MOLECULAR AND CELLULAR BIOLOGY, Jan. 1988, p. 186–195
0270-7306/88/010186-10$02.00/0

Vol. 8, No. 1

Structure and Expression of the Human L-*myc* Gene Reveal a Complex Pattern of Alternative mRNA Processing

FREDERIC KAYE, JAMES BATTEY, MARION NAU, BURKE BROOKS, ERIC SEIFTER, JACQUES DE GREVE, MICHAEL BIRRER, EDWARD SAUSVILLE, AND JOHN MINNA*

NCI-Navy Medical Oncology Branch, National Cancer Institute, Naval Hospital, and Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814

Received 26 August 1987/Accepted 22 October 1987

We analyzed in detail the structure of the L-*myc* gene isolated from human placental DNA and characterized its expression in several small-cell lung cancer cell lines. The gene is composed of three exons and two introns spanning 6.6 kilobases in human DNA. Several distinct mRNA species are produced in all small-cell lung cancer cell lines that express L-*myc*. These transcripts are generated from a single gene by alternative splicing of introns 1 and 2 and by use of alternative polyadenylation signals. In some mRNAs there is a long open reading frame with a predicted translated protein of 364 residues. Amino acid sequence comparison with c-*myc* and N-*myc* demonstrated multiple discrete regions with extensive homology. In contrast, other mRNA transcripts, generated by alternative processing, could encode a truncated protein with a novel carboxy-terminal end.

The L-*myc* gene was first isolated from a human small-cell lung cancer (SCLC) cell line, where it was detected by the presence of gene amplification and its sequence homology to c-*myc* and N-*myc* (42). Our initial studies showed that this gene was located in human chromosome region 1p32 and was frequently expressed with and without gene amplification in human SCLC cell lines. Given the abundant evidence of a role for deregulated c-*myc* or N-*myc* expression in several tumor types and their transforming activity in transfection assays (15, 31, 53, 56, 59), it seemed possible that deregulated expression of L-*myc* might be significant in the pathogenesis of SCLC.

C-*myc* and N-*myc* both have three exons, including an apparently untranslated exon 1, and several areas of nucleotide sequence homology (4, 29, 55). Comparison of their predicted amino acid sequences reveals several discrete, highly conserved domains. Although the biochemical function of c-*myc* or N-*myc* remains unknown, both genes code for nuclear phosphorylated proteins of short half-life and both bind to DNA (22, 23, 46). These similarities suggest that c-*myc* and N-*myc* are functionally related. Recently, c-*myc*, N-*myc*, and L-*myc* were found to have different patterns of expression in developing murine tissues, with L-*myc* expressed selectively in fetal lung, brain, and kidney and in adult lung but not other tissues examined (61). This was in contrast to the widely distributed expression of c-*myc* in many tissues and more similar to the restricted expression of N-*myc*. These findings suggested that the temporal and tissue-specific expression of N-*myc* and L-*myc* is important in mammalian development.

In this paper we describe the primary structure of the human L-*myc* gene and in so doing characterize its transcriptional unit, identify open reading frames, and examine its expression in SCLC. The topology of the L-*myc* gene is similar to those of c-*myc* and N-*myc*, with the major open reading frame sharing several areas of substantial amino acid sequence homology to the other *myc* family members interspersed with regions of unique sequence. In contrast to c-*myc* and N-*myc*, however, L-*myc* transcripts undergo a complex series of alternative mRNA processing and polyadenylation site selection in SCLC, giving rise to a family of several different mRNA species. These species not only predict the existence of a generic *myc* family protein but also offer the possibility of other proteins with novel carboxy-terminal sequences that lack exon 3.

MATERIALS AND METHODS

Cell lines and cell culture. SCLC cell line derivation and culture were as previously described (13).

Nucleic acids. Total RNA (14) and poly(A)-selected RNA (3) were extracted from the SCLC cell lines. The nuclear RNA fraction was obtained as previously described (51). The monoribosome- and polyribosome-associated fractions of total cytoplasmic RNA were obtained by centrifugation of a cytoplasmic preparation through a 10%-25% sucrose gradient and pooling of the appropriate fractions (25, 51).

Isolation and sequencing of L-*myc* cDNA and genomic clones. A cDNA library primed by oligo(dT)$_{12}$ to poly(A)-selected RNA from SCLC line NCI H209 (a cell line expressing but not amplified for L-*myc*) was constructed as described by Gubler and Hoffman (21), except that after synthesis of the second strand, the cDNA was methylated, *Eco*RI linkers were added, and the product was ligated into the *Eco*RI site of λgt11 (36, 60).

To obtain clones extending to the 5′ end of L-*myc* mRNA, a random-primed and specific-oligonucleotide-primed library was constructed as follows. Oligonucleotides complementary to sequence coordinates (see Fig. 2) 719 to 740, 739 to 758, and 752 to 773 were separately (approximately 2 μg each) mixed with 50 μg of dpN$_6$ (Collaborative Research, Inc., Waltham, Mass.) and 10 μg of poly(A)-selected RNA from NCI H510 and evaporated to dryness. The pellet was dissolved in 9 μl of H_2O and 1 μl of 0.1 M methyl mercury hydroxide (Alfa Products, Danvers, Mass.). After this mixture was heated to 42°C for 10 min, 2-mercaptoethanol was added to 0.016 M and first-strand synthesis was performed with Moloney murine leukemia virus reverse transcriptase (1,000 U in a 100-μl volume) and buffer supplied by the

* Corresponding author.

manufacturer (Bethesda Research Laboratories, Gaithersburg, Md.). Second-strand synthesis was performed as for the oligo(dT)$_{12}$-primed library. The double-stranded product was then passed over Sephadex G100 after phenol extraction, precipitated, and then blunt ended with T4 DNA polymerase in the presence of T4 polynucleotide kinase. The *Eco*RI sites were methylated, followed by attachment of *Eco*RI linkers, and size fractionated on a 5% polyacrylamide gel. The resulting cDNA was then ligated to λgt10.

Genomic L- *myc* clones were isolated from a partially *Mbo*I-digested human placental library constructed in Charon phage 28A (a generous gift from P. Leder). L-*myc*-specific clones were screened by using a previously characterized 1.9-kilobase (kb) genomic *Sma*I-*Eco*RI fragment from SCLC line NCI H378 (42) and isolated by plaque filter hybridization methods (18). The nucleotide sequence was determined with overlapping M13 phage clones (40) by using the dideoxynucleotide chain termination method (50). An M13 universal primer (Collaborative Research) or synthetic oligonucleotide primers (Applied Biosystems 380B DNA synthesizer) were used for sequencing.

RNA analysis. (i) Northern analysis. For Northern (RNA) blot analysis, total or poly(A)-selected RNA was transferred to nitrocellulose filters after electrophoresis through denaturing formaldehyde-containing gels (34). Hybridization of nitrocellulose filters with [^{32}P]dCTP nick-translated DNA fragments was performed as previously described (48).

(ii) S1 nuclease assay. A uniformly ^{32}P-labeled DNA probe was annealed with 20 μg of total RNA from an unamplified SCLC cell line designated NCI H209. Generation of the single-stranded probe in the M13 vector system, hybridization conditions, and digestion with S1 nuclease were as previously described (18). Protected species were identified by electrophoresis on 5% polyacrylamide–8 M urea gels and autoradiography.

