On the elucidation of a tumour suppressor role of 3p in lung cancer

Elst, Arja ter

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2006

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 4

Analysis of a new homozygous deletion in the tumour suppressor region at 3p12.3 reveals two novel intronic non-coding RNA genes

Debora Angeloni*
Arja ter Elst**
Ming Hui Wei*
Anneke Y. van der Veen**
Tineke Timmer**
Michael I. Lerman*
Charles H.C.M. Buys**

*Laboratory of Immunobiology, CCR, NCI-F, Frederick, Maryland, USA
**Department of Medical Genetics, University Medical Center Groningen, Groningen, The Netherlands
ABSTRACT

Homozygous deletions or loss of heterozygosity (LOH) at human chromosome 3p12 band are consistent features of lung and other common malignancies suggesting the presence of a tumour suppressor gene(s) (TSG) at this location. Only one gene, DUTT1 (Deleted in U Twenty Twenty) was so far cloned from the overlapping region deleted in several lung and breast cancer cell lines (U2020, NCI H2198, HCC38). DUTT1 is the human ortholog of the fly gene ROBO that has homology with NCAM proteins. Extensive analyses of DUTT1 in lung cancer did not reveal any mutations, suggesting that another gene(s) at this location could be associated with lung cancer initiation and/or progression. We report here the discovery in the SCLC cell line GLC20 of a new small, homozygous deletion nested in the known, overlapping, critical region. The deletion was PCR-characterised using several polymorphic markers and covered by three overlapping P1 phage clones. Fiber-FISH experiments using those clones defined precisely the genomic location and the size of the deletion (approx. 130 kb). Comparative genomic sequence analysis revealed short sequence elements highly conserved among mammalian genomes and in the chicken genome. The discovery of two EST clusters in the deletion led to the isolation of two non-coding RNA (ncRNA) genes. We further consider the possibility that these ncRNA and other highly conserved sequence elements may represent miRNA targets whose potential as novel 3p12 lung cancer TSG will be evaluated through subsequent mutation and functional studies.
INTRODUCTION

Loss of function of tumour suppressor genes (TSG) is a fundamental genetic change involved in the origin and pathogenesis of human tumours (Knudson, 1971; Marshall, 1991). TSG have a recessive mode of action, therefore both copies need to be inactivated to produce a phenotypic effect. That frequently happens when one allele is hit by a mutation or silenced through promoter region hypermethylation and the other is lost due to a large chromosomal deletion. In some instances both alleles might be inactivated by genetic loss as a consequence of a homozygous deletion. Sub-lethal homozygous deletions are frequently smaller than heterozygous counterparts, a characteristic that makes them a useful tool for localising tumour suppressor genes.

Cytogenetic and molecular deletion mapping studies have long implicated chromosome bands 3p25-26, 3p21.3 and 3p12-14 as harbouring tumour suppressor genes involved in multiple forms of human cancers including lung cancers (Whang-Peng et al., 1982; Kok et al., 1987; Kok et al., 1997; Zbar, 1989; Zabarovsky et al., 2002).

In 3p14.2 the FHIT gene is subject to homozygous deletions and alterations of the mRNA in many sporadic cancers (Huebner et al., 1998). Frequent allele loss at the FHIT locus has been found in low-grade breast cancer. 3p12 is a particularly significant region (Lerman and Minna, 2000), as demonstrated also by functional studies of Lott et al. (1998) and Lovell et al. (1999). Rabbitts et al. (1990) reported a homozygous deletion at the D3S3 locus in the U2020 cell line. It spans about 8 Mb and is flanked by the microsatellite markers D3S1284 and D3S1276 (Latif et al., 1992; Drabkin et al., 1992). Some other nested or overlapping homozygous deletions were reported in this region. Todd et al. (1997) reported a homozygous deletion that overlaps the U2020 region and is flanked by microsatellite markers D3S1254 and D3S1776. Another overlapping homozygous deletion was found in the breast cancer cell line HCC38. It spans about 5 Mb and is flanked by the microsatellite markers D3S2537 and D3S2527 (Sundaresan et al., 1998a). In the SCLC cell line NCI H219x a much smaller deletion was found that contains the microsatellite markers D3S1274, D3S2498, D3S4492 (telomere to centromere, Sundaresan et al., 1998a).

From this low gene-density region of chromosome 3, the DUTT1 (Deleted in U twenty twenty) gene was isolated (Sundaresan et al., 1998b). DUTT1 gene
expression is impaired by hypermethylation of the promoter in kidney and breast primary cancers but to a much lesser extent in lung cancers.

In this paper we describe a new, homozygous deletion at 3p12 discovered in the small cell lung cancer (SCLC) cell line GLC20. It affects exon 2 and flanking introns of the ROBO1/DUTT1 gene. It spans approximately 130 kb around the D3S1274 microsatellite marker and partly overlaps with the NCI-H219x deletion. In the second intron of DUTT1 we discovered two novel transcripts with the same orientation as DUTT1. Both are polyadenylated, show small ORFs without any known homologues or orthologues and likely do not encode proteins. Both are putatively noncoding (nc)RNA genes that will be evaluated for their role in lung cancer suppression. In the deleted region were also found some sequence elements highly conserved between the human and chicken genome (Bejerano et al, 2004; Hillier et al.2004).
MATERIALS AND METHODS

Human DNA
Signed, informed consent was obtained from all DNA donors in the study, according to the NCI institutional review board-approved protocol.

Cell Lines and DNA
Lung cancer cell lines NCI-H750, NCI-H2198 and NCI-H1450 were obtained from ATCC (Manassas VA). U2020 DNA was kindly provided by Dr. Pamela Rabbits (MRC, Cambridge, UK). GLC20 is a SCLC cell line established from a primary tumour biopsy (De Leij et al., 1985) and known to harbour a circa 440 kb homozygous deletion at 3p21.3 (Kok et al., 1994).

EST clones
EST clones were purchased from the I.M.A.G.E. Consortium (http://image.llnl.gov/).

