Gene Expression, Vol. 11, pp. 233–239
Printed in the USA.

1052-2166/04 $20.00 + .00

www.cognizantcommunication.com

A High Proportion of Chromosome 21 Promoter Polymorphisms Influence Transcriptional Activity

PAUL R. BUCKLAND,[1] SHAROL L. COLEMAN, BASTIAAN HOOGENDOORN, CAROL GUY, S. KAYE SMITH, AND MICHAEL C. O'DONOVAN

Department of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff, CF14 4XN, UK

We have sought to obtain an unbiased estimate of the proportion of polymorphisms in promoters of human genes that have functional effects. We carried out polymorphism discovery on a randomly selected group of 51 gene promoters mapping to human chromosome 21 and successfully analyzed the effect on transcription of 38 of the sequence variants. To achieve this, a total of 53 different haplotypes from 20 promoters were cloned into a modified pGL3 luciferase reporter gene vector and were tested for their abilities to promote transcription in HEK293t and JEG-3 cells. Up to seven (18%) of the 38 tested variants altered transcription by 1.5-fold, confirming that a surprisingly high proportion of promoter region polymorphisms are likely to be functionally important. The functional variants were distributed across the promoters of *CRYAA, IFNAR1, KCNJ15, NCAM2, IGSF5*, and *B3GALT5*. Three of the genes (*NCAM2, IFNAR1*, and *CRYAA*) have been previously associated with human phenotypes and the polymorphisms we describe here may therefore play a role in those phenotypes.

Key words: CRYAA; Cataracts; IFNAR1; Galactosyltransferase; KCNJ15; Down syndrome; NCAM2; B3GALT5; Hepatitis C

THE human genome contains fewer protein coding genes than expected, and this has stressed the potential importance of differential gene regulation in disease susceptibility and other inherited phenotypes (14, 23). Most molecular genetic studies of inherited traits have focused on DNA encoding protein, largely because our knowledge of regulatory sequence is elementary. The human genome contains thousands of common polymorphisms that change the amino acid sequence of protein. However, it contains several million polymorphisms in noncoding regions of the genome, and it is gradually becoming clear that the latter types of polymorphism are also a major genetic source of human phenotypic variation.

Many, or most, polymorphisms in noncoding regions presumably have no functional effect, while others may have major effects on gene function, including *cis*-acting influences on the control of gene expression. How to identify polymorphisms that influence the expression of genes is a key problem in human genetics (13). Sequence elements anywhere in the genome may influence the expression of genes; however, the principal elements that control transcription are within the core promoter and proximal promoter regions. Polymorphisms within these regions therefore have a priori the greatest likelihood of having a functional effect on transcription.

Recently, novel in vivo methods have been described that indirectly measure the effect of regulatory polymorphism by using a marker sequence variant in either the promoter (17) or an exon (3,18,24). While these methods have great utility, they do not determine which of the several sequence variants contained within any given haplotype actually exert functional effects. At present, this requires the use of in vitro reporter gene assays. Previously, reporter

Accepted January 28, 2004.
[1]Address correspondence to Paul R. Buckland, Department of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff, CF14 4XN, UK. Tel: (0)2920 744840; Fax: (0)2920 747839.

gene assays have been limited by low throughput and currently there is limited information available about the frequency and the size of effects of functionally relevant promoter polymorphisms.

