nature genetics

Regional copy number–independent deregulation of transcription in cancer

Nicolas Stransky[1,13], Céline Vallot[1,13], Fabien Reyal[1], Isabelle Bernard-Pierrot[1], Sixtina Gil Diez de Medina[1,2], Rick Segraves[3], Yann de Rycke[4], Paul Elvin[5], Andrew Cassidy[5], Carolyn Spraggon[5], Alexander Graham[5], Jennifer Southgate[6], Bernard Asselain[4], Yves Allory[2,7], Claude C Abbou[2,8], Donna G Albertson[3,9], Jean Paul Thiery[1,10,11], Dominique K Chopin[2,8,12], Daniel Pinkel[3] & François Radvanyi[1]

Genetic and epigenetic alterations have been identified that lead to transcriptional deregulation in cancers. Genetic mechanisms may affect single genes or regions containing several neighboring genes, as has been shown for DNA copy number changes. It was recently reported that epigenetic suppression of gene expression can also extend to a whole region; this is known as long-range epigenetic silencing. Various techniques are available for identifying regional genetic alterations, but no large-scale analysis has yet been carried out to obtain an overview of regional epigenetic alterations. We carried out an exhaustive search for regions susceptible to such mechanisms using a combination of transcriptome correlation map analysis and array CGH data for a series of bladder carcinomas. We validated one candidate region experimentally, demonstrating histone methylation leading to the loss of expression of neighboring genes without DNA methylation.

Changes to gene expression patterns are an important feature of cancer cells. These alterations are caused directly or indirectly by genetic or epigenetic events. DNA copy number changes are one of the key genetic mechanisms responsible for changes in gene expression. Comparative genomic hybridization on DNA microarrays now allow precise scanning of the entire genome for changes in DNA copy number[1,2].

Analyses of the expression of genes located in regions of gains or losses have shown that for a significant fraction of these genes, expression level varies consistently with gene copy number[3–7]. For example, comparisons of comparative genomic hybridization (CGH) array and transcriptome data have shown that 40%–60% of the genes in highly amplified regions are overexpressed[4,5,7].

Conversely, many recent studies have suggested that copy number changes are responsible for the gene expression biases inferred from transcriptome maps[8–13]. These maps are obtained by analyzing the transcriptome as a function of the position of genes on chromosomes. Two main approaches have been used to establish these transcriptome maps. In the first, levels of gene expression in tumor samples are compared with those in normal samples, and regions containing mostly upregulated or downregulated genes are identified[8–14]. The second approach involves trying to identify groups of neighboring genes with correlated expression levels. This approach was initially used in yeast and *Drosophila melanogaster*[15,16]. Building on this approach, we recently developed a new method for transcriptome analysis in human tumors: the transcriptome correlation map (TCM)[17]. All studies based on transcriptome maps have shown that the chromosomal regions of gene expression bias are mostly due to changes in copy number[8–12,14,17,18], with very few discrepancies reported[11,12]. However, such discrepancies are of particular interest, as they may result from important new mechanisms of tumor progression, such as long-range epigenetic mechanisms. Such mechanisms, involving DNA methylation and histone modification over a large chromosomal region, have recently been experimentally characterized in colon cancer[19], highlighting the biological existence of regions of transcriptome bias not due to copy number changes.

Here, by combining transcriptome correlation maps and genomic alterations assessed by array CGH for a series of 57 bladder tumors, we identified, in a systematic manner, those regions of transcriptome bias that were independent of DNA copy number changes. Moreover, comparison of these 'nongenetic' regions with the 'genetic' regions (regions of coordinated expression directly accounted for by DNA

[1]UMR 144 Centre National de la Recherche Scientifique (CNRS)/Institut Curie, 75248 Paris Cedex 05, France. [2]EMI Institut National de la Santé et de la Recherche Médicale (INSERM) 03-37, Hôpital Henri Mondor, 94010 Créteil Cedex, France. [3]Comprehensive Cancer Center, University of California San Francisco (UCSF), San Francisco, California 94143, USA. [4]Department of Biostatistics, Institut Curie, 75248 Paris Cedex 05, France. [5]Cancer and Infection Research Area, AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK. [6]Jack Birch Unit of Molecular Carcinogenesis, Department of Biology, University of York, Heslington YO10 5DD, York, UK. [7]Department of Pathology, Hôpital Henri Mondor, 94010 Créteil Cedex, France. [8]Department of Urology, Hôpital Henri Mondor, 94010 Créteil Cedex, France. [9]Cancer Research Institute, UCSF, San Francisco, California 94143, USA. [10]Department of Translational Research, Institut Curie, 75248 Paris Cedex 05, France. [11]Institute of Molecular and Cell Biology, Proteos, 138673 Singapore. [12]Deceased. [13]These authors contributed equally to this work. Correspondence should be addressed to F.R. (Francois.Radvanyi@curie.fr).

Received 9 August 2006; accepted 13 October 2006; published online 12 November 2006; corrected after print 27 February 2008; doi:10.1038/ng1923

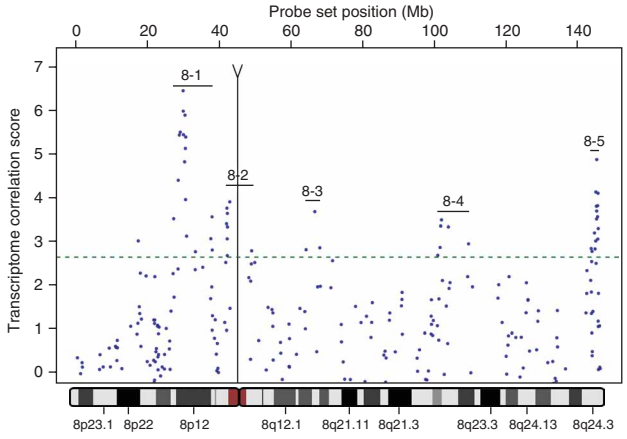

Figure 1 Transcriptome correlation map of chromosome 8 for the 57 bladder carcinomas. Each dot represents the transcriptome correlation score for a given gene, indicating the strength of the correlation between the expression profile of this gene and the expression profiles of the 14 neighboring genes. The transcriptome correlation map shows scores for the various genes as a function of their position along the chromosome. The dashed green line represents the significance threshold, calculated by taking the 499th quantile of the distribution of the transcriptome correlation scores obtained from randomly ordered datasets. Genes above the dashed line have a high transcriptome correlation score (>2.64) and are considered to be significantly correlated with their neighbors. Black lines above the dots indicate the five different TCM regions found on chromosome 8. These regions (8-1 to 8-5) contained 17, 8, 3, 6 and 15 genes and were 10.7, 7.0, 3.8, 8.6 and 2.0 Mb in size, respectively.

copy number changes) showed marked differences in terms of the proportion of downregulated genes and chromosomal location. We studied one of the copy number–independent regions in detail and found that epigenetic abnormalities were responsible for a coordinated loss of gene expression. Indeed, we showed that this loss of expression in a whole region was due to cancer-specific aberrant histone methylation, with no associated DNA methylation. Genes showing correlated expression in the nongenetic regions provide new candidates for involvement in tumor progression through epigenetic mechanisms.

RESULTS

Generation of transcriptome correlation maps

We determined mRNA levels for a series of 57 bladder tumor samples, using Affymetrix U95A or U95Av2 GeneChips. Of the 12,533 probe sets common to the U95A and U95Av2 chips, we retained 8,555 probe sets, each corresponding to a single gene, for the establishment of transcriptome correlation maps (see Methods). These maps highlight groups of neighboring genes with correlated expression patterns. We calculated the transcriptome correlation score for each gene using a moving window of 14 genes around each gene. This score is the sum of correlation values, for the tumor samples, between the RNA levels of the gene of interest and those of each of its neighbors (seven telomeric genes and seven centromeric genes). We established a significance threshold (the score above which a gene is considered to have an expression pattern correlated with that of its neighbors) using random data sets obtained by randomly ordering the 8,555 genes in the genome. We generated transcriptome correlation maps for all the chromosomes of the 57 bladder tumors (**Fig. 1** and **Supplementary**

