Guest Editorial

Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes

'In Nature's infinite book of secrecy, a little I can read.'

Antony and Cleopatra [Act I, Scene 2], William Shakespeare

Pathological mutations occurring within the extended consensus sequences of exon–intron splice junctions account for ~10 per cent of all inherited lesions logged in The *Human Gene Mutation Database* (HGMD®; http://www.hgmd.org)[1] and are frequently encountered in mutation screening studies.[2] Mutations residing in other intronic locations (including the canonical branchpoint sequence,[3] 5′-YURAY-3′), however, may often go undetected unless patient RNA can be analysed and the mutations in question induce aberrant splicing (eg exon skipping or cryptic splice site utilisation) that is readily distinguishable qualitatively or quantitatively from normal (and/or normal alternative) splicing. Indeed, introns probably represent a substantially larger mutational target than has hitherto been appreciated, on account of their containing a multiplicity of functional elements, including intron splice enhancers and silencers that regulate alternative splicing,[4,5] *trans*-splicing elements[6] and other regulatory elements, some of which may be deeply embedded within very large introns.[7]

In addition to pathological mutations *sensu stricto*, introns also harbour functional polymorphisms that can influence the expression of the genes that host them. Some of these intronic variants may also confer susceptibility to disease or otherwise modulate the genotype–phenotype relationship. For the reasons discussed above, it is very likely that such variants will have been seriously under-ascertained to date. Although most of these variants are single nucleotide polymorphisms (SNPs), others may be of the insertion/deletion type.[8] With the advent of genome-wide association studies (GWAS), an increasing number of potentially functional intronic variants are being identified.[9] In the majority of cases, however, it is unclear whether such variants are of direct functional significance, as opposed to simply being in linkage disequilibrium with another (as yet unidentified) functional SNP in the vicinity.[10] Even when GWAS studies deem a newly identified intronic polymorphism to be 'functional', it should be appreciated that such a term may often be ascribed solely on the basis of an observed association between a specific allele and a plasma protein level, enzymatic activity or a clinical/laboratory phenotype — even although in reality such associations cannot readily distinguish a bona fide functional SNP from a linkage disequilibrium effect.

As has been noted with pathological mutations, the vast majority of known functional intronic polymorphisms are located within the extended consensus sequences of exon–intron splice junctions.[2] Some intronic polymorphic variants do not occur within the splice junctions, however, but nevertheless still act so as to change the splicing phenotype as a consequence of their being located within an intron splice enhancer or branchpoint site, or by activating a cryptic splice site.[11,12] This is, from a biological point of view, a more interesting category of intronic SNP to study, since the

 © HENRY STEWART PUBLICATIONS 1479–7364. **HUMAN GENOMICS**. VOL 4. NO. 5. 284–288 JUNE 2010

mechanisms by which these variants exert their effects on the splicing phenotype are often unclear and may be quite subtle. In the pages of this issue, Millar *et al*.[13] report that a SNP, buried deep within intron 4 of the human growth hormone (*GH1*) gene, is of direct functional significance by virtue of its influence on the expression of this gene. This polymorphism therefore joins the ranks of the hitherto relatively small number of human intronic SNPs located outwith exon–intron splice junctions that have been shown by various methods of *in vitro* characterisation to be of direct functional significance. Table 1 lists some of the best characterised examples of such functional SNPs, most of which are located at least ~30 base pairs (bp) from the nearest splice site. These SNPs have been shown to influence either the transcriptional activity or the splicing efficiency of their host genes, or instead to alter the expression of alternative transcripts.

How should we go about increasing the number of identified functional intronic polymorphisms? One approach would be to employ exon-tiling microarrays to perform genome-wide scans to identify intronic SNPs responsible for inter-individual differences in the splicing phenotype.[11,14,15] Since currently available bioinformatics tools are inadequate to the task of predicting splicing consequences,[14] however, all SNPs identified in this way would have to be further validated using mini-gene constructs to determine the resulting splicing phenotype.[14] One feature that might prove helpful in identifying intronic SNPs is that such variants are often located within gene regions that are characterised by a reduced level of genetic variation.[16]

Precisely because we invariably adopt a gene-centric approach to screening introns for functional polymorphisms, we should be wary of the existence of overlapping genes, a not infrequent occurrence

Table 1. Selected examples of *in vitro* characterised human functional intronic polymorphisms located more than ~30 bp from the nearest splice site