We have recently developed a high throughput method for detecting functional polymorphisms (4) and have applied that methodology to more than half of all the experimentally proven human promoters in the Eukaryotic Promoter Database [(12); http://www.epd.isb-sib.ch/]. In this, the only large-scale published study of its kind that we are aware of, we previously reported that a high proportion (approximately 35%) of polymorphic promoters (12) yielded evidence for functional differences between haplotype pairs, and that around 10% of all examined promoters had functionally relevant polymorphisms. However, that estimate was based upon analysis of the relatively small proportion of possible promoters (<1%) that have been experimentally determined (12). Given the amount of experimental work required to define a promoter, it is likely that these promoters were in some way of atypical prior interest to geneticists. In this study, we report an analysis of promoters whose selection is relatively unbiased except by genome position. To compile this list of promoters, we chose chromosome 21 as it was one of the first chromosomes to be fully sequenced (10). We screened 51 promoters for sequence variation, and where variation was detected, we cloned the haplotypes into a reporter gene plasmid. Pairs of haplotypes were then analyzed for their ability to promote gene expression in human cell lines using a reporter gene assay system.

MATERIALS AND METHODS

Promoter Selection and Polymorphism Identification

DNA promoter sequences for 83 genes on chromosome 21 with proven [82] or predicted [1] open reading frames were identified at random based on the annotation of the published sequence on June 2000. Primers were designed to amplify 500–700 bp of the 5′ flanking sequence. In selecting the promoter sequence the specifications for the 3′ primer were, in order of priority, that: a) the start of transcription was included within the amplimer, b) it did not amplify an ATG within the 5′ UTR, and c) it was within 50 bp 3′ to the start of transcription to minimize effects of the 5′ UTR. The genes, promoter sequences, and primers used are listed in appendix 1, available at: http://www.uwcm.ac.uk/study/medicine/psychological_medicine/pub_data/Chr21.htm.

Approximately 5 μl of crude PCR product from each of 16 ethnically diverse individuals (eight subjects of Northern European origin, two of Zulu origin, two Chinese, two Australian Aborigines, one Warao, and one Kaingang; no further phenotypic data are available on these subjects) was screened for polymorphisms using DHPLC under conditions selected by DHPLCMelt (http://insertion.stanford.edu/melt.html) as described previously (16). Fragments yielding chromatograms indicating heteroduplex formation were sequenced using Big Dye Terminator chemistry (Applied Biosystems) to confirm and characterize polymorphisms. Minor allele frequency was estimated in pooled DNA from 100 local blood donors by automated sequencing.

Cloning

For high throughput we required a strategy of direct cloning into the luciferase expression vector pGL3 (Promega). To create a high-efficiency pGL3 T/A cloning vector, we first inserted an *Eco*RV site using standard procedures into the pGL3-basic multiple cloning site to create pGL3-basic-RV. This was cut with *Eco*RV (10 μg plasmid, 120 U *Eco*RV, BSA 100 μg/μl, 10 μl NEB buffer 3 in 100 μl) at 37°C for 18 h and purified using a Qiagen plasmid mini-prep kit. A 3′ T overhang was added by incubating 10 μg of cut vector at 75°C for 3 h in 10× PCR buffer (Qiagen) 20 μl, 25 mM $MgCl_2$ 8 μl, 100 M dTTP 2 μl, Taq DNA polymerase (Qiagen) 10 U in 200 μl final volume (19). Unpurified product was stored in single-use aliquots at −80°C.

DNA from heterozygous individuals was amplified using Expand High Fidelity DNA polymerase (Roche) to minimize misincorporation of nucleotides. PCR products to be cloned were purified using a Qiagen spin column. Amplimer (15 ng) and 50 ng of the prepared T-vector were ligated in 10 μl using T4 DNA ligase (Promega) at 4°C overnight or at RT for 1 h. After ligation, 1 μl of the mix was incubated at 37°C for 2 h with 20 U of *Eco*RV to cut any residual plasmid that lacked the additional 3′ T and had thus recircularized (this procedure was not carried out where the insert contained an *Eco*RV site). Approximately one tenth of the recut ligation mix (i.e., approximately 500 pg equivalent of vector) was used to transform 100 μl of SURE 2 supercompetent cells (Stratagene) (9). Colonies were picked using a sterile pipette tip, suspended in water (25 μl), and heated to 97°C for 5 min; 1.5 μl of the suspension was analyzed by PCR using the vector primers to measure the efficiency of the cloning (GL2 and RV3, Promega; annealing temperature 56°C). Correct orientation was determined using one gene-specific primer and one

plasmid-specific primer (RV3; Promega; annealing temp 58°C for all inserts).

