Structure and transcriptional regulation of the *Nat2* gene encoding for the drug-metabolizing enzyme arylamine N-acetyltransferase type 2 in mice

Sotiria BOUKOUVALA*, Naomi PRICE*, Kathryn E. PLANT† and Edith SIM*¹

*University of Oxford, Department of Pharmacology, Mansfield Road, Oxford OX1 3QT, U.K., and †Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, U.K.

Arylamine N-acetyltransferases (NATs) are polymorphic enzymes, well-known for their role in the metabolism of drugs and carcinogens. Mice have three NAT isoenzymes, of which NAT2 is postulated to be involved in endogenous, as well as xenobiotic, metabolism. To understand expression of the murine *Nat2* gene, we have analysed its structure and transcriptional regulation. We have accurately mapped the transcription initiation site 6.5 kb upstream of the coding region of the gene, adjacent to a recently described non-coding exon. Transcription was demonstrated to start from this region in embryonic and adult liver, spleen, submaxillary gland, kidney, brain, thymus, lung and placenta, but not in the heart. Database searches and analyses of cDNA by PCR suggested alternative splicing of the single 6.2 kb intron of *Nat2*, and determined the position of the polyadenylation signal at 0.44 kb downstream of the coding region of the gene. Examination of the 13 kb sequence flanking the coding and non-coding

exons of *Nat2* revealed a single promoter, located close to the transcription-initiation site, and indicated regions likely to harbour control elements. The *Nat2* promoter consists of an atypical TATA box and a Sp1 [SV40 (simian virus 40) protein 1] box identical with that found in many housekeeping gene promoters. Activity of the *Nat2* promoter was severely reduced by deletion or mutation of either of these two elements, whereas the region of the Sp1 box bound cellular protein and resisted DNase I digestion. Finally, the ability of the promoter region to bind cellular protein was reduced by competition with oligonucleotides bearing the Sp1 consensus sequence.

Key words: arylamine, arylamine N-acetyltransferase, drug metabolism, murine *Nat2* promoter, N-acetylation, xenobiotic-metabolizing-enzyme gene expression.

INTRODUCTION

Arylamine N-acetyltransferases (NATs; EC 2.3.1.5) and their homologues are found in many species. The enzymes catalyse the N-acetylation of arylamines and hydrazines from acetyl-CoA. Substrates include xenobiotics, although certain NAT isoenzymes have been proposed to have an endogenous role as well [1]. In humans, there are two NAT isoenzymes, NAT1 and NAT2, and their acetylation reactions may be affected by genetic polymorphisms in the *NAT1* and *NAT2* genes. There are well-established associations between 'slow acetylator' *NAT2* alleles and sensitivity to the toxic effects of drugs, while the *NAT1/NAT2* genotype may influence susceptibility to chemically induced carcinogenesis [2].

The two human NAT isoenzymes exhibit distinct tissue expression profiles and substrate specificities [2–4]. Whereas human NAT2 is a xenobiotic-metabolizing enzyme expressed predominantly in the liver and intestine, NAT1, which also acetylates arylamine carcinogens, is expressed in many tissues and is postulated to have an endogenous role. The folate catabolite pABGlu (*p*-aminobenzoylglutamate) has been identified as a potential endogenous human NAT1 substrate [5,6].

The mouse is a good model for studying the endogenous role of NAT. N-acetylation of pABGlu is carried out by murine NAT2 [7], the isoenzyme homologous with human NAT1 [7–9]. In contrast, murine NAT1, expressed in adult liver, appears to be an exclusively xenobiotic-metabolizing enzyme [10]. Mice also have a third NAT isoform, NAT3, of unknown function, although it may be expressed in spleen [10,11]. The three murine *Nat* genes, each showing polymorphisms [10,12], are clustered in a 57 kb

region of chromosome 8 [13] [MapViewer (mouse) at http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=mouse_chr.inf]. Although murine NAT2 has been detected in many tissues [9], differences in the amount of RNA, NAT2 protein and enzymic activity have been observed between tissues from embryonic [14], newborn [15] and mature mice [7–9,16]. It has also been demonstrated that renal expression of *Nat2* in mice may be subject to regulation by androgens [16].

All of the genes encoding NAT protein in mice and humans have an intronless open reading frame [1]. Non-coding exons have been described for the human *NAT2* gene through the comparison of genomic DNA and cDNA sequences [17] and, more recently, for the murine *Nat2* gene [13]. Genes with an intronless coding region account for < 5 % of all human genes [18] and often have upstream non-coding exons [19,20]. It is possible that efficient gene expression may require intron splicing, although the coding region may be contained in a single exon [21].

To understand how expression of the *Nat* gene family is regulated, we have investigated transcription of the mouse *Nat2* gene. A 14 kb region of murine genomic DNA harbouring the open reading frame of *Nat2* gene and part of an upstream non-coding exon [13] has been analysed to establish the structure of the gene and its transcriptional regulation.

EXPERIMENTAL

Materials

Chemicals were from Sigma-Aldrich (Poole, Dorset, U.K.) or BDH-Merck (Lutterworth, Leics., U.K.). Reagents for bacterial

Abbreviations used: AP1, activation protein 1; CREB, cAMP-response-element-binding protein; *Dhfr*, dihydrofolate reductase; DTT, dithiothreitol; EMSA, electrophoretic mobility-shift assay; ES cells, embryonic stem cells; EST, expressed sequence tag; *Hprt*, hypoxanthine phosphoribosyltransferase; NAT, arylamine N-acetyltransferase; pABA, *p*-aminobenzoic acid; pABGlu, *para*-aminobenzoylglutamate; RT-PCR, reverse transcription-PCR; Sp1, SV40 protein 1; SV40, simian virus 40; VAI, virus-associated RNA I.

¹ To whom correspondence should be addressed (e-mail edith.sim@pharm.ox.ac.uk).

Table 1 Primers used in RT-PCR, RNase protection assay and EMSA experiments

The primer annealing sites are shown relative to the 5' end of a 14 kb mouse genomic plasmid clone with GenBank® accession number AJ250123 [13].

