneural Geneatonal NEURAL-SPECIFIC EXPRESSION BY DUAL PROMOTER ELEMENTS*. Journal of Biological Chemistry, 1998. **273**(11): p. 6327-6333. PMID: 9497361

[21] Hinokio, Y., et al., *Beta-cell transcription factors and diabetes: no evidence for diabetes-associated mutations in the hepatocyte nuclear factor-3beta gene (HNF3B) in Japanese patients with maturity-onset diabetes of the young*. Diabetes, 2000. **49**(2): p. 302-305. PMID: 10868948

[22] Lee, S.-K. and S.L. Pfaff, *Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors*. Neuron, 2003. **38**(5): p. 731-745. PMID: 12797958

[23] Guo, L., et al., *Cloning, chromosome localization and features of a novel human gene, MATH2*. Journal of genetics, 2002. **81**(1): p. 13-17. PMID: 12357074

[24] Tamimi, R.M., et al., *NEUROD2andNEUROD3Genes map to human chromosomes 17q12 and 5q23–q31*

and mouse chromosomes 11 and 13, respectively. Genomics, 1997. **40**(2): p. 355-357. PMID: 9119405

[25] Noda, T., et al., *Transduction of NeuroD2 protein induced neural cell differentiation*. Journal of biotechnology, 2006. **126**(2): p. 230-236. PMID: 16730830

[26] Uittenbogaard, M., et al., *5' UTR of the neurogenic bHLH Nex1/MATH-2/NeuroD6 gene is regulated by two distinct promoters through CRE and C/EBP binding sites*. Journal of neuroscience research, 2007. **85**(1): p. 1-18. PMID: 17075921

[27] Malecki, M.T., et al., *Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus*. Nature genetics, 1999. **23**(3): p. 323. PMID: 10545951

Appendix

Table: 3

Gene Id	Gene	Protein
ENSP00000295108.3	NEUROD1	*neurogenic differentiation factor 1*
ENSP00000297142.3	NEUROD6	*neurogenic differentiation factor 6*
ENSP00000306754.4	NEUROD2	*neurogenic differentiation factor 2*
ENSP00000242994.3	NEUROD4	*neurogenic differentiation factor 4*
ENSP00000317333.3	NEUROG2	*neurogenin-2*
ENSP00000317580.4	NEUROG1	*neurogenin-1*
ENSP00000242462.4	NEUROG3	*neurogenin-3*
ENSP00000302216.3	ATOH1	*atonal homolog 1*
ENSP00000362777.3	ATOH7	*atonal homolog 7*
ENSP00000304676.3	ATOH8	*atonal homolog 8 isoform X2*
ENSP00000326391.2	BHLHA15	*class A basic helix-loop-helix 15*
ENSP00000476312.1	BHLHA15	*class A basic helix-loop-helix 15*
ENSP00000359371.2	BHLHE23	*class E basic helix-loop-helix 23*
ENSP00000480998.1	BHLHE23	*class E basic helix-loop-helix 23*
ENSP00000318799.1	BHLHE22	*class E basic helix-loop-helix 22*
ENSP00000275461.3	FER3	*fer3*
ENSP00000331040.3	OLIG2	*oligodendrocyte transcription factor 2*
ENSP00000371794.3	OLIG2	*oligodendrocyte transcription factor 2*
ENSP00000356708.2	OLIG3	*oligodendrocyte transcription factor 3*
ENSP00000476384.1	SCX	*basic helix-loop-helix transcription factor scleraxis*
ENSP00000246080.3	TCF15	*Transcription factor 15*
ENSP00000371785.1	OLIG1	*oligodendrocyte transcription factor 1*
ENSP00000334547.3	TAL2	*T-cell acute lymphocytic leukemia 2*
ENSP00000365687.3	PTF1A	*pancreas transcription factor 1 subunit alpha*
ENSP00000346582.5	TWIST1	*twist-related 1*
ENSP00000405176.2	TWIST2	*twist-related 2*
ENSP00000482581.1	TWIST2	*twist-related 2*
ENSP00000242261.5	TWIST1	*twist-related 1*
ENSP00000332293.4	ASCL2	*achaete-scute homolog 2*
ENSP00000322087.3	NHLH2	*helix-loop-helix 2*
ENSP00000358519.1	NHLH2	*helix-loop-helix 2*
ENSP00000302189.5	NHLH1	*helix-loop-helix 1*
ENSP00000237316.3	TCF21	*transcription factor 21*
ENSP00000356857.4	TCF21	*transcription factor 21*
ENSP00000266744.3	ASCL1	*achaete-scute homolog 1*
ENSP00000477638.1	HAND2	*HAND2 isoform 3*
ENSP00000294339.3	TAL1	*T-cell acute lymphocytic leukemia 1 isoform X1*
ENSP00000360951.1	TAL1	*T-cell acute lymphocytic leukemia 1 isoform X1*