PCR
PCR primers were from BioServe Biotechnologies (Laurel, MD). PCR reactions were performed in a total reaction volume of 12.5 µl, containing 100 ng of genomic DNA, 12.5 pmol of each primer, 200 µM dNTPs, 1.5 mM MgCl2. The PCR cycles were as follow (with the appropriate annealing temperature, Ta, indicated in Table 1): 95°C, 5 min; (95°C, 30 sec, Ta°C, 30 sec, 72°C, 30 sec) for 35 cycles; 72°C, 7 min. PCR products were run on 3% or 4% NuSieve 3:1 agarose gel (Cambrex, Baltimore, MD) and stained with Ethidium Bromide (SIGMA, St. Louis, MO).

Southern Blot
Southern blot was done according to Sambrook et al. (1998). Briefly, genomic DNA samples were digested overnight with EcoRI in the presence of 1% spermidine, precipitated with Na Acetate, re-suspended in TE buffer, load on 1% agarose gel and run overnight at 30V. The gel was then denatured and DNA transferred overnight in NaOH 0.4 N. The membrane was washed in 2X SSC and dried for three hrs. at 80°C in a vacuum oven. Hybridisation was carried out at 60°C overnight. Washes were done in 2X SSC, 1% - 0.1% SDS at 65°C.
Table 1. List of PCR primers (with annealing temperature) designed on ESTs around microsatellite markers, microsatellite markers, P1-clone ends and cDNA sequences of ROBO1.

<table>
<thead>
<tr>
<th>EST</th>
<th>Forward primer 5'-3'</th>
<th>Reverse primer 5'-3'</th>
<th>Ta (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D81026</td>
<td>GGAATATAAAGCTGGAGACTGTGGA</td>
<td>ACGACCTTACCCCATATGTAAT</td>
<td>62</td>
</tr>
<tr>
<td>H77734</td>
<td>ATCCAGTGGAGAGCTCCTTCCTTCT</td>
<td>CTTCAGGCTTATGAGCAAGGCC</td>
<td>64</td>
</tr>
<tr>
<td>T02864</td>
<td>TCTTTATAGCAGTGTGGGAATG</td>
<td>CAAAATATAAACCAGATCCACC</td>
<td>62</td>
</tr>
<tr>
<td>T02956</td>
<td>CAGGATGCTCAACCAGGAGGT</td>
<td>GTTCTTCTCAGGCTTATGTACG</td>
<td>62</td>
</tr>
<tr>
<td>T64897</td>
<td>TTACACAGGTCGCTGGACATAGAG</td>
<td>ATGTTCAAGGGTGACAGTGAATGAC</td>
<td>62</td>
</tr>
<tr>
<td>W91914</td>
<td>GCTTCTCAGTCATTTCCCATCCGT</td>
<td>GTTCCCTGTCCTCCTATATC</td>
<td>62</td>
</tr>
<tr>
<td>Z1019</td>
<td>GTGCAACCCCTTTATTCAGAATCCC</td>
<td>TAGTCTACCCGCTTGTCCTGCTC</td>
<td>64</td>
</tr>
<tr>
<td>H51703</td>
<td>CTGTAATACAGGGAATCTTCTC</td>
<td>CTTCCTTTGGTGCTGTTCAG</td>
<td>60</td>
</tr>
<tr>
<td>AW861295</td>
<td>GGGATAGCAGCTACACAGCAG</td>
<td>TCTACATCAGGCTCCCTGTCCT</td>
<td>62</td>
</tr>
<tr>
<td>Microsatellite</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3S1577</td>
<td>TCAAATGTGCGATCCG</td>
<td>TCCATTACATCCCCCTG</td>
<td>60</td>
</tr>
<tr>
<td>D3S3681</td>
<td>GTGAGACACCTTGGGCAAGGG</td>
<td>GGCAGCTATGCTGACGGG</td>
<td>60</td>
</tr>
<tr>
<td>D3S1604</td>
<td>CACCATTTGAAGAGCCTCA</td>
<td>AAATGGAGCCATAAATTGGA</td>
<td>62</td>
</tr>
<tr>
<td>D3S3</td>
<td>CAGAAGGACATATCCCATCCGG</td>
<td>GCAGTTTCTCAGCTTTCCT</td>
<td>55</td>
</tr>
<tr>
<td>D3S3049</td>
<td>AAAGACAAGGAGGCGTTCAG</td>
<td>TGGACATACCCGATCAAC</td>
<td>55</td>
</tr>
<tr>
<td>D3S1274</td>
<td>TTATACATCAGTCTCTGGAAACAC</td>
<td>TACTGTGCATATGGCTGGTA</td>
<td>62</td>
</tr>
<tr>
<td>D3S3507</td>
<td>TTCGCAGCCTTTATACCTACTTCCT</td>
<td>TGAATACAAGAAGACTGCAAC</td>
<td>60</td>
</tr>
<tr>
<td>D3S2563</td>
<td>ATATTTTACCGTCGTCGGTCG</td>
<td>TCGAGACCTGCTGTATGGGCTCTG</td>
<td>60</td>
</tr>
<tr>
<td>D3S2530</td>
<td>ACAGGGCAATGTTGCGAGCAGT</td>
<td>GATTTATCCATGGCCTGCTC</td>
<td>62</td>
</tr>
<tr>
<td>P1 Clone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1-97</td>
<td>CCTGATGTTGATATTCTGATGTTCT</td>
<td>GAGCTACAGTCCAGATGATIGG</td>
<td>61</td>
</tr>
<tr>
<td>P1-98</td>
<td>ATAGAATCTGTGATGCTGTCGAC</td>
<td>ATAGCATATATGGACAG</td>
<td>55</td>
</tr>
<tr>
<td>P1-80</td>
<td>GTGATGACATATTTGACGAG</td>
<td>TTTTAAAGGAACCACATGCCC</td>
<td>55</td>
</tr>
<tr>
<td>P1-97</td>
<td>TGCGATGCTGAAATTTCTGCTTCCAG</td>
<td>CAGGTGAGGAAATGGGGAAGATGAGA</td>
<td>64</td>
</tr>
<tr>
<td>P1-98</td>
<td>GTTCAGGTGTAGTTGGCTTCCGCTC</td>
<td>GGAGAGCATTTTTCCTCAGG</td>
<td>50</td>
</tr>
<tr>
<td>P1-80</td>
<td>GAGGTGGTGGATGTGAAAGTGC</td>
<td>ATGAGCAACCCAGATGAC</td>
<td>60</td>
</tr>
<tr>
<td>ROBO1 cDNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71-259</td>
<td>ATCCTCTCAGCCTCTGCCTCTGCT</td>
<td>ACACTCGACATGCTTTCTGCG</td>
<td>62</td>
</tr>
<tr>
<td>601-812</td>
<td>CACATGTTGAGGGGCCACG</td>
<td>TCAAGGGGCGIATCTGGTCTG</td>
<td>64</td>
</tr>
<tr>
<td>1021-1345</td>
<td>TCCGGCTTCTTCAGGAGAAG</td>
<td>TGGCTACCTTCAGGCAGATGC</td>
<td>64</td>
</tr>
<tr>
<td>1351-1490</td>
<td>TCCGGATGACCTACGAC</td>
<td>TATCCATCGAGGAGGACAGC</td>
<td>62</td>
</tr>
<tr>
<td>1510-1620</td>
<td>CGAGGAGGAAAGCCCTATGAG</td>
<td>AGTCACTGCGGCTACTTCAC</td>
<td>62</td>
</tr>
<tr>
<td>1623-1776</td>
<td>GCTCTTACGATCTCTCCTA</td>
<td>CGTGGATAGCCGCTGTTGAC</td>
<td>60</td>
</tr>
<tr>
<td>2341-2811</td>
<td>TAGTGGCCCATACAAAACCTG</td>
<td>GAGTCTGATATTTGCTGTTG</td>
<td>62</td>
</tr>
<tr>
<td>5291-5376</td>
<td>ATCTTCACCACACTTCCGTTG</td>
<td>TAGAAGGGAGATTTGGCACCC</td>
<td>60</td>
</tr>
<tr>
<td>6371-6609</td>
<td>CAAAACATGAAATGGGTAG</td>
<td>GCTCAATTTAAACATCCACTTG</td>
<td>62</td>
</tr>
</tbody>
</table>
Northern Blot
Northern blot was performed with Clontech Human Multiple Tissue Northern Blots 7760-1 (Clontech, Palo Alto, CA). Hybridisation was carried out at 42°C overnight. Washes were done first in 2X SSC, 1% SDS and then 2X SSC, 0.1% SDS at 65°C.