Identification of Allelic Pairs

At least six clones, with the insert correctly orientated, were picked to identify allele pairs. To limit the number of clones to be sequenced, allele pairs were identified by mixing the PCR products from pairs of clones and looking for heteroduplex formation by DHPLC under the same conditions originally required for polymorphism detection. Each mixture was heated to 97°C for 3 min and allowed to reanneal prior to analysis by DHPLC. The component plasmids of a mixture showing the signature of a heterozygote were isolated from the SURE cells using standard procedures, sequenced to confirm fidelity, and used to transfect eukaryotic cells in culture.

Cell Culture

The ability of each sequence to promote transcription of the luciferase gene was tested transiently in human cell lines HEK293t (human embryo kidney, a gift from GlaxoSmithKline) and JEG-3 (human placenta; ECACC). Cell lines were transfected with plasmid using lipofectamine following the manufacturer's (Gibco) protocol with modifications described below. Cell lines were cultured according to ECACC (European Collection of Cell Cultures; www.ecacc.org.uk) specifications at 37°C with 5% CO_2. Cells were seeded into black, clear bottom 96-well luminometric plates (Perkin Elmer) at approximately 60% of confluencey the day prior to transfection. Plates seeded with HEK293t were coated with lysine (Sigma) prior to seeding. Prior to transfection, all plasmids were quantitated fluorimetrically using Pico Green (Molecular Probes) and a TD-700 (Turner Designs) fluorimeter. HEK293t cells were transfected with 80 ng/well DNA, 1 μl/well lipofectamine; JEG3 cells were transfected with 80 ng/well DNA, 0.5 μl/well lipofectamine, 0.5 μl/well PLUS-reagent. To provide a control for transfection efficiency, cells were cotransfected with CMV-SPAP (an expression plasmid containing the gene-secreted placental alkaline phosphatase under the control of the cytomegalovirus promoter, a gift from GlaxoSmithKline). A ratio of approximately 1:4000 control to test DNA was used. Cell lines were transfected overnight in serum-free medium, which was replaced with complete heat-inactivated medium (PAA Laboratories) and incubated for a further 24 h.

SPAP activity was measured in the culture medium after transferring to a second 96-well plate using a phospha-light kit (Tropix) according to manufacturer's instructions. Luciferase activity in the remaining cells was measured in the original plate using a Luc Screen assaying kit (Tropix). Both plates were read on a TR717 (Berthold Technologies) single photon counting luminometer for 1–10 s per well.

RESULTS

We attempted to amplify by PCR 83 gene promoters (Appendix I; http://www.uwcm.ac.uk/study/medicine/psychological_medicine/pub_data/Chr21.htm). Of these 51 (61%) could be optimized using standard conditions. Those that failed to amplify generally had a high GC content. Reagents that facilitate amplification of high GC fragments (such as DMSO) were not used as we have previously found that these promote misincorporation of nucleotides and the amplimers prove difficult to clone due to frequent rearrangements despite the use of SURE cells (4). Clearly this precludes functional characterization. Of the 51 promoter sequences we were able to amplify, 30 (59%) contained one or more sequence variants in at least one individual. A total of 65 polymorphisms were found (Appendix II; http://www.uwcm.ac.uk/study/medicine/psychological_medicine/pub_data/Chr21.htm). Ten of the 30 polymorphic promoters proved impossible to clone, the principal reason being frequent rearrangement and partial deletion of the insert in the plasmid and not ligation failure (4). A total of 53 different haplotypes for the 20 cloned promoters were identified, representing 41 different sequence variants. The haplotypes found are listed in Table 1.

