Supplemental material

Bioinformatic analysis of break regions

We performed a bioinformatic analysis of all breaks reported in Supplemental Figure 4. Many of them (39 out of 64; 61%) mapped within single copy sequences. The remaining breaks were located within repeated elements. Six within SINEs, including Alu (three breaks) and mammalian interspersed repeat elements (MIR; three breaks); ten within LINEs, five within LTR elements, two within simple repeats, $(TTTTA)_n$ and $(TG)_n$, one in a Low Complexity sequence (CT-rich), and one on Mer91A. Microhomologies of one up to 5 nt, as well as anonymous insertions ranging from 2 to 53 nt, typical of Non-Homologous End-Joining (NHEJ) were found. Nine junctions were of blunt type (see Supplemental Figure 4).

Fusion genes generated by junctions

In GLC8 and STA-NB-3, we observed the juxtaposition of *NAG* (exons 37-41, and exons 41-44, respectively) to *BC035112* [last two exons (3-4), and two exons (2-3), respectively] with the same transcriptional orientation. In STA-NB-4 the 5' portion of *NAG* (13-42 exons) fused to the 3' portion of *FAM49A* (exons 3-12) in the same transcriptional orientation, while in STA-NB-10 the 5' of *FAM49A* (exon 1) was joined to the 3' of *SMC6* (exons 5-13) with the same transcriptional orientation. Additionally, we found an inversion of exon 2-3 and deletion of exon 1 within the *ALK* gene in STA-NB-15, but no resulting aberrant transcript was detected (data not shown).

Gene expression analysis, by RQ-PCR experiments, of genes interrupted by hsr insertions

MLLT10 and *LRRTM4*, were found broken by hsr insertion in SK-N-BE and STA-NB-3, respectively. They did not show an altered expression when compared to their average transcriptional levels in NB cell bearing undisrupted *MLLT10* and *LRRTM4* (Supplemental Figure 9 and Supplemental Table 9). *CDK3* (cyclin-dependent kinase 3), *EVPL* (envoplakin), *SRP68* (signal recognition particle 68kDa), and *GALR2* (galanin receptor 2), mapping within BAC RP11-449J21, internal to the duplicated region in STA-NB-15, displayed a doubled expression level with respect to the mean value of their transcriptional level in the other NB cell lines. Conversely, *ZACN* (zinc activated ligand-gated ion channel) and *EXOC7 (*exocyst complex component 7 isoform 2) were not upregulated (Supplemental Figure 9 and Supplemental Table 9). *P2RY1*, the only gene mapping within the fosmid G248P8527A1, displayed a splitting signal upstream and downstream to the hsr. It was significantly upregulated in GLC14 respect to GLC8, that did not show any hsr. However, we could not establish if the gene, relatively short with respect to the fosmid (~3 kb), was broken by the insertion.

Legends to Supplemental Figures and Tables

Supplemental Figure 1

Oligo-array CGH single-panel unaveraged plots of the log2 ratio values and the predicted DNA segmentations (DNA copy number changes) drawn as lines on the data points (NimbleGen platform; whole genome, 385 K; average oligo distance ~6.3kb). The plots contain data for all of the chromosomes in a single plot for each cell line, with the chromosomes or regions differentiated by color and by vertical dashed lines. The amplification of the 2p24.3 region is evident in all cell lines. NimbleGen array CGH data have been submitted to GEO database (http://www.ncbi.nlm.nih.gov/geo/query) under accession no. GSE22279.

Supplemental Figure 2

Details (chromosome 2, 5-32 Mb) of oligo-array CGH multi-panel unaveraged plots, showing high-copy number gains in the cell lines under investigation. The plots show the log2 ratios as black dots and the predicted segments as red lines.