(iii) Primer extension. A uniformly labeled DNA fragment was generated by priming of a 55-base-pair (bp) synthetic oligonucleotide template (coordinates 209 to 263) with a complementary oligonucleotide (coordinates 254 to 263) in the presence of ^{32}P-labeled deoxynucleotides as previously described (17). The radioactive antisense strand was then purified from a strand-separating gel (37) and allowed to anneal in the presence of 40 μg of total RNA from SCLC cell lines that either express or do not express L-*myc* mRNA. Hybridization conditions were as previously described for the S1 nuclease assay (18). The overnight hybridization mixture was then ethanol precipitated and suspended in reverse transcriptase buffer (Bethesda Research Laboratories) in the presence of 0.8 mM deoxyribonucleotides and 25 U of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories) at 37°C for 1 h. The reaction was terminated with phenol extraction and ethanol precipitation, followed by denaturing 6% polyacrylamide electrophoresis and autoradiography.

RESULTS

L-*myc* genomic and cDNA clones. A previously characterized 1.9-kb *Sma*I-*Eco*RI L-*myc* probe (42) was used to isolate a series of overlapping bacteriophage clones from a normal human placental genomic library (47). Restriction site mapping analysis showed that these clones contained the expected L-*myc*-specific 6.6- and 10-kb *Eco*RI fragments responsible for a previously reported restriction fragment length polymorphism (42), as well as genomic DNA flanking these fragments (Fig. 1).

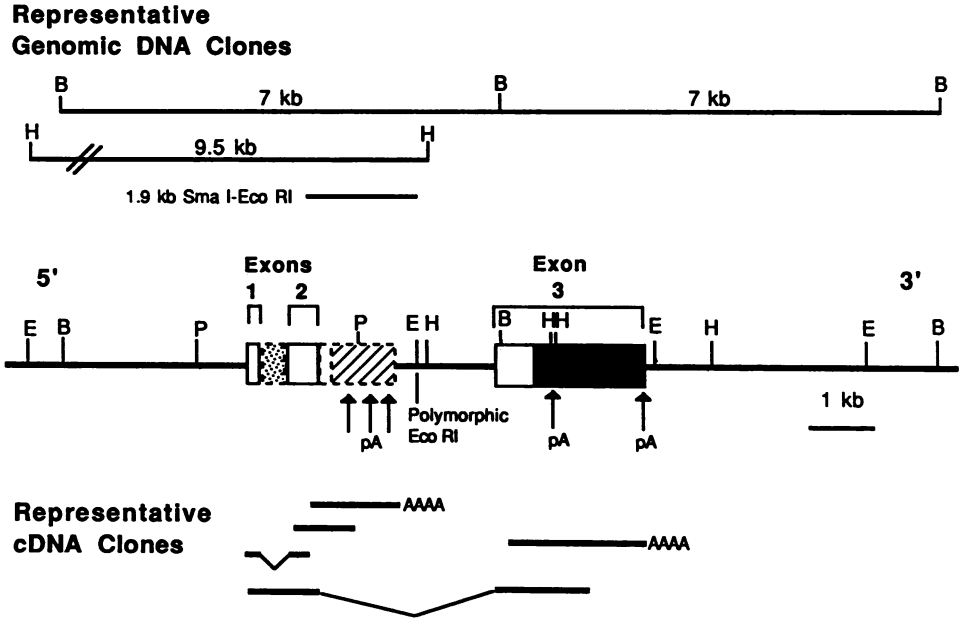

FIG. 1. Structural organization of the human L-*myc* locus, partial restriction enzyme map, and depiction of representative normal genomic DNA and SCLC cDNA clones. Boxes enclosed with solid lines, containing exons 1, 2, and 3, are identified. The stippled box between exons 1 and 2 represents intron 1, which is retained in many mature L-*myc* mRNA transcripts. The box enclosed with a dashed line adjacent to exon 2 represents that portion of intron 2 retained in mature mRNA which terminates at an alternative poly(A) (pA) signal in intron 2. The area filled in with hatched lines within this box represents predicted untranslated sequences, whereas the solid area of the exon 3 box similarly represents an untranslated region. Abbreviations: E, *Eco*RI; B, *Bam*HI; P, *Pvu*II; H, *Hin*dIII. Several additional *Pvu*II sites within exon 3 are not shown.

A series of cDNA clones was isolated from cDNA libraries from two SCLC cell lines that express the L-*myc* gene: NCI H510, a cell line with L-*myc* amplification, and NCI H209, a cell line without amplification. Specific oligonucleotides were used, in addition to a random hexamer primer, to initiate first-strand cDNA synthesis in one of the libraries (H510) in an effort to enrich for clones extending to the 5′ end of L-*myc* transcripts (see Materials and Methods). Figure 1 shows a comparison of the structures of the genomic and several representative cDNA clones isolated from these libraries. This comparison reveals a three-exon structure for the L-*myc* gene in genomic DNA, analogous to that previously reported for the c-*myc* and N-*myc* genes (4, 29, 55). In addition, many of the cDNA clones retained intron 1, portions of intron 2, or both, indicating a heterogeneous pattern of intron removal characteristic of alternative RNA processing.

Nucleotide sequence of the L-*myc* gene. The nucleotide sequence of the human L-*myc* gene, obtained from both placental genomic and SCLC cDNA clones, is shown in Fig. 2. The structural organization of this gene closely resembles the pattern established for the *myc* gene family, with three exons and two introns spanning approximately 6.6 kb. Comparison of the nucleotide sequences of placental DNA and our SCLC cDNA clones revealed no somatic mutations except for several isolated differences in the 3′ untranslated region, which may represent polymorphisms. Identification of the intron-exon splice boundaries was confirmed by our cDNA clones, and analysis of the nucleotide sequence revealed appropriate consensus and flanking sequences for the 5′ donor and 3′ acceptor of intron 1 (41). Similar consensus sequences are present for intron 2, except that two purines interrupt the stretch of flanking pyrimidines just adjacent to the 3′ acceptor of intron 2 at positions 4043 and 4044. A long open reading frame initiates near the 5′ end of exon 2 (nucleotide position 582) and extends for 364 predicted amino acid residues until a TAA stop codon within exon 3 at nucleotide position 4647 (Fig. 2). There are numerous stop codons in other reading frames of exons 2 and 3 and, as shown later, this translational frame generates a protein with significant homology to other *myc* proteins (4, 29, 55).

The L-*myc* gene undergoes alternative mRNA processing to generate several mRNA species. Our analysis of L-*myc* cDNA clones revealed a significant difference from those reported for c-*myc* and N-*myc*, namely, the presence of abundant mRNA transcripts which retain sequences from intron 1 or 2. We identified two major groups of cDNA clones which represent transcripts which either contain or have spliced out intron 1 but which have properly removed intron 2. These cDNA clones differ by only the presence or absence of the 363 bp encoding the first intron and do not disrupt the long open reading frame (nucleotides 582 through 4647) which codes for a protein similar in structure to those of c-*myc* and N-*myc*. Both forms terminate at a consensus AATAAA polyadenylation signal at nucleotide position 6572 in the 3′ untranslated region located 10 bp upstream from the poly(A) tail in exon 3. In addition, another polyadenylation signal is found in this exon at position 4795 in the 3′ untranslated region, but we found no evidence that this site is functional.