All but one putative promoter sequence (*TTC3*) displayed activity greater than that of the promoterless pGL3-basic-RV plasmid. However, since the design of the primers used for this promoter (June 2000), a number of alternative upstream starts of transcription have been predicted for this gene and it is probable that the sequence used in the present study does not represent the true promoter for this gene. Therefore, excluding the three variants we detected in the *TTC*3 putative promoter, we assessed a total of 38 sequence variants for functional effects.

We have previously suggested that three criteria be met for significant activity differences between promoter haplotypes in our system (12). First, the differences must be statistically significant; second, that statistically significant replication must be achieved with independent construct preparations; and third, that the ratio of the low- to high-expression haplotype be a minimum of 0.67 (representing a 1.5-fold difference in expression between haplotypes). The last cri-

TABLE 1
LIST OF GENES TESTED IN REPORTER GENE ASSAYS

Gene Symbol	Haplotype Designation	Activity Relative to Control		Relative Allele Activity		Haplotype
		HEK293T	JEG-3	HEK293T	JEG-3	
NCAM2	A	25	1.2	100	NA	C −358, G −300
	B	18	1	72*	NA	G −358, G −300
	C	12	1.1	50† (43, 57)	NA	C −358, A −300
IFNAR1	A	46	18	100	82	T −427, GT (5) −86
	B	45	22	98	100	C −427, GT (5) −86
	C	7	3.3	15† (2, 28)	15† (1, 29)	C −427, GT (13) −86
KCNJ15	A	76	4	100	100	G −106
	B	45	4	59† (49, 69)	97	T −106
B3GALT5	A	11.4	<1	100	NA	G −157
	B	7.5	<1	66† (55, 77)	NA	A −157
IGSF5	A	14	<1	100	NA	C −494, A −110
	B	9.5	<1	67* (57, 77)	NA	T −494, C −110
CRYAA	A	25	1.7	100	NA	A −549, G −464, G −426, del −420, A −295
	B	22	1.7	88	NA	A −549, C −464, G −426, del −420, A −295
	C	23	1.6	92	NA	G −549, G −464, T −426, del −420, A −295
	D	22	1.8	89	NA	G −549, G −464, G −426, A ins −420, A −295
	E	21	1.4	85	NA	G −549, C −464, G −426, del −420, A −295
	F	8	<1	29† (18, 40)	NA	A −549, G −464, G −426, del −420, G −295
C21ORF6	A	87	182	100	100	A ins −234, GAA del −222
	B	66	151	76	83	del −234, GAA ins −222
CCT8	A	21	894	84	85	G −443
	B	25	1050	100	100	C −443
SYNJ1	A	12	2.4	81	100	A −300
	B	15	1.8	100	75	G −300
PRKCBP2	A	12	1.8	100	100	A −371, C −161
	B	11	1.7	95	95	A −371, T −161
CRYZL1	A	190	110	93	96	G −312, T −209, C −23
	B	183	85	89	74	A −312, T −209, C −23
	C	206	115	100	100	G −312, T −209, A −23
ITSN1	A	31	1.9	100	68	A −384
	B	27	2.8	89	100	G −384
HLCS	A	148	24	93	89	G −430, T −426, G −353, G −182
	B	160	21	100	78	G −430, C −426, A −353, A −182
	C	141	22	88	82	G −430, T −426, A −353, G −182
	D	132	27	83	100	G −430, C −426, A −353, G −182
	E	133	26	83	96	A −430, T −426, A −353, G −182
TTC3	A	<1	<1	NA	NA	T −448, C −411, T −154
	B	<1	<1	NA	NA	C −448, T −411, C −154
DYRK1A	A	8.4	<1	100	NA	A −87
	B	7.9	<1	94	NA	G −87
DSCR4	A	86	259	82	98	C −640
	B	104	264	100	100	G −640
WRB	A	34	31	92	78	A −485, A −295, G −174
	B	37	40	100	100	G −485, G −295, A −174
	C	29	32	79	80	G −485, A −295, G −174
SH3BGR	A	10	8.3	100	100	A −402, A −370
	B	10	8.1	98	97	T −402, G −370
FTCD	A	17	3.3	88	100	G −345, A −292
	B	20	2.9	100	88	T −345, C −292
ITGB2	A	18	14	100	88	A −414, T −384
	B	18	16	98	100	G −414, G −384
	C	15	15	82	95	A −414, G −384