Primer name	Orientation	Sequence (5' → 3')	T _m (°C)	Specificity
PolyA-aR	Reverse	10860-AAATTAGCTGTAAGCCAATGCTG-10837	58	Region upstream of polyA1, polyA2–polyA5 and polyA6 putative polyadenylation signals of the <i>Nat2</i> gene respectively
PolyA-bR	Reverse	11131-ACTTACTCAATTGGTGTTCCTCC-11107	60	
PolyA-cR	Reverse	11386-GCCTAGGAATGTGCTTCTATGTGG-11362	63	
Ribo-1F	Forward	3351-TACATTTCCAGAGATCC-3368	52	
Ribo-1R	Reverse	3598-TCAGACTGTGTGTGCT-3581	54	Riboprobe spanning <i>Nat2</i> transcription initiation site
GS-1(F)	Forward	3293-CTCAGCAAAGAACGGAC-3310	62	
GS-1(R)	Reverse	3437-GGAAGAGCTCTTATTGGC-3420	56	GS1 EMSA probe
GS-2(F)	Forward	3137-GGAATGTGGTGACATG-3153	58	
GS-2(R)	Reverse	3293-GATATCATGATATTTTGC-3275	49	GS2 EMSA probe
GS-3(F)	Forward	2979-ACATAGGACAGTTGTG-2995	51	
GS-3(R)	Reverse	3138-CCTAATGATGCTTTCC-3121	54	GS3 EMSA probe
GS-4(F)	Forward	2818-ACTACATGCCACCAAGC-2834	57	
GS-4(R)	Reverse	2996-ACGACAACGTCTATG-2980	51	GS4 EMSA probe
GS-5(F)	Forward	2702-CCCTCAGAACCATGCAG-2718	60	
GS-5(R)	Reverse	2817-CGGGTGTGGCTCAGGTG-2801	67	

culture were from Becton Dickinson (Oxford, U.K.) and for mammalian cell culture from GibcoBRL (Paisley, U.K.). Molecular biology reagents, kits, enzymes, JM109 *Escherichia coli* cells and vectors were from Promega (Southampton, U.K.). ³²P-labelled nucleotide triphosphates were from Amersham Biosciences (Little Chalfont, Bucks., U.K.) and primers from Sigma-GenoSys (Cambridge, U.K.). The Balb/c mice were from Harlan Olac (Bicester, Oxon., U.K.).

Mammalian cell culture

The BNL.CL2 mouse embryonic liver cell line (American Type Culture Collection, Manassas, VA, U.S.A.) was maintained in an atmosphere of 10% CO₂ in air, using 90% (v/v) high-glucose (4.5 g/l) Dulbecco's modified Eagle's medium supplemented with 10% (v/v) foetal calf serum, 2 mM L-glutamine and antibiotics (0.1 mg/ml kanamycin, 0.1 mg/ml streptomycin, 100 units/ml penicillin).

PCR and RT (reverse transcription)-PCR

Total RNA was extracted from cultured cells or frozen tissue (60–100 mg) with the TRIzol® Reagent (Sigma), treated with RNase-free DNase I (1 unit/μg of RNA) and used for cDNA synthesis as previously described [10].

PCR was performed with either cDNA or plasmid DNA (50 ng) from a mouse genomic clone (GenBank® accession number AJ250123) [13]. The 50 μl reaction mixture contained 3 units of *Pfu* DNA polymerase, 200 μM of each dNTP, 0.25 μM of each primer (Table 1) and the appropriate buffer. Cycling conditions were 95 °C for 5 min, followed by 35 cycles of 30 s at 95 °C, 30 s at the appropriate annealing temperature, 45 s per 1 kb of target sequence at 72 °C, and a final extension at 72 °C for 5 min.

Reporter gene assays

PCR products were introduced to the *Mlu*I and *Xho*I sites of pGL3 *Photinus pyralis* luciferase reporter vectors (Basic or Promoter). Successful incorporation of each insert was confirmed by PCR, restriction digest and sequencing of selected clones. Each pGL3 construct (10 μg) was co-transfected into BNL.CL2 cells with an equal amount of pRL-TK *Renilla reniformis* luciferase reporter vector, used as an internal standard (ProFection Mammalian Transfection System, Promega).

Confluent cells were lysed and were subjected to reporter gene assay, using the Dual-Luciferase Reporter Assay System (Promega). Luminescence was measured on a LB Wallac 1250 luminometer (kindly made available by Professor G. Brownlee, Sir William Dunn School of Pathology, University of Oxford, Oxford, U.K.).

RNase protection assay

A ³²P-labelled antisense riboprobe was generated and prepared as described [22] using a PCR fragment amplified with primers Ribo-1F/Ribo-1R (Table 1), subcloned into the pGEM-T Easy vector, linearized at the *Nde*I site of the vector and used as template (1 μg in a 20 μl reaction) for T7 RNA polymerase in 10 mM DTT (dithiothreitol), 1 unit/μl RNase inhibitor, 500 μM of each rATP, rCTP and rGTP, 25 μM rUTP and 40 μCi of [α-³²P]rUTP (37 °C for 2 h). The generated riboprobe was purified from a denaturing polyacrylamide gel and was precipitated with ethanol [22]. The pellet was resuspended in 100 μl of R-loop buffer [80% (v/v) formamide, 40 mM Pipes, pH 6.4, 1 mM EDTA and 400 mM NaCl] and 1 μl (approx. 500 c.p.s.) was added to R-loop buffer (29 μl) containing 10 μg of purified total RNA, extracted from BNL.CL2 cells co-transfected with 15 μg of AJ250123 DNA and 5 μg of a pUC18 plasmid carrying the adenoviral VAI (virus-associated RNA I) gene (pVAI). Denaturation (80 °C for 10 min) was followed by overnight hybridization at 56 °C. Digestion with 40 μg/ml RNase A and 1000 units/ml RNase T₁ in 10 mM Tris/HCl (pH 7.5), 5 mM EDTA and 300 mM NaCl (18 °C for 2 h) was terminated with 10 μl of 10% (w/v) SDS and 5 μl of 10 mg/ml proteinase K (37 °C, 30 min). Following acid phenol/chloroform extraction and ethanol precipitation, the pellet was resuspended in 10 μl of formamide dye, denatured (80 °C, 10 min) and run on a 6% (w/v) denaturing polyacrylamide gel [22]. Controls lacking RNA or not subjected to RNase digestion were included in each set of reactions. A similarly generated riboprobe, hybridizing to the 3' end of the VAI transcript, was used as internal standard.

EMSA (electrophoretic mobility-shift assay)

BNL.CL2 cells were harvested in 40 mM Tris/HCl (pH 7.4), 1 mM EDTA and 150 mM NaCl, pelleted by centrifugation (250 g, 4 min) and resuspended in 500 μl of 40 mM Hepes (pH 7.9), 0.4 M KCl, 1 mM DTT, 10% (v/v) glycerol, 0.1 mM

Table 2 Primers used in reporter gene assays and DNase I footprinting

The primer annealing sites are shown relative to the 5' end of clone AJ250123. Primers marked with a (F) are forward, and those marked with a (R) are reverse. Restriction sites are shown in bold for *Mlu*I, in bold and italics for *Xho*I, underlined for *Nde*I, double underlined for *Not*I and shaded for *Xba*I. Constructs B1–B7 and P1–P6 cover the region from the *Nat2* non-coding exon [13] up to 3.2 kb upstream, constructs B8–B18 and P8–P18 span the 6.2 kb intron of *Nat2*, whereas constructs B19–B21 and P19–P21 extend 3.1 kb downstream of the *Nat2* coding exon. Primers DEL, SP1 and TATA were used for deletion or mutation of the *Nat2* promoter. Probe is the PCR-generated fragment used in DNase I footprinting assays.