In contrast, we identified a distinct set of cDNA clones which retain part of intron 2 and terminate at a functional polyadenylation site at nucleotide position 2383 within intron 2 or possibly at two other potential polyadenylation signals (positions 1460 and 2036) also present in intron 2. These species are considerably smaller, since they contain variable portions of intron 2 but lack all of exon 3. More importantly, mRNA species represented by these cDNA clones can no longer encode an L-*myc* protein that contains exon 3 protein-coding sequences. These RNA-processing alternatives result in a family of L-*myc* transcripts of differing sizes depending on both the polyadenylation site and the splicing pattern used. Therefore, we designated the family of mRNA species which have completely removed intron 2 and terminate at a polyadenylation signal in distal exon 3 as long-form transcripts. In contrast, those mRNA species which have not processed intron 2 and terminate early at polyadenylation signals in intron 2 are called short-form transcripts. The dotted lines in Fig. 2 enclose sequences from intron 1 or 2 which are retained with high frequency in L-*myc* mRNA.

To confirm that our cDNA clones were representative of mRNA expression in SCLC, we examined the pattern of expression of the L-*myc* gene in SCLC by using probes specific for various mRNA forms (Fig. 3). A probe generated exclusively from intron 1 (probe A) only detected the larger 3.9-kb band representing the L-*myc* transcript which has retained intron 1 and, on a longer exposure, this intron 1 probe also recognized the upper half of the heterogeneous short-form transcripts. Using either an exon 1 probe (data not shown) or an exon 2 probe (probe B), we recognized two mRNA species at 3.9 and 3.6 kb (long forms) and a heterogeneous population of short-form transcripts at approximately 2.2 kb. In contrast, an intron 2 probe (probe C) only detected the short-form transcripts, demonstrating that intron 2 remains only in the shorter 2.2-kb family of mRNA species. Finally, when a fragment from exon 3 was used as a probe (probe D), only the 3.9- and 3.6-kb bands were seen, even on an overexposed film, demonstrating that the 2.2-kb family of mRNAs lacks exon 3 sequences. The two long-form bands could be clearly separated by using a shorter exposure from a different experiment (probe E).

To establish that these transcripts are, in fact, not unstable nuclear precursors, we analyzed the monoribosome- and polyribosome-derived RNA fractions from NCI H209 cells for the presence of these mRNA species. Both long-form transcripts (spliced and unspliced intron 1) and the family of short-form transcripts (unspliced intron 2 terminating at a polyadenylation signal within intron 2) are enriched in the ribosome-associated fraction (Fig. 4).

Mapping of the initiation site and alternative processing of intron 1. To define precisely the 5′ boundary of the L-*myc* gene and to establish the prevalence of mRNA forms that included or lacked intron 1, we used S1 nuclease and primer extension analyses. We prepared two adjacent genomic fragments for S1 nuclease analysis (probes A and B in Fig. 5) that spanned 883 nucleotides and extended from the putative 5′-flanking region to the middle of intron 1. No region of upstream probe A was protected by mRNA from cells producing L-*myc* transcripts. In contrast, probe B revealed two major species of protection corresponding to L-*myc* mRNA transcripts which have either retained (405-bp protection) or spliced out (205-bp protection) intron 1, consistent with the alternative splicing observed in our cDNA clones.

Although this analysis suggested an exon 1 of approximately 205 bp, we could not exclude the possibility of an additional upstream exon. To resolve this question, a primer extension experiment was performed with mRNAs from two cell lines which produce L-*myc* (NCI H510 and NCI H209) and from a cell line which produces N-*myc* but no detectable L-*myc* (NCI N592) (Fig. 5). Extension from an oligonucleo-

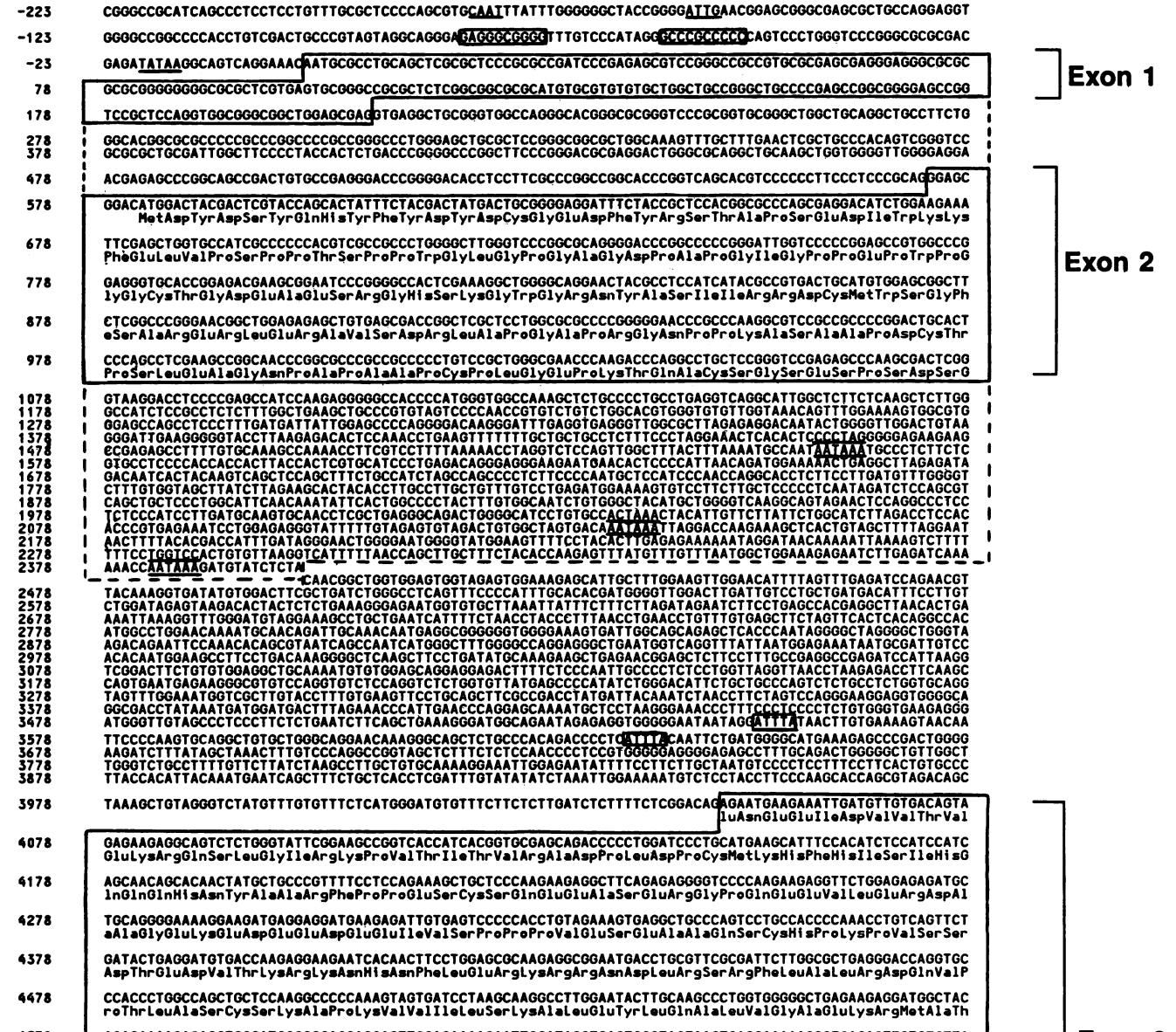

FIG. 2. Nucleotide sequence of the human L-*myc* gene. Solid boxes enclose the three exons of the L-*myc* gene, and dotted lines indicate sequences from introns 1 and 2 which are retained with high frequency in mature L-*myc* mRNA. Two Sp1-binding sites are enclosed in boxes at positions −50 and −73. A consensus TATAA (position −19 to −15) is underlined and, although no CCAAT elements are present, a CAAT sequence and its inverted repeat ATTG are also underlined at positions −181 and −156, respectively. Polyadenylation signals are indicated with a line above and below the sequence AATAAA. Recently, Shaw and Kamen (54) have described an AT-rich sequence in the 3′ untranslated region which confers selective mRNA instability, and a consensus ATTTA sequence has been recognized by comparative analysis of a number of distinct short-lived mRNAs. Analysis of the L-*myc* sequence identified two closely spaced ATTTA elements at positions 6478 and 6514 enclosed in a box in the distal 3′ untranslated region. Two other additional ATTTA pentamer sequences are also present in distal intron 2.