Gene	Disease/phenotype	Chromosomal location	Polymorphism, intronic location and dbSNP number	Consequences for gene expression or mRNA splicing	Reference
AGTR2	Predisposition to congenital anomalies of the kidney and urinary tract	Xq22-q23	IVS1, AS, A > G, −29 (rs1403543)	SNP occurs within branchpoint motif and alters splicing efficiency	Nishimura *et al.* (1999)[a]
BANK1	Susceptibility to systemic lupus erythematosus	4q23	IVS1, AS, T > C, −43 (rs17266594)	SNP occurs within branchpoint motif and risk allele alters expression of alternative transcripts	Kozyrev *et al.* (2008)[b]
CD244	Susceptibility to rheumatoid arthritis	1q23.1	IVS3, AS, T > C, −164 (rs6682654)	Risk allele associated with increased transcriptional activity	Suzuki *et al.* (2008)[c]
CD244	Susceptibility to rheumatoid arthritis	1q23.1	IVS5, DS, G > A, +526 (rs3766379)	Risk allele associated with increased transcriptional activity	Suzuki *et al.* (2008)[c]
COL1A1	Reduced bone density/ osteoporosis	17q21.33	IVS1, AS, G > T, −440 (rs1800012)	SNP occurs within Sp1-binding site; risk allele alters Sp1 binding and transcriptional activity	Mann *et al.* (2001)[d]

Continued

Table 1. Continued

Gene	Disease/phenotype	Chromosomal location	Polymorphism, intronic location and dbSNP number	Consequences for gene expression or mRNA splicing	Reference
CXCR3	Variation in immune cell response to chemokine-cytokine signals	Xq13	IVS1, DS, G > A, +234 (rs2280964)	Risk allele associated with reduced *CXCR3* gene expression	Choi *et al.* (2008)[e]
CYP2D6	Intermediate metaboliser (reduced expression of CYP2D6)	22q13.1	IVS6, DS, G > A, +39 (rs28371725)	Increased level (7.3-fold) of non-functional splice variant transcript lacking exon 6 and reduced level (2.9-fold) of functional transcript	Toscano *et al.* (2006)[f]
DRD2	Reduced DRD2 expression	11q22-q23	IVS1, DS, A > G, +3850 (rs2734836)	Risk allele associated with increased binding of transcriptional repressor (Freud-1) leading to reduced *DRD2* expression	Rogaeva *et al.* (2007)[g]
DRD2	Reduced DRD2 expression	11q23	IVS6, AS, C > A, −83 (rs 1076560)	Risk allele alters expression of alternative transcripts	Zhang *et al.* (2007)[h]
F2	Elevated prothrombin level/thrombosis	11p11-q12	IVS13, AS, A > G, −59	Risk allele influences splicing efficiency	von Ahsen & Oellerich (2004)[i]
FGFR2	Susceptibility to breast cancer	10q26	IVS2, DS, T > C, +12912 (rs2981578)	Risk allele alters binding affinity for transcription factors Oct-1/Runx2, leading to increased *FGFR2* expression	Meyer *et al.* (2008)[j]
FOXP3	Susceptibility to psoriasis	Xp11.23	IVS1, DS, A > C, +2882 (rs3761548)	Risk allele causes loss of binding of E47 and c-Myb, leading to reduced *FOXP3* transcription	Shen *et al.* (2010)[k]
GFPT1	Reduced GFPT1 expression	2p13	IVS1, DS, T > C, +36 (rs6720415)	SNP occurs within GC box and risk allele decreases transcriptional activity	Kunika *et al.* (2006)[l]
GSK3B	Risk of Parkinson's disease	3q13.3	IVS5, AS, T > C, −157 (rs6438552)	Risk allele associated with increased level of *GSK3B* transcripts lacking exons 9 and 11	Kwok *et al.* (2005)[m]
IRF4	Risk of childhood acute lymphoblastic leukaemia in males	6p25-p23	IVS4, DS, C > T, +386 (rs12203592)	Risk allele increases IRF4 promoter activity/ expression	Do *et al.* (2010)[n]

