

Computational Analysis of Microsatellites in Human Insulin Promoter Factor 1 Gene

Allam Appa Rao and Suresh B Mudunuri

International Center for Bioinformatics, Department of Computer science and Systems Engineering, Andhra University
College of Engineering, Visakhapatnam-530003, India
Corresponding author: E-mail: sureshverma@gmail.com

Received April 20, 2008; Accepted May 15, 2008; Published May 25, 2008

Citation: Allam AR, Suresh BM (2008) Computational Analysis of Microsatellites in Human Insulin Promoter Factor 1 Gene. *J Proteomics Bioinform S1: S001-S004*. doi:10.4172/jpb.s1000001

Copyright: © 2008 Allam AR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Human Insulin Promoter Factor 1 (IPF-1) gene plays an important role in the embryonic development of pancreas and in the transcriptional regulation of insulin production. Mutations in this gene are known to cause pancreatic agenesis and diabetes mellitus. A detailed bioinformatic study of all the known mutations from HGMD database in the IPF-1 gene revealed interesting information. The information of all the experimentally proven mutations were collected and analyzed using bioinformatic tools IMEx, SMART, PSIPRED and software programs developed by us. We tried to find out whether the presence of microsatellites in the IPF-1 gene has any significance in the generation of these mutations. Our analysis revealed that the InsCCG243 (Proline insertion) mutation, known to inhibit the insulin production, is due to microsatellite polymorphism. We analyzed 9 known mutations (excluding the silence mutations) and found out except one (R197H), all the other mutations (C18R, Q59L, Pro63fsdel, D76N, G212R, E224, P239Q, InsCCG243) fall outside the domain region. The mutation falling in the domain region seems to be inducing a change in the secondary structure and resulting in change or absence of protein function. We report that 4 out of these 9 mutations fall inside the microsatellite tracts and thus indicating a positive role of microsatellites in mutagenesis.

Keywords: Mutagenesis; Microsatellites; IPF-1; Diabetes Mellitus; Pancreatic agenesis; Secondary Structure

Introduction

Endocrine pancreas is made up of 4 types of cells: beta cells (secreting insulin), alpha cells (secreting glucagon), delta cells (secreting somatostatin) and PP cells (secreting pancreatic polypeptide). Insulin Promoter Factor 1 (IPF-1) or Somatostatin Transcription Factor (STF-1) (other synonyms: IDX1 and PDX1) is a homeodomain containing protein, known to play a key role in the transcription of endocrine pancreas specific genes in adults such as insulin, glucose transporter 2 (GLUT-2) and glucokinase in beta-cells and somatostatin in gamma-cells (Ohlsson, H et al, 1993; Petersen, H.V et al, 1994; Schwartz, P.T et al, 2000). Apart from gene regulation, IPF-1 is also found to be responsible for the development of the pancreas (Sander, M. and German, M.S, 1997). The IPF-1 is also required for the expression of FGFR1 signaling components in beta cells to main-

tain proper glucose sensing, insulin processing and glucose homeostasis (Hart, A.W et al, 2000). Mutations in IPF-1 are known to be involved in several disorders, including pancreatic agenesis and diabetes mellitus (Kim, S.K et al, 2002). Altered regulation of the expression of insulin gene leads to abnormal beta-cell function that leads to diabetes.

Apart from genes, the human genome also consists of a large number of nucleotide repeat units of size 1-6 bp repeated tandemly called Microsatellites or Simple Sequence Repeats (SSRs) or Short Tandem Repeats (STRs) (Schlotterer, C, 2000). Microsatellites are found in all the known genomes, spanning from prokaryotes, eukaryotes and viruses and are widely distributed both in coding and non-coding regions ((Toth, G et al, 2000; Sreenu, V.B. et al, 2007).