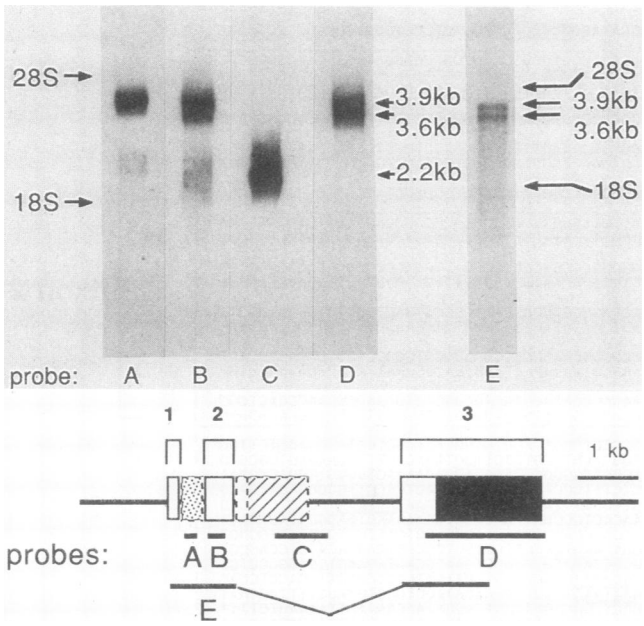

FIG. 3. Analysis of RNA expression with nonoverlapping probes revealed distinct L-*myc* mRNA species. Lanes A to D: Poly(A)-selected RNA (1 μg) from NCI H510 was size fractionated and transferred to nitrocellulose. Duplicate lanes were then hybridized with the following probes. Probe A: Two different oligonucleotides spanning nucleotide coordinates 209 to 263 and 518 to 572 (both contained within intron 1) were uniformly labeled by priming with a complementary nanomer oligonucleotide in the presence of ^{32}P-labeled deoxyribonucleotides (17). Probe B: A 200-bp *Sma*I-*Sma*I fragment from the 5′ region of exon 2. Probe C: A 1-kb *Bgl*II-*Eco*RI fragment from intron 2. Probe D: A 2-kb cDNA insert containing only exon 3 sequences. For lane E, 10 μg of total RNA was size fractionated in a similar experiment but hybridized with probe E (a 1.9-kb cDNA insert containing sequences from exon 1, intron 1, exon 2, and portions of exon 3). A light exposure is presented to demonstrate clearly the two distinct long-form mRNA transcripts of L-*myc* (see the text).

tide primer (originating from nucleotide coordinates 209 to 263) mapped the 5′ boundary of the L-*myc* gene to the same location previously implicated by S1 nuclease analysis. This site is also within 5 nucleotides of our most 5′ cDNA clones. Taken together, these data establish this location as a transcription initiation site of the L-*myc* locus (nucleotide 1 in Fig. 2). A consensus TATAA sequence is found 15 bp upstream from this site, and a CAAT sequence and its inverted repeat ATTG are found at nucleotide positions −181 and −156, respectively. In addition, this region is flanked by two consensus Sp1-binding sites (9 of 10 nucleotides match) (27) at positions −50 and −73. Similar GC-rich regions with Sp1-binding sites have been identified in an increasing number of eucaryotic genes, including housekeeping genes (19, 39) and, more recently, proto-oncogenes (5, 26). Such elements have been associated with 5′ heterogeneity of gene transcription (52). Although we observed several discrete intermediate bands on longer exposures of the S1 nuclease and primer extension experiments, further investigation is required to determine whether these bands also represent heterogeneous transcription start sites (Fig. 5).

Predicted proteins encoded by L-*myc*. Inspection of the predicted amino acid sequence of the long form of the L-*myc* protein readily identified numerous discrete regions of extensive homology with the c-*myc* and N-*myc* protein products (Fig. 6). Flanking these regions of extensive homology, the three *myc* genes sharply diverge, with frequent insertions and deletions. There are two additional upstream ATGs present in exon 1 of L-*myc*. The first ATG is found only in our most 5′ cDNA clones at position 2 and is followed by a stop codon in frame at position 98. The other exon 1 ATG is found at position 128 and is in frame with the long-form open reading frame if intron 1 has been spliced. Translation from this upstream open reading frame would, therefore, add an additional 30 amino-terminal residues. If intron 1 has not been removed, however, the open reading frame commencing at position 128 is not in frame with the exon 2 ATG, terminating at a stop codon at position 620. Neither of the two exon 1 ATGs, however, conforms to the requirements for a strong consensus translational start site (30) in that both contain pyrimidines at the −3 position. In contrast, the sequence CGGACATGG (coordinates 577 to 585) flanking the exon 2 ATG is more favorable, suggesting that this is the predominant initiation site for protein translation.

Since we also demonstrated that a substantial amount of steady-state L-*myc* mRNA transcripts terminate in intron 2 (by use of an alternative polyadenylation site), we expect that a protein that arises from such a short-form mRNA will

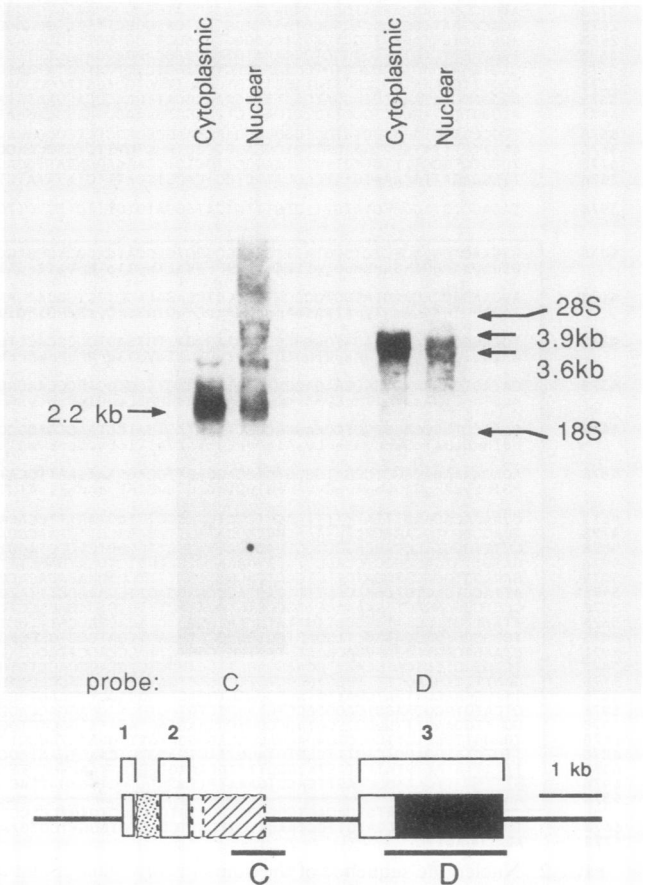

FIG. 4. All L-*myc* mRNA species are enriched in the monoribosome- and polyribosome-associated fractions. RNA (2 μg) isolated from either a nuclear preparation (nuclear) or the mono- and polyribosome fractions (cytoplasmic) of cell line NCI H209 were size fractionated and transferred to nitrocellulose. Probe C is specific for the short-form transcripts, whereas probe D detects exclusively long-form species.