Primer name	Sequence (5' → 3')	<i>T_m</i> (°C)	Construct (pGL3)
REP1(F)	418-GCCCAACATGGTGA ACGCGT TTTACTATTAGG-452	76	B1, P1
REP2(R)	988-CGGGTGCACTGG CTCGAG TTCTGTTTGAGGACTG-954	83	
REP3(F)	938-GATTCAGGTA ACTACGCGT CCTCAAAACAG-968	72	B2, P2
REP4(R)	1464-ACTGTTTACAT CTCGAG TCAGATCCACAAG-1434	70	
REP5(F)	1424-ACGGGTAA ACGCGT GGATCTGATTCAAG-1452	75	B3, P3
REP6(R)	1919-TGGGAATCCTGATAG CTCGAG ACTCTCCGATG-1888	78	
REP7(F)	1850-TCTCCATAGCT ACGCGT ACTGCTGGACAGG-1879	76	B4, P4
REP8(R)	2316-CTATGCAGTGCAAA CTCGAG GCCTGGCATTGG-2285	82	
REP9(F)	2283-TTCCAATGCC ACGCGT CTAGTTTTCAGCTG-2311	78	B5, P5
REP10(R)	2821-TAGTCGGGTGTGG CTCGAG TGAGGAGAAGTGAG-2789	80	
REP11(F)	2720-GAGGTATTGCTGTGC ACGCGT CACAGTG-2747	78	B6, P6
REP12(R)	3286-TGATATTTTGCATT CTCGAG ATAGTTTG-3256	69	
REP17(F)	3305-ACGGACTAGAGA ACGCGT TCACCGGGTTAAG-3336	77	probe
PROM-1(F)	3255-GCAAACCTATC ACGCGT AATGCAAAAATATC-3285	65	B7
EXON-1(R)	3620-AAAGAACTTGGGC CTCGAG AGTCAGGACTG-3590	70	B7, probe
RA-1(F)	3663-ACTTTCAGTAC ACGCGT GATTCTGACAA-3691	74	B8, P8, B9
RA-1(R)	4776-ACAGTAGTCAGA CTCGAG GGAGGATTACAC-4747	73	B8, P8
REP20(R)	4142-CCAAAGTACAGA CTCGAG TTTCATGATATCAG-4112	69	B9
RA-2(F)	4733-TATCGGCCCTA ACGCGT GAATCTCTCTTC-4761	79	B10, P10, B11
RA-2(R)	5772-TGTAATAACAC CTCGAG TCACACGAAACC-5743	71	B10, P10
REP22(R)	5226-AGTTTGGTCAGC CTCGAG TTTCATTCATG-5197	73	B11
RA-3(F)	5728-GACAGATGTGA ACGCGT TCGTGTGAGTC-5757	75	B12, P12, B13
RA-3(R)	6880-GCATATTATGG CTCGAG AACAGGCTATTTC-6851	71	B12, P12
REP24(R)	6247-TTGATGGCACAC CTCGAG TGATGCACCTGTTC-6216	80	B13
RA-4(F)	6843-TATAAGAGGAAT ACGCGT GTTCAGGAGCCCA-6872	74	B14, P14, P15
RA-4(R)	8009-CCACTAAAGGAT CTCGAG TAGTATTTGATTC-7979	66	B14, P14
REP26(R)	7319-CATTTTAAGAAAA CTCGAG TTTCATACATATG-7288	66	B15
RA-5(F)	7830-CACTCCCTGGT ACGCGT TTCATCTCTATTTTC-7862	76	B16, P16, B17
RA-5(R)	8939-ATAATGATGGGG CTCGAG GGAAATTCCTGTG-8909	76	B16, P16
REP28(R)	8443-AGGAGTAGTTAT CTCGAG CGTACATTATGAG-8413	67	B17
RA-6(F)	8838-TCTCTGTGAGCT ACGCGT GAATCTATAGTATTC-8870	70	B18, P18
INTRON-1(R)	9850-TGGTTTC CTCGAG GCAAGAAAAACAATCC-9823	66	
RA-7(F)	11177-TATATATCTAA ACGCGT TTTCAAGTGG-11205	65	B19, P19
RA-7(R)	12183-AATGTCACAATT CTCGAG TTAATGACATTAC-12153	66	
RA-8(F)	12132-ATCTTTAGATA ACGCGT GGGTAATGTCATTAAG-12166	70	B20, P20
RA-8(R)	13280-TGTGGTTTCT CTCGAG AATTAAGGGGAAC-13249	74	
RA-9(F)	13194-GCAGATAAAGAA ACGCGT GACCAAGTCCCAATG-13224	77	B21, P21
RA-9(R)	14286-ATTCTTGGATT CTCGAG GGGTTGAGAATTG-14256	74	
DEL(F)	3409-CAACTTGAACAGCCCATATGAGCTCTTCC-3437	73	DEL
DEL(R)	3392-TAGTTTTTTTTAAACCATATGAAGATCTC-3362	65	
SP1(F)	3388-ACTAAGCATAAGCGCGCGCACTTGAACA-3419	80	SP1(mt)
SP1(R)	3419-TGTTCAAGTTGCGCGCGCGCTATGCTTAGTT-3388	80	
TATA(F)	3366-TCCTTGAGATGGTCTAGAAAAAACTAAGC-3395	67	TATA(mt)
TATA(R)	3395-GCTTAGTTTTTTCTAGACCATCTCAAGGA-3366	67	

PMSF and 0.1 % (v/v) aprotinin. Lysates were produced by quick freeze–thawing of cells in a dry ice/ethanol bath and stored at -70°C after centrifugation for 5 min in a microcentrifuge (15000 g, 4°C). PCR-generated probes (3.5 pmol) were end-labelled with 10 μCi of [γ - ^{32}P]dATP, and unincorporated label was removed by gel filtration [22]. Cell extracts (2–4 μg of total protein) were incubated [30 min, room temperature (20°C)] with 35 fmol of radioactive probe in 10 μl of binding buffer containing 4 % (v/v) glycerol, 1 mM MgCl_2 , 0.5 mM EDTA, 0.5 mM DTT, 70 mM NaCl, 10 mM Tris/HCl (pH 7.5) and 50 $\mu\text{g}/\text{ml}$ poly(dI-dC)·(dI-dC). The products were separated on a 4 % (w/v) non-denaturing polyacrylamide gel [22]. Control reactions either lacked cell extract or contained 1.75 pmol of either non-specific or specific non-radioactive competitor. Unlabelled double-stranded oligonucleotides (1.75 pmol), bearing the consensus sequences for AP2 (activation protein 2), Sp1 [SV40 (simian virus 40)

protein 1], AP1 (activation protein 1; c-Jun), Oct-1 (octamer motif recognizing transcription factor 1), CREB (cAMP-response-element-binding protein) and NF- κB (nuclear factor κB) (Promega) were also used in competition assays.