Reporter gene activity corrected by SPAP is given relative to the promoterless pGL3 basic vector. Relative allele activity is expressed as a percentage of the highest expressing haplotype. Where the difference is significant according to our full criteria, 95% confidence intervals for the difference are given in parentheses. NA denotes levels of reporter gene activity not high enough for analysis.
*$p < 0.005$ based on a comparison with the most active haplotype.
†$p < 0.001$ based on a comparison with the most active haplotype.

terion was chosen because 1) homozygosity for the high-expression haplotype is conceptually equivalent to triploidy for the low-expression haplotype and 2) our empirical replication studies suggest that changes of this magnitude are highly robust, and almost always replicate in independent experiments using independent preparations of plasmid. Polymorphisms in six promoter sequences (*NCAM2, IFNAR1, KCNJ15, B3GALT5, IGSF5, CRYAA*) met these criteria in at least one cell line (Table 1). This figure represents 12% of the promoters we were able to analyze for the presence of sequence variants in our screening set, although if adjusted for the proportion of promoters we were unable to clone (33%), this figure rises to 18%. Of the promoter sequences with polymorphisms that we were able to clone ($n = 19$), six promoters with functional variants represent approximately 30% of the total.

Cell adhesion molecule, neural, 2 (*NCAM2*): In HEK cells, the C/G variant at −358 has a statistically significant effect (haplotype A compared with B) but falls short of our threshold of 1.5-fold difference. However, the G allele at position −300 drives expression twice as strongly as the A allele (haplotype A compared with C).

Interferon, alpha, beta, and omega, receptor 1 (*IFNAR1*): The T/C polymorphism at −427 has no effect on transcription (haplotype A compared with B) but the GT repeat at −86 has a major effect in both cell lines, the 5 repeat allele having a sevenfold higher activity than the 12 repeat allele.

Potassium channel, inwardly rectifying, subfamily J, member 15 (*KCNJ15*): In HEK cells, the G at position −106 increases activity of the promoter by 1.7-fold compared with T at that position.

B3 galactosyltransferase 5 (*B3GALT5*): The G allele at −157 was analyzed and at this position gives 1.5-fold higher activity than the A allele in HEK cells.

LOC150084 (*IGSF5*): Two polymorphisms were found in the screening set but only two different haplotypes were identified and so the effects of each variant could not be determined individually. In HEK cells, the C −494, A −110 haplotype had approximately twice the activity of the T −494, C −110 haplotype.

Alpha A crystallin (*CRYAA*): Five sequence variants were found represented in six haplotypes. Five of the haplotypes (A, B, C, D, and E) had similar activity showing that the variants at −549, −464, −426, and −420 had no effect on transcriptional activity. However, a comparison of haplotypes A and F reveals that a G at position −295 lowers the activity by over threefold compared with an A at that position. Analyses in each cell line were qualitatively similar, although the level of expression in the JEG cell line was below our threshold for analysis.

Viewed from the perspective of the individual polymorphisms, of the 38 variants we were able to examine, 7 were associated with altered function, giving a point estimate of 18% for the proportion of promoter polymorphisms that, in our assay system, have functional effects.