Supplemental Figure 3

Examples of co-hybridization FISH experiments on studied cell lines. Each Figure reports the line under study and the probes used in the experiment. (a) and (b) show two examples of heterogeneity of dmin composition. This heterogeneity supports the different level of amplification detected by array CGH. Heterogeneity, because of chromosome compaction, could not be assessed in hsr amplifications, as shown in (c) and (d).

Supplemental Figure 4

All the 32 successfully sequenced junctions (involving 64 sequences) are partially reported in this Figure. Each junction is identified by a Roman figure (see Figure 5). Junction sequences are aligned with the corresponding normal sequence. Inserted nucleotides and micro-homologies at the junctions are indicated in underlined and in purple bold italics, respectively. STA-NB-10 junctions refer to the STA-NB-10/dmin subclone. Junctions I, III, and VI were also sequenced in the NB10/hsr subclone, and turned out to be identical (Supplemental Figure 3). SK-N-BE junction I and II showed an intervening 241 bp segment at the junction point. BLAST analysis indicated that the segment mapped within the amplicon (chr2:16,457,956-16,458,197). A PCR experiment performed using a forward and a reverse primer located upstream and downstream this fragment

yielded a single PCR product, corresponding to the expected reference sequence on chromosome 2 (data not shown). This result clearly indicated that the segment's move was due to duplication and not to excision from the original location within the duplicon (Figure 5). Colors of sequences refer to segments' color in Figure 4.

Supplemental Figure 5

Proposed models to explain the genesis of *MYCN*-containing dmin/hsr in the investigated tumor cell lines. In (a), the episome model (excision-circularization-amplification) is shown as perfectly applicable to simple amplicons, composed by a single amplified segment, tandemly arrayed "head to tail", such as in GLC14, STA-NB-4, and SK-N-BE.
In (b), for other cell lines, with amplicons made of multiple, non colinear, often inverted, segments (such as GLC8 or STA-NB-10), the episome might be obtained as a result of multiple internal rearrangements of an ancestral excised amplicon. In (c) and (d), amplicons with coamplified segments mapping on separate regions of the genome, within the same (STA-NB-3, STA-NB-13, STA-NB-15) (c) or different (STA-NB-8) (d) chromosomes, might be originated consequently to an excision event from a chromosome involved in intra- or inter-chromosomal rearrangements, respectively.

Supplemental Figure 6

Examples of M-FISH and FISH experiments, using WCP probes, performed in order to map the insertion sites of hsr regions. *MYCN* was detected by the BAC RP11-480N14. In all cases, except STA-NB-3, the hsr localized within chromosomes other than chromosome 2.

Supplemental Figure 7

Partial chromatograms and schematic diagrams showing the fusion junctions and fusion transcripts, respectively, obtained accordingly to RT-PCR results, generated consequently to the juxtaposition of truncated genes at amplicon breakpoints. Purple arrows indicate the breakpoints. Nucleotide micro-homologies at the fusion junctions are underlined in purple.

Supplemental Figure 8

(a) Organization of the amplicons (solid bars) on chromosomes 2, inserted as a track on UCSC genome browser. The Figure reports also the RefSeq genes of the region. The interval 7-19 Mb is zoomed in b).

Supplemental Figure 9

Relative mRNA expression levels determined by quantitative PCR in NB and SCLC lines and control tissue samples (samples with gene amplification are marked in violet) (relative scale was rescaled to an average expression level of 1)

Supplemental Figure 10

RQ-PCR expression analysis of genes mapping within the domains of hsr insertion sites in cell lines STA-NB-15 (a), SK-N-BE (b), STA-NB-3 (c) and GLC (c). In (a), (b), and (c), the total brain (white column; value = 1) is always used as calibrator. The red column represents the ratio of a specific gene versus the value of the same gene in the other two NB cell lines. In figure (d) the yellow and the green bars display GLC14 and GLC8 expression level, respectively, using lung as calibrator (white bar; value = 1).