Probe labelling
The W91914 and H51703 cDNAs were excised with EcoRI/Pacl from the modified polylinker of the pT7T3D vector (Pharmacia, North Peapack, NJ) and radioactive probes were prepared by 32P-labeling with random primers (Rediprime DNA Labelling System, Amersham, Arlington Heights, IL).

P1 library screening
The P1 Human Library (Genome Systems, St. Louis, MO) was screened initially with PCR primers designed on the EST W91914 and three clones were isolated: P1-97, P1-98, P1-99. A second round of PCR, with primers designed on the Sp6-end of P1-98, produced the fourth clone P1-80.

Fiber Fluorescence In Situ Hybridisation (fiber-FISH)
Preparations for fiber FISH analysis were obtained essentially according to Giles et al. (1997). GLC20 cells in culture were spun down by centrifugation at 1200 rpm for 10 min and resuspended in distilled water to a concentration of a few, (1 to 5) 10exp5, cells per ml. A volume of 100 µl was pipetted onto a microscope slide coated with a 5% 3-aminopropylethoxy-silane in acetone. Coating was achieved by incubating the slide in that solution for 30 min, after which they were washed with distilled water, allowed to air dry and stored at 4°C until use. The cell suspension was spread over the coated slide using the edge of a coverslip and dried with a hairdryer. Exposure of chromatin threads from the nuclei was obtained by applying two drops of 50 µl from a 0.5% SDS, 50mM EDTA, 0.2M TrisHCl, pH 7.0 lysis solution on 24mm x 60mm cover slips. The microscope slides were then put upside down (with the cells under) on top of the coverslips, slides were then turned and kept like that for 30 sec (with the coverslip up). Coverslips were then gently slid off and preparations were dried with a hairdryer.

Bicolor FISH analysis was performed on these preparations using differentially labelled P1 phages and routine FISH procedures, essentially as described by Driesen et al. (1991).
Novel non-coding RNAs

Sequencing
sequencing reactions were done automatically (ABI 373 Stretch Automated DNA Sequencer, Applied Biosystems, Foster City CA).

Computational analysis
WWW-based Servers and Databases were used to analyse genomic, cDNA and protein sequences. Global and pairwise sequence alignments by BLAST at http://www.ncbi.nlm.nih.gov/BLAST/ and BLAT: http://genome.ucsc.edu/cgi-bin/hgBlat
Multiple sequence alignment with ClustalW at:
http://www.ebi.ac.uk/clustalw/index.html Search of CpG island with CpG Plot at:
http://www.ebi.ac.uk/emboss/cpgplot/index.html?
Analysis of non-redundant sets of gene-oriented clusters: at
Genome browsers and annotations by The Human Genome Browser Gateway:
http://genome.ucsc.edu/cgi-bin/hgGateway, Acembly: at:
http://www.ncbi.nih.gov/IEB/Research/Acembly/index.html and Ensembl at:
Prediction of complete gene structures in genomic sequences (exons, introns, promoters
and poly-adenylation signals) by:
GenScan at: http://genome.dkfz-heidelberg.de/cgi-bin/GENSCAN/genscan.welcome.pl
Analysis of protein features: Psort at http://psort.nibb.ac.jp and Pfam at
Search of possible miRNA target sites by: http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi
Analysis of known miRNA sequence at: The miRNA Registry,
http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml
RESULTS

Discovery of a new homozygous deletion in 3p12.3
We performed PCR experiments on genomic DNA from a panel of lung cancer cell lines with primers designed on seven ESTs (namely T02956, H77734, D81026, T02864, T64897, W91914 and Z41019, data not shown, summary in Table 2A) that map between the genetic markers D3S1274 (alias AFM154xa7, Z16684) and D3S1604 (alias AFM316vc1, Z24325). Among the cell lines analysed, only U2020 and NCI H2198 were known to harbour a homozygous deletion at 3p12. All markers, except T02956 and D81026, were found deleted in the U2020 cell line (data not shown, summary in Table 2A). W91914 was found deleted also in the small-cell lung cancer cell lines H2198 and GLC20 (Fig. 1A), the latter not known to harbour deletions in this region of 3p. This serendipitous result was confirmed by Southern blotting experiments (Fig.1B). The deletion was named 'GLC20-3p12 deletion.