DISCUSSION

We have sought to determine if human promoters frequently harbor polymorphisms of functional importance. To avoid biases that may be introduced by selecting promoters that have been identified by others as of being of particular interest, we carried out polymorphism discovery on a randomly selected group of 51 gene promoters mapping to human chromosome 21. Whether selection for a common chromosomal location introduces a different type of bias is unknown. Allowing for dropouts at the PCR and cloning stage, we estimate that in our panel of subjects, around 18% of gene promoters on chromosome 21 have sequence variants that affect the transcriptional activity of the gene by an amount equivalent to that produced by an additional copy of the gene. However, the proportion of polymorphic promoters is clearly dependent upon the number of individuals studied, and therefore perhaps a more useful estimate for geneticists is our estimate that functional differences between alleles or haplotypes were found for 18% of the individual polymorphisms studied. This estimate is based upon observation that for five of the six promoters with functionally different haplotypes we were able to identify a single sequence variant responsible for a 1.5 change in transcriptional activity; for the other gene, LOC150084 (*IGSF5*), the contribution of the two sequence variants could not be distinguished and we therefore include both as functional.

In our previous study based upon a more biased series of gene promoters from the eukaryotic promoter database, allowing for dropouts, around 10% of promoters had functional sequence variants (12). As that study was based upon screening half the number of subjects for sequence variants as the present study, that lower estimate is roughly consistent with the present findings. It is more difficult to compare the estimates of the proportion of polymorphisms with functional consequences because, in our earlier study, the effects of individual SNPs in haplotypes could not as readily be distinguished as in the present study.

We have previously proposed a conservative threshold for defining true promoter activity as 10 times the background activity found with the promoterless pGL3-basic plasmid in any cell line (5). This was based upon our observation that randomly selected sequences may occasionally show activity up to this level. One sequence selected for the gene *TTC* is probably not a gene promoter at all and was selected on the basis of incorrect annotation of the human genome sequence. *DYRK1A* had maximum activity approximately eight times that of the pGL3-basic-RV,

and therefore we cannot be confident we examined a true DYRK1A promoter. However, this sequence is, in the most current database entries, still adjacent to the putative start of transcription and therefore is still very likely to be a promoter (5).

Although the genes for this study were selected in a random fashion to provide relatively unbiased estimates of general importance of the extent to which promoter polymorphisms are functional, some of our data regarding specific promoters are likely to be of considerable interest to specific research groups. These data are briefly discussed below.

NCAM2 is a neural cell adhesion molecule that has been proposed to be involved in the variable phenotypes seen in Down syndrome (21). It has also recently been shown that changes in its expression affects odorant receptor (OR) growth (1). Olfactory sensory neurons (OSNs) individually express only one of the hundreds of OR genes, and each projects with precision to topographically defined convergence sites in the glomeruli of the olfactory bulb. It has recently been shown that normally regulated levels of NCAM2 are important for accurate OR-specific axon segregation of some OSNs. Overexpressing transmembrane-bound NCAM2 results in co-convergence events close to the correct target glomeruli. By contrast, overexpression of Gpi-anchored NCAM2 results in axons that can bypass the correct target before co-converging on glomeruli located at a distance. This suggests that the functional promoter variant we have found may affect development of OSNs and possibly other neural systems.

IFNAR1 (and *IFNAR2*, which we were unable to study) belongs to the class II cytokine receptor family, and forms a heterodimer when bound to interferon-α or -β. This results in phosphorylation of the receptors as well as Jak and Stat proteins. Activated Jak-Stat proteins then initiate the transcription of *IFN*-inducible genes that are responsible for antiproliferative, immunomodulatory, and antiviral effects. Thus, *IFNAR1* is an important gene in many fields of research. Epstein and Epstein showed that fibroblasts with trisomy 21 had an exaggerated response to interferon and it is possible that this is due to increased gene dosage of *IFNAR1* (6). Matsuyama et al. studied the same GT repeat polymorphism that we found, and have shown that hepatitis C patients with the 12 repeat allele have lower response to interferon than those with the 5 allele repeat (20). We found that the 5 repeat allele has sevenfold greater expression than the 12 repeat allele, suggesting that the latter may lower response to interferon by lowering the expression levels of *IFNAR1*.