Continued

Table 1. Continued

Gene	Disease/phenotype	Chromosomal location	Polymorphism, intronic location and dbSNP number	Consequences for gene expression or mRNA splicing	Reference
LTA	Susceptibility to myocardial infarction	6p21.3	IVS1, AS, G > A, −198 (rs909253)	Risk allele associated with increased transcriptional activity	Ozaki *et al.* (2002)[o]
NLRP3	Susceptibility to food-induced anaphylaxis	1q44	IVS7, AS, C > T, −202 (rs4612666)	Risk allele increases enhancer activity by 20%	Hitomi *et al.* (2009)[p]
SCG3	Association with obesity	15q21	IVS1, DS, G > A, +190 (rs16964476)	Risk allele alters transcriptional activity	Tanabe *et al.* (2007)[q]
TH	Risk of essential tension	11p15.5	IVS12, DS, T > C, +127 (rs2070762)	Risk allele associated with increased transcriptional activity	Wang *et al.* (2008)[r]
USF1	Association with familial combined hyperlipidaemia	1q22-q23	IVS7, AS, G > A, −100 (rs2073658)	SNP alleles exhibit differential binding to nuclear proteins. USF1-regulated genes are differentially regulated, depending on the identity of the rs2073658 allele	Naukkarinen *et al.* (2005)[s] Naukkarinen *et al.* (2009)[t]

Abbreviations: AS, acceptor splice site; DRD2, dopamine D2 receptor; DS, donor splice site; IVS, intron (number)
Nucleotide numbering relative to specified splice site.
rs numbers are provided courtesy of dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/). For the sake of simplicity, only SNPs have been included in Table 1 (thus, for example, functional intronic microsatellite polymorphisms would require a separate treatment).

References to table

a. Nishimura, H., Yerkes, E., Hohenfellner, K., Miyazaki, Y. *et al.* (1999), 'Role of the angiotensin type 2 receptor gene in congenital anomalies of the kidney and urinary tract, CAKUT, of mice and men', *Mol. Cell* Vol. 3, pp. 1–10.

b. Kozyrev, S.V., Abelson, A.K., Wojcik, J., Zaghlool, A. *et al.* (2008), 'Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus', *Nat. Genet.* Vol. 40, pp. 211–216.

c. Suzuki, A., Yamada, R., Kochi, Y., Sawada, T. *et al.* (2008), 'Functional SNPs in *CD244* increase the risk of rheumatoid arthritis in a Japanese population', *Nat. Genet.* Vol. 40, pp. 1224–1229.

d. Mann, V., Hobson, E.E., Li, B., Stewart, T.L *et al.* (2001), 'A *COL1A1* Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality', *J. Clin. Invest.* Vol. 107, pp. 899–907.

e. Choi, J.W., Park, C.S., Hwang, M., Nam, H.Y. *et al.* (2008), 'A common intronic variant of *CXCR3* is functionally associated with gene expression levels and the polymorphic immune cell responses to stimuli', *J. Allergy Clin. Immunol.* Vol. 122, pp. 1119–1126.

f. Toscano, C., Klein, K., Blievernicht, J., Schaeffeler, E. *et al.* (2006), 'Impaired expression of *CYP2D6* in intermediate metabolizers carrying the *41 allele caused by the intronic SNP 2988G>A: Evidence for modulation of splicing events', *Pharmacogenet. Genomics* Vol. 16, pp. 755–766.

g. Rogaeva, A., Ou, X.M., Jafar-Nejad, H., Lemonde, S. *et al.* (2007), 'Differential repression by freud-1/CC2D1A at a polymorphic site in the dopamine-D2 receptor gene'. *J. Biol. Chem.* Vol. 282, pp. 20897–20905.

h. Zhang, Y., Bertolino, A., Fazio, L., Blasi, G. *et al.* (2007), 'Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory', *Proc. Natl. Acad. Sci. USA* Vol. 104, pp. 20552–20557.

i. von Ahsen, N. and Oellerich, M. (2004), 'The intronic prothrombin 19911A>G polymorphism influences splicing efficiency and modulates effects of the 20210G>A polymorphism on mRNA amount and expression in a stable reporter gene assay system', *Blood* Vol. 103, pp. 586–593.

j. Meyer, K.B., Maia, A.T., O'Reilly, M., Teschendorff, A.E. *et al.* (2008), 'Allele-specific up-regulation of *FGFR2* increases susceptibility to breast cancer', *PLoS Biol.* Vol. 6, p. e108.

k. Shen, Z., Chen, L., Hao, F., Wang, G. *et al.* (2010), 'Intron-1 rs3761548 is related to the defective transcription of *Foxp3* in psoriasis through abrogating E47/c-Myb binding', *J. Cell. Mol. Med.* Vol. 14, pp. 226–241.

