Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 5648–5652, June 1993
Medical Sciences

Identification of a human achaete-scute homolog highly expressed in neuroendocrine tumors

(basic helix–loop–helix protein/trinucleotide repeat/medullary thyroid carcinoma/small cell lung cancer)

DOUGLAS W. BALL*†, CHRISTOPHER G. AZZOLI*, STEPHEN B. BAYLIN*‡, DAVID CHI§, SHENSHEN DOU§, HELEN DONIS-KELLER§, ARUNTHATHI CUMARASWAMY‡, MICHAEL BORGES‡, AND BARRY D. NELKIN‡

*Division of Endocrinology and Metabolism and ‡The Oncology Center, Johns Hopkins University School of Medicine, Baltimore, MD 21231; and §Division of Human Molecular Genetics, Department of Surgery, Washington University School of Medicine, St. Louis, MO 63110

Communicated by Victor A. McKusick, February 25, 1993

ABSTRACT **Basic helix–loop–helix transcription factors of the achaete-scute family are instrumental in *Drosophila* neurosensory development and are candidate regulators of development in the mammalian central nervous system and neural crest. We report the isolation and initial characterization of a human achaete-scute homolog that is highly expressed in two neuroendocrine cancers, medullary thyroid cancer (MTC) and small cell lung cancer (SCLC). The human gene, which we have termed human achaete-scute homolog 1 (hASH1), was cloned from a human MTC cDNA library. It encodes a predicted protein of 238 aa that is 95% homologous to mammalian achaete-scute homolog (MASH) 1, a rodent basic helix–loop–helix factor. The 57-residue basic helix–loop–helix domain is identical to that in the rodent gene, and the basic and helical regions, excluding the loop, are 72–80% identical to *Drosophila* achaete-scute family members. The proximal coding region of the hASH1 cDNA contains a striking 14-copy repeat of the triplet CAG that exhibits polymorphism in human genomic DNA. Thus, hASH1 is a candidate locus for disease-causing mutations via triplet repeat amplification. Analysis of rodent–human somatic cell hybrids permitted assignment of hASH1 to human chromosome 12. Northern blots revealed hASH1 transcripts in RNA from a human MTC cell line, two fresh MTC tumors, fetal brain, and three lines of human SCLC. In contrast, cultured lines of non-SCLC lung cancers and a panel of normal adult human tissues showed no detectable hASH1 transcripts. Expression of hASH1 may provide a useful marker for cancers with neuroendocrine features and may contribute to the differentiation and growth regulation of these cells.**

Cancers with neuroendocrine features such as small cell lung cancer (SCLC) and the calcitonin-secreting tumor, medullary thyroid carcinoma (MTC), frequently lose their characteristic endocrine phenotype as tumor progression occurs (1–3). To understand the evolution of endocrine cancers, it is important to define, at a molecular level, both the regulation of neuroendocrine phenotypic features in the normal cellular precursors of these tumors and the alterations in these processes that occur during tumor progression.

In the complex regulation of neuroendocrine phenotypic expression at a transcriptional level, there is increasing evidence for involvement of basic helix–loop–helix (bHLH) transcriptional enhancer factors. bHLH proteins may have particular importance in the control of polypeptide hormone synthesis and secretion. Our laboratory (4) and others (5) have demonstrated that bHLH recognition elements form a constitutive enhancer in the human calcitonin gene. Other polypeptide hormones including insulin, gastrin, and secretin also appear to utilize bHLH enhancer factors that are restricted to their differentiated host tissues (6–8). Although a role for bHLH factors in controlling the proliferation of neuroendocrine cells in humans remains uncertain, neural developmental pathways in *Drosophila* may provide useful clues. In this organism, the four bHLH factors of the achaete-scute complex (achaete, scute, asense, and lethal of scute) coordinately regulate the capacity of ectodermal cells to become developing neuroblasts in the peripheral sensory nervous system and the central nervous system (9). Loss-of-function mutations of these genes suppress the formation of sensory organs, whereas ectopic expression of achaete and scute leads to the supernumerary sense organs associated with the hairy-wing mutation (10). The four achaete-scute proteins contain a highly homologous bHLH domain that permits heterodimer formation with the widely expressed daughterless protein (11, 12) as well as inhibitory interactions with the truncated bHLH protein extramachrochaetae (13).