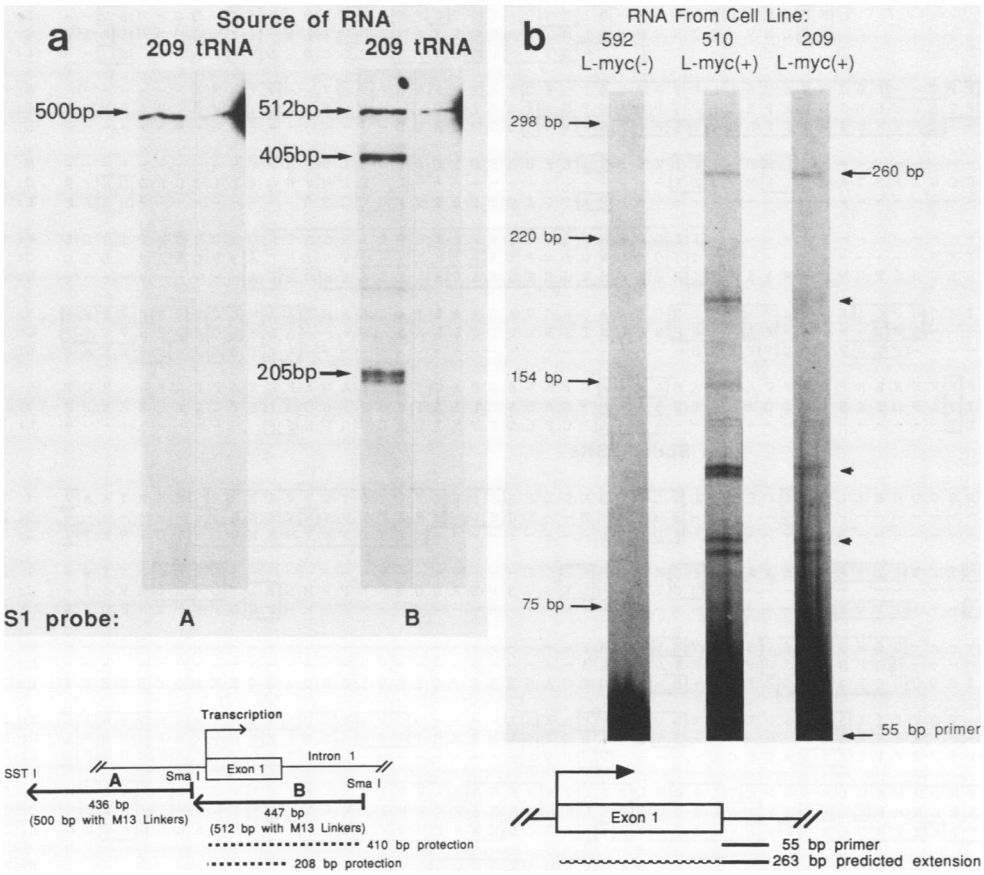

FIG. 5. (a) S1 nuclease analysis of the 5′ end of the L-*myc* gene. Protection of a 447-bp *Sma*I-*Sma*I fragment (probe B) by RNA isolated from an L-*myc*-producing cell line, NCI H209, revealed a major species at 405 bp (transcripts which have retained intron 1) and a closely spaced doublet at 205 bp (transcripts which have removed intron 1), whereas a tRNA control offered no protection. An upstream *Sst*I-*Sma*I fragment (probe A) was not protected. This maps the 5′ boundary of the first exon to within 5 nucleotides of our most 5′ cDNA clones, establishing an exon 1 of 208 bp. The lower band of the doublet at 205 bp may represent a population of mRNA species with an exon 1 approximately 10 bp smaller than that predicted by our longest cDNA clones. Although initiation of transcription from a start site slightly downstream from that proposed in Fig. 2 appears more favorable given the location of the TATAA consensus sequence (nucleotide position −15), we do not have direct evidence to support this hypothesis. Finally, this experiment demonstrated that large amounts of steady-state transcripts with an unspliced intron 1 exist. (b) Primer extension analysis maps the initiation site. A 55-bp uniformly labeled oligonucleotide primer (coordinates 209 to 263) was allowed to anneal with 40 μg of total RNA either from two cell lines (NCI H510 and NCI H209) that express L-*myc* or from a cell line (NCI N592) that produces no detectable L-*myc*. Full-length extension of approximately 260 bp in the presence of reverse transcriptase (see Materials and Methods) confirmed the 5′ boundary of the L-*myc* gene. Several intermediate bands (arrowhead) were also seen. Since this is a highly GC-rich region with potential for secondary structure formation, such bands may only represent strong stop sites for reverse transcriptase. Mapping of these smaller bands with the faint intermediate signals seen with the S1 nuclease assay revealed that they differ by approximately 20 bp. Further investigation, therefore, will be required to determine whether these signals represent heterogeneous start sites for transcription.

retain the amino-terminal domain encoded by exon 2 yet will have truncated the carboxy-terminal domain encoded by exon 3. This predicted protein would span 206 amino acids and contain 41 novel carboxy-terminal residues before reaching a TGA termination codon at position 1200 (Fig. 7). Thus, the presence of alternative splicing and alternative termination patterns in L-*myc* suggests the existence of a second type of *myc* protein that differs greatly from the L-*myc* long form.

DISCUSSION

The *myc* gene family. This work demonstrates that the structural organization of the *myc* gene family is highly conserved, with three exons and two introns (2, 4, 29, 55). In addition, the presence in L-*myc* of a predominant translational start site near the 5′ end of exon 2, with an apparently untranslated exon 1, resembles the organization of the other *myc* genes (4, 29, 55). Attempts to explain the evolutionary conservation of c-*myc* exon 1 have been controversial (12, 16), although recent evidence suggesting the possibility of additional open reading frames in this region is provocative (6). Sequence analysis of exon 1 and intron 1 of the L-*myc* gene has shown no significant homology with analogous regions of the c-*myc* and N-*myc* genes. This region of the L-*myc* gene, however, is highly G and C rich and may participate in the formation of very stable stem-and-loop structures, as has been previously predicted for both c-*myc* and N-*myc* (4, 29, 55). In addition, the exon 1-intron 1 boundary is a region of intragenic transcriptional pausing that may be important in the regulation of the c-*myc* gene (7, 44) and has been associated with clustering of somatic mutations in Burkitt lymphoma (45). These observations have led to the hypothesis that changes in this region are associated with alterations in the regulation of c-*myc* expres-