B3GALT5 is a galactosyltransferase, a membrane-bound protein, and one of a family of 14 β Gal-T genes (11). It is probably involved in the synthesis of type 1 Lewis antigens in gastrointestinal and pancreatic cancers (15) but its role in these and other processes has yet to be fully established.

IGSF5 is a predicted protein designated LOC150084 with similar sequence to the immunoglobulin superfamily. Since the primers for this promoter were designed, two additional upstream exons have been predicted (April 10, 2003 revision). Given that the sequence we examined has 11 times the promoter activity of the negative control, it seems likely that there are multiple promoters for this gene.

CRYAA: The importance of α-crystallins in the maintenance of lens transparency was demonstrated by the work of Brady et al., who showed that mice homozygous for a targeted disruption of the α A-crystallin gene developed cataracts and had cytoplasmic inclusion bodies containing the small heat shock protein α B-crystallin (2). Since then a number of missense and nonsense mutations have been described in the literature [reviewed in (8)] that can lead to either recessive or dominant cataracts. Our finding of a threefold difference between allelic variants of the promoter region suggests that the polymorphism involved may be a candidate for predisposition to the development of cataracts.

KCNJ15: The potassium channel inward rectifiers (Kir) constitute a family of potassium channels that maintain the resting membrane potential close to the equilibrium potential for potassium ions. *KCNJ15* is in the Down syndrome critical region (7). The murine equivalent of *KCNJ15* (*mkir4.2*) is developmentally regulated and found in tissues that may be affected in Down syndrome—heart, thymus, thyroid gland, and perichondrium—although there is no evidence that it is involved in the etiology of Down syndrome (22).

In summary, we have sought to obtain an unbiased estimate of the proportion of polymorphisms in the promoters of human genes with functional effects. Although our analysis is conservative, our data suggest that around 18% of variants in proximal promoters are functional, and provide strong empirical support for affording the promoters of genes high priority in genetic analyses. Of those promoter variants with functional effects, three (*NCAM2, IFNAR1*, and *CRYAA*) have been previously associated with human phenotypes. The polymorphisms we describe here may therefore play a role in those phenotypes.

ACKNOWLEDGMENT

This work was funded by the Medical Research Council, UK.