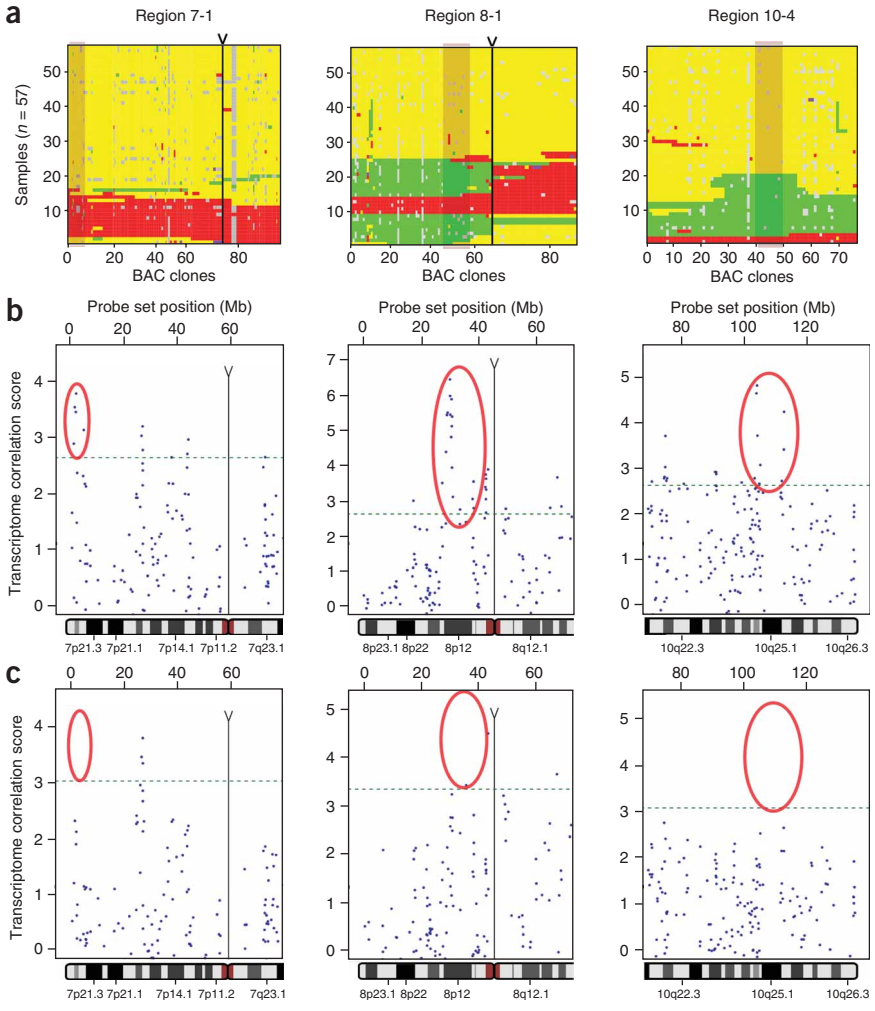

Figure 2 Comparison of CGH array data and transcriptome correlation maps: identification of regions of correlation due to DNA copy number changes. (**a**) CGH BAC array results for 57 bladder carcinomas, analyzed with a segmentation algorithm (see Methods) for chromosomes 7, 8 and 10 (yellow, no change in DNA copy number; green, loss; red, gain; purple, amplification). The shading indicates which BAC clones are located near the correlated genes from the three TCM regions on the second row. (**b**) Transcriptome correlation map for chromosomes 7, 8 and 10 for 57 bladder carcinomas. The correlated genes of regions 7-1, 8-1 and 10-4 are circled. (**c**) Transcriptome correlation maps for a subset of bladder carcinoma samples presenting no DNA copy number changes around the genes from the three TCM regions. For the three regions, when we excluded the tumors presenting genomic alterations (14, 29 and 20 tumors for region 7-1, 8-1 and 10-4, respectively) from the data set, we did not observe any more correlated genes.

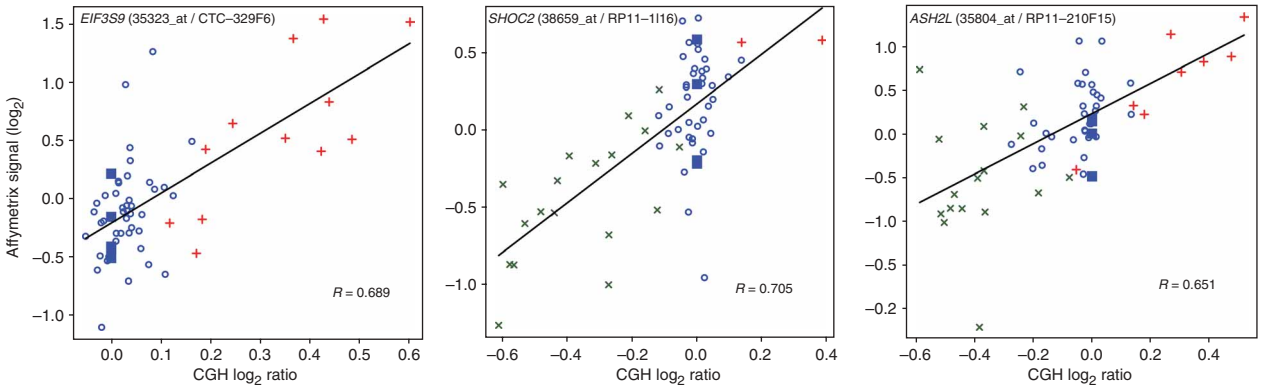

Figure 3 Scatter plots of Affymetrix signal versus CGH $\log_2$ ratio for the *EIF3S9*, *SHOC2* and *ASH2L* genes. These three genes with correlated expression patterns belong to regions 7-1, 10-4 and 8-1, respectively. Probe set intensities for *EIF3S9* (35323_at), *SHOC2* (38659_at) and *ASH2L* (35804_at) were correlated with the CGH $\log_2$ ratio of the nearest BAC clones, CTC-329F6, RP11-1l16 and RP11-210F15. The symbols on the scatter plot denote gains (+), losses (×) or unchanged status (open circle), as assessed by the segmentation algorithm, or normal samples (filled square). For region 7-1, *EIF3S9* expression was upregulated owing to an increase in gene copy number ($R = 0.689$); for region 10-4, we observed a downregulation of *SHOC2* expression due to DNA losses ($R = 0.705$); for region 8-1, in which both gains and losses occurred, we observed both upregulation due to DNA gains and downregulation due to DNA losses ($R = 0.651$).

Fig. 1 online). We found that 900 probe sets (corresponding to 900 single genes) had transcriptome correlation scores above the threshold. These 900 probe sets included 830 that could be grouped into 136 regions of correlated genes (as defined in Methods) (**Supplementary Table 1** online). These regions, for our bladder tumor data set, contained up to 27 correlated genes (median of five genes) and were 24 kb–23.5 Mb (median of 1.3 Mb) in size.

Copy number–dependent regions of correlation

We investigated whether the regions of correlated expression were due to underlying DNA copy number changes by combining the transcriptome and CGH array data of the 57 bladder tumor samples. For each of the 136 regions of correlation, we recalculated transcriptome correlation scores, excluding tumor samples presenting a DNA copy number change in this region. We thus obtained 136 new transcriptome correlation maps. For each of these transcriptome correlation maps, we determined a new threshold, taking into account the new tumor data set. We found that 528 of the initial total of 830 genes with correlated expression patterns located in the regions of correlation were not conserved (the correlation disappeared when samples with DNA copy number changes were removed from the data set). In 51 (37.5%) of the 136 regions of correlation, 100% of the correlated genes disappeared, and in 82 regions (60% of the regions of correlation), more than 60% disappeared. These 82 regions that were mostly not conserved after recalculation, in which the correlation was essentially due to DNA copy number changes, are described hereafter as 'DNA copy number–dependent'. They contained 448 genes from the initial set of 8,555 genes, 388 of which no longer correlated after recalculation. Three examples of DNA copy number–dependent regions of correlation are shown in **Figure 2**. These three regions presented very different patterns of genomic alterations: only gains for region 7-1, mainly losses for region 10-4 and both gains and losses for region 8-1 (**Fig. 2a**). **Figure 2b** shows the transcriptome correlation maps for each of these three regions. These regions of correlation disappeared when we generated a new TCM for each region after excluding tumors with genomic alterations (**Fig. 2c**). The correlated expression in these regions was therefore due to DNA copy number changes. The most likely explanation for this is that the level of expression of the correlated genes was linked to gene copy number. Indeed, for one correlated gene located in each of the three regions (*EIF3S9* in region 7-1, *SHOC2* in region 10-4 and *ASH2L* in region 8-1), we observed a strong correlation between expression and gene copy number (**Fig. 3**). We observed this relationship between mRNA levels and gene copy number for all the correlated genes of the three regions (data not shown).

Copy number–independent regions of correlation

In addition to the 82 regions of correlation entirely or partly accounted for by DNA copy number changes, we identified regions retained on transcriptome correlation maps after recalculation in which the correlation could not be accounted for by DNA copy number changes. In nine regions of correlation, 100% of the correlated genes remained correlated after recalculation, and in 28 regions of correlation, more than 60% of the genes remained correlated. We classified these 28 regions of correlation as copy number–independent. These regions contained 191 of the initial set of 8,555 genes, 149 of which remained correlated after recalculation. A list of the copy number–independent regions and the correlated genes they contain after recalculation is given in **Table 1**. We observed the inclusion of additional correlated genes after TCM recalculation (one or two per region) for 5 of the 28 nongenetic regions. Three of the nongenetic regions and the corresponding DNA losses and gains are shown in **Figure 4**. As expected, the levels of mRNA for each of the correlated genes in these regions were poorly correlated with gene copy number (the mean correlation coefficient was 0.189 for region 1-6, 0.206 for region 5-3 and –0.095 for region 7-2) (data not shown). The identification of these regions of correlation not due to DNA copy number changes could not be accounted for by insufficient coverage of the genome by the BAC set used in the CGH array, because (i) the mean distance between a correlated gene and the nearest BAC was not significantly different ($P = 0.284$, Wilcoxon rank sum test) for the 448 correlated genes from copy number–dependent regions (mean distance = 490 kb) and the 191 correlated genes from copy number–independent regions (mean distance = 442 kb); (ii) similar proportions of genes from copy number–independent (11.0%) and copy number–dependent regions (11.4%) were located on the BACs used for CGH array (χ^2, $P = 0.99$) and (iii) the BACs corresponding to copy number–dependent or –independent regions were of similar mean size (Student's *t*-test, $P = 0.94$).