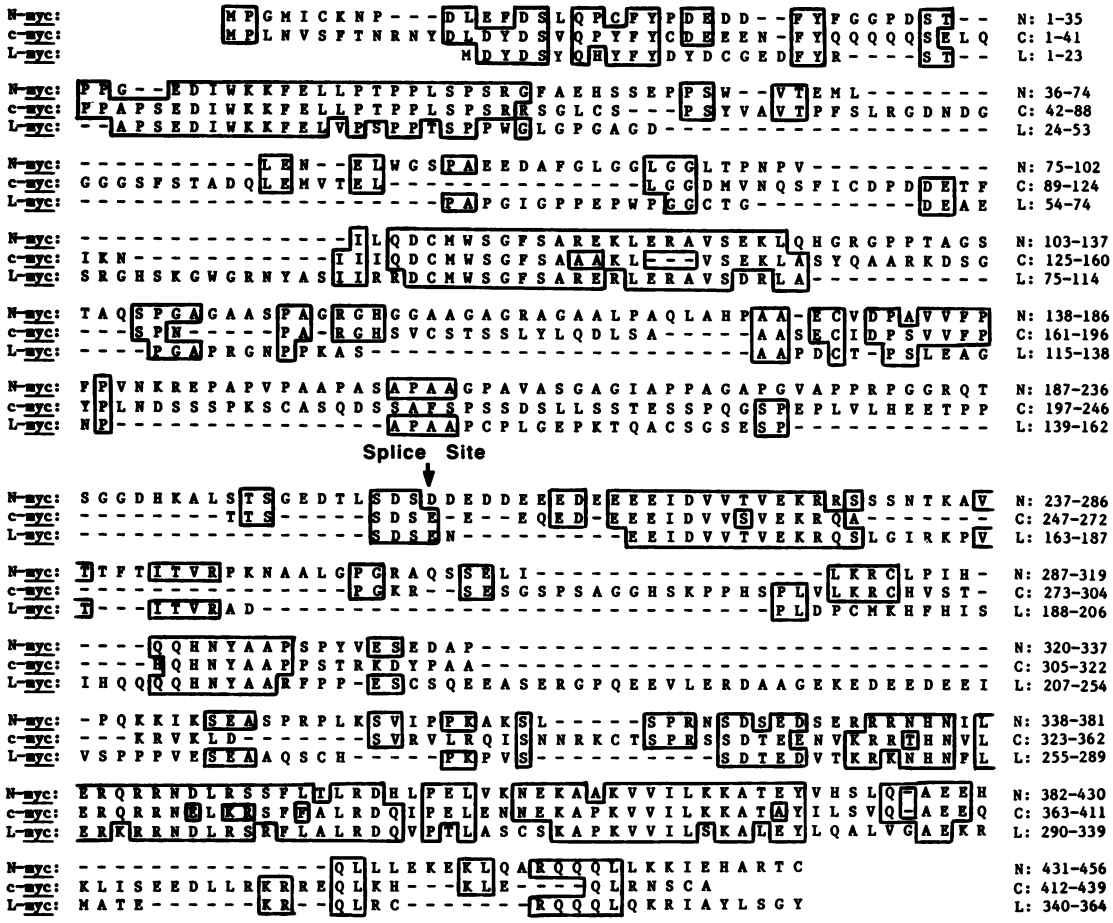

FIG. 6. Comparative analysis of the predicted *myc* gene products shows multiple domains with striking amino acid homology. The amino acid coordinates for the different *myc* gene proteins are shown in a column to the right. Amino acid alignment was performed by the computer program PrtAln (58) and by visual inspection.

sion. Whereas there are no obvious conserved sequences between c-*myc* and L-*myc* in the vicinity of this splice boundary, it is reasonable to test whether this region participates in similar regulatory processes in SCLC. The abundance of mRNA transcripts that contain intron 1 permitted us to sequence this region in cDNA from SCLC line H510, which is amplified for L-*myc*. We found no mutations in the expressed gene in this region in comparison with the normal placental DNA sequence. It will be important to analyze this region similarly in other SCLC lines activated for L-*myc* expression.

Comparison of the three homologous *myc* family proteins

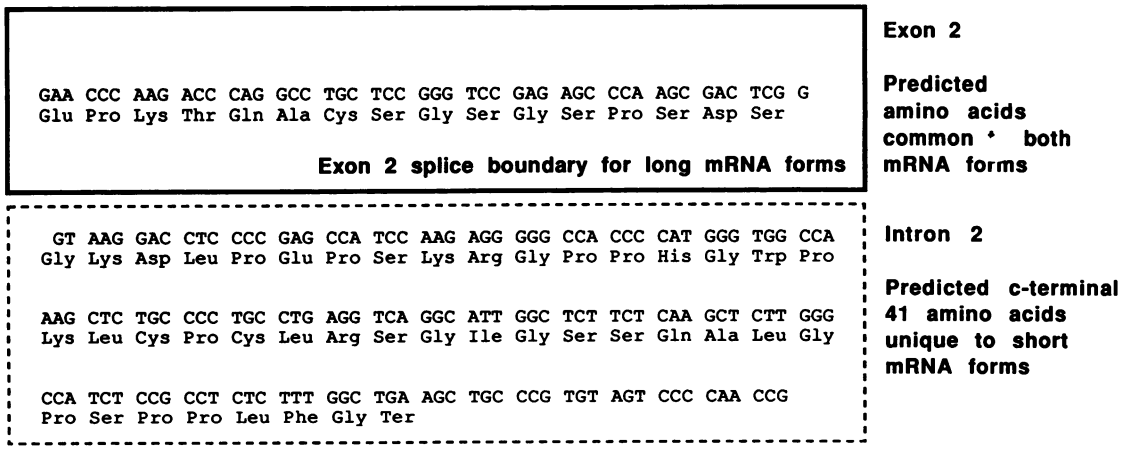

FIG. 7. The primary structure of the short-form L-*myc* transcript predicts a novel carboxy-terminal open reading frame.

showed that L-*myc* has a putative translation product of 364 amino acids, whereas c-*myc* (439 residues) and N-*myc* (456 or 464 residues if counted from a more upstream start site) contain larger translation products. Despite these differences, several regions of distinct amino acid sequence homology could be observed between the three proteins (Fig. 6). Recently, specific domains of the c-*myc* and v-*myc* proteins that are involved in in vitro transformation and nuclear localization have been described (24, 57). These analyses identified two discrete areas within the c-*myc* protein (c-*myc* amino acids 105 to 143 and 320 to 439) which were critical for complementation of *ras* genes in transforming normal rat embryo cells (57). These two critical domains contain specific regions of amino acid homology with both L-*myc* and N-*myc*. It will be of interest to test the biologic activity of *myc* genes with small deletions that only involve the areas of conserved sequence to further delineate specific sites of functional activity.

In contrast, large deletions within the most amino-terminal regions of the protein (c-*myc* amino acids 6 to 104) were required to abrogate transformation activity, whereas even large deletions in the middle of the protein (c-*myc* amino acids 144 to 320) did not affect full transforming activity. As shown in Fig. 6, there are several highly conserved amino acid sequences within these regions which can be deleted yet still allow retention of cotransformation activity. This suggests the presence of other, as yet undescribed, functions for these regions.