l. Kunika, K., Tanahashi, T., Kudo, E., Mizusawa, N. *et al.* (2006), 'Effect of +36T>C in intron 1 on the glutamine: fructose-6-phosphate amidotransferase 1 gene and its contribution to type 2 diabetes in different populations', *J. Hum. Genet.* Vol. 51, pp. 1100–1109.

m. Kwok, J.B., Hallupp, M., Loy, C.T., Chan, D.K. *et al.* (2005), '*GSK3B* polymorphisms alter transcription and splicing in Parkinson's disease', *Ann. Neurol.* Vol. 58, pp. 829–839.

n. Do, T.N., Ucisik-Akkaya, E., Davis, C.F., Morrison, B.A. *et al.* (2010), 'An intronic polymorphism of *IRF4* gene influences gene transcription *in vitro* and shows a risk association with childhood acute lymphoblastic leukemia in males', *Biochim. Biophys. Acta* Vol. 1802, pp. 292–300.

o. Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A. *et al.* (2002), 'Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction', *Nat. Genet.* Vol. 32, pp. 650–654.

p. Hitomi, Y., Ebisawa, M., Tomikawa, M., Imai, T. *et al.* (2009), 'Associations of functional *NLRP3* polymorphisms with susceptibility to food-induced anaphylaxis and aspirin-induced asthma', *J. Allergy Clin. Immunol.* Vol. 124, pp. 779–785.

q. Tanabe, A., Yanagiya, T., Iida, A., Saito, S. *et al.* (2007), 'Functional single-nucleotide polymorphisms in the secretogranin III (*SCG3*) gene that form secretory granules with appetite-related neuropeptides are associated with obesity', *J. Clin. Endocrinol. Metab.* Vol. 92, pp. 1145–1154.
r. Wang, L., Li, B., Lu, X., Zhao, Q. *et al.* (2008), 'A functional intronic variant in the tyrosine hydroxylase (*TH*) gene confers risk of essential hypertension in the Northern Chinese Han population', *Clin. Sci.* Vol. 115, pp. 151–158.
s. Naukkarinen, J., Gentile, M., Soro-Paavonen, A., Saarela, J. *et al.* (2005), 'USF1 and dyslipidemias: Converging evidence for a functional intronic variant', *Hum. Mol. Genet.* Vol. 14, pp. 2595–2605.
t. Naukkarinen, J., Nilsson, E., Koistinen, H.A., Söderlund, S. *et al.* (2009), 'Functional variant disrupts insulin induction of USF1: Mechanism for USF1-associated dyslipidemias', *Circ. Cardiovasc. Genet.* Vol. 2, pp. 522–529.

in our complex genome. Thus, for example, the functional SNP rs4988235, located 13.9 kilobases upstream of the lactase (*LCT*) gene and associated with adult-type hypolactasia, actually resides deep within intron 13 of the minichromosome maintenance complex component 6 (*MCM6*) gene.[17–19] In addition, since disease-associated intronic SNPs that play a role in long-range gene regulation have also recently been identified,[20,21] we should be aware that some SNPs may influence the expression of remote genes at distance, rather than the expression of those genes which actually host them. These caveats notwithstanding, new techniques such as chromosome conformational capture[22] and chromatin immunoprecipitation followed by deep sequencing (ChIP-seq)[23] promise greatly to increase the number of functional intronic polymorphisms identified, thereby potentially pinpointing the locations of a whole new lexicon of intron-located regulatory elements, which will increase our understanding of intron structure and function.

David N. Cooper
Institute of Medical Genetics, School of Medicine
Cardiff University, Heath Park Cardiff, CF14 4XN
UK
Tel: +44 2920 744062; E-mail: cooperDN@cardiff.ac.uk