DNase I footprinting assay

A probe, generated by PCR with primers REP17(F)/EXON-1(R) (Table 2), was ^{32}P -labelled as above, with 20 units of T4 polynucleotide kinase (37°C , 1 h). The probe was digested with *Xho*I, to remove the radioactive label on the side of primer EXON-1(R), and eluted in 100 μl of Tris/HCl (pH 7.4), after purification with the QIAquick Gel Extraction Kit (Qiagen, Crawley, West Sussex, U.K.). DNase I footprinting assays used the Core Footprinting System (Promega). Binding reactions (except control) contained 5 μl of HeLa nuclear extract (Promega) and 50 ng/ μl

Figure 1 Detection of 31 kDa murine NAT2 protein in BNL.CL2 cell lysates

Western blot analysis was performed with antiserum 184, specific for murine NAT2 [8]. Lane 1 is 10 µg of recombinant NAT2 protein with a hexahistidine tag, used as reference. Lanes 2–6 are, respectively, 10, 25, 50, 75 and 100 µg of total protein contained in BNL.CL2 cell lysates. The position of the 31 kDa band is indicated.

poly(dI-dC) · (dI-dC). Products digested with DNase I (0.15 unit, 3 min) were separated on a 6% (w/v) denaturing polyacrylamide gel and were autoradiographed [22].

NAT activity assay and Western blotting

BNL.CL2 cells were harvested in 20 mM sodium phosphate (pH 7.4) and 145 mM NaCl, pelleted by centrifugation (250 g, 4 min) and resuspended in 20 mM Tris/HCl (pH 7.5), 1 mM EDTA and 1 mM DTT (1 ml/75 cm² flask). Cells were lysed by sonication on ice (three lots of 5 s at 10 s intervals). Aliquots (stored at –70 °C) were thawed, centrifuged (14 000 g for 10 min at 4 °C), and the supernatant was used either (within 1 h) to measure NAT2 activity with 0.2 mM pABA (*p*-aminobenzoic acid) or for SDS/PAGE and Western blotting with antiserum 184, as previously described [10]. Recombinant murine NAT2 protein, used as a standard in Western blot analysis, was expressed from a pET28b plasmid clone carrying the entire *Nat2* coding region and a hexahistidine tag, as described for other *nat* genes [23].

RESULTS AND DISCUSSION

Characterization of the BNL.CL2 mouse embryonic liver cell line

As a prerequisite to studying *Nat2* gene structure and transcriptional regulation, expression of NAT2 protein and *Nat2* mRNA was investigated in the mouse embryonic liver cell line BNL.CL2. NAT2 protein was detected in BNL.CL2 cell lysates by Western blot analysis with the isoenzyme-specific antiserum 184 [8] (Figure 1). In the same lysates, NAT2-specific enzymic activity was 1.2 nmol of pABA N-acetylated/min per mg of protein. RT-PCR showed that *Nat2* is the only *Nat* gene transcribed in BNL.CL2 cells (Figures 2a–2c) and confirmed (Figure 2d) that the *Nat2* transcript contains part of a non-coding exon, located approx. 6.3 kb upstream of the *Nat2* coding region, as has previously been found in murine ES (embryonic stem) cells [13]. The expression of the *Nat2* gene, and not the other *Nat* genes, in embryonic liver is consistent with observations on the ontogeny of the murine *Nat* genes. Murine *Nat2* has been identified as being expressed early in development [7,10,24], whereas the other *Nat* genes are not.

Mapping of the mouse *Nat2* gene transcription initiation site in BNL.CL2 cells

A 247 nt PCR-amplified cloned riboprobe was used to accurately map the transcription-initiation site of murine *Nat2* by RNase protection assay. The riboprobe was designed to extend 202 nt upstream of the primer-annealing site at the beginning of a RT-PCR product with GenBank® accession number AJ251710, known to contain part of the *Nat2* non-coding exon [13]. Following RNase treatment, the flanking plasmid linkers and at least

Figure 2 *Nat* gene expression in BNL.CL2 cells, detected by RT-PCR

The primers used are described in [13]. (a)–(c) Amplification from genomic DNA (lane 1) and cDNA (lane 2), as well as from the product of a reaction lacking reverse transcriptase (lane 3). Lane 4 is the PCR control with no DNA. (a) RT-PCR with *Nat1*-specific primers Mus12 and Mus13, showing absence of *Nat1* transcript. (b) RT-PCR with *Nat2*-specific primers mNAT2 1 and mNAT2 910, left, demonstrating *Nat2* expression. Amplification of the *Gapdh* gene (encoding for glyceraldehyde-3-phosphate dehydrogenase) with primers GAPDH-S and GAPDH-AS [10], right, was used as a positive control for all RT-PCR reactions presented in this Figure. (c) RT-PCR with *Nat3*-specific primers Mus12 and Mus15, showing absence of *Nat3* transcript. (d) Amplification from BNL.CL2 cDNA with primers NCE-F (specific for the *Nat2* non-coding exon) and mNAT2 910 (specific for the *Nat2* coding region) in lane 1, as well as mNAT2 1 and mNAT2 910 in lane 2. Lane 3 shows amplification from the product of a reaction lacking reverse transcriptase (primers mNAT2 1 and mNAT2 910), whereas lanes 4 and 5 are the PCR negative control for each primer set. Amplification in lane 1 confirmed transcription of the *Nat2* non-coding exon in BNL.CL2 cells. Lanes M are 1 µg of 1 kb DNA ladder (GibcoBRL).

37 nt at the 5' end of the *Nat2* riboprobe were digested, while 210 nt at the 3' end were protected via hybridization to the *Nat2* mRNA expressed in transfected BNL.CL2 cells (Figure 3a). The transcription initiation site of the *Nat2* gene can, therefore, be assigned to 6462 bp upstream of the *Nat2* coding region (Figure 3b). Initiation of transcription from this genomic region was supported further by computational NIX-analysis (TSSW/Promoter prediction programme) available from the UK Human Genome Mapping Project (<http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/index.html>).

Mapping of the mouse *Nat2* gene polyadenylation signal

Analysis of the sequence up to 1000 bp downstream of the mouse *Nat2* coding region indicated the presence of six putative polyadenylation signals (Figure 4a), located 193 bp (polyA1), 438 bp (polyA2), 442 bp (polyA3), 512 bp (polyA4), 513 bp (polyA5) and 767 bp (polyA6) downstream of the stop codon of the gene. Putative signals polyA2 and polyA3, as well as polyA4 and polyA5 are overlapping, and all four of them are located within a short region of 75 bp. RT-PCR with a *Nat2*-specific forward primer and reverse primers annealing just upstream of sequences polyA1, polyA2–polyA5 or polyA6 (Figure 4a) demonstrated that the *Nat2* transcript is terminated before sequence polyA6, but after sequences polyA2–polyA5 (Figure 4b). Polyadenylation signal polyA3 is more likely to dominate over its three nearby signals, as it has a sequence identical with the consensus AATAAA. Cleavage and polyadenylation of the *Nat2* transcript would, therefore, be expected to occur approx. 10–30 bp downstream of polyA3 [25].