Recently, two rodent achaete-scute homologs, mammalian achaete-scute homolog (MASH) 1 and MASH-2 were cloned from a sympathoadrenal cell line library by using a PCR-based approach (14). MASH-1 transcripts were detected in several neural tissues in the fetal rat, including brain, adrenal medulla, and cervical ganglion, and fetal lung but not in adult brain or a variety of other adult rodent tissues. MASH-1 expression could be induced by nerve growth factor in the PC-12 rodent pheochromocytoma cell line. Also of significance to the potential involvement of MASH-1 in neuroendocrine cells, a single rat MTC cell line was reported positive for MASH-1 expression. In light of these findings and our data (4) indicating that bHLH enhancer sites are important for calcitonin gene expression, we became interested in the potential of achaete-scute homologs to regulate the differentiation of human neuroendocrine tissues and tumor cells. We have identified and characterized a human achaete-scute-like factor¶ and find that it is abundantly expressed not only in MTC but also in SCLC, a lung tumor with neuroendocrine features. These observations significantly expand the range of mammalian tissues known to express achaete-scute homologs and suggest a broader regulatory role for these factors. In addition, we find that this human achaete-scute homolog (hASH) contains a polymorphic trinucleotide repeat within its protein coding sequence, making it a candidate locus for inherited disease.

MATERIALS AND METHODS

Cell Culture and Tissue Specimens. Culture conditions for human TT and DMS53 cells (4), NCI-H60, H82, H209, and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bHLH, basic helix–loop–helix; MTC, medullary thyroid cancer; SCLC, small cell lung cancer; MASH, mammalian achaete-scute homolog; hASH1, human achaete-scute homolog 1; CEPH, Centre d'Etude du Polymorphisme Humain.

†To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. L08424).

H249 cells (15), NCI-H169 (16), and U1752 cells (17) have been described. IMR-90 human fibroblasts (National Institute of Aging Cell Repository) were grown in Eagle's modified minimal essential medium with 10% (vol/vol) fetal calf serum. Human medullary thyroid cancer and pheochromocytoma specimens obtained at surgical resection were rapidly frozen in liquid nitrogen and stored up to 2 years at −70°C prior to use.

Northern Blot Analyses. To survey initially for achaete-scute homolog expression in human cells, Northern blots containing the indicated amounts of total RNA or poly(A)-enriched RNA were probed with a 1.3-kb *Hin*dIII–*Xba* I insert from pNj1 containing a full-length rodent MASH-1 cDNA insert (generously provided by D. Anderson, California Institute of Technology), labeled by random priming, to survey achaete-scute homolog expression. Subsequent studies employed a human cDNA insert probe described below. RNA isolation, poly(A)$^+$ selection, and hybridization conditions were as described (18). A Northern blot containing 2 μg of poly(A)$^+$ RNA from various normal adult tissues and poly(A)$^+$ RNA from human fetal brain were purchased from Clontech.

Library Screening. A random hexamer-primed cDNA library was constructed from TT-cell poly(A)$^+$ RNA and cloned into λ Zap II (Stratagene) as described (19). A total of 400,000 phage plaques in the XL-1 Blue strain of *Escherichia coli* (Stratagene) was screened using the random-labeled MASH-1 insert (10^6 cpm/ml). Hybridization of duplicate nitrocellulose filter lifts was performed at 55°C in Blotto [1% SDS/powdered milk (5 mg/ml)/PEG (60 mg/ml)/5 × SSPE/10% (vol/vol) formamide/sonicated salmon sperm DNA (0.2 mg/ml) (1 × SSPE = 0.18 M NaCl/10 mM sodium phosphate, pH 7.4/1 mM EDTA)]. Washes were performed in 2 × standard saline citrate (SSC)/0.1% SDS at room temperature and then in 0.5 × SSC/0.1% SDS at 55°C. Plaques that continued to hybridize through three rounds of screening were subjected to *in vivo* excision according to the manufacturer's protocol (Stratagene).

DNA Sequencing and Analyses. A series of overlapping cDNA clones was obtained in pBluescript SK +/− (Stratagene) and sequenced on both strands using the Sequenase Version 2.0 kit (United States Biochemical). An *Eco*RI insert from a plasmid designated pHASH4.1 containing a 1.3-kb human achaete-scute homolog cDNA was used for further hybridization studies. Comparisons with existing nucleotide and amino acid sequences from GenBank/EMBL Release 71 and NBRF-PIR Release 32 were analyzed using Wilbur & Lipman and Needleman–Wunsch algorithms (DNAStar, Madison, WI).