References

1. Stenson, P.D., Mort, M., Ball, E.V., Howells, K. *et al.* (2009), 'The Human Gene Mutation Database: 2008 update', *Genome Med.* Vol. 1, p.13.
2. Krawczak, M., Thomas, N.S., Hundrieser, B., Mort, M. *et al.* (2007), 'Single base-pair substitutions in exon-intron junctions of human genes: Nature, distribution, and consequences for mRNA splicing', *Hum. Mutat.* Vol. 28, pp. 150–158.
3. Královicová, J., Lei, H. and Vorechovský, I. (2006), 'Phenotypic consequences of branch point substitutions', *Hum. Mutat.* Vol. 27, pp. 803–813.
4. Wang, X., Wang, K., Radovich, M., Wang, Y. *et al.* (2009), 'Genome-wide prediction of *cis*-acting RNA elements regulating tissue-specific pre-mRNA alternative splicing', *BMC Genomics* Vol. 10 (Suppl. 1), p. S4.
5. Tress, M.L., Martelli, P.L., Frankish, A., Reeves, G.A. *et al.* (2007), 'The implications of alternative splicing in the ENCODE protein complement', *Proc. Natl. Acad. Sci. USA* Vol. 104, pp. 5495–5500.
6. Gingeras, T.R. (2009), 'Implications of chimaeric non-co-linear transcripts', *Nature* Vol. 461, pp. 206–211.
7. Solis, A.S., Shariat, N. and Patton, J.G. (2008), 'Splicing fidelity, enhancers, and disease', *Front. Biosci.* Vol. 13, pp. 1926–1942.
8. Wilkins, J.M., Southam, L., Mustafa, Z., Chapman, K. *et al.* (2009), 'Association of a functional microsatellite within intron 1 of the *BMP5* gene with susceptibility to osteoarthritis', *BMC Med. Genet.* Vol. 10, p. 141.
9. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B. *et al.* (2009), 'Finding the missing heritability of complex diseases'. *Nature* Vol. 461, pp. 747–753.
10. McCauley, J.L., Kenealy, S.J., Margulies, E.H., Schnetz-Boutaud, N. *et al.* (2007), 'SNPs in multi-species conserved Sequences (MCS) as useful markers in association studies: A practical approach', *BMC Genomics* Vol. 8, p. 266.
11. Kwan, T., Benovoy, D., Dias, C., Gurd, S. *et al.* (2008), 'Genome-wide analysis of transcript isoform variation in humans', *Nat. Genet.* Vol. 40, pp. 225–231.
12. Coulombe-Huntington, J., Lam, K.C., Dias, C. and Majewski, J. (2009), 'Fine-scale variation and genetic determinants of alternative splicing across individuals', *PLoS Genet.* Vol. 5, p. e1000766.
13. Millar, D.S., Horan, M., Chuzhanova, N.A. and Cooper, D.N. (2010), 'Characterisation of a functional intronic polymorphism in the human growth hormone (*GH1*) gene', *Hum. Genomics*, Vol. 4, pp. 289–301.
14. Hull, J., Campino, S., Rowlands, K., Chan, M.-S. *et al.* (2007), 'Identification of common genetic variation that modulates alternative splicing', *PLoS Genet.* Vol. 3, p. e99.
15. Nembarware, V., Lupindo, B., Schouest, K., Spillane, C. *et al.* (2008), 'Genome-wide survey of allele-specific splicing in humans', *BMC Genomics* Vol. 9, p. 265.
16. Lomelin, D., Jorgenson, E. and Risch, N. (2010), 'Human genetic variation recognizes functional elements in noncoding sequence', *Genome Res.* Vol. 20, pp. 311–319.
17. Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D. *et al.* (2002), 'Identification of a variant associated with adult-type hypolactasia', *Nat. Genet.* Vol. 30, pp. 233–237.
18. Olds, L.C. and Sibley, E. (2003), 'Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: Functional role as a *cis* regulatory element', *Hum. Mol. Genet.* Vol. 12, pp. 2333–2340.
19. Lewinsky, R.H., Jensen, T.G., Møller, J., Stensballe, A. *et al.* (2005), 'T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*', *Hum. Mol. Genet.* Vol. 14, pp. 3945–3953.
20. Ragvin, A., Moro, E., Fredman, D., Navratilova, P. *et al.* (2010), 'Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to *HHEX, SOX4*, and *IRX3*', *Proc. Natl. Acad. Sci. USA* Vol. 107, pp. 775–780.
21. Jowett, J.B., Curran, J.E., Johnson, M.P., Carless, M.A. *et al.* (2010), 'Genetic variation at the FTO locus influences RBL2 gene expression', *Diabetes* Vol. 59, pp. 726–732.
22. Dostie, J. and Dekker, J. (2007), 'Mapping networks of physical interactions between genomic elements using 5C technology', *Nat. Protoc.* Vol. 2, pp. 988–1002.
23. Visel, A., Blow, M.J., Li, Z., Zhang, T. *et al.* (2009), 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature* Vol. 457, pp. 854–858.