ENSP00000352565.4	HAND1	*heart and neural crest derivatives expressed isoform*
ENSP00000264824.3	LYL1	*lyl-1*
ENSP00000321445.4	MSC	*musculin*
ENSP00000231121.2	HAND1	*Heart and neural crest derivatives expressed 1*
ENSP00000300057.4	MESP1	*mesoderm posterior 1*
ENSP00000342392.3	MESP2	*mesoderm posterior 2*
ENSP00000345420.4	ASCL4	*achaete-scute homolog 4*
ENSP00000333097.6	FIGLA	*factor in the germline alpha*
ENSP00000469019.2	ASCL5	*achaete-scute homolog 5*
ENSP00000472681.1	ASCL5	*achaete-scute homolog 5*
ENSP00000281047.3	MSGN1	*mesogenin-1*
ENSP00000435770.1	ASCL3	*achaete-scute homolog 3*
ENSP00000375248.1	BHLHA9	*class A basic helix-loop-helix 9*
ENSP00000228644.3	MYF5	*myogenic factor 5*
ENSP00000455444.1	TCF24	*transcription factor 24*
ENSP00000296096.5	TCF23	*transcription factor 23*
ENSP00000204517.6	TFAP4	*transcription factor AP-4*
ENSP00000250003.3	MYOD1	*myoblast determination 1*
ENSP00000228641.3	MYF6	*myogenic factor 6*
ENSP00000436313.1	ARNTL	*ARNTL isoform*
ENSP00000433571.1	ARNT	*aryl hydrocarbon receptor nuclear translocator 1 X4*
ENSP00000241651.4	MYOG	*myogenin*
ENSP00000420752.1	KIAA2018	*basic helix-loop-helix domain-containing KIAA2018*
ENSP00000367972.3	ID4	*DNA-binding inhibitor ID-4*
ENSP00000234091.4	ID2	*DNA-binding inhibitor ID-2*
ENSP00000379585.1	ID2	*DNA-binding inhibitor ID-2*
ENSP00000385465.2	ID2	*DNA-binding inhibitor ID-2*
ENSP00000365273.3	ID1	*DNA-binding inhibitor ID-1*
ENSP00000365280.3	ID1	*DNA-binding inhibitor ID-1*
ENSP00000363689.5	ID3	*DNA-binding inhibitor ID-3*
ENSP00000489102.1	ID3	*DNA-binding inhibitor ID-3*
ENSP00000385581.3	ARNT	*aryl hydrocarbon receptor nuclear translocator 1 X4*
ENSP00000385915.1	ARNT	*aryl hydrocarbon receptor nuclear translocator 1 X3*
ENSP00000491476.1	MYCN	*N-myc proto-oncogene*
ENSP00000385897.3	ARNT	*aryl hydrocarbon receptor nuclear translocator 1 X1*
ENSP00000374357.4	ARNT	*aryl hydrocarbon receptor nuclear translocator 1 X4*
ENSP00000384517.1	ARNT	*aryl hydrocarbon receptor nuclear translocator 1 X3*

(a) Homo sapiens

Gene Id	Gene	Protein
ENSMUSP00000040364.4	NEUROD1	*neurogenic differentiation factor 1*
ENSMUSP00000041373.6	NEUROD2	*neurogenic differentiation factor 2*

ENSMUSP00000051379.3	NEUROD4	*neurogenic differentiation factor 4*
ENSMUSP00000047016.8	NEUROD6	*neurogenic differentiation factor 6*
ENSMUSP00000050484.4	NEUROG1	*neurogenin-1*
ENSMUSP00000029587.7	NEUROG2	*neurogenin-2*
ENSMUSP00000054054.1	NEUROG3	*neurogenin-3*
ENSMUSP00000098903.4	ATOH1	*protein atonal homolog 1*
ENSMUSP00000039801.3	ATOH7	*protein atonal homolog 7*
ENSMUSP00000036981.7	ATOH8	*protein atonal homolog 8*
ENSMUSP00000055493.7	BHLHA15	*class A basic helix-loop-helix protein 15*
ENSMUSP00000104506.1	BHLHE23	*class E basic helix-loop-helix protein 23*
ENSMUSP00000026120.6	BHLHE22	*class E basic helix-loop-helix protein 22*
ENSMUSP00000058994.3	FER3	*fer3-like protein*
ENSMUSP00000061408.5	OLIG1	*oligodendrocyte transcription factor 1*
ENSMUSP00000036797.8	OLIG2	*oligodendrocyte transcription factor 2*
ENSMUSP00000057106.5	OLIG3	*oligodendrocyte transcription factor 3*
ENSMUSP00000043668.7	SCX	*basic helix-loop-helix transcription factor scleraxis*