Chromosomal Localization and Analysis of Trinucleotide Repeat Polymorphism. Genomic DNA from a panel of hybrid rodent–human somatic cell lines (20) was amplified using hASH1-specific PCR primers to determine the chromosomal location of the gene. The sequence of the forward (sense strand) primer was CAGCCTGTTTCTTTGCCACGG; the sequence of the reverse primer was TTGCTTGGGCGCTGACTTGTG. PCRs were performed in a total volume of 5 μl containing PCR buffer (10 mM Tris·HCl, pH 8.3/50 mM KCl/1.5 mM $MgCl_2$/5 mM NH_4Cl), 50 ng of genomic DNA, all four dNTPs (each at 800 μM), 2.0 μM forward primer, 2.0 μM reverse primer, 10% (vol/vol) dimethyl sulfoxide, and 0.25 unit of *Taq* polymerase. Reaction mixtures were incubated at 94°C for 30 sec at 63°C for 30 sec, and at 72°C for 1 min for 25 cycles. A trinucleotide repeat sequence detected in the hASH1 cDNA was analyzed for polymorphism in genomic DNA, obtained from a set of parents from the CEPH (Centre d'Etude du Polymorphisme Humain) reference pedigree collection, using PCR primers specific for sequences flanking the repeat. The forward (sense strand) primer sequence was AGCCCTTCCTGCCGCCCGCA; the reverse primer sequence was GGCGCTGACTTGTGACCGCC. PCRs were performed in a total volume of 5 μl containing PCR buffer, 50 ng of genomic DNA, all four dNTPs (each at 800 μM), 0.25 μM 5′-[^{32}P]dATP-end-labeled reverse primer, 1.5 μM unlabeled reverse primer, 2.0 μM forward primer, 10% dimethylsulfoxide, and 0.25 unit of *Taq* polymerase.

RESULTS

Isolation and Characterization of hASH1. To screen initially for the presence of MASH expression in human neuroendocrine cells, we used a rodent MASH-1 cDNA insert to probe Northern blots from two calcitonin-producing tumor cell lines, the TT line of MTC and the DMS53 line of SCLC. Both tumor cell RNAs contained 3.0-kb transcripts that hybridized to the rodent cDNA under stringent conditions (data not shown). Southern blot analyses of human genomic DNA at slightly reduced stringency with the same probe also revealed significant hybridization (data not shown). Based on evidence for hybridizing human achaete-scute-like transcripts from these cells, we proceeded to screen a cDNA library to isolate a human homolog of rodent MASH-1.

A random hexamer-primed cDNA library prepared from TT cell RNA was screened with the rodent MASH-1 probe. Of 400,000 recombinant plaques screened, 5 clones hybrid-