Mutation	Change	Disease Phenotype
C18R, Exon 1	TGC (Cys) -> CGC (Arg)	Diabetes mellitus, Type II
Q59L, Exon 1	CAG (Gln) -> CTG (Leu)	Diabetes mellitus, Type II
Pro63fsdel, Exon 1	Induces Stop Codon	Pancreatic agenesis
D76N, Exon1	GAC (Asp) -> AAC (Asn)	Diabetes mellitus, Type II
R197H, Exon 2	CGC (Arg) -> CAC (His)	Diabetes mellitus, Type II
G212R, Exon 2	GGG (Gly) -> AGG (Arg)	Diabetes mellitus, Type II
E224K, Exon 2	GAG (Glu) -> AAG (Lys)	Diabetes mellitus, Type II
P239Q, Exon 2	CCG (Pro) -> CAG (Gln)	Diabetes mellitus, Type II
InsCCG243, Exon2	Induces extra Proline	Diabetes mellitus, Type II

Table 1: List of genes collected from HGMD, their changes and phenotypes

712 CTG CCG CCG CCG CCG --- CCC CCC GGA (Normal)
 CTG CCG CCG CCG CCG CCG CCC CCC GGA (Mutant)
 238 L P P P P P P P P G

Figure 1: Alignment of the normal and mutant sequence depicting the change in the amino acid sequence. The colored site is the CCG insertion that induces an extra proline in the sequence.

Mutations in these microsatellite regions occur at much higher rate when compared with those in the rest of the genome (Ellegren, H. 2000).

Microsatellites are known to be highly polymorphic due to the high rate of mutations in their tracts (Jarne, P. and Lagoda, P.J.L. 1996). These mutations can be either in the form of increase / decrease of repeat units or in the form of single nucleotide substitutions/deletions/insertions and other events (Fan, H. and Chu, J.Y. 2007). Increase or decrease of repeat units of microsatellites in coding regions might lead to shift in reading frames there by causing changes in protein product (Li, Y.C et al, 2004) and in non-coding regions are known to effect the gene regulation (Martin P et al, 2005). Point mutations (Substitutions and Indels) are also found to occur at a higher rate in microsatellites than elsewhere (Sibly, R.M. et al, 2003). Microsatellite mutations with in or near certain genes are known to be responsible for some human neurodegenerative diseases (Tautz, D et al, 1994). So, we made a brief study to check whether the mutations in this IPF-1 gene has any relation with these microsatellite repeats and the study revealed interesting results.

Methods

All the experimental proved mutations of the IPF-1 gene, that are falling inside the coding regions and eventually leading to phenotypic differences were collected from the Human Gene Mutation Database (HGMD) (Stenson PD et al, 2003). Table 1 gives the list of all the 9 mutations considered for analysis. The mutations do not include silent mutations, which do not induce any change in the amino acid sequence, and all the 9 mutations that produce a disease phenotype (Hani, E. H et al, 1999; Macfarlane, W. M et al, 1999; Weng, J et al, 2001; Cockburn, B. N et al, 2004).

Results and Discussion

Out of the nine mutations, four are falling in the microsatellite regions. Two of the four mutations seems to be a result of ‘Strand Slippage Replication’, which is a predominant mutation mechanism of microsatellites. Strand Slippage Replication (also known as DNA Slippage or Slipped Strand Mispairing) occurs during DNA replication, which results in the mispairing of one or more repeat units by forming a loop out at the mismatch site (Fan, H. and Chu, J.Y,

2007). This causes a decrease/increase of repeat units in the microsatellite tract, thus making the microsatellite tract highly polymorphic.

The mutation InsCCG243 is a clear indication of microsatellite polymorphism inserting an extra repeat unit of CCG. The actual site of mutation is a microsatellite tract of (CCG)₄. (See Figure. 1). This results in an extra proline in the amino acid sequence and thereby inhibiting the insulin expression to a significant extent (Hani, E. H et al, 1999). Another mutation, Pro63fsdel refers to a point mutation (deletion) of a nucleotide 'C' from the 63rd codon leading to a frame shift, thereby inducing a stop codon.