Characterization of the deletion breakpoints
The genomic location and extension of the GLC20-3p12 deletion were defined through PCR experiments with primers designed on several microsatellite genetic markers (Table 2B, Fig. 2). The deletion was found to be located around the D3S1274 marker, both in NCI-H2198 and GLC20 cell lines (Fig. 2).
To address the issues of how large the deletion is, we looked for genomic clones that would represent it entirely and performed a PCR screening of a Human P1 Library (Genome Systems, St. Louis, MO).

With PCR primers designed on the W91914 EST three clones were isolated, identified as P1-97, P1-98 and P1-99. A fourth clone, identified as P1-80, was necessary to cover the deletion in the GLC20 cell line. The fourth P1 clone was isolated after a second PCR screening of the same P1 library, using primers designed on the Sp6-end of P1-98 (screening data not shown). Both ends of each P1 clone were sequenced. PCR reactions performed with primers designed on both ends of each P1 clone (Fig. 3, Table 2C) allowed the construction of a contig as follows (schematically represented in Fig. 5): P1- 97 contains the proximal boundary of the deletion and is centromeric to P1-98 that overlaps only for a small region with P1-80. (P1-99 overlaps extensively with P1-97). P1- 80 contains the telomeric boundary of the deletion.
Table 2. Results of PCR experiments of ESTs, markers, P1-clones and ROBO1 sequences.

A) ESTs designed around the microsatellite markers D3S1274 and D3S1604. EST W91914 was found homozygously deleted in the NCI-H2198 small cell lung cancer cell line.


C) P1-clone ends used to build up the contig.

D) CDNA sequences of the ROBO1/DUTT1 gene (nucleotides position as in GenBank Z95705). Exon 2 was found deleted in the small cell lung cancer cell line GLC20.

<table>
<thead>
<tr>
<th></th>
<th>U2020</th>
<th>H2198</th>
<th>H1450</th>
<th>GLC20</th>
<th>P197</th>
<th>P198</th>
<th>P199</th>
<th>P180</th>
</tr>
</thead>
<tbody>
<tr>
<td>A EST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D81026</td>
<td>+</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>H77734</td>
<td>-</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>T02864</td>
<td>-</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>T02956</td>
<td>+</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>T64897</td>
<td>-</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>W91914</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Z1019</td>
<td>-</td>
<td>+</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>H51703</td>
<td>-</td>
<td>-</td>
<td>nd</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>AW861295</td>
<td>-</td>
<td>-</td>
<td>nd</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B Microsatellite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3S1577</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D3S3681</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D3S1604</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D3S3</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D3S3049</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D3S1274</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>D3S3057</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D3S2663</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D3S2530</td>
<td>nd</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C P1 Clone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1-97</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P1-98</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>P1-80</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>P1-97</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P1-98</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>P1-80</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>D ROBO1 cDNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71-259</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>601-812</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1021-1345</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1351-1490</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1510-1620</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1623-1776</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2341-2811</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5291-5376</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6371-6609</td>
<td>-</td>
<td>nd</td>
<td>nd</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Chapter 4

Other ESTs (Fig. 4 and Fig. 5) were positioned in the contig by PCR. EST AW861295 was also shown by PCR to be homozygously deleted in GLC20-3p12. It was mapped on P1-98 (Fig. 4) but not investigated any further because a BLAST search did not retrieve other overlapping EST clones.

To corroborate the PCR data and verify the contig location with respect to the centromeric and telomeric ends of 3p, the P1 clones were used for a fiber FISH experiment on normal DNA and DNA from GLC20 cell line (Fig. 6), using referral clones indicated as P1-26 and P1-27, previously isolated by Latif et al., (1992). Seeding in BLAT the sequences of P1-97 Sp6-end and P1-80 T7-end (deposited under GenBank Accession number DQ100613 and DQ100614 respectively), we were able to retrieve their exact genomic location (respectively: 78.918,087-78.918,453 and 79.117,734-79.118,091; Human Genome Browser, May 2004, hg 17 assembly, http://genome.ucsc.edu/cgi-bin/hgGateway) and measure the extension of the deletion, which must be less than 200.005 bp (i.e. the distance between the two anchoring markers P1-80 T7-end and P1-97 Sp6-end). By means of Fiber-

Figure 1. A novel homozygous deletion at 3p12.3 in the small-cell lung cancer cell lines GLC20 and H2198.

A) Genomic PCR with W91914 primers on normal (CEPH) DNA (lane 1 and 2), H740 (lane 3), H1450 (lane 4), GLC20 (lane 5), U2020 (lane 6), NCI-H2198 (lane 7), 100-bp ladder (lane 8, Invitrogen). 3% NuSieve agarose gel (Cambrex). B) Southern blot performed with 10 μg of genomic DNA digested with EcoRI. Probe: EST W91914. Lane 1: GLC20, lane 2 and 3: normal individuals, lane 4: molecular marker (Invitrogen). The two normal individuals differ for an EcoRI polymorphism.
Fluorescent In Situ Hybridisation experiments it was subsequently found to be approximately 110 to 130 kb (see legend of Fig. 6)
The genomic sequence of the deleted region contained between the two anchoring markers, was deposited under provisional GenBank Accession number bankit729386.