Evidence for alternative splicing in SCLC. We have presented evidence for multiple forms of L-*myc* mRNA transcribed from a single gene. These species were generated by alternative RNA splicing of introns 1 and 2 and by alternative polyadenylation site utilization. In addition, we have demonstrated that these mRNA species are poly(A)$^+$, contain potential open reading frames, and are significantly enriched in the ribosome-associated fraction of the cytoplasm. These observations support the notion that these are mature mRNAs which may encode a truncated protein with a novel carboxy-terminal end. Such a protein would lack sequences predicted to be important for nuclear localization and, therefore, would be expected to serve a different functional role.

Posttranscriptional mRNA processing in eucaryotic cells requires endonucleolytic cleavage and polyadenylation of the nuclear mRNA precursor about 10 to 30 bp downstream from a AATAAA consensus (8), occurring almost simultaneously with the removal of intervening sequences (1). In this regard, c-*myc* lacks an analogous AATAAA sequence whereas N-*myc* contains such a polyadenylation signal within intron 2. This suggests a search for alternative processing of N-*myc* in SCLC lines that express this gene.

Despite the rapidly growing number of eucaryotic genes now known to undergo complex posttranscriptional processing, little is known about the specific mechanisms that control these events (9, 33). *cis* elements, such as consensus sequences flanking the intron-exon boundaries (41) and lariat branchpoints (28, 49), are essential for efficient processing. Thus, it will be of interest to study the role that the two purines that interrupt a stretch of pyrimidines just adjacent to the 3′ acceptor sequence of L-*myc* intron 2 may play in generating short-form transcripts. Recently, soluble *trans* factors, induced during muscle-specific differentiation, have also been identified to be involved in alternative splicing (10). Neuron-specific alternative RNA processing has been described for a number of genes, including the calcitonin/CGRP gene (32) and possibly the pp60$^{c\text{-}src}$ gene (38), raising the questions of whether this is an important mechanism for gene regulation in neuroendocrine cells such as SCLC and whether alternative processing of L-*myc* occurs in other tissue types as well.

Biologic properties of the L-*myc* gene. Each of the *myc* gene family members can complement a mutated *ras* gene in transforming primary rat embryo fibroblast cells after cotransfection (2, 31, 59). In addition, L-*myc* expression is seen at specific periods in developing lungs, kidneys, and central nervous systems of mice (61). These data are consistent with the idea that the L-*myc* protein is involved in tissue-specific regulation of cell growth and that alterations in the expression of L-*myc* may participate in malignant transformation. Human SCLC is a unique tumor, in which amplification of all three *myc* family genes may be found in different tumors (11, 35, 43). It is of interest, however, that SCLC cell lines that produce L-*myc* tend to be relatively slow growing, exhibit a classic small-cell morphology, and express neuroendocrine markers, as compared with SCLC cell lines that express large amounts of c-*myc*, tend to have a variant morphology and rapid doubling time, and have lost many neuroendocrine features (13, 20). L-*myc* deregulation, therefore, may be important in one of the initial steps in the progression of events which leads to malignant change in the precursor of SCLC. Finally, the L-*myc* primary transcript is often processed into a number of different mature mRNAs which could be translated into structurally distinct proteins. It will be important to explore the role that these predicted protein products may serve in mediating normal or aberrant cellular functions.

ACKNOWLEDGMENTS

We are grateful to Nancy Shaper and Lisa Sultzman for advice on constructing the oligonucleotide-directed cDNA library, Adi Gazdar for the SCLC cell lines used in this study, Joe Fedorko for help with oligonucleotide synthesis, and Kelly Clemons and Sylvia Stephenson for excellent technical assistance. We thank Ron Depinho and Fred Alt for helpful discussions during the course of this work.

J. De Greve was supported by a grant from the Fogarty International Foundation.

LITERATURE CITED

1. **Abelson, J.** 1979. RNA processing and the intervening sequence problem. Annu. Rev. Biochem. **48**:1035–1069.
2. **Alt, F. W., R. DePinho, K. Zimmerman, E. Legouy, K. Hatton, P. Ferrier, A. Tesfaye, G. Yancopoulous, and P. Nisen.** 1986. The human *myc* gene family. Cold Spring Harbor Symp. Quant. Biol. **51**:931–941.
3. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1413.
4. **Battey, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder.** 1983. The human c-*myc* oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. Cell **34**:779–787.
5. **Bender, T., and W. M. Kuehl.** 1986. Murine *myb* proto-oncogene messenger RNA:cDNA sequence and evidence for 5′ heterogeneity. Proc. Natl. Acad. Sci. USA **83**:3204–3208.
6. **Bentley, D. L., and M. Groudine.** 1986. Novel promoter upstream of the human c-*myc* gene and regulation of c-*myc* expression in B-cell lymphomas. Mol. Cell. Biol. **6**:3481–3489.
7. **Bentley, D. L., and M. Groudine.** 1986. A block to elongation is largely responsible for decreased transcription of c-*myc* in differentiated HL-60 cells. Nature (London) **321**:702–706.
8. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**: 349–359.