Properties of copy number–dependent and –independent regions

We identified two types of regions of correlated genes: those in which the correlation resulted mainly from changes in gene copy number and those in which the correlation was mainly copy number–independent. We then investigated whether these two different types of regions had different properties in terms of size, expression level and genomic location. The median sizes of the copy number–dependent and copy number–independent regions were not significantly different (1.33 Mb and 1.27 Mb, respectively; $P = 0.83$, Wilcoxon rank sum test). The number of genes significantly underexpressed or overexpressed in tumors with respect to normal urothelium is shown in **Table 2** for these two types of regions. The ratio of underexpressed to overexpressed genes for DNA copy number–independent regions differed significantly from that for DNA copy number–dependent regions (χ^2, $P = 0.015$) and that for all 8,555 genes (χ^2, $P = 0.010$). By contrast, this ratio was not significantly different between copy number–dependent regions and all the 8,555 genes considered together (χ^2, $P = 0.65$). Therefore, the proportion of downregulated genes was significantly higher for the copy number–independent regions than for the copy number–dependent regions or all genes considered together.

We next investigated whether the chromosomal locations of the two types of regions differed by comparing the proportions of genes within these regions located close to the telomeres (within 10 Mb of the telomere) and those further away from the telomeres. We found that 7.9% of the genes within DNA copy number–independent regions

Table 1 Description of the 28 copy number–independent regions

TCM region (identification number)	Cytobands	Number of correlated genes in original TCM region	Number of remaining correlated genes in new TCM region	Correlated genes in the new TCM region	Region size (kb)
1-6	1p31.1	5	5 (100%)	*CRYZ*, *RABGGTB*, *ZZZ3*, *USP33*, *FUBP1*	3,242
4-2	4q13.3	6	6 (100%)	*IL8*, *CXCL6*, *CXCL1*, *CXCL2*, *CXCL5*, *CXCL3*	296
6-2	6p21.32	12	12 (100%)	*HLA-DRB1*, *HLA-DQA2*, *HLA-DOB*, *TAP2*, *PSMB8*, *TAP1*, *PSMB9*, *HLA-DMB*, *HLA-DMA*, *HLA-DOA*, *HLA-DPA1*, *HLA-DPB1*	497
6-7	6q23.3–6q24.1	7	7 (100%)	*MAP7*, *MAP3K5*, *PEX7*, *IFNGR1*, *HEBP2*, *C6orf80*, *CITED2*	3,029
7-2	7p15.2	3	3 (100%)	*SCAP2*, *HOXA1*, *HOXA2*, *HOXA4*, *HOXA5*	472
12-4	12q24.23–12q24.31	3	3 (100%)	*GCN1L1*, *COX6A1*, *SFRS9*	334
14-1	14q11.2	9	9 (100%)	*LRP10*, *ACIN1*, *PABPN1*, *EFS*, *AP1G2*, *DHRS2*, *PCK2*, *WDR23*, *PSME1*	1,260
19-2	19p13.13	2	2 (100%)	*PRDX2*, *FARSLA*	125
19-3	19p13.12–19p13.11	12	12 (100%)	*WIZ*, *CYP4F3*, *CYP4F12*, *CYP4F11*, *CHERP*, *SIN3B*, *MYO9B*, *FLJ22709*, *NR2F6*, *FCHO1*, *B3GNT3*, *INSL3*, *SLC5A5*, *PIK3R2*	2,732
2-7	2q31.1–2q31.2	6	5 (83.3%)	*HOXD4*, *HOXD3*, *HOXD1*, *HNRPA3*, *NFE2L2*, (*PRKRA*)	2,281
6-5	6q16.1	5	4 (80%)	*KIAA0776*, *C6orf111*, (*CCNC*), *ASCC3*, *PREP*	8,755
19-1	19p13.3	10	8 (80%)	*STK11*, *ATP5D*, (*NDUFS7*), *RPS15*, *MBD3*, *UQCR*, *CSNK1G2*, *BTBD2*, *LSM7*, (*SPPL2B*)	1,133
6-3	6p21.1	9	7 (77.8%)	(*TNRC5*), *PPP2R5D*, *MEA*, *CUL7*, *C6orf108*, (*ABCC10*), *C6orf109*, *POLR1C*, *MAD2L1BP*	707
1-4	1p34.1	8	6 (75%)	*ATP6V0B*, (*B4GALT2*), *PRNPIP*, *RPS15A*, *UROD*, *PRDX1*, *AKR1A1*, (*PIK3R3*)	2,108
17-7	17q21.32	4	3 (75%)	*HOXB2*, *HOXB3*, *HOXB5*, (*HOXB7*)	64
1-12	1q23.3	3	2 (66.7%)	*HSPA6*, *FCGR3A*, *FCGR2B*	57
2-5	2q23.3–2q24.2	9	6 (66.7%)	(*RIF1*), (*FNBP3*), *GPD2*, *PKP4*, *BAZ2B*, *MARCH7*, *LY75*, *ITGB6*, (*RBMS1*)	8,808
5-4	5q15–5q23.1	12	8 (66.7%)	*CAST*, *KIAA0433*, (*PJA2*), (*MAN2A1*), *SRP19*, *U2AF1L2*, *DCP2*, *YTHDC2*, *PGGT1B*, *FEM1C*, *TMED7*, *APG12L*, (*AP3S1*), (*TNFAIP8*)	22,662
12-2	12q13.2–12q13.3	6	4 (66.7%)	*ERBB3*, (*PA2G4*), (*RPL41*), *MBC2*, *SMARCC2*, *TMEM4*	230
17-8	17q21.33–17q22	3	2 (66.7%)	*WDR50*, (*TOM1L1*), *COX11*	3,701
19-8	19q13.32	6	4 (66.7%)	*DMWD*, (*SYMPK*), *IRF2BP1*, *PPP5C*, (*CALM3*), *PRKD2*, *SLC1A5*	891
23-1	Xq13.1	3	2 (66.7%)	*EDA*, *DLG3*, (*MLLT7*)	1,480
23-2	Xq22.1–Xq22.2	3	2 (66.7%)	(*BEXL1*), *TCEAL4*, *TCEAL1*	413
1-1	1p36.33–1p36.32	5	3 (60%)	(*DVL1*), *PRKCZ*, (*RER1*), *PLCL4*, *TNFRSF14*, *ARHGEF16*	2,095
2-3	2p16.1–2p14	5	3 (60%)	(*RPS27A*), *FANCL*, (*XPO1*), *MDH1*, *RAB1A*	9,855
3-2	3p22.3	5	3 (60%)	(*CTDSPL*), (*VILL*), *PLCD1*, *ACAA1*, *MYD88*	277
3-5	3p21.31	10	6 (60%)	(*UQCRC1*), (*ARIH2*), *IMPDH2*, *GPX1*, *RHOA*, *AMT*, *UBE1L*, *MST1R*	1,288
5-3	5q12.1–5q13.2	20	12 (60%)	(*PDE4D*), *ERCC8*, (*KIF2*), *HSA9761*, (*FLJ36754*), *KIAA0073*, (*TRIM23*), *SFRS12*, *CDK7*, *TAF9*, *RAD17*, (*OCLN*), *FLJ40092*, *LOC166994*, *SMN2*, (*SMA4*), *LOC153561*, *BIRC1*, (*MRPS27*), (*BTF3*)	14,528

For each region, the number of correlated genes in the original TCM region is indicated, together with the number of remaining genes in this region after recalculation of the TCM, excluding the samples presenting genetic alterations. Retained correlated genes are shown; those disappearing after recalculation are in parentheses, and those significantly correlated only after recalculation are underlined. Region size corresponds to the original TCM region size.

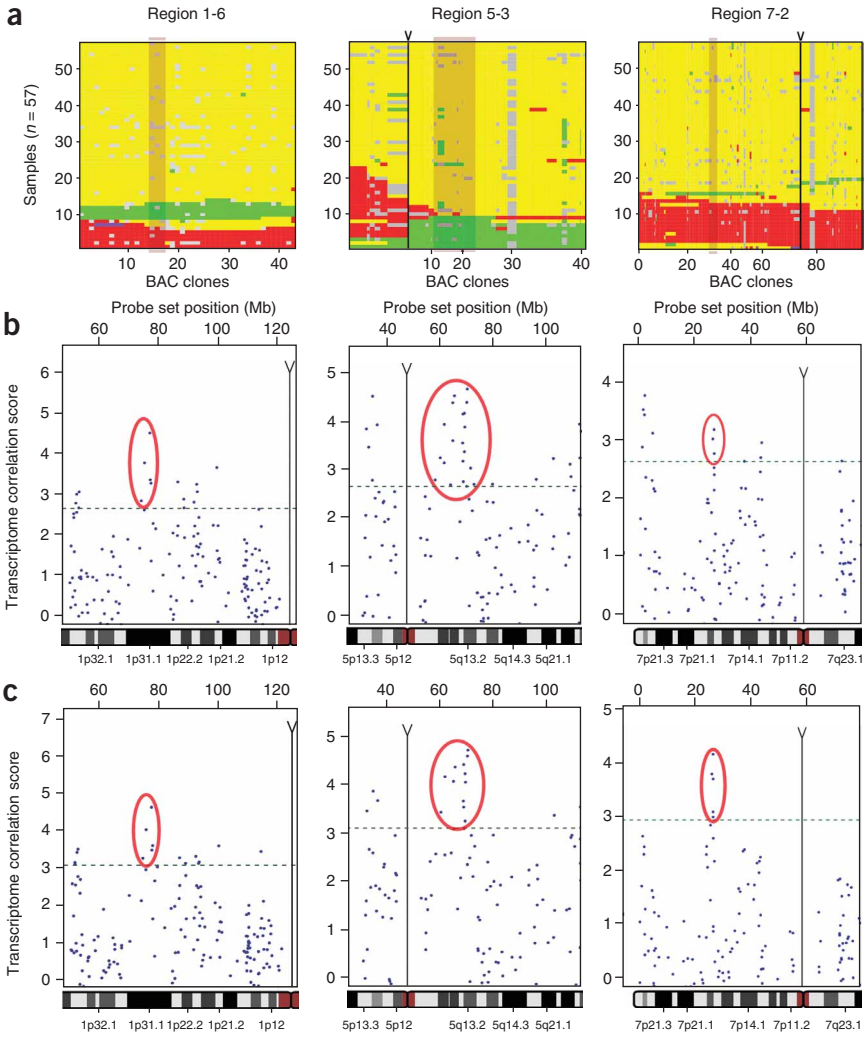

Figure 4 Comparison of CGH array data and transcriptome correlation maps: identification of regions of correlation not due to DNA copy number changes. (**a**) CGH BAC array results for 57 bladder carcinomas, analyzed with a segmentation algorithm (see Methods) for chromosomes 1, 5 and 7 (yellow, no change in DNA copy number; green, loss; red, gain; purple, amplification). The shading indicates which BAC clones are located near the correlated genes from the three TCM regions on the second row. (**b**) Transcriptome correlation map for chromosomes 1, 5 and 7 for 57 bladder carcinomas. The correlated genes of regions 1-6, 5-3 and 7-2 are circled. (**c**) Transcriptome correlation maps for a subset of bladder carcinoma samples presenting no DNA copy number changes around the genes from TCM regions 1-6, 5-3 and 7-2. The expression profiles of the genes from these TCM regions are still significantly correlated (genes in the ellipses) if only samples with an unchanged BAC status in these three regions are considered.