Analysis of the *Nat2* transcript in murine tissues

RT-PCR was used to investigate expression of the non-coding exon in other tissues of the Balb/c mouse (Figure 5). The generated

Figure 3 Mapping of the transcription-initiation site of murine *Nat2* gene by RNase protection assay

(a) Total RNA was extracted from BNL.CL2 cells that were co-transfected with DNA from mouse genomic clone AJ250123 and pVAL plasmid (internal standard). Incubation with 32 P-labelled VAI-specific (left photo) or *Nat2*-specific (right photo) riboprobe was followed by RNase treatment, the products of which were subjected to gel electrophoresis and autoradiography for 48 h. Portions of the riboprobes escaping digestion are evident in lanes 1 (VAI-specific riboprobe) and 8 (*Nat2*-specific riboprobe). Lanes 2 and 9 show the products of mock-transfected BNL.CL2 cells, assayed with each riboprobe. RNase-digested and intact riboprobes are shown in lanes 3 and 5 (VAI-specific riboprobe), and 4 and 7 (*Nat2*-specific riboprobe) respectively. Lane 6 is a miscellaneous riboprobe. Numbers with arrows indicate the size of representative DNA markers (lanes M). (b) Summary of the results obtained by RNase protection assay with the *Nat2*-specific riboprobe. The numbers are nucleotide positions relative to the beginning of the *Nat2* coding region. The box represents the previously characterized portion of the *Nat2* non-coding exon [13]. The hatched bar represents the 358 nt full-length riboprobe, consisting of flanking linkers of vector origin and a 247 nt portion matching the sequence of genomic clone AJ250123. The protected portion (dotted bar) of the *Nat2* riboprobe is 210 nt in size and determines the position of the transcription-initiation site (right-facing arrow) of the mouse *Nat2* gene.

Figure 4 Identification of the mouse *Nat2* gene polyadenylation signal

(a) The relative positions of six putative polyadenylation signals (sequences polyA1–polyA6, identified by numbers 1–6 in boxes), located downstream of the *Nat2* coding region (shaded box). The primers used for RT-PCR are indicated with single-headed horizontal arrows. The nucleotide position of reverse primers polyA-aR, polyA-bR and polyA-cR (Table 1), annealing just upstream of sequences polyA1, polyA2–polyA5 and polyA6, respectively, is shown relative to the stop codon (TAG) of the *Nat2* gene. Signal polyA3, responsible for polyadenylation of the *Nat2* transcript, is highlighted in grey. (b) RT-PCR was performed with the *Nat2*-specific forward primer mNAT2 1 and reverse primers polyA-aR (lanes 1–4), polyA-bR (lanes 5–8) and polyA-cR (lanes 9–12). Control reactions (lanes 13–16) were performed with mNAT2 1 and the *Nat2*-specific reverse primer mNAT2 910 [13]. Genomic DNA (lanes 1, 5, 9 and 13) and cDNA (lanes 2, 6, 10 and 14) from 129/Ola mouse ES cells was used as template, as well as the product of a RT-lacking enzyme (lanes 3, 7, 11 and 15). Lanes 4, 8, 12 and 16 are PCR negative controls, and lanes M1 and M2 are 1 μ g of 1 kb (GibcoBRL) and λ /EcoRI + HindIII (Promega) size markers respectively. The expected 1011 and 1282 bp products were amplified from genomic DNA and cDNA with the first and second primer set respectively, whereas the 1537 bp product expected with the third set of primers was amplified only from genomic DNA, indicating that the *Nat2* transcript is cleaved between primers polyA-bR and polyA-cR. Control amplifications in lanes 13–16 provided the expected 910 bp product from genomic DNA and cDNA.

Figure 5 Example of expression of the *Nat2* non-coding exon in tissues of the Balb/c mouse

RT-PCR was performed with cDNA template from liver (lanes 1–3), spleen (lanes 4–6), heart (lanes 7–9) and placenta (lanes 10–12). Products of 0.8 kb in lanes 1, 4, 7 and 10 were amplified with primers NCE-F and mNAT2 691 [13], designed to detect the *Nat2* non-coding exon. Products of 0.9 kb in lanes 2, 5, 8 and 11 were amplified with primers mNAT2 1 and mNAT2 910 [13], specific for the *Nat2* coding region. Lanes 3, 6, 9 and 12 are the corresponding amplification products from a RT without enzyme. Lanes M1 and M2 are 1 kb (GibcoBRL) and λ /EcoRI + *Hind*III (Promega) size markers respectively.

Figure 6 Alignment of mouse *Nat2* genomic sequence (clone AJ250123) with ESTs deposited in dbEST electronic database

The exact positions of representative ESTs are shown relative to the *Nat2* translation-initiation codon (+1). ESTs marked with an asterisk end after polyadenylation signal polyA3. The accession number and tissue origin of each EST are listed on the right of the figure. NCE, non-coding exon.

products from liver, spleen, submaxillary gland, kidney, brain, thymus, lung and placenta were shown to contain the *Nat2* non-coding exon and had a sequence identical with that of transcript AJ251710 from mouse ES cells [13]. In contrast, amplification with primers designed to detect the non-coding exon was not evident in the heart, although a positive reaction was obtained with primers specific for the *Nat2* coding region (Figure 5). It seems likely that the *Nat2* transcript, although present in the heart, does not contain the upstream non-coding exon.

Further evidence supporting alternative splicing of the murine *Nat2* transcript came from the dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST>). Representative ESTs (expressed sequence tags) aligning to the sequence of mouse genomic clone AJ250123 are shown in Figure 6. ESTs with GenBank[®] accession numbers BF164333 and AI006867 partially overlapped with the sequence of the *Nat2* non-coding exon, previously identified in ES cells [13]. The donor splice site activated during processing of transcript AJ251710 from ES cells [13] is located at position 6172 bp upstream of the *Nat2* coding region, whereas the donor splice sites for transcripts BF164333 and AI006867 are located at 172 bp upstream and 175 bp downstream of that position respectively. The acceptor splice site is invariably localized 6 bp upstream of the *Nat2* coding region, whereas five of the identified ESTs (Figure 6) end 15–23 bp after polyadenylation signal polyA3, consistent with the RT-PCR results described above.

The *Nat2* gene extends from a transcription-initiation site, approx. 6.5 kb upstream, to a functional polyadenylation signal, about 0.44 kb downstream of the intronless coding region. Splice

sites for the 6.2 kb intron, separating the non-coding and coding exons of *Nat2*, were RT-PCR-characterized previously in ES cells [13], as well as in the present study in adult tissues and in BNL. CL2 embryonic liver cells of the mouse. A 1.4 kb RT-PCR product, amplified from several murine tissues (e.g. lane 1 in Figure 5), may represent an alternative *Nat2* transcript, possibly matching thymic EST AI006867, while the presence of a third alternative *Nat2* transcript is supported by EST BF164333 of lung tumour origin. The heart provides a further variation, in which the *Nat2* transcript is present, but lacks the upstream non-coding exon and the 6.2 kb intron, as demonstrated by RT-PCR.