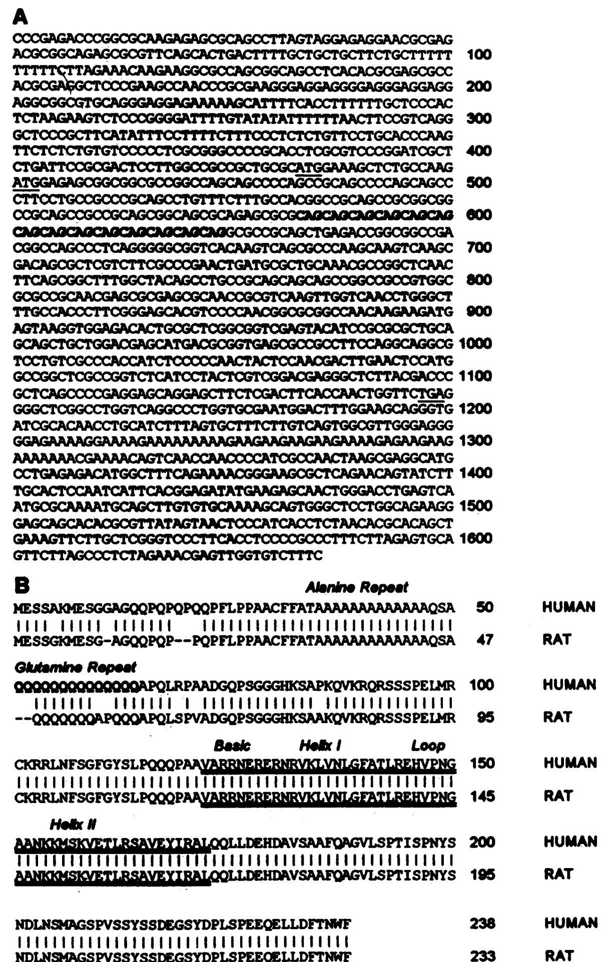

FIG. 1. (*A*) hASH1 cDNA sequence is shown with two candidate translation start sites and one terminator codon (underlined) and a CAG trinucleotide repeat (boldface type). (*B*) Predicted hASH1 amino acid sequence, compared to rat MASH-1 (14), with conserved residues (vertical bars), the human polyglutamine tract (boldface type), and the conserved bHLH domain (double underlined) indicated.

```
                     Basic       Helix          Loop              Helix                  Identity
                                                                                          bHLH   bHH
                     ----------- ------------- ---------------- -----------------------
hASH1                VARRNERERNRVKLVNLGFATLREHVP--------NG------AANKKMSKVETLRSAVEYIRALQQLLDE   --     --
MASH-1               :::::::::::::::::::::::::::--------::------::::::::::::::::::::::::::::  100    100
MASH-2               ::::::::::::::::::::QA::Q::--------H:------G:::L::::::::::::::::::R::A:   86     89

achaete (T5)         :I:::A:::::::Q::N::SQ::Q:I:AAVIADLS::RRGIGPG:::L:::S::KM::::::R::KV:H:    56     72
scute   (T4)         :Q:::A:::::::Q::NS::R::Q:I:QSIITDLTK:--G-GRGPH::I:::D:::I::::::S::D:V:D   56     74
lethal  (T3)         :::::A:::::::Q::N::VN::Q:L:QTVVNSLS::----GRGSS::L:::D:::I::::::G::DM::D   60     76
asense  (T8)         :::::A:::::::Q::N:::L:::KI:EEVSEAFEAQ--GAGRG:S::L:::::::M::::::S:EK::GF   58     80

Achaete-scute        V*RRNARERNRVKQVNNGF**LRQH*P                G**KK*SKV*TLR*AVEYIR*LQ*****
 Consensus                E       L  LS     EK                  A        K       E
```

FIG. 2. bHLH domain of hASH1 is shown in comparison to other related proteins. Colons indicate conserved residues; identities listed under "bHH" indicate sequence identity excluding the loop. Sequence data were obtained from NBRF-PIR Release 32 and GenBank/EMBL Release 71. "Achaete-scute Consensus" indicates residues common to all seven known *Drosophila* and mammalian achaete scute homologs.

ized strongly through three rounds of plaque purification. Dideoxynucleotide sequencing revealed that 4 of the 5 clones were overlapping and had substantial homology to the rat MASH-1 sequence. The overlapping sequence data from these 4 clones defined a 1.6-kb cDNA with an open reading frame of 714 bp, predicting a 238-aa peptide that we designated hASH1 (Fig. 1). Two in-frame ATG sequences are present at positions 433 and 451. The context of the downstream ATG more closely matches the consensus initiator sequence of Kozak (21). A TGA stop codon is present at position 1147. Comparison at the amino acid level indicated 95% overall identity with rat MASH-1 and 40% identity with MASH-2. The hASH1 amino acid sequence exhibits 100% conservation of the MASH-1 bHLH domain, as shown in Fig. 2. As expected, hASH1 also shares a high degree of conservation with MASH-2 across this domain. Parallel comparisons with the four *Drosophila* achaete-scute proteins versus a variety of other mammalian and *Drosophila* bHLH proteins reinforce the relatedness of mammalian and *Drosophila* achaete-scute homologs suggested by Johnson *et al.* (14). Excluding the nonconserved loop sequence, all four of the *Drosophila* achaete-scute peptides share 72–80% amino acid identity with hASH1 across this region (Fig. 2), whereas no other bHLH protein has >51% identity to hASH1. Thirty-one of the bHLH residues are absolutely conserved among the seven known mammalian and *Drosophila* achaete-scute family members, with particular conservation of the basic and helix II domains. Significantly, all three mammalian achaete-scute proteins diverge from the *Drosophila* family outside of the bHLH region. The hASH1 protein shares 32–52% overall identity with the *Drosophila* proteins, comparable to MASH-1.

The proximal 500 bp of the hASH1 cDNA is unusually G+C-rich (61%), with a relatively high ratio of CpG to GpC dinucleotides (0.7), and sites for the methylation-sensitive rare-cutting enzymes *Not* I (1), *Bss*HII (3), *Nae* I (3), and *Nar* I (3), suggesting that the proximal hASH1 cDNA may overlap with a CpG island (22).

Presence of a Trinucleotide Repeat in hASH1. A striking feature of the hASH1 coding sequence, shown in Fig. 1*A*, is a 14-copy repeat of the trinucleotide CAG (bp 583–624), coding for glutamine. In contrast to hASH1, the rodent homolog MASH-1 contains an interrupted series of glutamines and no trinucleotide repeat. Trinucleotide repeats of this length are relatively uncommon in eukaryotic coding sequences. Amplification of C + G-rich trinucleotide repeat sequences is associated with at least four human diseases, fragile X syndrome, myotonic dystrophy, Kennedy disease, and Huntington disease (for review, see refs. 23 and 24). The normal function of the CAG repeat sequence in hASH1 is unclear. However, glutamine-rich regions have been implicated as transcriptional activator domains in Sp1 (25), antennapedia (26), and other transcriptional regulatory proteins. Thirteen alanine residues, located proximal to the glutamine repeat, are shared by the rat homolog. This pattern of alanine residues adjacent to glutamine-rich segments is also reminiscent of the Sp1 transactivation domain (25).