were located within 10 Mb of a telomere but that the percentage of genes close to telomeres was much higher for genes within DNA copy number–dependent regions (35.9%). This difference was highly significant (χ^2, $P \leq 10^{-12}$). When we considered all 8,555 genes, we found 19% to be close to the telomeres. This percentage differed significantly from the percentages for copy number–independent and –dependent regions (χ^2, $P = 0.00014$ and $P \leq 10^{-15}$, respectively). We can thus conclude that DNA copy number–dependent regions are significantly more likely to be located near the telomeres, whereas DNA copy number–independent regions are significantly more likely to be located far from the telomeres.

Delineation of a copy number–independent region on 3p22.3

We studied one copy number–independent region in detail to validate our bioinformatic approach biologically. As we were looking for regions affected by epigenetic alterations leading to the abolition of expression, we searched for regions containing mostly downregulated genes. We focused on region 3-2, located on chromosome 3p22.3 (**Table 1**), a region of relatively small size for which all the genes and flanking genes have a corresponding probe set on the U95Av2 chip. The region obtained by TCM analysis, taking all the tumors into account, contained five correlated genes (*CTDSPL*, *VILL*, *PLCD1*, *ACAA1*, *MYD88*) and one noncorrelated gene (*DLEC1*) (**Fig. 5a**, left). The absence of correlation between the expression of *DLEC1* and of its neighbors could not be interpreted, as the signals for the probe set corresponding to this gene were considered 'absent' in all samples in Affymetrix assays. After excluding the 18 tumors presenting a genetic alteration in this region, three of the five correlated genes remained correlated (*PLCD1*, *ACAA1*, *MYD88*), one (*VILL*) had a correlation score close to the threshold (transcriptome correlation score = 2.87 with a threshold of 2.99) and the remaining gene was no longer correlated (*CTDSPL*) (**Fig. 5a**, right). The lower panel of **Figure 5a** summarizes the correlation status of the different genes of this region after TCM recalculation. A comparison of the Affymetrix expression data for tumor samples with data for normal tissues is shown in **Figure 5b** for these genes (except for *DLEC1*, for which the Affymetrix data could not be interpreted). Four genes (*CTDSPL*, *VILL*, *PLCD1*, *ACAA1*) were significantly downregulated in tumors. We compared the levels of expression obtained with U95Av2 chips for *VILL*, *PLCD1*, *ACAA1* and *MYD88* with quantitative RT-PCR measures for ten bladder tumor samples and four normal samples, to confirm the Affymetrix data. For *VILL*, *PLCD1* and *ACAA1*, the levels of expression measured by quantitative RT-PCR were strongly correlated with

Table 2 Properties of copy number–dependent and –independent regions

	Genes within copy number–independent regions	Genes within copy number–dependent regions	All genes
Total number	191	448	8,555
Underexpressed	30 (15.7%)	45 (10.0%)	951 (11.1%)
Overexpressed	13 (6.8%)	53 (11.8%)	997 (11.6%)
Close to telomeres	15 (7.9%)	161 (35.9%)	1,623 (19%)
Far from telomeres	176 (92.1%)	287 (64.1%)	6,932 (81%)

U95Av2 results, giving correlation coefficients of $R = 0.787$, 0.903 and 0.610 (**Fig. 5c** and data not shown). However for *MYD88*, quantitative RT-PCR analysis did not validate the Affymetrix data, as no correlation was found between RT-PCR and Affymetrix data ($R = -0.0918$) (data not shown). As the Affymetrix signals for the probe set corresponding to *DLEC1* were not detectable, we determined the level of expression of this gene by quantitative RT-PCR, which showed that the expression of this gene was correlated with that of its neighbors *VILL*, *PLCD1* and *ACAA1* but not with that of *MYD88* (**Fig. 5d** and data not shown). In addition, *DLEC1* was expressed less strongly in tumors than in normal tissues (data not shown). The region containing the four neighboring coordinately downregulated genes (*VILL*, *PLCD1*, *DLEC1*, *ACAA1* (**Fig. 5e**)) was thus identified as a potentially interesting candidate region for further experimental studies.

Correlated genes in region 3-2 are under epigenetic control

We first carried out quantitative PCR analysis to exclude the possibility that the coordinated downregulation of the four contiguous genes of the 3-2 region was due to a small genomic deletion undetected by CGH analysis (**Supplementary Fig. 2** online). Functional studies were carried out to determine the mechanisms underlying the coordinated downregulation of *VILL*, *PLCD1*, *DLEC1* and *ACAA1* in region 3-2. We used the TCCSUP bladder tumor cell line for these studies, as this cell line shows downregulation of these genes but not of the flanking genes *CTDSPL*, *MYD88* and *OXSR1*, with respect to normal urothelium. This cell line also has no genomic alteration at 3p22.3 (data not shown). We investigated whether the genes of region 3-2 were coordinately downregulated by an epigenetic mechanism by treating TCCSUP cells with 5-aza-deoxycytidine, a demethylating agent, and/or with trichostatin A (TSA), a deacetylase inhibitor. Normal human urothelial cells grown in culture (NHU cells) were used as a control. We reexpressed the four contiguous genes (*VILL*, *PLCD1*, *DLEC1* and *ACAA1*) by treating cells with TSA or TSA and 5-aza-deoxycytidine; after either treatment, expression increased by a maximum of 22-fold for *VILL*, 33-fold for *PLCD1*, 74-fold for *DLEC1* and fourfold for *ACAA1* (**Fig. 6a**). Treatment with 5-aza-deoxycytidine alone caused no significant reexpression of this group of genes, except for *VILL*, which showed a fivefold increase in expression. The levels of expression of the flanking genes *CTDSPL*, *MYD88* and *OXSR1* were not increased by either treatment alone or by the combined treatment. These results show that expression in the region including *VILL*, *PLCD1*, *DLEC1*