Unlike the donor splice site, which can vary between different *Nat2* transcripts, the acceptor splice site of the 6.2 kb intron of the gene is fixed. It is located only 6 bp upstream of the translation-initiation codon and bears all the characteristics of the consensus [26], including the typical TAG/G sequence at the intron/exon junction, a branch point (TTTAAAC) 87 bp upstream, and a 75 % pyrimidine-rich stretch adjacent to the splice site. The position of the *Nat2* polyadenylation signal also appears to be invariable, as suggested by RT-PCR and EST analysis.

The presence of upstream non-coding exons may be associated with differential use of more than one promoter in a tissue- or developmental-specific manner [27]. The results described here demonstrate the use of a transcription initiation site, located in the beginning of the *Nat2* non-coding exon, in a wide range of tissues, whereas in the heart, the non-coding exon appears not to be transcribed. Use of alternative transcription-initiation sites is compatible with previous primer-extension analyses of adult kidney and liver tissue [16], which identified multiple sites adjacent to the start of the *Nat2* coding region.

Identification of regions bearing transcription-regulatory elements for the mouse *Nat2* gene

Using the primers shown in Table 2, 21 partially overlapping sequences were amplified from the 14 kb mouse genomic plasmid clone AJ250123. The generated DNA fragments, covering the full length of the clone, apart from coding and non-coding exons (Figure 7a), were subcloned into the pGL3-Basic reporter vector, which lacks indigenous transcription-control elements (constructs B1–B21). Promoter activity in dual-luciferase reporter assays was detected only with construct B7 (Figure 7b), which extends from 133 bp upstream to 232 bp within the *Nat2* non-coding exon (position 3255–3620 of clone AJ250123) and spans the transcription-initiation site of the gene.

Figure 7 Identification of regions bearing transcription regulatory elements of the mouse *Nat2* gene

(a) Diagram showing the relative position of DNA fragments 1–21 (grey lines), amplified from clone AJ250123 (black line), using the primers listed in Table 2. Each fragment was subcloned into pGL3-Basic (constructs B1–B21) or pGL3-Promoter (constructs P1, P3–P6, P8, P10, P12, P14, P16 and P18–P21) reporter vector. The position of *Nat1* (hatched box) and *Nat2* (black box) coding region, as well as of *Nat2* non-coding exon (dotted box) is indicated relative to the beginning of clone AJ250123. (b)–(c) Luciferase activities generated with constructs B1–B21 (b) or constructs P1–P21 (c). Each construct was co-transfected with pRL-TK vector into BNL-CL2 cells and the cell lysates produced were subjected to dual-luciferase reporter assay. The luminescence produced by *P. pyralis* luciferase is expressed as a percentage relative to the luminescence produced by *R. reniformis* luciferase (relative luminescence). Vectors pGL3-Promoter, bearing a SV40 promoter, and pGL3-Control, bearing a SV40 promoter and a SV40 enhancer, were used as positive controls for *P. pyralis* luciferase activity. Vector pGL3-Basic, lacking insert, was used to estimate background *P. pyralis* luciferase activity. Mock-transfected cells (MT) were used as negative control for luciferase activity. Results in (c) are means $\pm$ S.D. of measurements of relative luminescence, using lysates from multiple transfections with the same construct ($n = 2-6$).

Clone AJ250123 was also examined for regions likely to harbour transcriptional activators or repressors. Of the DNA fragments shown in Figure 7a, 14 were subcloned into the pGL3-Promoter reporter vector, which carries its own SV40 promoter (constructs P1, P3–P6, P8, P10, P12, P14, P16 and P18–P21). On average, levels of luciferase activity were higher with constructs P1–P6, covering the region upstream of the *Nat2* non-coding exon, compared with the activity measured with constructs P8, P10, P12, P14, P16 and P18, which span the 6.2 kb intron of the *Nat2* gene. Constructs P19–P21, covering the region downstream of *Nat2*, produced intermediate levels of luciferase activity (Figure 7c). Construct P4 provided the highest, and construct P10 the lowest, luciferase activity, suggesting that their inserts are likely to contain enhancer and repressor elements respectively (Figure 7c).

Identification of the mouse *Nat2* basal promoter

The 365 bp insert of construct B7, producing high luciferase activity in reporter-gene assays (Figure 7b), was subjected to TRANSFAC database analysis (<http://transfac.gbf.de/TRANSFAC/>), which revealed two sequences of interest: (i) TTTAAAA at position 3378 of clone AJ250123, with a core similarity of 0.928 to the classical TATA box [28], and (ii) GGGGCGGAGC at position 3400 of clone AJ250123, with a core similarity of 1.000 to the Sp1 box of the *Hprt* (hypoxanthine phosphoribosyltransferase) and *Dhfr* (dihydrofolate reductase) gene promoters [29]. These were also the only basal promoter elements predicted on the 14 kb AJ250123 clone by computational NIX analysis (TSSW/Promoter prediction program).

Figure 8 Deletion or mutation of candidate promoter elements, located close to the mouse *Nat2* transcription-initiation site

(a) Diagram showing the strategies used to ablate the putative TATA (white box) and Sp1 (black box) elements contained in construct B7 (black horizontal line). The numbers are positions on clone AJ250123. Primer pairs PROM-1(F)/DEL(R) and DEL(F)/EXON-1(R) (Table 2) were used to amplify portions of the B7 insert flanking the candidate promoter elements. The generated fragments were ligated at the *NdeI* sites of the DEL primers, creating a 57 bp deletion of the entire promoter region (insert DEL). Similarly, primer pairs PROM-1(F)/TATA(R) and TATA(F)/EXON-1(R) (Table 2) were used to introduce a *XbaI* site within the putative TATA box, creating two point mutations [insert TATA(mt)]. Finally, primer pairs PROM-1(F)/SP1(R) and SP1(F)/EXON-1(R) (Table 2) were used to introduce a *NotI* site within the putative Sp1 box, creating three point mutations [insert SP1(mt)]. The three inserts were sequenced and ligated to the *MluI* and *XhoI* sites of pGL3-Basic vector, and were then subjected to dual-luciferase reporter assay. (b) Comparison between the relative luciferase activities generated with construct B7 (intact promoter) and the mutated constructs DEL, SP1(mt) and TATA(mt). Relative luminescence was calculated as described in Figure 7, and the results for each mutated construct were expressed as a percentage of those determined for construct B7. MT are mock-transfected cells and pGL3-Basic is vector lacking insert.

The two elements were confirmed to constitute an active promoter by deletion of the region bearing both candidate sequences, as well as by mutation of either the putative TATA box or the putative Sp1 box separately (Figure 8a). Deletion of both elements caused a 97% decrease, whereas mutation of the TATA box caused a 90% decrease and mutation of the Sp1 box caused a 92% decrease in the relative luminescence produced by the intact promoter (Figure 8b). Since these elements are adjacent to the transcription-initiation site of the *Nat2* gene, they are expected to regulate its expression *in vivo*.