This results in an incomplete protein and has been proved to be responsible for pancreatic agenesis (Stoffers, D. A. et al, 1997). The mutation Pro63fsdel also falls in a mononucleotide microsatellite region (C)₆. *The deletion of 'C' refers to a deletion of a repeat unit of a mononucleotide microsatellite tract indicating a strand slippage replication. (See Figure 2)*

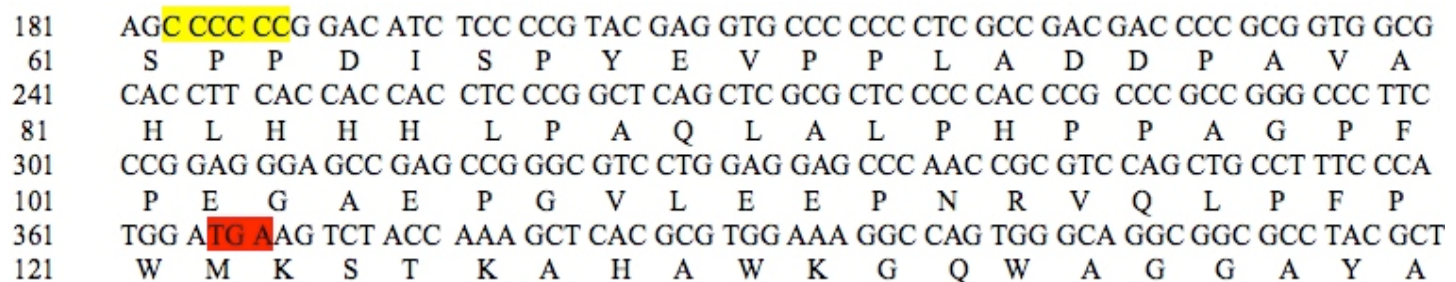


Figure 2: The Yellow colored region is the mononucleotide microsatellite tract (C repeating 6 times). A deletion of a C from the tract is indicative of slippage strand replication where a repeat unit 'C' is deleted leading to a premature termination of the protein. Red color indicates the stop codon.

Further more, most of the mutations in the IPF-1 gene fall in GC rich regions. It is well known fact that the rate of occurrence of mutations in GC rich regions is less when compared to AT rich regions because of the strong triple bond between G and C. But, interestingly, 7 of the 9 mutations are changing either C or G or both and 4 out of these 7 mutations fall inside the microsatellite regions. This indicates that micro satellites play a relatively positive role in the mutagenesis.

Conclusion

Microsatellites are known for their higher rate of mutations and are known to be associated with various diseases. So, we analyzed the IPF-1 mutations and their possible association with the microsatellites. The IPF-1 mutations from HGMD database are mapped on to the microsatellite tracts

As discussed earlier, point mutations also occur at a higher rate in the microsatellite regions than the rest of the genome, we also looked for point mutations inside the microsatellite tracts in the IPF-1 gene. Two mutations G212R and P239Q fall in the microsatellite tracts (GCG)₃ [Range:626-634] and (CCG)₄ [Range:714-728] and suggest their role in mutation generation.

It is also observed that none of these mutations except one (R197H) does not fall in the homeodomain region (HOX Range: 146-208) that is very essential for the actual function of the protein. But, all these mutations despite of lying outside the domain region effect the function of this protein leading to diabetes or pancreatic agenesis. The R197H mutation falls inside the domain region and also seems to be effecting the helix formation at that position. The sequence submitted to PSIPRED server indicates that the Arginine at that position is likely to form a helix and the mutation might change the secondary structure of the domain leading to malfunction of the protein.

of IPF-1 and the results seem to indicate that microsatellites play an important role in the mutagenesis of IPF-1 gene leading to Pancreatic agenesis and Type 2 Diabetes Mellitus. Extending this work on a large scale by analyzing large number of genes might give a better evidence of the role of microsatellites in generating mutations.