hASH1 Maps to Human Chromosome 12. Physical localization of the hASH1 gene was achieved by PCR amplifica-

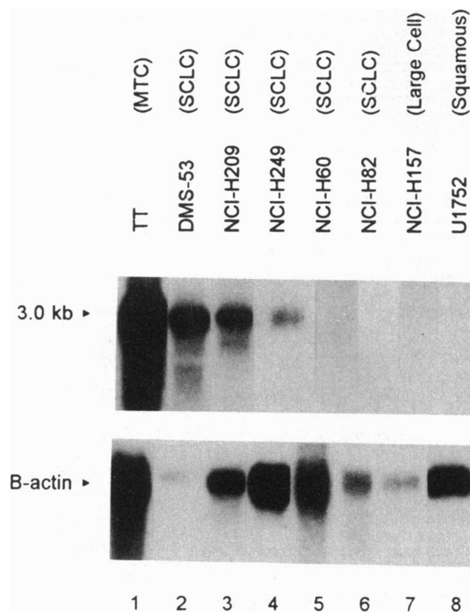

FIG. 3. hASH1 expression in TT cells and in human lung cancer cell lines. Estimated loading of poly(A)$^+$ RNA and exposure times were as follows: Lanes: 1, 3.5 μg, 14 h; 2, 4 μg, 24 h; 3, 1 μg, 24 h; 4, 4 μg, 24 h; 5, 2 μg, 72 h; 6, 4 μg, 24 h; 7, 3 μg, 48 h; 8, 1 μg, 16 h. All B-actin exposures were 6 h except lane 4, which was 15 h.

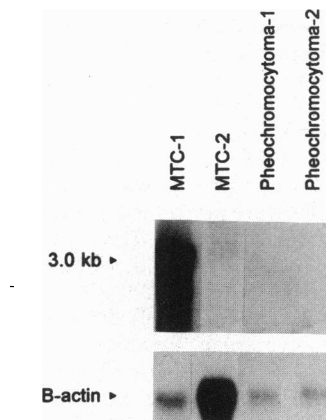

FIG. 4. hASH1 expression in human MTC and pheochromocytoma tumor specimens. Estimated loading of poly(A)$^+$ RNA and exposures are as follows. Lanes: MTC-1, 5 μg, 48 h; MTC-2, 2.5 μg, 72 h; pheochromocytomas 1 and 2, 2.5 μg, 72 h. B-actin exposures were 6 h.

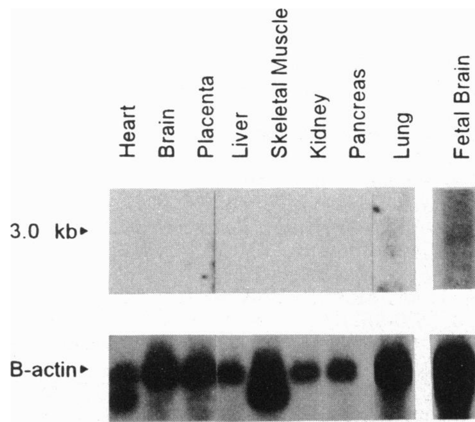

FIG. 5. hASH1 expression in normal adult tissues and fetal brain. Poly(A)$^+$ RNA (2 μg or 5 μg for fetal brain) was loaded per lane and exposed for 48 h (72 h for fetal brain). B-actin exposures were 3 h.

tion of a portion of the gene sequence using genomic DNA from a panel of rodent–human somatic cell lines representing the entire human DNA complement (mainly as monochromosome hybrids) (20). The amplified product included the trinucleotide repeat elements and flanking unique sequences. Only hybrids containing human chromosome 12 (e.g., GM10792 and GM10868; NIGMS Mutant Cell Repository, Camden, NJ) yielded PCR products of the expected length (188 bp). In addition, the repeat element was found to be polymorphic when tested against a set of parents from the CEPH reference pedigree collection. Thus far, nine alleles have been observed, with one allele dominant, at a frequency of 65%. The heterozygosity of the polymorphism in the CEPH families is 51%. Subregional localization and incorporation of the hASH1 locus within a genetic linkage map of chromosome 12 will be described elsewhere (D.C., S.D., and H.D.-K., unpublished data).