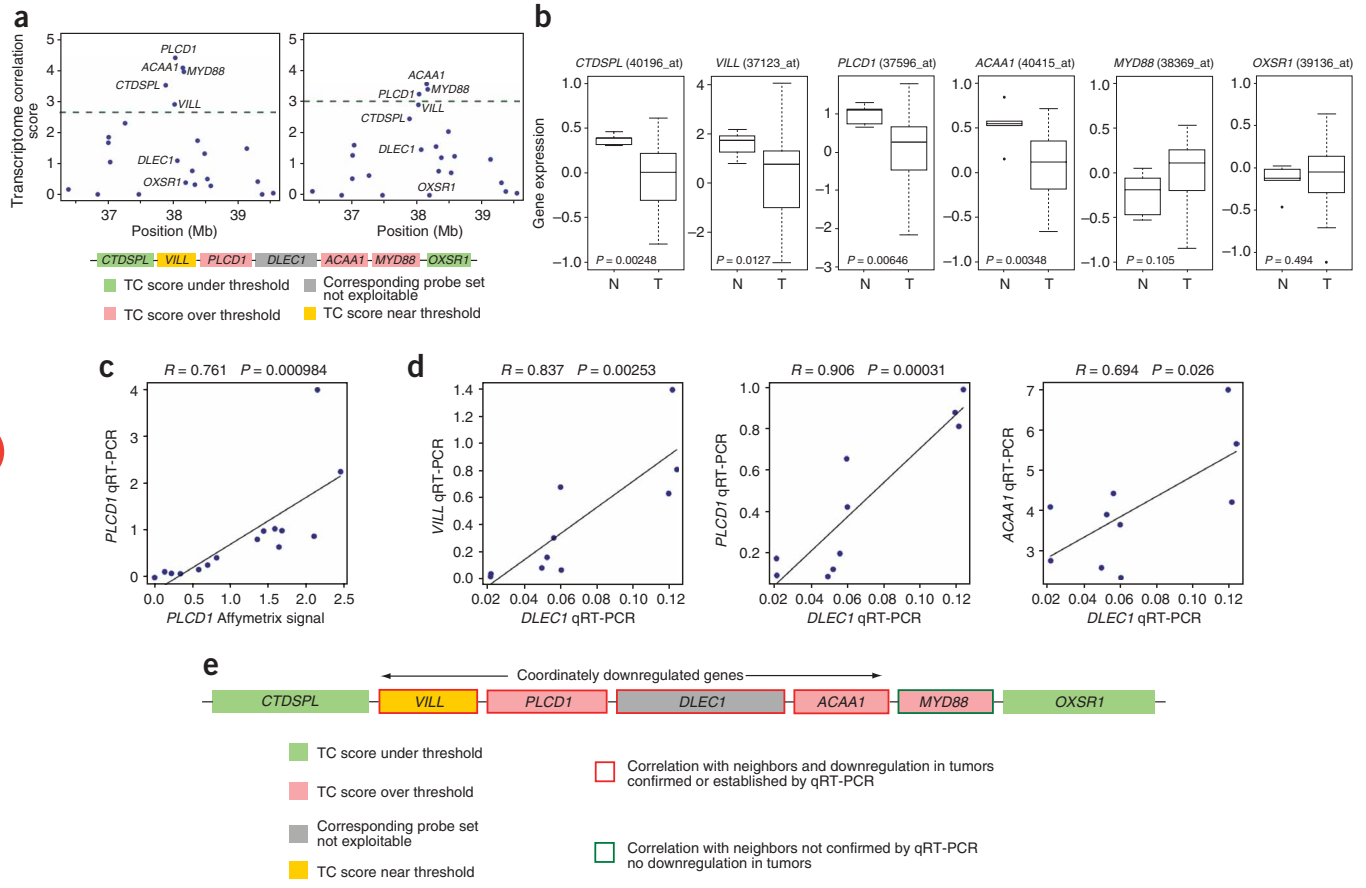

Figure 5 Delimitation of the 3-2 region by transcriptome and genome analyses. (**a**) Right: transcriptome correlation maps (TCM) of the 3-2 region, calculated for all 57 tumor samples. For the sake of clarity, negative transcriptome correlation scores were arbitrarily set to zero. Left: transcriptome correlation maps of the 3-2 region calculated for only the 39 samples without genetic alterations in this region. Lower panel: schematic representation of region 3-2 and the new TCM data, established with only the 39 samples without genetic alterations for this region (not to scale). (**b**) Box plots representing the distribution of $\log_2$ Affymetrix values of the genes in the five normal and 39 tumor samples; *P* values were obtained using a Wilcoxon rank-sum test. The two flanking noncorrelated genes *CTDSPL* and *OXSR1* are also shown. (**c**) Scatter plot of *PLCD1* expression in Affymetrix microarrays versus its expression as measured by quantitative RT-PCR (qRT-PCR). (**d**) Scatter plots representing the level of expression of *DLEC1* versus that of its three neighbors, *VILL*, *PLCD1* and *ACAA1*, for ten tumor samples. (**e**) Schematic representation of region 3-2 (not to scale), taking into account both the new TCM data and qRT-PCR data.

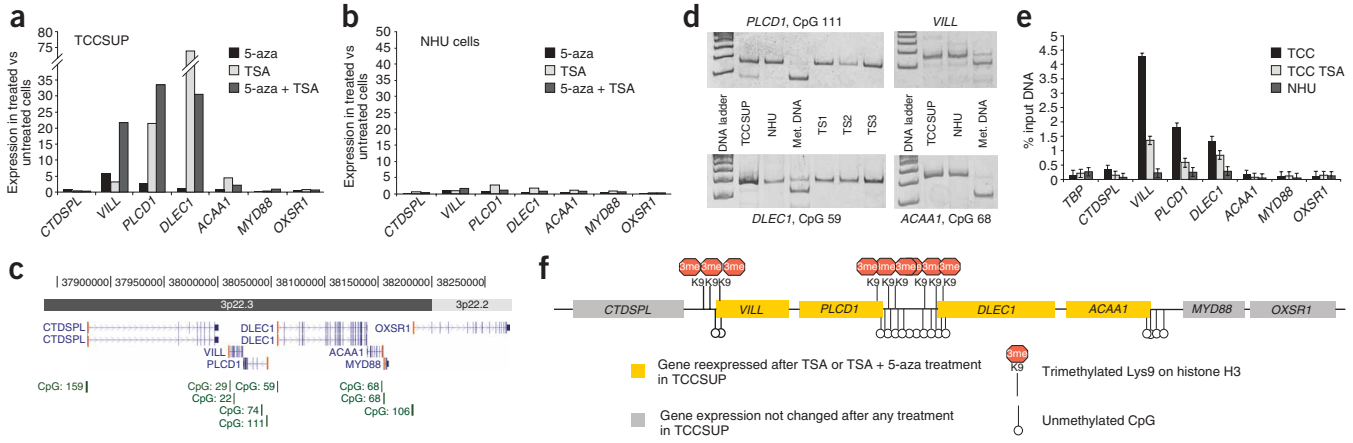

Figure 6 Experimental validation of the 3-2 region. (**a**,**b**) Comparison of the mRNA levels of the genes in region 3-2 and their neighbors in TCCSUP cells (**a**) and in NHU normal human urothelial cells (**b**) before and after treatment with 5-aza-deoxycytidine, TSA or 5-aza-deoxycytidine + TSA. (**c**) UCSC genome browser map of the 3-2 region showing the base position, the chromosome bands, the RefSeq genes and the localization of CpG islands in green, with the number of CpG dinucleotides for each CpG island. The transcription start site is indicated for each gene by an orange vertical bar. (**d**) COBRA analysis of the promoters of genes reexpressed in the 3-2 region after treatment with TSA or 5-aza-deoxycytidine + TSA for TCCSUP DNA, NHU DNA, DNA methylated by *Sss*I CpG methylase ('Met DNA') (for *VILL*, *PLCD1*, *DLEC1* and *ACAA1*), and also for three tumor samples, TS1–TS3 (for *PLCD1* and *DLEC1*). The *VILL* PCR product (132 bp) was digested with *Bst*UI (99+33 bp), the *PLCD1* PCR product (127 bp) was digested with *Hin*pI (41+86 bp), the *DLEC1* PCR product (134 bp) was digested with *Bst*UI (31+103 bp) and the *ACAA1* PCR product (158 bp) was digested with *Hin*pI (105+53 bp). When a studied CpG is methylated, the COBRA PCR product is digested by the restriction enzyme used. (**e**) ChIP analysis of the promoters of region 3-2, with an antibody against trimethyl H3K9. The bar chart shows the amount of immunoprecipitated target DNA expressed as a percentage of total input DNA. *TBP* was used as a negative control. The error bars indicate the mean variation between two independent experiments. (**f**) Summary of experimental validation of region 3-2 (not to scale).

and *ACAA1*, previously identified by a combined genomic and transcriptomic approach (**Fig. 5e**), is controlled by an epigenetic mechanism. We did not observe any reexpression in NHU cells (**Fig. 6b**), indicating that the epigenetic mechanism was not induced by culture conditions.

We investigated the possible involvement of DNA or histone methylation through combined bisulfite restriction analysis (COBRA) of DNA methylation for the various promoters and through a chromatin immunoprecipitation (ChIP) assay with an antibody against the trimethyl residue on Lys9 of histone 3 (H3K9). Three of the four coordinately downregulated genes, *PLCD1*, *DLEC1* and *ACAA1*, have a CpG island in their promoters, around the transcription start site (**Fig. 6c**). *VILL* presents only some CpG dinucleotides around its transcription start site. We studied the methylation status of the three CpG islands and one of the CpG dinucleotides closest to the *VILL* transcription start site in TCCSUP and NHU cells and in three tumor samples (TS1-3) underexpressing *VILL*, *PLCD1*, *DLEC1* and *ACAA1*. We observed little to no methylation in the TCCSUP tumor cell line and did not observe any methylation in the tumor samples (**Fig. 6d** and data not shown). As expected, the CpG islands within the *CTDSPL*, *MYD88* and *OXSR1* promoters did not show any DNA methylation in either the TCCSUP cells or tumor samples (data not shown).

We used ChIP followed by quantitative PCR to study the histone H3 modifications on Lys9 in the promoters of the various genes within or flanking the 3-2 region. We found that trimethyl groups were present only on the promoters of *VILL*, *PLCD1* and *DLEC1* in TCCSUP cells (**Fig. 6e**). No such mark was present in any of the promoters studied in NHU cells. As expected, the trimethyl modification was decreased by the TSA treatment of TCCSUP cells (**Fig. 6e**), which also led to the reexpression of the genes of the 3-2 region (**Fig. 6a**). **Figure 6f** summarizes the epigenetic modifications of region 3-2: absence of DNA methylation in the promoters of *VILL*, *PLCD1*, *DLEC1* and *ACAA1* and presence of H3K9 trimethylation in the promoters of *VILL*, *PLCD1* and *DLEC1*.

DISCUSSION

The two major contributions of this work are the development of a systematic approach for identifying candidate regions controlled by epigenetic mechanisms in cancer and the characterization of one of these regions. In a region controlled by an epigenetic mechanism, the genes would be expected to show correlated expression independent of any genomic alteration. Therefore, we used a combination of our recently developed TCM approach[17] and CGH array analysis results to search for regions controlled by epigenetic mechanisms: we first looked for regions containing genes with correlated expression profiles, and then, for each region identified, we investigated whether the correlation persisted if only tumors without genomic alteration in this region were taken into account. In addition to the 28 copy number–independent regions, we identified several mixed regions containing both copy number–dependent and copy number–independent correlated genes after recalculation and showing significant clustering of the copy number–independent genes (data not shown). This suggests that expression of the correlated genes in these regions may be controlled by nongenetic mechanisms, genomic alterations or both (one mechanism controlling each allele).