Thus identified, the *Nat2* promoter consists of a TATA and a Sp1 box. The TATA box has an atypical sequence, also found in the promoters of somatostatin and tyrosine hydroxylase genes

[30]. The Sp1 box binds a transcription factor known to initiate expression of TATA-less housekeeping genes [29]. Although the Sp1 box of the *Nat2* promoter is capable of driving high levels of reporter-gene expression only in the presence of an intact TATA box, it has a sequence identical with the Sp1 boxes of murine *Hprt* and *Dhfr* housekeeping promoters [29], and could serve to maintain constitutive expression of the *Nat2* gene.

Binding of transcription factors to the mouse *Nat2* promoter region

To assess the protein-binding capacity of the *Nat2* promoter region, EMSA was performed with five PCR-generated probes (GS1–GS5; Table 1), spanning the 735 bp sequence from 49 bp

Figure 9 EMSA of the mouse *Nat2* promoter region

BNL.CL2 whole-cell protein extract was incubated with GS1 probe, the relative position of which on clone AJ250123 is shown schematically (top). The position of primers GS-1(F) and GS-1(R) (Table 1), used for the generation of probe GS1, is indicated, whereas the black and dotted box represent the TATA and the Sp1 box respectively. On the autoradiogram (bottom), lane 1 is probe only and lane 2 is the product of incubation of 32 P-labelled probe with cell extract. Lane 3 shows specific competition with excess of unlabelled GS1 probe, whereas lane 4 shows non-specific competition with an unlabelled probe, generated by PCR with primers mNAT2 516 and mNAT2 691 [13]. Lanes 5–10 are the products of competition with oligonucleotides bearing recognition sites for the AP2 (activation protein 2), Sp1, AP1, Oct-1 (octamer motif recognizing transcription factor 1), CREB and NF- κ B transcription factors respectively. P is the band of the unbound probe, and A–E are shifted bands of protein-bound probe. Interference of the unlabelled competitor with protein binding to the labelled probe is indicated by decrease in the intensity of the shifted bands.

downstream to 686 bp upstream of the *Nat2* transcription initiation site. Incubation of the 144 bp GS1 probe, containing the *Nat2* promoter, with BNL.CL2 cell extract produced five shifted bands (bands A–E in Figure 9). Protein binding was specific, as these bands were not affected by competition with unlabelled non-specific competitor DNA, but were completely removed by specific competition with unlabelled GS1 probe (lanes 2–4 in Figure 9). Moreover, competition with unlabelled Sp1-specific oligonucleotide changed the relative intensity of the shifted bands (lane 6 in Figure 9), indicating the presence of an Sp1-recognition site within the sequence of GS1 probe. Other oligonucleotide competitors (lanes 5 and 7–10 in Figure 9) did not affect the banding pattern on the autoradiogram.

Although binding of cellular protein to probes GS2–GS5 (Table 1) was less prominent, it suggested the presence of AP1- and CREB-recognition sites within the sequence of probes GS3 and GS5. This finding was also supported by computational TRANSFAC analysis (<http://transfac.gbf.de/CYTOMER/>), which further predicted ubiquitous expression of the *Nat2* gene, consistent with the experimental results of previous studies [8–10, 13–15].

A PCR-generated probe, covering most of the non-coding exon and a portion of its upstream region, including the *Nat2* promoter (position 3305–3620 on clone AJ250123), was used in DNaseI footprinting assays with nuclear extract from HeLa cells. The

Figure 10 DNase I footprinting of the region spanning the promoter of the mouse *Nat2* gene

HeLa nuclear protein extract was incubated with a 315 bp 32 P-labelled probe, amplified with primers REP17(F) and EXON-1(R) (Table 2). On the autoradiogram, lane 1 is DNase I-digested probe only, whereas lane 2 is the digestion product of probe incubated with nuclear protein. Lane M is 32 P-labelled 10 bp DNA ladder (GibcoBRL). The protein-bound region of the probe, which was protected from DNase I digestion, is indicated on the left of the autoradiogram. The relative position on clone AJ250123 of the TATA box, the *Nat2* transcription-initiation site and the Sp1 box is presented schematically on the right.

region around 90–95 bp from the 5' end of the probe was substantially protected from DNase I digestion (Figure 10), demonstrating protein binding to this position. The protected region overlaps with the Sp1 box of the *Nat2* promoter, confirming further that this is a functional regulatory element, capable of transcription-factor binding.

The perfect nucleotide-sequence match between the consensus Sp1 box [29] and the characterized element of the *Nat2* promoter, supports strongly the prediction of the nature of the specific protein-factor binding the exact sequence in the genomic region, although supershift analysis with anti-Sp1 antibodies, as well as EMSA or DNase I footprinting with pure Sp1 protein, would serve to confirm specific binding of the Sp1 transcription factor to the *Nat2* promoter.

Transcriptional regulation could explain the observed tissue- and developmental-specific variation in expression of the NAT2 isoenzyme. The possible endogenous role of murine NAT2 in folate metabolism, as well as its developmental expression in the neural tube and in the cardiovascular system, is an area where understanding of transcriptional regulation of the *Nat2* gene may be important. Recent studies have demonstrated that deletion of the *Nat2* gene leads to phenotypically normal mice [31], whereas mice overexpressing the human equivalent of murine *Nat2* gene

have developmental abnormalities [4]. Maintaining different levels of *Nat2* expression in various tissues, possibly via alternative splicing or differential utilization of more than one promoter, may have important implications in both the developing and the adult organism and also in cancer [32].

We are grateful to Professor N. Proudfoot for help and advice, as well as to S. Murphy, J. Mellor and L. Wakefield for helpful discussion. Work was funded by the Wellcome Trust and Action Research (SPARKS). S.B. was in receipt of a Bodossaki Foundation graduate scholarship.