hASH1 Expression in Neuroendocrine Tissues. We studied expression of hASH1 in a variety of cultured cell lines, tumor specimens, and normal human tissues by using Northern blot analyses. The TT line of human MTC, from which the hASH1 gene was cloned, exhibited an abundant 3.0-kb transcript (Fig. 3). Two fresh MTC tumors also had detectable hASH1 expression (Fig. 4). Tumor MTC-1 exhibited a signal comparable in intensity to the TT line, whereas the transcript from tumor MTC-2 was only faintly visible after 72 h. In addition, the calcitonin-producing SCLC line DMS53 also produced an intense transcript (Fig. 3). Two classic SCLC lines, NCI-H209 and H249, also had significant hASH1 expression, whereas NCI-H82, a variant SCLC line that lacks neuroendocrine features such as L-dopa decarboxylase activity (16), had no demonstrable hASH1 expression. NCI-H60, a classic SCLC line with low to intermediate L-dopa decarboxylase, lacked readily detectable hASH1 transcripts. In addition, two non-SCLC lines that lack neuroendocrine features, the NCI-H157 line of large cell lung cancer and the U1752 line of squamous cell lung carcinoma, also had no detectable hASH1 transcripts. Interestingly, two fresh pheochromocytoma tumor specimens lacked apparent hASH1 expression (Fig. 4). A panel of poly(A)$^+$ RNA from adult heart, brain, placenta, liver, skeletal muscle, kidney, pancreas, and lung failed to show any distinct transcripts (Fig. 5), nor did cultured IMR-90 human fibroblasts (data not shown). A 72-h exposure of 5 μg of poly(A)$^+$ RNA from midgestation human fetal brain revealed a very faint hybridizing signal for hASH1 (Fig. 5).

DISCUSSION

bHLH proteins form a large family of transcriptional regulatory proteins, including tissue-specific transcriptional activators, widely expressed "partner factors," and a number of transcriptional inhibitors. The mammalian achaete-scute homologs may constitute an important structural and functional subclass within the larger family of bHLH proteins. In analogy to the *Drosophila* achaete-scute complex, the expression of rodent MASH-1 appears to be restricted to several regions of the developing rodent nervous system including portions of the central nervous system, peripheral sensory nervous system, and neural crest (27). In these tissues, MASH-1 expression generally precedes the appearance of markers of neuronal differentiation such as tyrosine hydroxylase and neurofilaments and, like achaete-scute, appears to be extinguished shortly prior to terminal neuronal differentiation (9, 14, 15). To date, the transcriptional targets for mammalian achaete-scute homologs are unknown, as are the effects of overexpression or inactivation of these proteins.

We have now isolated a human achaete-scute homolog, hASH1, that is expressed in a subset of tumors manifesting distinct neuroendocrine properties. Our findings extend an initial report of expression of the rodent homolog, MASH-1, in a single cultured line of rat MTC (14). The presence of hASH1 mRNA in human MTC tumors and cell lines has potential significance for understanding the function of this protein in neuroendocrine cell development. MTC arises from thyroid parafollicular C-cells, which are derived from embryonic neural crest (28). Although the expression pattern of achaete-scute homologs in normal C-cells is still unknown, it is reasonable to expect that hASH1 may be expressed in embryonic C-cell precursors migrating from neural crest, in analogy to the demonstration of MASH-1 in migrating rat sympathoadrenal precursors by Lo *et al.* (27). In addition, hASH1 may be a candidate transactivating factor for the human calcitonin gene. This polypeptide hormone is the predominant secreted product of C-cells and MTC. An essential constitutive enhancer site in the 5′ flanking region of the human calcitonin gene contains paired E-box motifs, both capable of binding bHLH proteins (4, 5).

The detection of hASH1 expression in SCLC may have particular importance for understanding both the evolution of this common form of lung cancer and the role of achaete-scute homologs in neuroendocrine cell differentiation. In the series of lung cancer cell lines that we investigated, the expression of this transcription factor appears to correlate with the presence of characterized endocrine features of SCLC such as L-dopa decarboxylase activity (15), dense core secretory granules, and the production of polypeptide hormones such as calcitonin (29). Lung cancer cell lines that are relatively or absolutely deficient in these endocrine markers (NCI-H60, NCI-H82, NCI-H157, and U1752) also lacked detectable hASH1 expression. The presence of this transcription factor in endocrine-rich SCLC tumors suggests a potential role for mammalian achaete-scute homologs outside of classic neural and neural crest-derived tissues.