We studied in detail one of the copy number–independent regions of correlation: region 3-2. Bioinformatic analysis of transcriptomic and genomic data followed by complementary analysis by quantitative RT-PCR showed that this region was 130 kb long and consisted of four consecutive genes with correlated expression profiles downregulated in tumors: *VILL*, *PLCD1*, *DLEC1* and *ACAA1*. Using a bladder tumor cell line showing the downregulation of these four genes and no deletion of the 3-2 region, we showed that TSA or combined TSA and 5-aza-deoxycytidine treatments upregulated the expression of these four genes by a factor of 4 to 75, strongly suggesting that histone

modifications were responsible for the observed downregulation. ChIP experiments showed that, for three of the four genes (*VILL*, *PLCD1* and *DLEC1*), there was a histone modification in the silenced region (trimethylation of Lys9 of histone 3, H3K9) that was decreased, as expected, by TSA treatment. Moreover, *VILL* reexpression in response to TSA is weaker than that of *PLCD1* and *DLEC1*, possibly because its H3K9 trimethyl mark remains 5.9 times higher than that in NHU cells after TSA treatment, versus only 2.2 and 2.9 times higher for *PLCD1* and *DLEC1*, respectively. The absence of H3K9 trimethylation for *ACAA1* and the reexpression of this gene after TSA treatment suggest that other histone modifications are involved in *ACAA1* silencing. In addition to H3K9 trimethylation, it would be interesting to look for other histone marks[20] in the 3-2 region. The absence of reexpression in response to 5-aza-deoxycytidine treatment, confirmed by DNA methylation analyses using COBRA (**Supplementary Note**), strongly suggests that DNA methylation is not involved in epigenetic control for *PLCD1*, *DLEC1* and *ACAA1*. For *VILL*, the study of one CpG dinucleotide close to the transcription start site did not show any significant methylation. The effect of 5-aza-deoxycytidine on *VILL* may correspond to an effect of this molecule on histone methylation, as previously described[21]. The genes of this 3-2 region are expressed in normal resting urothelium samples and in normal urothelial cells proliferating in culture. As these normal cells proliferate as rapidly as tumor cell lines (doubling time of about 24 h), the downregulation of the genes of the 3-2 region is not associated with proliferation but is likely to be tumor specific. Moreover, the downregulation of expression of this region is probably involved in tumor progression, as two of the four genes within this region have been identified as potential tumor suppressor genes: *DLEC1* has tumor-suppressive properties in cancer cell lines of various origins[22], and *PLCD1* knockout in mice induces skin tumors[23].

Using the bioinformatic approach described here, we were able to isolate a 130-kb region containing four genes downregulated by an epigenetic mechanism involving histone modifications in tumors. Until recently, the epigenetic inactivation of gene expression in cancer was thought to act locally, affecting individual genes. Others[19] recently reported the inactivation of all ten genes in an entire chromosomal band by both histone modifications and DNA methylation. DNA methylation was thought to be a key factor in the silencing of this region. By contrast, the active mechanism in the 3-2 region seems to be different: this region is much smaller and there seems to be no DNA methylation in promoters. Despite differences in the inactivation mechanism, our results and those of ref. 19 suggest that long-range epigenetic silencing may be a common phenomenon in cancer.

Some of the other regions we identified may also be regulated by chromatin modifications. These regions include the three HOX clusters (regions 2-7, 7-2 and 17-7) and the major histocompatibility complex locus (region 6-2). Indeed, in normal cells, the expression of the genes within these loci has already been shown to be influenced by chromatin modifications[24–26] occurring during development and differentiation. Other nonexclusive, nongenetic mechanisms may also be involved in the regulation of gene expression in copy number–independent regions. These mechanisms include (i) DNA methylation[27], (ii) changes in the nuclear location of chromosomal regions[28] and (iii) changes affecting a transcription factor common to the correlated genes of a given region. Until the study reported in ref. 19, methylation was considered to affect only individual genes[29]. We found that most of the genes known to be affected by promoter methylation in bladder cancer were not located in regions of correlated expression, suggesting that methylation may not have a central role in the copy number–independent regions. Changes in the nuclear location of a chromosomal region may also account, at least partly, for some of the regions we identified; some regions of the nucleus favor gene expression, whereas others may inhibit expression[28]. Correlated expression and histone modifications have been reported for neighboring genes in normal tissues[30,31]. Thus, it will be essential to distinguish the tumor-specific regions from the normal ones. In this study, we used NHU cells to address this question for the 3-2 region. For systematic sorting, we would need to construct a TCM with transcriptomes from normal urothelia in various conditions (at various developmental stages or during wound repair or senescence), as some of the regions we identified may be associated with such conditions. For example, it has recently been reported that genes associated with wound repair are more strongly expressed in tumors with a poor prognosis[32].

Here we describe a new approach for identifying new regions and genes involved in tumor progression. This approach makes it possible to identify clusters of genes with expression profiles likely to be affected by nongenetic mechanisms. The systematic experimental study of such regions should make it possible to estimate the importance of histone modifications and DNA methylation in the long-range inactivation of genes in tumors. In light of the results obtained in this study, the TCM approach for identifying new candidate regions and genes involved in tumor progression could be applied to other tumor types for which transcriptome and CGH data are available for the same samples, potentially leading to the identification of new mechanisms of tumor progression. These advances could also lead to the identification of new groups of tumors susceptible to treatments targeting DNA methylation or histone modifications.

METHODS

Affected individuals and tissue samples. We analyzed the gene expression profiles and genomic alterations of 57 urothelial bladder carcinomas. These carcinomas were obtained from 53 affected individuals included between 1988 and 2001 in the prospective database set up in 1988 at the Department of Urology of Henri Mondor Hospital. The tumor samples came from 16 Ta, 9 T1, 6 T2, 13 T3 and 13 T4 tumors (**Supplementary Table 2** online). The flash-frozen tumor samples were stored at −80 °C immediately after transurethral resection or cystectomy. All tumor samples contained >80% tumor cells, as assessed by hematoxylin and eosin staining of histological sections adjacent to the samples used for transcriptome and genome analyses. Five normal urothelial samples obtained as described in ref. 33 were also used for transcriptome analysis. All subjects provided informed consent and the study was approved by the ethics committee of Henri Mondor Hospital.

RNA and DNA extraction. RNA and DNA were extracted from the samples by cesium chloride density centrifugation[34,35]. The concentration, integrity and purity of each RNA sample were determined with the RNA 6000 LabChip Kit (Agilent Technologies) and an Agilent 2100 Bioanalyzer. DNA purity was also assessed by determining the ratio of absorbances at 260 and 280 nm. DNA concentration was determined with a Hoechst dye-based fluorescence assay[36].

Affymetrix microarray data. We used the Human Genome U95A and U95Av2 arrays (Affymetrix) containing almost 12,500 probe sets. Details of the methods for RNA amplification, cDNA probe labeling and hybridization can be obtained from the Affymetrix website. Microarrays were scanned and the intensities for each probe set were calculated using Affymetrix MAS 5.0 default settings. The mean intensity of the probe sets for each array was set to a constant target value (500) by linearly scaling the array signal intensities. Log-transformed signal values were normalized by removing chip-specific and probe set–specific effects (the mean signal for all probe sets across one chip and the mean signal for one probe set across all chips, respectively).

CGH arrays. The 57 bladder tumor samples analyzed on Affymetrix DNA microarrays were also analyzed on CGH microarrays. These arrays consist of 2,385 BAC clones covering the human genome with an average resolution of 1.3 Mb (HumArray 2.0)[37]. The arrays were obtained from the UCSF Cancer Center Array CGH Core Facility. Probes were labeled and hybridization was carried out as described in ref. 2. Images were analyzed with SPOT 2.0 software[38]. Poor-quality spots were removed in a pre-processing step: only spots with a reference signal intensity (and 4,6-diamidino-2-phenylindole (DAPI) signal intensity) above 25% of the background reference signal (the DAPI signal) were considered reliable. Spots located in zones of spatial bias (abnormally high $\log_2$ ratios measured in some areas of the array, generally due to an edge or corner effect) were ignored[39], as were triplicates with $\log_2$ ratio s.d. > 0.1. For subsequent transcriptome and CGH correlations, we used the $\log_2$ ratio variable calculated by SPOT 2.0. Each BAC was assigned 'gain', 'normal' or 'loss' status, using the GLAD segmentation algorithm[40].

Generation of transcriptome correlation maps. We calculated transcriptome correlation scores and identified neighboring genes with correlated expression as follows. Regions containing neighboring genes with correlated expression patterns (transcriptome correlation map) were identified in the series of 57 bladder tumor transcriptomes obtained using Affymetrix U95A or U95Av2 GeneChips, as described in ref. 17. In total, 12,533 probe sets are common to both the Affymetrix U95A and U95Av2 chips. The complete U95 probe set list was filtered as previously described[17] to avoid redundancy and to ensure that a single probe set was retained for each identified gene. From the initial list of probe sets common to U95A and U95Av2, we retained 8,555 probe sets specific for a unique gene.