Supplemental Table 1

The Table lists all BAC and fosmid clones (fosmids in Italics) used in FISH experiments in order to characterize the amplicons (Supplemental Table 1a) or to precisely map the insertion sites of hsr (Supplemental Table 1b). The results are detailed only for clones used to characterize the amplicons. Because of growth difficulty of the STA-NB-10/hsr cell line, not all FISH experiments performed on STA-NB-10/dmin were performed in the STA-NB-10/hsr.

Supplemental Table 2

The Table lists all primer pairs utilized to define amplicon breakpoints (a). MYCNB primer pair, internal to *MYCN* was always used as internal control. In (b) are reported the primers used to amplify and sequence the junction regions reported in Supplemental Figure 4.

Supplemental Table 3

Detail of STA-NB-8 (3A), STA-NB-10/hsr (3B), and STA-NB-15 (3C) Affymetrix 6.0 SNP array analyses, focused on amplified genomic segments (in yellow). The analysis was performed using the Partek Genomic Suite software. Affymetrix CGH data have been submitted to Array Express database (http://www.ebi.ac.uk/microarray) under accession no. E-MEXP-2735.

Supplemental Table 4

SNP (column E) and Copy Number (column D) results on Affymetrix platform of the region chr2:14,000,000-20,000,000 of the STA-NB-10/dmin, containing the amplified region chr2:15,481,998-18,677,524 (Genome Wide Human SNP Array 6.0). The LOH state is reported in column 5: 1 stands for LOH, 0 for normal. “null” refers to non-polymorphic oligos. In all amplified

regions, heterozygous SNP loci are misclassified as “1” (LOH) because one allele is considerably overrepresented and masks the other single-copy allele. The analysis was performed using the Affymetrix Genotyping Console.

Supplemental Table 5

The Table lists all BAC and fosmid clones used in FISH experiments performed to precisely map the insertion sites of hsr.

Supplemental Table 6

Lists of genes mapping within each amplicon. Many amplicons are defined at sequence level (position in column 3 and 4). Other amplicon boundaries (intervals) were defined by RQ-PCR as reported in Supplemental Table 2.

Supplemental Table 7

Gene expression analyses: the Table lists all primers used in RQ-PCR experiments.

Supplemental Table 8

Expression analysis results obtained by RQ-PCR in NB cell lines (8a) and in SCLC cell lines (8b). Gene expression levels were calculated using, as calibrator, the total brain RNA in NB cell lines and a pool of normal lung RNA in SCLC cell lines. Expression values of amplified genes in each cell line are underscored. Statistically significant *p* values are in bold. (*= interrupted gene by one or more breakpoints; N.D.= not detected)

Supplemental Table 9

RQ-PCR expression analysis of genes mapping within the domains of hsr insertion sites in cell lines STA-NB-15 (a), SK-N-BE (b), STA-NB-3 (c) and GLC (c). In (a), (b), and (c), the total brain (white column; value = 1) is always used as calibrator. The red column represents the ratio of a specific gene versus the value of the same gene in the other two NB cell lines. In figure (d) the yellow and the green bars display GLC14 and GLC8 expression level, respectively, using lung as calibrator (white bar; value = 1).

Supplemental Table 10

Regions surrounding the breakpoint.

Supplemental Table 11

Consensus sequence of DNA motifs linked to chromosomal instability. Regions listed in Supplemental Table 9 were searched for the presence of these motifs using the Fuzznuc algorithm. Results are listed in Supplemental Tables 11-15.

Supplemental Table 12

Fuzznuc results for the region chr2:77040740-77440739 in BED format.

Supplemental Table 13

Fuzznuc results for the region chr3:153832048-154232047 in BED format.

Supplemental Table 14

Fuzznuc results for the region chr4:1-400000 in BED format.

Supplemental Table 15

Fuzznuc results for the region chr10:21796997-22196996 in BED format.

Supplemental Table 16

Fuzznuc results for the region chr17:71359024-71759023 in BED format.