Figure 2. PCR experiments to investigate the position of several microsatellite markers with respect to the GLC20-3p12 homozygous deletion and the four P1 clones that cover the deletion. Microsatellites are ordered (top to bottom) from the more telomeric to the more centromeric. 4% NuSieve agarose gel stained with Ethidium Bromide. M.M.: molecular weight marker (100 bp ladder or 1 Kb ladder, Invitrogen). PCR primers and conditions are listed in Table 1B.
Figure 3. PCR experiments with primers designed on the ends of each P1 clone were performed to build a P1 clone contig and to find out which clone would cover the deletion boundaries. The small-cell lung cancer cell lines GLC20, U2020 and NCI-H2198 were investigated. 4% NuSieve agarose gel stained with Ethidium Bromide. M.M.: molecular weight marker (100 bp ladder or 1 Kb ladder, Invitrogen). PCR primers and conditions are listed in Table 1C.
Figure 4. PCR experiments with primers designed on the ESTs AW861295 and H51703. Both ESTs are deleted in the GLC20-3p12 homozygous deletion and represented in the P1 contig. 4% NuSieve agarose gel stained with Ethidium Bromide. M.M.: molecular weight marker (100 bp ladder or 1 Kb ladder, Invitrogen). PCR primers and conditions are listed in Table 1A.
**Figure 5.** Drawing that summarises all results obtained with PCR and fiber-FISH experiments, showing the genomic location of the GLC20-3p12 homozygous deletion. Coloured rectangles: sequenced ends of P1 clones. Coloured circles: positive results of specific PCR experiments indicating ‘anchoring points’ of sequences to one another, used to build up the contig.

**Figure 6.** Fiber - Fluorescent In Situ Hybridisation with P1 clones. The yellow spots result from merging of green and red labelling and show regions of overlapping between two clones. The referral clone P1-27 is centromeric to P1-97. Therefore the contig is oriented with P1-97 as the most proximal and P1-80 as the most distal (telomeric) clone. A) P1-97 and P1-98 showing an overlap of an estimated size of about 20 kb when hybridised to fibers from an EBV-transformed lymphoblastoid cell line. B) P1-98 and P1-80 hardly showing an overlap when hybridised to fibers from the same cell line. C) size comparison between the part of P1-97 hybridising to fibers from the cell line GLC20 and the cohybridised cosmid cosD8 (Kok et al., 1995) showing that the part of P1-97 extending beyond the deletion is somewhat larger than the 40 kb cosmid insert. D) P1-97 and P1-80 hybridised together with the more centromeric P1-27 to fibers from the lymphoblastoid cell line, showing the orientation of P1-97 and P1-80 with respect to the centromere and the 3p telomere. Since cohybridisation of P1-80 and cosD8 to fibers from GLC20 gave a picture very similar to C (i.e. both P1 phages extend about 45 kb beyond the deletion) the size of the deletion will be about 3 x the P1 insert size of 80 kb (cf. Fig. 3), minus the overlaps of together about 20 kb, minus the extending P1 parts of about 2 x 45 kb. The deletion size will thus be approx. 130 kb.
Status of the *DUTT1* gene in the GLC20-3p12 homozygous deletion

Using a set of PCR primers (Table 2D, Fig.7) designed on the exonic sequence of the *ROBO1/DUTT1* gene, we investigated to what extent the GLC20-3p12 homozygous deletion affects the genomic structure of this gene in the GLC20 cell line. We found that the exon corresponding to cDNA residues 1021-1345 of *DUTT1* (variant 2, GenBank Z95705, NM_133631) is lost (Fig. 7) due to the small deletion that spreads across exon 2 and is comprised entirely between introns 1 and 2 (Fig. 5). These data were confirmed by RT-PCR on *DUTT1* cDNA prepared from GLC20 cells. The deletion causes the loss of amino acids 19-128 (as described for the cell line H219x by Sundaresan et al., 1998b). These data were also confirmed by PCR on YAC clones 912A11, 15HC9 and 35AH8 (data not shown). The coverage of the new deletion in terms of BAC clones was determined *in silico* with electronic PCR using the P1-ends sequence as probes (Fig. 5). When we compared our experimental results with the Human Genome Browser database (May 2004, hg 17 assembly, http://genome.ucsc.edu/), we found a perfect match for the location of all markers. The new homozygous deletion in GLC20-3p12 is nested into but much smaller than the 8 Mb deletion of the U2020 cell line or the 5 Mb of the HCC38 cell line. It partly overlaps with H219x whose exact size was not determined but that is known to be internal to *DUTT1* (Sundaresan et al., 1998a).

The new deletion harbours two previously unknown transcripts – Computational analysis

Several ESTs (some of which adenylated) map in the region hit by the new deletion in 3p12.3. We focused on two clusters and picked one from each, namely W91914 and H51703, for further analysis. Several SAGE entries support these ESTs that originated from different cDNA libraries. Seeding the sequence of W91914 and H51703 in BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) against the human EST database, we were able to assemble two separate clusters and analyse their sequence with GenScan (http://genes.mit.edu/GENSCAN.html) looking for predicted exons.

Cluster of W91914

The cluster corresponding to the initial probe W91914 comprises now eight ESTs (namely: W91914, W94988, H90477, H90421, AI078492, AA639329, AA777646, CR740005) isolated from Fetal Liver and Spleen, Head and Neck carcinoma and
Figure 7. Exon 2 of the DUTT1 gene (cDNA 1021-1345, as in GenBank Z95705) is homozygously deleted in the GLC20-3p12 deletion. PCR experiments with primers (listed in Table 1D) designed on some exons at the 5'-end of DUTT1. The GLC20-3p12 homozygous deletion bridges across exon 2 to introns 1 and 2. 4% NuSieve agarose gel stained with Ethidium Bromide. M.M.: molecular weight marker (100 bp ladder or 1 Kb ladder, Invitrogen).