The transcriptome correlation map was generated by calculating a transcriptome correlation score (transcriptome correlation score) for each gene. This score was calculated as the sum of the correlations between the expression levels of this gene and of the $2n$ neighboring genes (n telomeric and n centromeric), for our 8,555-gene data set. A threshold was defined, as previously described, by creating 100 random data sets of the same size by randomly ordering the 8,555 probe sets within the genome. To avoid local correlations, the probe sets were ordered randomly with only one constraint: that no correlation between two probe sets from the same chromosome should be calculated. A threshold was calculated from the transcriptome correlation scores obtained in these 'random TCMs' with a P value < 0.002 (see the statistics section of the Methods). The appropriate number ($2n$) of neighboring genes required to calculate the transcriptome correlation score for each gene of the data set was determined by calculating the total number of genes with a score above the threshold as a function of $2n$ for several values ranging from 2–30. Above $2n = 14$, the number of genes with scores above the threshold reached a plateau (data not shown). Therefore, we used the 14 neighboring genes to calculate the transcriptome correlation score. A region of correlated genes was defined as a group of genes with a significant transcriptome correlation score (indicating correlation) with no more than $n = 7$ noncorrelated genes between two correlated genes, corresponding to half the window size. Consequently, groups of genes with significant transcriptome correlation scores separated by more than seven noncorrelated genes were considered to constitute independent TCM regions. The TCM analysis of our bladder cancer data set identified 900 genes with a transcriptome correlation score over the significance threshold, constituting 206 regions of correlation. We excluded 70 regions containing only one gene from subsequent analysis.

Cell culture. NHU cells were cultured as a monolayer in keratinocyte serum-free medium supplemented with recombinant epithelial growth factor (rEGF), bovine pituitary extract and cholera toxin[41]. The TCCSUP bladder cancer cell line was cultured in DMEM F-12 Glutamax medium supplemented with 10% FCS and HEPES (10 mM). Normal and tumor cells were added to six-well plates at a density of 8×10^5 cells/well. They were treated the next day with 300 nM trichostatin A (TSA) (Calbiochem) for 16 h, 5 μM 5-aza-deoxycytidine (Calbiochem) for 48 h, or 5 μM 5-aza-deoxycytidine for 24 h followed by 300 nM TSA for 16 h.

Quantitative RT-PCR. We used 1 μg of total RNA for reverse transcription, with random hexamers (20 pmol) and 200 U Moloney murine leukemia virus (MMLV) reverse transcriptase. The levels of mRNA for various genes showing differential expression were determined by real-time PCR using 18S rRNA as an internal standard. PCR was performed in an ABI PRISM 7700 real-time thermal cycler, using the SYBR Green kit (Applied Biosystems) for the genes of interest and the TaqMan kit for 18S rRNA. The sequences of the various primers used are shown in **Supplementary Table 3** online. PCR conditions were as follows: denaturation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s and annealing/extension at 60 °C for 1 min. A dissociation curve was produced at the end of each PCR run to check that only one fragment was amplified. Each sample was studied in duplicate and a negative control (an RNA sample not subjected to reverse transcription) was included in every assay. C_T (the threshold cycle) was defined as the cycle number at which the fluorescence generated by the incorporation of SYBR green exceeded a predetermined threshold. For a sample, the level of mRNA for a gene of interest was therefore calculated as follows: $2^{(C_T(18S\ rRNA)-C_T(gene\ mRNA))}$.

Methylation analysis. We investigated the methylation status of the six promoters (*CTDSPL*, *PLCD1*, *DLEC1*, *ACAA1*, *MYD88* and *OXSR1*) by combined bisulfite restriction analysis and COBRA[42]. Briefly, we treated 2 μg of genomic DNA with sodium bisulfite (Acros Organics) for 16 h, converting the nonmethylated cytosines to thymines. The DNA was purified and an aliquot of 1 μl was used as the template in each PCR: initial incubation at 94 °C for 4 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at T_m for 30 s and extension at 72 °C for 30 s using Biolabs Taq Polymerase (Ozyme). The PCR products were then digested for 16 h with a restriction enzyme containing a cytosine in its restriction site. The various primer sets, T_m values and enzymes are listed in **Supplementary Table 3** online.

ChIP. ChIP assays[43] were carried out on three 150 cm^2 dishes for NHU and TCCSUP cells. Chromatin was prepared using an enzymatic kit (Active Motif). An extract of the original chromatin was kept as an internal standard (input DNA). The complexes were immunoprecipitated with 5 μg of an antibody specific for trimethyl-histone H3(Lys9) (Upstate Biotechnology). We included a control antibody for each ChIP assay. The amount of immunoprecipitated target was measured by real-time PCR, in duplicate, using the ABI PRISM 7900HT Sequence Detection System. The primers used for amplification are listed in **Supplementary Table 3** online. For each sample and each promoter, an average C_T value was obtained for immunoprecipitated material and for the input chromatin. The amount of immunoprecipitated material was defined as $2^{(C_T(input\ DNA)-C_T(immunoprecipitated\ DNA))}$.

Quantitative PCR for DNA copy number assessment. Quantitative PCR was performed on seven DNA tumor samples presenting downregulation of the genes of the 3-2 region at 3p22.3 but no genetic alteration in this region, as assessed by CGH array. This analysis was used to determine whether certain losses might have gone undetected in CGH array analysis. We measured the DNA copy number of the 3-2 region in 10-ng DNA samples and compared it with that of the *GAPDH* gene, which is located in a region not reported to be genetically altered in bladder cancer. Two primer sets were designed, one amplifying a fragment of the *PLCD1* gene, located in the 3-2 region and the other amplifying *GAPDH* (**Supplementary Table 3**). The relative DNA copy number N of the region in a given sample was then defined as $N = 2^{(C_T(GAPDH)-C_T(region\ 3-2))}$. We compared the results for N in tumors with results from a normal blood sample.

Statistics. Transcriptome correlation scores were calculated using Spearman's rank correlation analysis of the gene expression profiles for the 57 samples. The significance threshold for transcriptome correlation scores was set as the 499th quantile of the distribution of randomly generated transcriptome correlation scores. Genes showing significant differential expression (upregulation or downregulation) were identified by comparing the 57 tumor samples with the five normal urothelium samples using Wilcoxon rank sum tests and a P value of 0.05. The properties of the copy number–dependent and copy number–independent regions were compared using χ^2 tests. The Affymetrix signal and CGH $\log_2$ ratio were compared using Pearson correlation tests. The box plots (**Fig. 5b**) represent the data distribution. The horizontal line in the middle box represents the median, the two hinges of the middle box represent

the first and third interquartiles and the whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range from the box.

Public databases and software. The public databases used in this study were the University of California Santa Cruz (UCSC) Human Genome reference sequence and the annotation database from the May 2004 freeze (hg17), the National Center for Biotechnology Information (NCBI) *Homo sapiens* genome view build 35 and the *Homo sapiens* UniGene Build #184. The genomic positions of Affymetrix probe sets were obtained using a BLAT program[44] for sequence-matching searches of probe set–specific target sequences against the UCSC reference sequence. The genomic positions of the CGH BAC clones were obtained by looking for their accessions, STS or BAC End IDs in the UCSC annotation database tables and in NCBI genome view tables. Based on BAC positions and probe set positions, we identified lists of probe sets located within BACs or between two BACs.

Statistical analysis and numerical calculations were carried out with R 2.1.1 (ref. 45). Gene and BAC positions were determined from public databases, and the data were formatted with HKIS software.

URLs. Protocols relevant to Affymetrix microarrays can be obtained from http://www.affymetrix.com/. UCSC Human Genome reference sequence and annotation database were obtained from http://hgdownload.cse.ucsc.edu/goldenPath/hg17/chromosomes/ and http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/. The NCBI *Homo sapiens* genome view and UniGene annotation databases were obtained from ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/ARCHIVE/BUILD.35.1 and ftp://ftp.ncbi.nih.gov/repository/UniGene/Homo_sapiens/. Information about the R project can be found at http://www.R-project.org/, and information about HKIS can be found at http://isoft.free.fr/hkis/.