Although the cellular origins of SCLC remain to be definitively proven, data from several groups (for review, see ref. 1) suggest an ultimate derivation of this cancer not from neural crest but rather from endoderm-derived bronchial epithelia, possibly via an immature pulmonary endocrine cell. This concept is supported by both clinical data indicating that SCLC can evolve into tumors that have the non-SCLC phenotype (30) and gene insertion experiments indicating phenotypic plasticity in lung cancer cell lines. For example, coexpression of the c-*myc* and v-Ha-*ras* oncogenes can alter the phenotype of several classic SCLC cell lines toward that of non-SCLC (31), whereas v-Ha-*ras* gene insertion leads to partial endocrine differentiation of the DMS53 line (32). If SCLC is indeed understood to originate in endodermally derived bronchial epithelium, the current study provides an indication that achaete-scute homologs may be expressed

outside of the nervous system or neural crest. hASH1 might thus act as a regulator of neuroendocrine differentiation in normal and neoplastic cells of diverse lineage. In this context, hASH1 expression may provide a useful marker for delineating the control of neuroendocrine phenotypic features in cultured SCLC and may have potential as a clinical marker for differentiated SCLC tumors.

The absence of detectable hASH1 transcripts in two clinically benign pheochromocytomas is somewhat surprising. Pheochromocytomas arise from neural crest-derived chromaffin cells of the adrenal medulla. Although the migrating precursors of rodent adrenal medulla and sympathetic ganglia express MASH-1 peptide, there is apparently no MASH-1 expression in the adult rodent adrenal medulla (27). However, MASH-1 transcripts are readily apparent in the rodent PC-12 pheochromocytoma cell line and are further induced by treatment with nerve growth factor (14). Examination of additional benign and malignant sporadic pheochromocytomas, as well as inherited pheochromocytomas, is necessary for a firm conclusion regarding patterns of hASH1 expression in these disorders.

An outstanding structural feature of the hASH1 cDNA is a multiply repeated copy of the trinucleotide CAG. C + G-rich trinucleotide repeats of this size are frequently polymorphic (33) and may exhibit meiotic instability through an incompletely understood process termed "dynamic mutation" (23, 34). This process is exemplified by the progressive increase in CGG triplet number observed in the *FMR-1* gene in kindreds with the fragile X syndrome (35). Genomic instability at a CAG repeat in the human androgen receptor is associated with Kennedy disease, an X chromosome-linked form of spinal and bulbar muscle atrophy that also results in mild androgen resistance (36). Normal individuals have a range of 17–26 CAG repeats in the androgen receptor proximal coding sequence, whereas individuals affected with Kennedy disease have a repeat number ranging from 40 to 52 (36).

The possibility that structural alterations in the hASH1 CAG repeat may also play a role in human disease, including the endocrine tumors shown to express the factor, awaits further study. The hASH1 CAG repeat resembles the human androgen receptor trinucleotide repeat in terms of its size, spatial orientation within the coding sequence, polymorphism, and potential to code for a transcriptional activation domain. Although glutamine-rich sequences are known in several genes to function as transcriptional activation domains (25, 26), it is unclear whether amplification of a polyglutamine tract would enhance or impair the transactivating function of hASH1 or the androgen receptor, and in turn how such alterations would result in disease. The hASH1 CAG repeat sequence, like the human androgen receptor repeat, is unique to the human gene and is not conserved in its rodent counterpart. Precise subchromosomal localization and linkage analysis of hASH1 will be useful to demonstrate whether dynamic mutation of this motif is associated with human disease.

We thank David Anderson (California Institute of Technology) for MASH-1 and MASH-2 probes, Olive Pettengill and George Sorenson (Dartmouth College) for the DMS53 cell line, and Robert Udelsman, Mack Mabry, and Robert Casero (Johns Hopkins University) for contributing tissue and RNA samples. This work was supported in part by a Physician-Scientist Award (K12-DK01298) and a Daland Fellowship from the American Philosophical Society (D.W.B.), by American Cancer Society Grant PDT417H (S.B.B. and B.D.N.), and by National Institutes of Health Grants RO1 CA47480 (B.D.N.) and RO1 CA53524 (H.D.-K.).