ENSMUSP00000028068.2	PTF1A	*pancreas transcription factor 1*
ENSMUSP00000030489.2	TAL1	*T-cell acute lymphocytic leukemia protein 1 isoform X1*
ENSMUSP00000124983.1	TAL1	*T-cell acute lymphocytic leukemia protein 1 isoform X1*
ENSMUSP00000125202.1	TAL1	*T-cell acute lymphocytic leukemia protein 1 isoform X1*
ENSMUSP00000030124.3	TAL2	*T-cell acute lymphocytic leukemia protein 2*
ENSMUSP00000040089.5	TWIST1	*twist-related protein 1*
ENSMUSP00000007949.3	TWIST2	*twist-related protein 2*
ENSMUSP00000113012.1	TWIST2	*twist-related protein 2*
ENSMUSP00000057489.3	NHLH1	*helix-loop-helix protein 1*
ENSMUSP00000064355.4	NHLH2	*helix-loop-helix protein 2*
ENSMUSP00000142746.1	NHLH2	*helix-loop-helix protein 2*
ENSMUSP00000143362.1	NHLH2	*helix-loop-helix protein 2*
ENSMUSP00000086511.5	TCF15	*transcription factor 15*
ENSMUSP00000151767.1	TCF21	*transcription factor 21*
ENSMUSP00000053178.7	TCF21	*transcription factor 21*

ENSMUSP00000006818.2	TCF23	*transcription factor 23*
ENSMUSP00000138827.1	TCF24	*transcription factor 24*
ENSMUSP00000020243.7	ASCL1	*achaete-scute homolog 1*
ENSMUSP00000113012.1	ASCL2	*achaete-scute homolog 2*
ENSMUSP00000009392.4	ASCL2	*achaete-scute homolog 2*
ENSMUSP00000037702.1	ASCL3	*achaete-scute homolog 3*
ENSMUSP00000137650.1	ASCL4	*achaete-scute homolog 4*
ENSMUSP00000137746.1	ASCL5	*achaete-scute homolog 5*
ENSMUSP00000046999.2	HAND1	*heart and neural crest derivatives expressed transcript 1*
ENSMUSP00000124951.2	HAND1	*heart and neural crest derivatives expressed transcript 1*
ENSMUSP00000044983.3	HAND2	*dHand protein*
ENSMUSP00000046010.4	LYL1	*protein lyl-1*
ENSMUSP00000027062.5	MSC	*musculin*
ENSMUSP00000032760.5	MESP1	*mesoderm posterior protein 1*
ENSMUSP00000103017.1	MESP2	*mesoderm posterior protein 2*
ENSMUSP00000032070.3	FIGLA	*factor in the germline alpha*
ENSMUSP00000005862.7	TFAP4	*transcription factor AP-4*
ENSMUSP00000072330.1	MYOD1	*myoblast determination protein 1*
ENSMUSP00000000445.1	MYF5	*myogenic factor 5*
ENSMUSP00000047529.3	MYF6	*myogenic factor 6*
ENSMUSP00000147764.1	ARNT	*aryl hydrocarbon receptor nuclear translocator-like protein 1*
ENSMUSP00000046235.7	ARNT	*aryl hydrocarbon receptor nuclear translocator-like protein 1*
ENSMUSP00000147989.1	ARNT	*aryl hydrocarbon receptor nuclear translocator-like protein 1*
ENSMUSP00000147823.1	ARNT	*aryl hydrocarbon receptor nuclear translocator-like protein 1*
ENSMUSP00000027730.4	MYOG	*myogenin*
ENSMUSP00000092019.4	ID1	*DNA-binding protein inhibitor ID-1*
ENSMUSP00000105449.1	ID1	*DNA-binding protein inhibitor ID-1*
ENSMUSP00000008016.2	ID3	*DNA-binding protein inhibitor ID-3*
ENSMUSP00000021810.1	ID4	*DNA-binding protein inhibitor ID-4*
ENSMUSP00000045993.7	MYCN	*N-myc proto-oncogene protein*
ENSMUSP00000114225.1	MYCN	*N-myc proto-oncogene protein*
ENSMUSP00000054158.3	MYC	*protein S-Myc*

ENSMUSP00000050516.1	BHLHA9	*class A basic helix-loop-helix protein 9*
ENSMUSP00000055001.1	MSGN1	*mesogenin-1*
ENSMUSP00000085694.4	USF3	*basic helix-loop-helix domain-containing protein USF3*

(b) Mus musculus

Summary of the Gene Ontology annotation