Accession codes. The raw and normalized Affymetrix microarray data files are available online at http://www.ebi.ac.uk/arrayexpress/index.html under accession number E-TABM-147 and http://microarrays.curie.fr/publications/oncologie_moleculaire/bladder_TCM/. The raw and normalized CGH data sets are also available at http://microarrays.curie.fr/publications/oncologie_moleculaire/bladder_TCM/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank C. Rouveirol for discussions, P. Hupé for his GLAD algorithm expertise and the Institut Curie Bioinformatics Service headed by E. Barillot for support. We also thank J. Sappa from Alex Edelman & Associates for careful reading of the manuscript and the UCSF Cancer Center Array CGH Core for providing the BAC arrays. This article is dedicated to the memory of D. Chopin, whose commitment to cancer research was of paramount importance for the initiation of this work. This work was supported by the CNRS, the Institut Curie, AstraZeneca, the Canceropole Ile de France and the Ligue Nationale Contre le Cancer. N.S., C.V., F. Reyal, I.B.-P., S.G.D. de M. and F. Radvanyi are members of the Equipe Oncologie Moléculaire, labellisée par La Ligue Nationale Contre le Cancer. N.S. was supported by a fellowship from the French Ministry of Education and Research and a fellowship from the Association pour la Recherche sur le Cancer. C.V. was supported by a fellowship from the French Ministry of Education and Research and F. Reyal by a fellowship from the Ligue Nationale Contre le Cancer.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Solinas-Toldo, S. *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosom. Cancer* **20**, 399–407 (1997).
2. Pinkel, D. *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20**, 207–211 (1998).
3. Hughes, T.R. *et al.* Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.* **25**, 333–337 (2000).
4. Hyman, E. *et al.* Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.* **62**, 6240–6245 (2002).
5. Pollack, J.R. *et al.* Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. USA* **99**, 12963–12968 (2002).
6. Masayesva, B.G. *et al.* Gene expression alterations over large chromosomal regions in cancers include multiple genes unrelated to malignant progression. *Proc. Natl. Acad. Sci. USA* **101**, 8715–8720 (2004).
7. Heidenblad, M. *et al.* Microarray analyses reveal strong influence of DNA copy number alterations on the transcriptional patterns in pancreatic cancer: implications for the interpretation of genomic amplifications. *Oncogene* **24**, 1794–1801 (2005).
8. Crawley, J.J. & Furge, K.A. Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression microarray data. *Genome Biol.* **3**, RESEARCH0075 (2002).
9. Zhou, Y. *et al.* Genome-wide identification of chromosomal regions of increased tumor expression by transcriptome analysis. *Cancer Res.* **63**, 5781–5784 (2003).
10. Kano, M. *et al.* Expression imbalance map: a new visualization method for detection of mRNA expression imbalance regions. *Physiol. Genomics* **13**, 31–46 (2003).
11. Midorikawa, Y. *et al.* Distinct chromosomal bias of gene expression signatures in the progression of hepatocellular carcinoma. *Cancer Res.* **64**, 7263–7270 (2004).
12. Furge, K.A., Dykema, K.J., Ho, C. & Chen, X. Comparison of array-based comparative genomic hybridization with gene expression-based regional expression biases to identify genetic abnormalities in hepatocellular carcinoma. *BMC Genomics* **6**, 67 (2005).
13. Yang, X.J. *et al.* A molecular classification of papillary renal cell carcinoma. *Cancer Res.* **65**, 5628–5637 (2005).
14. Fujii, T. *et al.* A preliminary transcriptome map of non-small cell lung cancer. *Cancer Res.* **62**, 3340–3346 (2002).
15. Cohen, B.A., Mitra, R.D., Hughes, J.D. & Church, G.M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* **26**, 183–186 (2000).
16. Spellman, P.T. & Rubin, G.M. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1**, 5 (2002).
17. Reyal, F. *et al.* Visualizing chromosomes as transcriptome correlation maps: evidence of chromosomal domains containing co-expressed genes–a study of 130 invasive ductal breast carcinomas. *Cancer Res.* **65**, 1376–1383 (2005).
18. Yi, Y., Mirosevich, J., Shyr, Y., Matusik, R. & George, A.L., Jr. Coupled analysis of gene expression and chromosomal location. *Genomics* **85**, 401–412 (2005).
19. Frigola, J. *et al.* Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat. Genet.* **38**, 540–549 (2006).
20. Turner, B.M. Cellular memory and the histone code. *Cell* **111**, 285–291 (2002).
21. Nguyen, C.T. *et al.* Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine. *Cancer Res.* **62**, 6456–6461 (2002).
22. Daigo, Y. *et al.* Molecular cloning of a candidate tumor suppressor gene, DLC1, from chromosome 3p21.3. *Cancer Res.* **59**, 1966–1972 (1999).
23. Nakamura, Y. *et al.* Phospholipase Cdelta1 is required for skin stem cell lineage commitment. *EMBO J.* **22**, 2981–2991 (2003).
24. Chambeyron, S. & Bickmore, W.A. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
25. Huebert, D.J. & Bernstein, B.E. Genomic views of chromatin. *Curr. Opin. Genet. Dev.* **15**, 476–481 (2005).
26. Gialitakis, M. *et al.* Coordinated changes of histone modifications and HDAC mobilization regulate the induction of MHC class II genes by trichostatin A. *Nucleic Acids Res.* **34**, 765–772 (2006).
27. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
28. van Driel, R., Fransz, P.F. & Verschure, P.J. The eukaryotic genome: a system regulated at different hierarchical levels. *J. Cell Sci.* **116**, 4067–4075 (2003).
29. Zardo, G. *et al.* Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat. Genet.* **32**, 453–458 (2002).
30. Hurst, L.D., Pal, C. & Lercher, M.J. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299–310 (2004).
31. Sproul, D., Gilbert, N. & Bickmore, W.A. The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet.* **6**, 775–781 (2005).
32. Chang, H.Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, E7 (2004).
33. Diez de Medina, S.G. *et al.* Decreased expression of keratinocyte growth factor receptor in a subset of human transitional cell bladder carcinomas. *Oncogene* **14**, 323–330 (1997).
34. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. & Rutter, W.J. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299 (1979).
35. Coombs, L.M. *et al.* Simultaneous isolation of DNA, RNA, and antigenic protein exhibiting kinase activity from small tumor samples using guanidine isothiocyanate. *Anal. Biochem.* **188**, 338–343 (1990).
36. Labarca, C. & Paigen, K. A simple, rapid, and sensitive DNA assay procedure. *Anal. Biochem.* **102**, 344–352 (1980).

37. Snijders, A.M. *et al.* Genome-wide-array-based comparative genomic hybridization reveals genetic homogeneity and frequent copy number increases encompassing CCNE1 in fallopian tube carcinoma. *Oncogene* **22**, 4281–4286 (2003).
38. Jain, A.N. *et al.* Fully automatic quantification of microarray image data. *Genome Res.* **12**, 325–332 (2002).
39. Neuvial, P. *et al.* Spatial normalization of array-CGH data. *BMC Bioinformatics* **7**, 264 (2006).
40. Hupé, P., Stransky, N., Thiery, J.P., Radvanyi, F. & Barillot, E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* **20**, 3413–3422 (2004).
41. Southgate, J., Hutton, K.A., Thomas, D.F. & Trejdosiewicz, L.K. Normal human urothelial cells in vitro: proliferation and induction of stratification. *Lab. Invest.* **71**, 583–594 (1994).
42. Xiong, Z. & Laird, P.W. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25**, 2532–2534 (1997).
43. O'Neill, L.P. & Turner, B.M. Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76–82 (2003).
44. Kent, W.J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
45. R development core team *R: a Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, Vienna, 2006).

Corrigendum: Regional copy number–independent deregulation of transcription in cancer

Nicolas Stransky, Céline Vallot, Fabien Reyal, Isabelle Bernard-Pierrot, Sixtina Gil Diez de Medina, Rick Segraves, Yann de Rycke, Paul Elvin, Andrew Cassidy, Carolyn Spraggon, Alexander Graham, Jennifer Southgate, Bernard Asselain, Yves Allory, Claude C Abbou, Donna G Albertson, Jean Paul Thiery, Dominique K Chopin, Daniel Pinkel & François Radvanyi

Nature Genetics **38, 1386–1396 (2006); published online 12 November 2006; corrected after print 27 February 2008**

In the version of this article initially published, the horizontal dashed lines representing the threshold value in the panels in row b of Figures 2 and 4 were incorrectly placed. The errors have been corrected in the HTML and PDF versions of this article. See below for the corrected version of the figures.

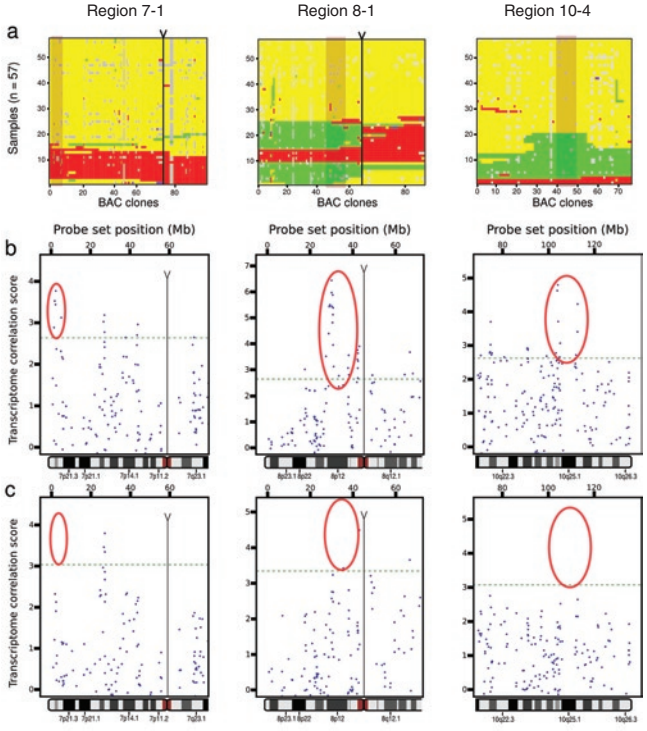

Figure 2

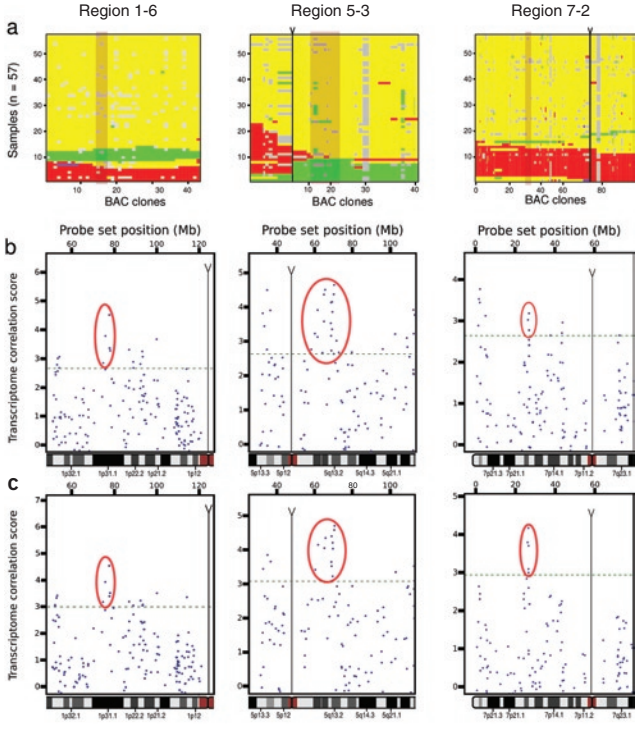

Figure 4