1. Mabry, M., Nelkin, B. D., Falco, J. P., Barr, L. F. & Baylin, S. B. (1991) *Cancer Cells* **3**, 53–58.
2. Johnson, B. E., Battey, J., Linnoila, I., Becker, K. L., Makuch, R. W., Snider, R. H., Carney, D. N. & Minna, J. D. (1986) *J. Clin. Invest.* **78**, 525–532.
3. Baylin, S. B. & Mendelsohn, G. (1982) in *Tumor Cell Heterogeneity: Origins and Implications*, eds. Owens, A. H., Coffey, D. S. & Baylin, S. B. (Academic, New York), pp. 9–27.
4. Ball, D. W., Compton, D., Nelkin, B. D., Baylin, S. B. & de Bustros, A. (1992) *Nucleic Acids Res.* **20**, 117–123.
5. Peleg, S., Abruzzese, R. V., Cote, G. J. & Gagel, R. F. (1990) *Mol. Endocrinol.* **4**, 1750–1757.
6. Moss, L. G., Moss, J. B. & Rutter, W. J. (1988) *Mol. Cell. Biol.* **8**, 2620–2627.
7. Wang, T. C. & Brand, S. J. (1990) *J. Biol. Chem.* **265**, 8908–8914.
8. Wheeler, M. B., Nishitani, J., Buchanan, A. M. J., Kopin, A. S., Chey, W. Y., Chang, T.-A. & Leiter, A. B. (1992) *Mol. Cell. Biol.* **12**, 3531–3539.
9. Campuzano, S. & Mododell, J. (1992) *Trends Genet.* **8**, 202–208.
10. Balcells, L., Mododell, J. & Ruiz-Gomez, M. (1988) *EMBO J.* **7**, 3899–3906.
11. Caudy, M., Vassin, H., Brand, M., Tuna, R., Jan, L. Y. & Jan, Y. N. (1988) *Cell* **55**, 1061–1067.
12. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56**, 777–783.
13. Ellis, H. M., Spann, D. R. & Posakony, J. W. (1990) *Cell* **61**, 27–38.
14. Johnson, J. E., Birren, S. J. & Anderson, D. J. (1990) *Nature (London)* **346**, 858–861.
15. Carney, D. N., Gazdar, A. F., Bepler, G., Guccion, J. G., Marangos, P. J., Moody, T. W., Zweig, M. H. & Minna, J. D. (1985) *Cancer Res.* **45**, 2913–2923.
16. Casero, R. A., Celano, P., Ervin, S. J., Porter, C. W., Bergeron, R. J. & Libby, P. R. (1989) *Cancer Res.* **49**, 3829–3833.
17. Bergh, J., Nilsson, K., Zech, L. & Giovanella, B. (1981) *Anticancer Res.* **1**, 317–322.
18. de Bustros, A., Baylin, S. B., Levine, M. A. & Nelkin, B. D. (1986) *J. Biol. Chem.* **261**, 8036–8041.
19. Chiu Yen, R.-W., Vertino, P. M., Nelkin, B. D., Yu, J. J., El-Deiry, W., Cumaraswamy, A., Lennon, G. G., Trask, B. J., Celano, P. & Baylin, S. B. (1992) *Nucleic Acids Res.* **20**, 2287–2291.
20. Dubois, B. L. & Naylor, S. L. (1993) *Genomics* **16**, in press.
21. Kozak, M. (1986) *Cell* **44**, 283–292.
22. Bird, A. P. (1986) *Nature (London)* **321**, 209–213.
23. Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. G. & Nelson, D. L. (1992) *Science* **256**, 784–789.
24. Huntington's Disease Collaborative Research Group (1993) *Cell* **72**, 971–983.
25. Courney, A. J. & Tjian, R. (1988) *Cell* **55**, 887–898.
26. Mitchell, P. J. & Tjian, R. (1989) *Science* **245**, 371–378.
27. Lo, L. C., Johnson, J. E., Wuenschell, C. W., Saito, T. & Anderson, D. J. (1991) *Genes Dev.* **5**, 1524–1537.
28. Le Douarin, N. (1980) *Curr. Top. Dev. Biol.* **16**, 31–85.
29. Edbrooke, M. R., Parker, D., McVey, J. H., Riley, J. H., Sorenson, G. D., Pettengill, O. S. & Craig, R. S. (1986) *EMBO J.* **4**, 715–724.
30. Brereton, H. D., Mathews, M. M., Costa, J., Kent, C. H. & Johnson, R. E. (1978) *Ann. Intern. Med.* **88**, 805–806.
31. Falco, J. P., Baylin, S. B., Lupu, R., Borges, M., Nelkin, B. D., Jasti, R., Davidson, N. E. & Mabry, M. (1990) *J. Clin. Invest.* **85**, 1740–1745.
32. Mabry, M., Nakagawa, T., Baylin, S. B., Pettengill, O., Sorenson, G. & Nelkin, B. D. (1989) *J. Clin. Invest.* **84**, 194–199.
33. Riggins, G. J., Lokey, L. K., Chastain, J. L., Leiner, H. A., Sherman, S. L., Wilkinson, K. D. & Warren, S. T. (1992) *Nature Genet.* **2**, 186–190.
34. Richards, R. I. & Sutherland, G. I. (1992) *Cell* **70**, 709–712.
35. Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K. L., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R. & Richards, R. I. (1991) *Science* **252**, 1711–1714.
36. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. (1991) *Nature (London)* **352**, 77–79.