A Genscan analysis (http://genes.mit.edu/GENSCAN.html) of this mRNA, predicts a putative exon with ORF of 49 amino acids, but the lack of homologues or orthologues in other species suggests that this transcript unlikely codes for a protein. Acembly (http://www.ncbi.nih.gov/IEB/Research/Acembly/index.html) annotated the gene as a single-exon transcript of 3.1 kb. Interestingly, the alignment of BC017743 sequence with the chicken genome produces two hits on Gallus gallus chromosome 1, one of which shares 86% identity over 353 residues. Given the more and more recognised importance of microRNAs (miRNAs) in cancer biology (Xu et al., 2004; McManus, 2003), we looked in the BC017743 sequence for possible miRNA target sites, which would suggest the existence of a posttranscriptional control mechanism acting on this transcript. The DIANA algorithm (http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi) predicted two possible target sites in the sequence (Fig. 8A), both scoring in the high-confidence range. Both miRNAs belong to a group of miRNA genes that were experimentally identified and their expression analysed by Northern blot. They have a similar sequence originating from different loci: has-miR-17-5p derives from a locus on chromosome 13 and has-miR-106a from a locus on chromosome X (Mourelatos et al., 2002; Dostie et al., 2003; Kasashima et al., 2004; Suh et al., 2004).

Cluster of H51703
The cluster assembled around the sequence of H51703 is now represented by seventeen ESTs (H51703, T69773, T70759, T84499, H40323, H40377, R83269, AA668381, AA669442, BI598464, BM993003, BX105987, BE062088, BF746150, BE061843, BF746204, CA440361) isolated from Liver and Spleen, Hypothalamus, Lung carcinoma and Metastatic Lung Chondrosarcoma libraries. Again, it is significant that this cluster aligns (see the Human Genome Browser, May 2004, http://genome.ucsc.edu/cgi-bin/hgGateway, position: chr3: 78,485,247-79,099,496) with a full-length transcript, namely BC043430 (isolated by The Mammalian Gene Collection Program, Imanishi et al., 2004). Genscan analysis (http://genes.mit.edu/GENSCAN.html) does not predict any exon and the possible, small ORFs do not show similarities with known homologues or orthologues.
Acembly (http://www.ncbi.nih.gov/IEB/Research/Acembly/index.html) annotated this transcript as a putative single-exon gene of about 1.9 kb, with an ORF of 46 amino acids, without known similarities. However, two bovine ESTs, AV601957 and CR454939, align in correspondence of BC043430, with 92% and 81% identity respectively (in a region free of repetitive DNA). The DIANA algorithm (http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi) predicted four putative miRNA target sites in the BC043430 sequence (Fig. 8B). Those miRNAs were previously identified through homology with known miRNAs (Houbaviy et al., 2003; Weber, 2005).

![Figure 8](image)

**Figure 8.** Human microRNAs that putatively hit target sequences on (A) BC017743 and (B) BC043430 transcripts (adapted from a window of the DIANA algorithm results (http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi)).

<table>
<thead>
<tr>
<th>Number</th>
<th>miRNA ID</th>
<th>Energy</th>
<th>Hit location</th>
<th>Target Annotation</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hsa-miR-17-5p</td>
<td>-33.6</td>
<td>2038</td>
<td>bc1.</td>
<td>5’ --------ctgtaatctagcactttg 3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3’ tggacgtgacat------tcgtgaaac 5’</td>
</tr>
<tr>
<td>2</td>
<td>hsa-miR-106a</td>
<td>-35.6</td>
<td>2037</td>
<td>bc1.</td>
<td>5’ --------ctgtaatctagcacttt 3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3’ cgatggacgtgacat------tcgtgaa 5’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number</th>
<th>miRNA ID</th>
<th>Energy</th>
<th>Hit location</th>
<th>Target Annotation</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hsa-miR-17-5p</td>
<td>-34.3</td>
<td>438</td>
<td>bc.</td>
<td>5’ --------ctgtaatcccgcactttg 3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3’ tgatggacgtgacat------tcgtgaaac 5’</td>
</tr>
<tr>
<td>2</td>
<td>hsa-miR-106a</td>
<td>-32.2</td>
<td>437</td>
<td>bc.</td>
<td>5’ ------ctgtaatcccgcaacttt 3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3’ gacgtgacat------tcgtgaa 5’</td>
</tr>
<tr>
<td>3</td>
<td>hsa-miR-106b</td>
<td>-33.6</td>
<td>437</td>
<td>bc.</td>
<td>5’ ctgc-ctgtaatcccgcaaccttt 3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3’ gacgtgacaa------tcgtgaa 5’</td>
</tr>
<tr>
<td>4</td>
<td>hsa-miR-34b</td>
<td>-27.8</td>
<td>347</td>
<td>bc.</td>
<td>5’ -------ctatattgtgtgactgcct 3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3’ gttagtcgat--tag-tgtgacgga 5’</td>
</tr>
</tbody>
</table>
Conservation analysis of the deleted sequence
Aligning the sequence of the GLC20-3p12 deletion with the chicken genome (Bejerano et al., 2004; Hillier et al., 2004, Chicken Genome Browser Gateway, February 2004 assembly), we identified seven blocks of highly conserved sequence. Four blocks contain Alu repeats. Two of those show 93% identity between human and chicken and find homology with the 7SL chicken gene on chicken chromosome 5. Two others blocks, almost identical to the previous ones, hit chicken chromosome 2 and perhaps represent another copy of the 7SL gene, located on this other chromosome. Finally, three highly conserved regions hit chicken chromosome 1 (numbers refer to the human sequence - Human Genome Browser, May 2004 hg 17 assembly): Seq5: 69 bp (79000588 - 79000654), 84% identity hum/chick Seq6: 353 bp (78954642 - 78954984), 86% identity hum/chick Seq7: 243 bp (79044143 - 79044377), 81% identity hum/chick. These regions do not contain repetitive elements. Interestingly seq6 aligns inside the non-coding transcript BC017743. Understanding whether seq5 and seq7 are actually transcribed and therefore possibly encode functional RNA, would require further experimentation.

Expression studies
To independently verify whether the two clusters of ESTs are indeed transcribed, we performed Northern blot experiments using W91914 and H51703 cDNA as probes. In our conditions, the W91914 probe lighted two bands of about 2.4 and 3 kb in all tissues tested (Clontech MTN blot, Palo Alto, CA) except kidney and pancreas, (Fig.9A). In the same conditions, the H51703 probe lighted a band of about 2 kb in pancreas, one of about 3 kb in skeletal muscle and one of about 1.35 kb in liver (Fig.9B). RT-PCR experiments performed with primers designed on W91914 and H51703 also confirmed that these sequences are actually expressed. H51703 was amplified from all cDNA samples prepared from various organs (Clontech, Palo Alto CA), (Fig.9C). W91914 was amplified from fetal lung cDNA (Clontech, Palo Alto CA), Fig.9D. By means of RT-PCR experiments using primers designed across splice sites, all cDNA samples were tested to be free of genomic DNA (data not shown).
Figure 9. Expression studies of ESTs W91914 and H51703, chosen as representative of two clusters of ESTs located in the region affected by the GCL20-3p12 homozygous deletion.