9. **Breitbart, R. E., A. Andreadis, and B. Nadal-Ginard.** 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu. Rev. Biochem. **56**:467–495.
10. **Breitbart, R. E., and B. Nadal-Ginard.** 1987. Developmentally induced, muscle-specific *trans* factors control the differential splicing of alternative and constitutive troponin T exons. Cell **49**:793–803.
11. **Brooks, B. J., Jr., J. Battey, M. M. Nau, A. F. Gazdar, and J. D. Minna.** 1987. Amplification and expression of the *myc* gene in small cell lung cancer. Adv. Viral Oncol. **7**:155–172.
12. **Butnick, N. Z., C. Miyamoto, R. Chizzonite, B. R. Cullen, G. Ju, and A. M. Skalka.** 1985. Regulation of the human c-*myc* gene:5′ noncoding sequences do not affect translation. Mol. Cell. Biol. **5**:3009–3016.
13. **Carney, D. N., A. F. Gazdar, G. Bepler, J. G. Guccion, P. J. Marangos, T. W. Moody, M. H. Zeig, and J. D. Minna.** 1985. Establishment and identification of small cell lung cancer cell lines having classic and variant features. Cancer Res. **45**:2913–2923.
14. **Chirgwin, J., A. Przybyla, R. Macdonald, and W. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
15. **Dalla-Favera, R., S. Martinotti, R. Gallo, J. Erikson, and C. Croce.** 1983. Translocation and rearrangements of the c-*myc* oncogene locus in human undifferentiated B-cell lymphomas. Science **219**:963–967.
16. **Darveau, A., J. Pelletier, and N. Sonenberg.** 1985. Differential efficiencies of *in vitro* translation of mouse c-*myc* transcripts differing in the 5′ untranslated region. Proc. Natl. Acad. Sci. USA **82**:2315–2319.
17. **Dattagupta, N., D. Rabin, G. Michaud, and P. M. Rae.** 1987. A simple method for generation of high specific activity oligonucleotide probes. Biotechniques **5**:38–43.
18. **Davis, L. G., M. D. Dibner, and J. F. Battey.** 1986. Basic methods in molecular biology. Elsevier/North Holland Publishing Co., Amsterdam.
19. **Dynan, W. S., S. Sazer, R. Tjian, and R. T. Schimke.** 1986. Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. Nature (London) **319**:246–248.
20. **Gazdar, A. F., D. N. Carney, M. M. Nau, and J. D. Minna.** 1985. Characterization of variant subclasses of cell lines derived from small cell lung cancer having distinctive biochemical, morphological, and growth properties. Cancer Res. **45**:2924–2930.
21. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
22. **Hann, S. E., H. D. Abrams, L. R. Rohrschneider, and R. N. Eisenman.** 1983. Proteins encoded by the v-*myc* and c-*myc* oncogenes: identification and localization in acute leukemia virus transformants and bursal lymphoma cell lines. Cell **34**:789–798.
23. **Hann, S. R., and R. N. Eisenman.** 1984. Proteins encoded by the human c-*myc* oncogene: differential expression in neoplastic cells. Mol. Cell. Biol. **4**:2486–2497.
24. **Heaney, M. L., J. Pierce, and J. T. Parsons.** 1986. Site-directed mutagenesis of the *gag-myc* gene of avian myelocytomatosis virus 29: biological activity and intracellular localization of structurally altered proteins. J. Virol. **60**:167–176.
25. **Huang, M.-T., and A. P. Grollman.** 1972. Effects of aurintricarboxylic acid on ribosomes and the biosynthesis of globin in rabbit reticulocytes. Mol. Pharmacol. **8**:111–127.
26. **Ishii, S., J. T. Kadonaga, R. Tjian, J. N. Brady, G. T. Merlino, and I. Pastan.** 1986. Binding of the Sp1 transcription factor by the human Harvey *ras* 1 proto-oncogene promoter. Science **232**:1410–1413.
27. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**:20–23.
28. **Keller, E. B., and W. A. Noon.** 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc. Natl. Acad. Sci. USA **81**:7417–7420.
29. **Kohl, N. E., E. Legouy, R. A. DePinho, P. D. Nisen, R. K. Smith, C. E. Gee, and F. W. Alt.** 1986. Human N-*myc* is closely related in organization and nucleotide sequence to c-*myc*. Nature (London) **319**:73–77.
30. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic RNAs. Nucleic Acids Res. **12**:857–870.
31. **Land, H., L. F. Parada, and R. A. Weinberg.** 1983. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature (London) **304**:596–601.
32. **Leff, S. E., R. M. Evans, and M. G. Rosenfeld.** 1987. Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. Cell **48**:517–524.
33. **Leff, S. E., M. G. Rosenfeld, and R. M. Evans.** 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. Annu. Rev. Biochem. **55**:1091–1117.
34. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4743–4751.
35. **Little, C. D., M. M. Nau, D. N. Carney, A. F. Gazdar, and J. D. Minna.** 1983. Amplification and expression of the c-*myc* oncogene in human lung cancer cell lines. Nature (London) **306**:194–196.
36. **Littman, D. R., Y. Thomas, P. J. Madden, L. Chess, and R. Axel.** 1985. The isolation and sequence of the gene encoding T8: a molecule defining functional classes of T lymphocytes. Cell **40**:237–246.
37. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Martinez, R., B. Mathey-Prevot, A. Bernards, and D. Baltimore.** 1987. Neuronal pp60$^{c\text{-}src}$ contains a 6-amino-acid insertion relative to its non-neuronal counterpart. Science **237**:411–414.
39. **McGrogan, M., C. C. Simonsen, D. T. Smouse, P. J. Farnham, and R. T. Schimke.** 1985. Heterogeneity at the 5′ termini of mouse dihydofolate reductase mRNAs. Evidence for multiple promoter regions. J. Biol. Chem. **260**:2307–2314.
40. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
41. **Mount, S.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
42. **Nau, M. M., B. J. Brooks, J. Battey, E. Sausville, A. F. Gazdar, I. R. Kirsch, O. W. McBride, V. Bertness, G. F. Hollis, and J. D. Minna.** 1985. L-*myc*: a new *myc*-related gene amplified and expressed in human small cell lung cancer. Nature (London) **318**:69–73.
43. **Nau, M. M., B. J. Brooks, D. N. Carney, A. F. Gazdar, J. F. Battey, E. A. Sausville, and J. D. Minna.** 1986. Human small cell lung cancers show amplification and expression of the N-*myc* gene. Proc. Natl. Acad. Sci. USA **83**:1092–1096.
44. **Nepveau, A., and K. B. Marcu.** 1986. Intragenic pausing and anti-sense transcription within the murine c-*myc* locus. EMBO J. **5**:2859–2865.
45. **Pelicci, P. G., D. M. Knowles, I. Magrath, and R. Dalla-Favera.** 1986. Chromosomal breakpoints and structural alteration of the c-*myc* locus differ in endemic and sporadic forms of Burkitt lymphoma. Proc. Natl. Acad. Sci. USA **83**:2984–2988.
46. **Ramsay, G., L. Stanton, M. Schwab, and J. M. Bishop.** 1986. The human proto-oncogene N-*myc* encodes nuclear proteins that bind DNA. Mol. Cell. Biol. **6**:4450–4457.
47. **Ravetch, J. V., U. Siebenlist, S. Korsmeyer, T. Waldmann, and P. Leder.** 1981. Structure of the human immunoglobulin M locus: characterization of embryonic and rearranged J and D genes. Cell **27**:583–591.
48. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
49. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell **38**:317–331.
50. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci.

USA **74**:5463–5467.

51. **Sausville, E. A., A.-M. Lebacq-Verheyden, E. R. Spindel, F. Cuttitta, A. F. Gazdar, and J. F. Battey.** 1986. Expression of the gastrin-releasing peptide gene in human small cell lung cancer. Evidence for alternative processing resulting in three distinct mRNAs. J. Biol. Chem. **261**:2451–2457.
52. **Sazer, S., and R. T. Schimke.** 1986. A re-examination of the 5′ termini of mouse dihydrofolate reductase RNA. J. Biol. Chem. **261**:4685–4690.
53. **Schwab, M., J. Ellison, M. Busch, W. Rosenau, H. E. Varmus, and J. M. Bishop.** 1984. Enhanced expression of the human gene N-*myc* consequent to amplification of DNA may contribute to malignant progression of neuroblastoma. Proc. Natl. Acad. Sci. USA **81**:4940–4945.
54. **Shaw, G., and R. Kamen.** 1986. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell **46**:659–667.
55. **Stanton, L. W., M. Schwab, and J. M. Bishop.** 1986. Nucleotide sequence of the human N-*myc* gene. Proc. Natl. Acad. Sci. USA **83**:1772–1776.
56. **Stewart, T. A., P. K. Pattengale, and P. Leder.** 1984. Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes. Cell **38**:627–637.
57. **Stone, J., T. de Lange, G. Ramsay, E. Jakobovits, J. M. Bishop, H. Varmus, and W. Lee.** 1987. Definition of regions in human c-*myc* that are involved in transformation and nuclear localization. Mol. Cell. Biol. **7**:1697–1709.
58. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acids and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.
59. **Yancopoulous, G. D., P. D. Nisen, A. Tesfaye, N. E. Kohl, M. P. Goldfarb, and F. W. Alt.** 1985. N-*myc* can cooperate with *ras* to transform normal cells in culture. Proc. Natl. Acad. Sci. USA **82**:5455–5459.
60. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80**:1195–1198.
61. **Zimmerman, K. A., G. D. Yancopoulous, R. G. Collum, R. K. Smith, N. E. Kohl, K. A. Denis, M. M. Nau, O. N. Witte, D. Toran-Allerand, C. E. Gee, J. D. Minna, and F. W. Alt.** 1986. Differential expression of *myc* family genes during murine development. Nature (London) **319**:780–783.