Letters to the Editor


Letter

We believe that the first 47 bp of the com1 (candidate of metastasis-1) cDNA reported by Ree et al. (1) in Fig. 4 on page 4678 is not from human tissue. Instead, we think it is the Marathon adaptor that is found in many Clontech kits for generating cDNAs and ends of cDNAs.

Recently, we were trying to interpret sequences of wound-induced transcripts in plants that were isolated using a Clontech cDNA subtractive library procedure (2). While performing BLAST searches at the NCBI website (3), we noticed that a 44-bp sequence (CTAAT-ACGACTCATAAGGGTCTCAAGCGCCCAGGAGGT) appeared in many of our clones and repeatedly drew matches from the same GenBank entries. One of these matches was the com1 gene (GenBank accession number AF135266.1) reported by Ree et al. (1).

We originally thought that the matches might have some biological significance, but then we realized that the match was, in fact, a commonly used Clontech adaptor. An adaptor is a short oligonucleotide that is ligated to the ends of cDNAs so that they may be incorporated into a vector cloning site. Usually, adaptors consist of several restriction sites, one blunt end (for ligation to cDNA), and one cohesive end (for ligation to a vector). Adaptors are frequently used in the construction of cDNA libraries and in generating cDNA ends using rapid amplification of cDNA ends PCR.

The com1 gene reported by Ree et al. (1) was independently discovered by another research group, who called it human p8 (4). Our recognition of the adaptor contamination seems to resolve a discrepancy between these two reported sequences. Both Vasseur et al. (4) and Ree et al. (1) found the same amino acid sequence. However, similarity between the 5'-untranslated regions of the two reported cDNA sequences (GenBank accession number AF069073.1 for p8 and GenBank accession number AF135266.1 for com1) does not begin until nucleotide 48. Ree et al. (1) reported that “the full-length com1 cDNA was subsequently determined by 5'-rapid amplification (Marathon: Clontech, Hampshire, United Kingdom) of cDNA generated from nonlactating mammary gland mRNA.” The Marathon 5' rapid amplification of cDNA ends protocol, which can be found online,2 uses the 44-bp Marathon cDNA adaptor. The 5'-untranslated region reported by Ree et al. (1) has a 43/44 similarity to the Clontech Marathon adaptor beginning at nucleotide 4 of com1. We considered the possibility that this similarity was just coincidence, and therefore we searched for this sequence in the Human Genome Project and other eukaryotic genome projects (using the genomic blast tools found on the NCBI BLAST homepage), but we found no matches. Thus, it appears that GenBank accession number AF069073.1 for p8 (4) may be more accurate.

p8/com1 expression is activated in response to cellular injury (5) and during tumor development (1, 6), and it may be an indicator of breast cancer cells with metastatic potential (1, 7–9). Fortunately, to our knowledge, no biological significance has yet been associated with the contaminated region. However, as Bratland et al. (8) mention, proving the transcriptional activating mechanism “demands comprehensive experiments on the still uncharacterized 5'-flanking region of the human com1 gene.” This will be difficult if the incorrect sequence of the 5' end of the cDNA is used to find the 5' end of the gene. Correct resolution of the 5' end will be important for resolving transcriptional activation sites, promoter regulatory mechanisms, and gene organization of p8/com1.

After looking further into adaptor sequence contamination (by BLAST and literature searches involving many commercial adaptors), we have found hundreds of other contaminated sequences that appear in peer-reviewed articles.3 We emphasize that the problem is not unique to com1 and involves many different organisms. Nevertheless, we think that researchers studying humans and particularly those studying human cancer should be especially aware of this type of mistake because cancer researchers rely more heavily (out of necessity) on tissue and/or DNA sources gathered by other laboratories. For example, it is ethically and functionally much more difficult to obtain tissue from a cancerous human brain to generate a cDNA library than it would be to do so from a plant, mouse, or fly. Our findings suggest that many sequence editing mistakes may be due to lack of awareness of cDNA synthesis protocols used before sample acquisition. It is important that sequence editing mistakes are properly recognized and corrected so that they do not hinder future research discoveries. Contamination screening tools such as VecScreen on the NCBI website can help in this regard, although such tools are not normally optimized for finding adaptors.

We write this account with humility because we also discovered that a publication from our laboratory contained a similar error (2). We thank the staff at the University of Nebraska-Lincoln DNA & Genetics Core Facility and Sophia Clotho for answering our questions and Juan Iovanna for his correspondence regarding p8.

Jeffrey Scott Coker
Eric Davies
Botany Department
Box 7612
North Carolina State University
Raleigh, North Carolina 27695

References


Received 2/13/02; accepted 5/13/02.

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


Reply

Coker and Davies (1) have made the important finding that the first 47 bp of com1 cDNA (2) are from a commonly used adaptor linker and are not a part of the 5’-flanking region of this cDNA. com1 cDNA has also been reported as p8 cDNA (GenBank accession number AF069073) by Vasseur et al. (3), with a 5’-untranslated region that does not contain the adaptor sequence. The p8 sequence should therefore be considered more accurate.

We appreciate the thorough analysis by Coker and Davies (1) on our original submission (GenBank accession number AF135266). In accordance with the GenBank policy, we have now removed the linker sequence from the revised entry.

To gain new insight into the regulatory mechanisms of metastasis, one should identify molecular properties that may characterize bone marrow micrometastatic cells and analyze the fluctuation of their gene expression profile, which may correlate with their ability for distant growth in specific organs (4). Our identification of com1 cDNA resulted from an experimental approach to this aspect of metastasis biology (2).

Characterization of biological properties specific for metastatic lesions is technically challenging due to the low number of target cells usually available for this purpose. We cloned a 3’-nontranslated com1 PCR fragment from metastatic tumors that developed within the spinal cord of athymic nude rats after arterial injection (into the left cardiac ventricle) of bone marrow micrometastatic carcinoma cells obtained from a breast cancer patient clinically devoid of metastatic disease (i.e., stage II disease; Ref. 2). The ideal source for achieving the “correct” full-length cDNA sequence would have been, indeed, a brain metastasis from a breast cancer patient with metastatic disease, which is the clinical condition that corresponds pathophysiologically to the experimental model used in our study. We concluded, however, that it would be nearly impossible to acquire a tissue specimen representing the clinical situation because the majority of such patients are given therapeutic irradiation without diagnostic biopsy or other preceding surgical procedures.

The com1 PCR fragment matched expressed sequence tags found in several databases at the time of analysis. Because these expressed sequence tags revealed some slight differences in their sequences, which are common features of such fragments, the missing 5’ end of com1 cDNA was amplified from a commercially available cDNA library from nonlactating mammary gland mRNA (2). This was considered to represent a “common lineage” phenotype of breast gland-derived neoplasms. By instead generating a cDNA library from the experimental metastatic lesions procured from the animal model, cloning of the contaminated sequence might have been avoided, depending on the technical procedure used. However, this approach was not chosen because of both qualitative (host cell contamination) and quantitative limitations regarding the availability of target cells.

As a clinical oncologist, I am concerned about scientific precision toward translational medicine, in particular, and I am therefore grateful to Coker and Davies for their contribution. I also thank Eivind Hovig for his help in updating the GenBank entry.

Anne Hansen Ree
Department of Tumor Biology and Department of Oncology
The Norwegian Radium Hospital
0310 Oslo, Norway

References

Jeffrey Scott Coker and Eric Davies


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/14/4164

Cited articles
This article cites 9 articles, 5 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/14/4164.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/62/14/4164.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.