DISCUSSION

Detection of homozygous deletions and microsatellite mapping to identify allele loss and deletions are still the most powerful method to localise putative TSG. Interstitial deletions in the low gene-density chromosome region 3p12 were reported in lung and other malignancies (Daly et al., 1991; Ganly et al., 1992; Pandis et al., 1993). Rabbitts et al. (1990) reported a homozygous deletion at the locus D3S3 in the U2020 cell line. Chen et al. (1994) reported a homozygous deletion in this same region in breast cancer. Sundaresan et al. (1998a) found in lung and breast cancer two more overlapping homozygous deletions harbouring in this region.

The U2020 region has been strongly suspected to harbour a tumour suppressor gene since the mid 1990s when karyotype analysis had shown that deletions in 3p12 are the only evident abnormality in cells cultured from normal bronchial epithelial cells of lung cancer patients (Sundaresan et al., 1995). Sanchez et al. (1994) showed that the introduction of two centromeric fragments of 3p (encompassing 3p12-q24 and 3p14-q11) into a highly malignant renal cell carcinoma (RCC) cell line resulted in a dramatic suppression of tumour growth in athymic nude mice, suggesting that a locus in this region controls the growth of RCC cells by inducing rapid cell death in vivo. Lott et al. (1998) and Lovell et al. (1999) showed that a fragment of human chromosome 3 overlapping with U2020 deletion mediates rapid cell death and tumour growth suppression of RCC cells in vivo, whereas a deletion of this region is associated with immortalisation of human uroepithelial cells (Vieten et al., 1998).

Detailed mapping showed that the U2020 deletion is about 8 Mb in size (Drabkin et al., 1992; Latif et al., 1992) and harbours the smaller deletions HCC38 and H219x (Sundaresan et al., 1998a). In the U2020 region, Sundaresan et al. (1998b) identified and cloned a gene that is disrupted by these deletions: DUTT1 (also known as ROBO1, independently isolated by Kidd et al., 1998). DUTT1/ROBO1 is an integral membrane protein. It is an axon guidance/cell adhesion receptor whose best-characterised function deals with regulating the decision by axons to cross the central nervous system midline. DUTT1 is widely expressed as an about 8 Kb mRNA and two transcript variants are known that differ for their 5’ terminus (start site and first exon). The H219x deletion hits exon 2 (encoding the first Ig domain) of DUTT1-short variant. Introduction of this mutation in the mouse germ line (Xian et al., 2001) generates animals that in the homozygous
state frequently die at birth of respiratory failure due to functional immaturity of lungs. Survivors acquire bronchial epithelial abnormalities similar to those involved in early stages of lung cancer and die in the first year of life (Xian et al., 2001). Heterozygous mice grow normally but develop spontaneously lymphomas and carcinomas in their second year of life with a 3-fold increase in incidence compared with controls (Xian et al., 2004). Invasive lung adenocarcinoma is by far the predominant carcinoma. In addition to the mutant allele, loss of heterozygosity analysis indicates that these tumours retain the structurally normal allele but with substantial methylation of the gene's promoter (Xian et al., 2004).

Dallol et al., (2002) found that DUTT1 promoter is hypermethylated in 19% of primary invasive breast carcinomas, 18% of primary clear cell renal cell carcinomas and in 4% of primary NSCLC tumours. In addition, 80% of breast and 75% of clear cell renal cell carcinomas (CC-RCC) showing DUTT1 hypermethylation also showed allelic loss at 3p12 suggesting that DUTT1/ROBO1 is a classic tumour suppressor gene requiring inactivation of both alleles to elicit tumourigenesis. However, an extensive mutation analysis of DUTT1 in lung, breast and kidney cancer did not retrieve inactivating mutations (Dallol et al., 2002).

Here we report the identification in the lung cancer cell line GLC20 of a new homozygous deletion at 3p12.3 that spans about 110 to 130 kb (perhaps the smallest described so far in this region) and hits the second exon of DUTT1. Moreover, using molecular biology and bioinformatics methods, we identified in this region two novel putative genes that reside in the second intron of DUTT1 and therefore are also hit by the deletion. Both transcripts do not show obvious splicing signals however in our hands each one decorated slightly different bands in Northern blots. Both show a polyA tail and very small ORFs that comparative genome analyses suggest are unlikely to encode proteins. On the bases of these characteristics they are possibly mRNA-like non-coding RNAs (Tupy et al., 2005; Erdmann et al., 2000). Transcribed by RNA polymerase II, in absence of protein products, these types of RNAs serve as riboregulators or regulators of expression of related genes (Tupy et al., 2005; Numata et al., 2003; Erdmann et al., 2002). Based on their sequence, BC017743 and BC043430 seemingly lack homologues or orthologues but this should not be a discrediting criterion as it was reported for other known genes and ncRNAs (Conrad et al., 2002; Weber et al., 2005). Also, some biologically important ncRNA families show that, inside each group, conservation of secondary structure has a higher significance compared to conservation of primary
sequence (Weinberg and Ruzzo, 2004).