Acknowledgment

This work was supported by IIT up gradation grants of AUCE (A).

References

1. Cockburn BN, Bermano G, Boodram LLG, Teelucksingh S, Tsuchiya T, et al. (2004) Insulin promoter factor-1 mu-

- tations and diabetes in Trinidad: identification of a novel diabetes-associated mutation (E224K) in an Indo-Trinidadian family. *J Clin Endocr Metab* 89: 971-978. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat Genet* 24: 400-402. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 3. Fan H, Chu JY (2007) A brief review of short tandem repeat mutation. *Genomics Proteomics Bioinformatics* 5: 7-14. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 4. Hani EH, Stoffers DA, Chevre JC, Durand E, Stanojevic V, et al. (1999) Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus. *J Clin Invest* 104: R41-R48 » [CrossRef](#) » [Pubmed](#)
 5. Hart AW, Baeza N, Apelqvist A, Edlund H (2000) Attenuation of FGF signalling in mouse beta-cells leads to diabetes. *Nature* 408: 864-8. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 6. Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 11: 424-429. » [CrossRef](#) » [Google Scholar](#)
 7. Kim SK, Selleri L, Lee JS, Zhang AY, Gu X, et al. (2002) Pbx1 inactivation disrupts pancreas development and in *Ipf1*-deficient mice promotes diabetes mellitus. *Nature Genet* 30: 430-435. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 8. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, et al. (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 32: D142-4. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 9. Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* 21: 991-1007. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 10. Macfarlane WM, Frayling TM, Ellard S, Evans JC, Allen LIS, et al. (1999) Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. *J Clin Invest* 104: R33-R39. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 11. Martin P, Makepeace K, Hill SA, Hood DW, Moxon ER (2005) Microsatellite instability regulates transcription factor binding and gene expression. *PNAS* 102: 3800-3804. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 12. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404-405. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 13. Mudunuri SB, Nagarajaram HA (2007) IMEx: Imperfect Microsatellite Extractor. *Bioinformatics* 23:1181-1187. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 14. Ohlsson H, Karlsson K, Edlund T (1993) IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J* 12: 4251-59. » [Pubmed](#) » [Google Scholar](#)
 15. Petersen HV, Serup P, Leonard J, Michelsen BK, Madsen OD (1994) Transcriptional regulation of the human insulin gene is dependent on the homeodomain protein STF1/IPF1 acting through the CT boxes. *PNAS* 91:10465-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 16. Sander M, German MS (1997) The beta cell transcription factors and development of the pancreas. *J Mol Med* 75: 327-40. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 17. Schlotterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365-371. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 18. Schwartz PT, Perez VB, Rivera A, Moratalla R, Vallejo M (2000) Pancreatic homeodomain transcription factor IDX1/IPF1 expressed in developing brain regulates somatostatin gene transcription in embryonic neural cells. *J Biol Chem* 275:19106-14 » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 19. Sibly RM, Meade A, Boxall N, Wilkinson MJ, Corne DW, et al. (2003) The structure of interrupted human AC microsatellites. *Mol Biol Evol* 20: 453-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 20. Sreenu VB, Kumar P, Nagarajaram HA (2007) Simple sequence repeats in mycobacterial genomes. *J Biosci* 32: 3-15. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 21. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, et al. (2003) The Human Gene Mutation Database (HGMD®): 2003 Update. *Hum Mutat* 21: 577-581. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 22. Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF (1997) Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nature Genet* 15:106-110. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 23. Tautz D, Schlotterer C (1994) Simple sequences. *Curr Opin Genet Dev* 4: 832-837. » [Pubmed](#) » [Google Scholar](#)
 24. Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10: 967-981. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
 25. Weng J, Macfarlane WM, Lehto M, Gu HF, Shepherd LM, et al. (2001) Functional consequences of mutations in the MODY4 gene (IPF1) and coexistence with MODY3 mutations. *Diabetologia* 44: 249-58. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)