REFERENCES

1. Alenius, M.; Bohm, S. Differential function of RNCAM isoforms in precise target selection of olfactory sensory neurons. Development 130:917–927; 2003.
2. Brady, J. P.; Garland, D.; Duglas-Tabor, Y.; Robison, W. G., Jr.; Groome, A.; Wawrousek, E. F. Targeted disruption of the mouse alpha A-crystallin gene induces cataract and cytoplasmic inclusion bodies containing the small heat shock protein alpha B-crystallin. Proc. Natl. Acad. Sci. USA 94:884–889; 1997.
3. Bray, N. J.; Buckland, P. R.; Williams, N. M.; Williams, H. J.; Owen, M. J.; O'Donovan, M. C. A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. Am. J. Hum. Genet. 73:152–161; 2003.
4. Coleman, S. L.; Hoogendoorn, B.; Guy, C.; Smith, S. K.; O'Donovan, M. C.; Buckland, P. R. A streamlined approach to functional analysis of promoter polymorphisms. Biotechniques 33:412–418; 2002.
5. Coleman, S. L.; Buckland, P. R.; Hoogendoorn, B.; Guy, C.; Smith, K.; O'Donovan, M. C. Experimental analysis of the annotation of promoters in the public database. Hum. Mol. Genet. 11:1817–1821; 2002.
6. Epstein, L. B.; Epstein, C. J. Localization of the gene AVG for the antiviral expression of immune and classical interferon to the distal portion of the long arm of chromosome 21. J. Infect. Dis. 133(Suppl.):A56–A62; 1976.
7. Gosset, P.; Ghezala, G. A.; Korn, B.; Yaspo, M.-L.; Poutska, A.; Lehrach, H.; Sinet, P.-M.; Creau, N. A new inward rectifier potassium channel gene (KCNJ15) localized on chromosome 21 in the Down syndrome chromosome region 1 (DCR1). Genomics 44:237–241; 1997.
8. Graw, J.; Loster, J. Developmental genetics in ophthalmology. Ophthalmic Genet. 24:1–33; 2003.
9. Greener, A. *E.* coli SURE strain: Clone "unclonable" DNA. Strategies 3:5–6; 1990.
10. Hattori, M.; Fujiyama, A.; Taylor, T. D.; Watanabe, H.; Yada, T.; Park, H. S.; Toyoda, A.; Ishii, K.; Totoki, Y.; Choi, D. K.; et al. The DNA sequence of human chromosome 21. Nature 405:311–319; 2000.
11. Hennet, T. The galactosyltransferase family. Cell. Mol. Life Sci. 59:1081–1095; 2002.
12. Hoogendoorn, B.; Coleman, S.; Guy, C.; Smith, K.; Bowen, T.; Buckland, P. R.; O'Donovan, M. C. Functional analysis of human promoter polymorphisms. Hum. Mol. Gen. 12:2249–2254; 2003.
13. Hudson, T. J. Wanted: Regulatory SNPs. Nat. Genet. 33:439–440; 2003.
14. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409:860–921; 2001.
15. Isshiki, S.; Togayachi, A.; Kudo, T.; Nishihara, S.; Watanabe, M.; Kubota, T.; Kitajima, M.; Shiraishi, N.; Sasaki, K.; Andoh, T.; Narimatsu, H. Cloning, expression, and characterization of a novel UDP-galactose: Beta-N-acetylglucosamine beta1,3-galactosyltransferase (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal and pancreatic epithelia and tumor cells derived therefrom. J. Biol. Chem. 274:12499–12507; 1999.
16. Jones, A. C.; Austin, J.; Hansen, N.; Hoogendoorn, B.; Oefner, P. J.; Cheadle, J. P.; O'Donovan, M. C. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. Clin. Chem. 45:1133–1140; 1999.
17. Knight, J. C.; Keating, B. J.; Rockett, K. A.; Kwiatkowski, D. P. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat. Genet. 33:469–475; 2003.
18. Lo, H. S.; Wang, Z.; Hu, Y.; Yang, H. H.; Gere, S.; Buetow, K. H.; Lee, M. P. Allelic variation in gene expression is common in the human genome. Genome Res. 13:1855–1862; 2003.
19. Marchuk, D.; Drumm, M.; Saulino, A.; Collins, F. S. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. 19:1154; 1991.
20. Matsuyama, N.; Mishiro, S.; Sugimoto, M.; Furuichi, Y.; Hashimoto, M.; Hijikata, M.; Ohta, Y. The dinucleotide microsatellite polymorphism of the IFNAR1 gene promoter correlates with responsiveness of hepatitis C patients to interferon. Hepatol. Res. 25:221–225; 2003.
21. Paoloni-Giacobino, A.; Chen, H.; Antonarakis, S. E. Cloning of a novel human neural cell adhesion molecule gene (NCAM2) that maps to chromosome region 21q21 and is potentially involved in Down syndrome. Genomics 43:43–51; 1997.
22. Thiery, E.; Gosset, P.; Damotte, D.; Delezoide, A. L.; de Saint-Sauveur, N.; Vayssettes, C.; Creau, N. Developmentally regulated expression of the murine ortholog of the potassium channel KIR4.2 (KCNJ15). Mech. Dev. 95:313–316; 2000.
23. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; et al. The sequence of the human genome. Science 291:1304–1351; 2001.
24. Yan, H.; Yuan, W.; Velculescu, V. E.; Vogelstein, B.; Kinzler, K. W. Allelic variation in human gene expression. Science 297:1143; 2002.