Up to now, very few cases of independent transcription units embedded inside gene introns were described. The genes EV12A and EV12B are encoded by one intron of the human neurofibromatosis type 1 (NF1) gene and are transcribed in the opposite direction. However their products are functionally unrelated to NF1. The Ach transporter gene (transcribed from the first intron of the rat ChAT gene, Bejanin et al., 1994) and the Saitohin gene (transcribed from intron 9 of the human tau gene, Conrad et al., 2002) are different as both their respective products are transcribed in the same orientation and also are functionally related to the longer gene inside which they are harboured. The casistics is bound to increase as it was recently reported (Reis et al., 2005) that 233,303 clusters of ESTs are totally contained within intronic regions. Both BC017743 and BC043430 are transcribed in the same direction of DUTT1 suggesting that their product could be co-regulated and possibly related to DUTT1 function. The fact that both are transcribed in the same orientation as DUTT1 suggests also that the hypermethylation that affects DUTT1 promoter in some cancers (Dallol et al., 2002) might concomitantly deregulate their own transcription. It will be interesting to verify whether both transcripts are actually under translational control by the miRNAs whose putative target sites were predicted in their sequence. Since the loss of BC017743 and BC043430 was found associated with a tumour phenotype, one might argue that both transcripts exert a tumour suppressor action, possibly mediated by the miRNA action. In fact, several cases are now known (Calin et al., 2005; Xu et al., 2004; McManus, 2003) of changes in the expression level of miRNAs that may affect the control of cell growth or survival and therefore are involved with cancer onset or progression. Noteworthy, hsa-mir-17-5p originates from 13q31.3 (http://www.ensembl.org/Homo_sapiens/contigview?highlight=&chr=13&vc_start=90700860&vc_end=90900943), a region whose LOH has been involved with breast cancer progression (Eiriksdottir et al., 1998). Therefore one might speculate that LOH at the miRNA locus (e.g. hsa-mir-17-5p) might obtain the same effect of LOH at the miRNA target site (conceivably BC017743 and BC043430 in the case of lung, breast or kidney cancer). Moreover, hsa-mir-17-5p was shown to be down-regulated following differentiation induced by TPA treatment of HL-60 promyelocytic leukemia cells (Kasashima et al., 2004). This suggests that miRNA-induced control of cell differentiation might consist of two distinct mechanisms: I) a mechanism of gene silencing through up-regulation of miRNAs, II) a mechanism of gene activation through termination of miRNA-regulated gene silencing.
Similarly, it would be interesting to investigate whether LOH at 11q23 in lung cancer involves not only loss of the TSLC1 gene that resides there (Kuramochi et al., 2001) but also loss of the locus encoding *hsa-mir-34b*, the one that putatively hits BC043430 (as well as other targets in the genome probably: it was shown, for example, that reduced levels of *hsa-mir-15a* and *hsa-mir-16* are a trait shared by different forms of lymphomas and leukemias; Calin et al., 2002; Eis et al., 2005). A situation possibly symmetrical to what described in the GLC20-3p12 deletion, is given by the accumulation of the ncRNA BIC (Eis et al., 2005). BIC transcript is polyadenylated, has short putative ORFs that are not conserved. Most likely BIC does not encode a protein. However, a phylogenetically conserved region of BIC was indeed shown to encode a miRNA, mir-155, whose accumulation is strongly correlated with an aggressive B cell neoplasm (Eis et al., 2005). The GLC20-3p12 deletion might also harbour miRNA loci. Recently, a list of computationally identified human miRNA genes was reported, some of whom are encoded at 3p12 (Berezikov et al., 2005). One of them (cand893 HS3, 78.768.573-78.768.661 R, whose closest experimentally identified miRNA is *M. musculus mmu-mir 297*, Houbaviy et al., 2003) is harboured inside the deletion. Further studies are necessary to investigate whether DUTT1 itself is a target for this miRNA, as it was suggested that intronic miRNAs might exert a role in regulating fast cell transitions in response to external stimuli, perhaps modifying the expression of genes that harbour them, bypassing protein synthesis (Ying and Lin, 2004).

Thanks to the recently published draft of the chicken genome, we were able to analyse the degree of conservation of the non-coding sequence around the exon 2 of *DUTT1*. This region contains at least three blocks of extremely conserved, non-repetitive sequence, spanning respectively 353, 69 and 243 bp, with a degree of identity from 81 to 86% between the human and chicken genome, whose functional meaning was not investigated in this work.

In conclusion, it is interesting to note that the two smallest, partly overlapping homozygous deletions described at 3p12.3 (that is H219x, Sundaresan et al., 1998a, and the one described here) remove, besides one exon of *DUTT1*, two ncRNA genes. BC043430 and BC017743 are the first ncRNA transcripts/genes found in a homozygous cancer deletion affecting the 3p12.3 region. The presence of two highly conserved intronic ncRNAs of the *ROBO1/DUTT1* gene, along with several short non-coding regions highly conserved between human and chicken genomes
Novel non-coding RNAs (Bejerano et al., 2004; Hillier et al., 2004), suggests that the effect of the 3p12 deletion might be complex. Moreover, it must be mentioned that another element of complexity in determining the full tumour phenotype is introduced by the presence in GLC20 cells of another homozygous deletion, at 3p21.3 (De Leij et al., 1985; Kok et al., 1997). In fact, with respect to the short arm of chromosome 3, it was suggested for several types of tumour that multiple deletions, coexisting but with different 3p locations, most likely have a synergistic effect in driving tumourigenesis (Van den Berg et al., 1997; Senchenko et al., 2003).

In further studies the two ncRNA genes, BC043430 and BC017743, need to be evaluated as novel 3p12 lung, breast, and kidney TSG through mutation and functional studies. Similarly, the highly conserved, short sequence elements will be evaluated as potential ncRNA loci.

ACKNOWLEDGEMENTS

DAA, MHW and MIL were funded in toto with funds from the National Cancer Institute, National Institutes of Health. The content of the publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organisations imply endorsement by the U.S. Government. The study was also supported by grants to CHCMB from the Dutch Cancer Society (Grants RUG 94-834 and RUG 2000-2317) and from the J.K. de Cock Stichting (Grant 99-07). We are grateful to Andrei Kouranov (Department of Genetics, School of Medicine, University of Pennsylvania, Philadelphia, USA) for his kind help with interpreting the DIANA algorithm results. Special thanks to each and every LIB members.
REFERENCES


Novel non-coding RNAs


Novel non-coding RNAs


the 3p12 gene, DUTT1/Robo1, predisposes mice to lung adenocarcinomas and lymphomas with methylation of the gene promoter. Cancer Res. 64:6432-7.