REFERENCES

- Butcher, N. J., Boukouvala, S., Sim, E. and Minchin, R. F. (2002) Pharmacogenetics of the arylamine N-acetyltransferases. *Pharmacogenomics J.* **2**, 30–42.
- Hein, D. W., Doll, M. A., Fretland, A. J., Leff, M. A., Webb, S. J., Xiao, G. H., Devanaboyina, U.-S., Nangju, N. A. and Feng, Y. (2000) Molecular genetics and epidemiology of the *NAT1* and *NAT2* acetylation polymorphisms. *Cancer Epidemiol. Biomark. Prev.* **9**, 29–42.
- Windmill, K. F., McKinnon, R. A., Zhu, X., Gaedigk, A., Grant, D. M. and McManus, M. E. (1997) The role of xenobiotic metabolising enzymes in arylamine toxicity and carcinogenesis: Functional and localisation studies. *Mutat. Res.* **376**, 153–160.
- Sim, E., Pinter, K., Mushtaq, A., Upton, A., Sandy, J., Bhakta, S. and Noble, M. (2003) Arylamine N-acetyltransferases: a pharmacogenetic approach to drug metabolism and endogenous function. *Biochem. Soc. Trans.* **31**, 615–619.
- Minchin, R. F. (1995) Acetylation of *p*-amino benzoylglutamate, a folic acid catabolite, by recombinant arylamine N-acetyltransferase and U937 cells. *Biochem. J.* **307**, 1–3.
- Ward, A., Summers, M. E. and Sim, E. (1995) Purification of recombinant human N-acetyltransferase type 1 expressed in *E. coli* and characterization of its potential role in folate metabolism. *Biochem. Pharmacol.* **49**, 1759–1767.
- Payton, M., Smelt, V., Upton, A. and Sim, E. (1999) A method for genotyping murine arylamine N-acetyltransferase type 2 (*NAT2*): a gene expressed in preimplantation embryonic stem cells encoding an enzyme acetylating the folate catabolite *p*-aminobenzoylglutamate. *Biochem. Pharmacol.* **58**, 779–785.
- Stanley, L. A., Mills, I. G. and Sim, E. (1997) Localisation of polymorphic N-acetyltransferase (*NAT2*) in tissues of inbred mice. *Pharmacogenetics* **7**, 121–130.
- Chung, J. G., Levy, G. N. and Weber, W. W. (1993) Distribution of 2-aminofluorene and *p*-aminobenzoic acid N-acetyltransferase activity in the tissues of C57BL/6J rapid and B6.A-*Nat*⁶ slow acetylator congenic mice. *Drug Metab. Dispos.* **21**, 1057–1063.
- Boukouvala, S., Price, N. and Sim, E. (2002) Identification and functional characterisation of novel polymorphisms associated with the genes for arylamine N-acetyltransferases in mice. *Pharmacogenetics* **12**, 385–394.
- Fretland, A. J., Doll, M. A., Gray, K., Feng, Y. and Hein, D. W. (1997) Cloning, sequencing, and recombinant expression of *NAT1*, *NAT2*, and *NAT3* derived from the C3H/HeJ (rapid) and A/HeJ (slow) acetylator inbred mouse: functional characterization of the activation and deactivation of aromatic amine carcinogens. *Toxicol. Appl. Pharmacol.* **142**, 360–366.
- Martell, K. J., Vatsis, K. P. and Weber, W. W. (1991) Molecular genetic basis of rapid and slow acetylation in mice. *Mol. Pharmacol.* **40**, 218–227.
- Fakis, G., Boukouvala, S., Buckle, V., Payton, M., Denning, C. and Sim, E. (2000) Chromosomal localisation and mapping of the genes for murine arylamine N-acetyltransferases (*NATs*), enzymes involved in the metabolism of carcinogens: identification of a novel upstream non-coding exon for murine *Nat2*. *Cytogenet. Cell Genet.* **90**, 134–138.
- Stanley, L. A., Copp, A. J., Pope, J., Rolls, S., Smelt, V., Perry, V. H. and Sim, E. (1998) Immunochemical detection of arylamine N-acetyltransferase during mouse embryonic development and in adult mouse brain. *Teratology* **58**, 174–182.
- McQueen, C. A. and Chau, B. (2003) Neonatal ontogeny of murine arylamine N-acetyltransferases: implications for arylamine genotoxicity. *Toxicol. Sci.* **73**, 279–286.
- Estrada-Rodgers, L., Levy, G. N. and Weber, W. W. (1998) Characterisation of a hormone response element in the mouse N-acetyltransferase 2 (*Nat2*^{*}) promoter. *Gene Expression* **7**, 13–24.
- Ebisawa, T. and Deguchi, T. (1991) Structure and restriction fragment length polymorphism of genes for human liver arylamine N-acetyltransferases. *Biochem. Biophys. Res. Commun.* **177**, 1252–1257.
- Rampazzo, A., Pivotto, F., Occhi, G., Tiso, N., Bortoluzzi, S., Rowen, L., Hood, L., Nava, A. and Danieli, G. A. (2000) Characterisation of C14orf4, a novel intronless human gene containing a polyglutamine repeat, mapped to the ARVD1 critical region. *Biochem. Biophys. Res. Commun.* **278**, 766–774.
- Mummidi, S., Ahuja, S. S., McDaniel, B. L. and Ahuja, S. K. (1997) The human CC chemokine receptor 5 (*CCR5*) gene: Multiple transcripts with 5'-end heterogeneity, dual promoter usage and evidence for polymorphisms within the regulatory regions and noncoding exons. *J. Biol. Chem.* **272**, 30662–30671.
- Sosinsky, A., Glusman, G. and Lancet, D. (2000) The genomic structure of human olfactory receptor genes. *Genomics* **70**, 49–61.
- Palmiter, R. D., Sandgren, E. P., Avarbock, M. R., Allen, D. D. and Brinster, R. L. (1991) Heterologous introns can enhance expression of transgenes in mice. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 478–482.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, New York.
- Sinclair, J. C., Delgoda, R., Noble, M. E. M., Jarman, S., Goh, N. K. and Sim, E. (1998) Purification, characterization, and crystallization of an hydroxyarylamine O-acetyltransferase from *Salmonella typhimurium*. *Protein Expression Purif.* **12**, 371–380.
- Mitchell, M. K., Futscher, B. W. and McQueen, C. A. (1999) Developmental expression of N-acetyltransferases in C57BL/6 mice. *Drug Metab. Dispos.* **27**, 261–264.
- Proudfoot, N. J. and Whitelaw, E. (1988) Termination and 3' end processing of eukaryotic RNA. In *Transcription and Splicing* (Hames, B. D. and Glover, D. M., eds.) pp. 97–129, IRL Press, New York.
- Smith, C. W. J., Patton, J. G. and Nadal-Ginard, B. (1989) Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* **23**, 527–577.
- Geiger, A., Decaux, J. F., Burcelin, R., Le Cam, A., Salazar, G., Charron, M. J., Girard, J. and Kervran, A. (2000) Structural and functional characterisations of the 5'-flanking region of the mouse glucagon receptor gene: comparison with the rat gene. *Biochem. Biophys. Res. Commun.* **272**, 912–921.
- Beebe, T. and Burke, J. (1990) *Gene Structure and Transcription*, IRL Press, Oxford.
- Dynan, W. S. (1986) Promoters for housekeeping genes. *Trends Genet.* **2**, 196–197.
- Tavianini, M. A., Hayes, T. E., Magazin, M. D., Minth, C. D. and Dixon, J. E. (1984) Isolation, characterisation and DNA sequence of the rat somatostatin gene. *J. Biol. Chem.* **259**, 11798–11803.
- Cornish, V. A., Pinter, K., Boukouvala, S., Johnson, N., Labrousse, C., Payton, M., Priddle, H., Smith, A. J. H. and Sim, E. (2003) Generation and analysis of mice with a targeted disruption of the arylamine N-acetyltransferase type 2 gene. *Pharmacogenomics J.* **3**, 169–177.
- Adam, P. J., Berry, J., Loader, J. A., Tyson, K. L., Craggs, G., Smith, P., De Belin, J., Steers, G., Pezzella, F., Sachsenmeier, K. F. et al. (2003) *NAT-1* is highly expressed in breast cancers and conveys enhanced growth resistance to etoposide *in vitro*. *Mol. Cancer Res.*, in the press.

Received 3 June 2003/24 July 2003; accepted 6 August 2003

Published as BJ Immediate Publication 6 August 2003, DOI 10.1042/BJ20030812