Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors

Giuseppe Zardo[1,2]*, Maarit I. Tiirikainen[1,2]*, Chibo Hong[1,2], Anjan Misra[1–3], Burt G. Feuerstein[1–3], Stanislav Volik[1,4], Colin C. Collins[1,4], Kathleen R. Lamborn[1,2], Andrew Bollen[1,5], Daniel Pinkel[1,3], Donna G. Albertson[1,4] & Joseph F. Costello[1,2]

*These authors contributed equally to this work.

Aberrant methylation of CpG islands and genomic deletion are two predominant mechanisms of gene inactivation in tumorigenesis, but the extent to which they interact is largely unknown. The lack of an integrated approach to study these mechanisms has limited the understanding of tumor genomes and cancer genes. Restriction landmark genomic scanning (RLGS; ref. 1) is useful for global analysis of aberrant methylation of CpG islands, but has not been amenable to alignment with deletion maps because the identity of most RLGS fragments is unknown. Here, we determined the nucleotide sequence and exact chromosomal position of RLGS fragments throughout the genome using the whole chromosome of origin of the fragments[2] and *in silico* restriction digestion of the human genome sequence. To study the interaction of these gene-inactivation mechanisms in primary brain tumors, we integrated RLGS-based methylation analysis with high-resolution deletion maps from microarray-based comparative genomic hybridization (array CGH; ref. 3). Certain subsets of gene-associated CpG islands were preferentially affected by convergent methylation and deletion, including genes that exhibit tumor-suppressor activity, such as *CISH1* (encoding *SOCS1*; ref. 4), as well as genes such as *COE3* that have been missed by traditional non-integrated approaches. Our results show that most aberrant methylation events are focal and independent of deletions, and the rare convergence of these mechanisms can pinpoint biallelic gene inactivation without the use of positional cloning.
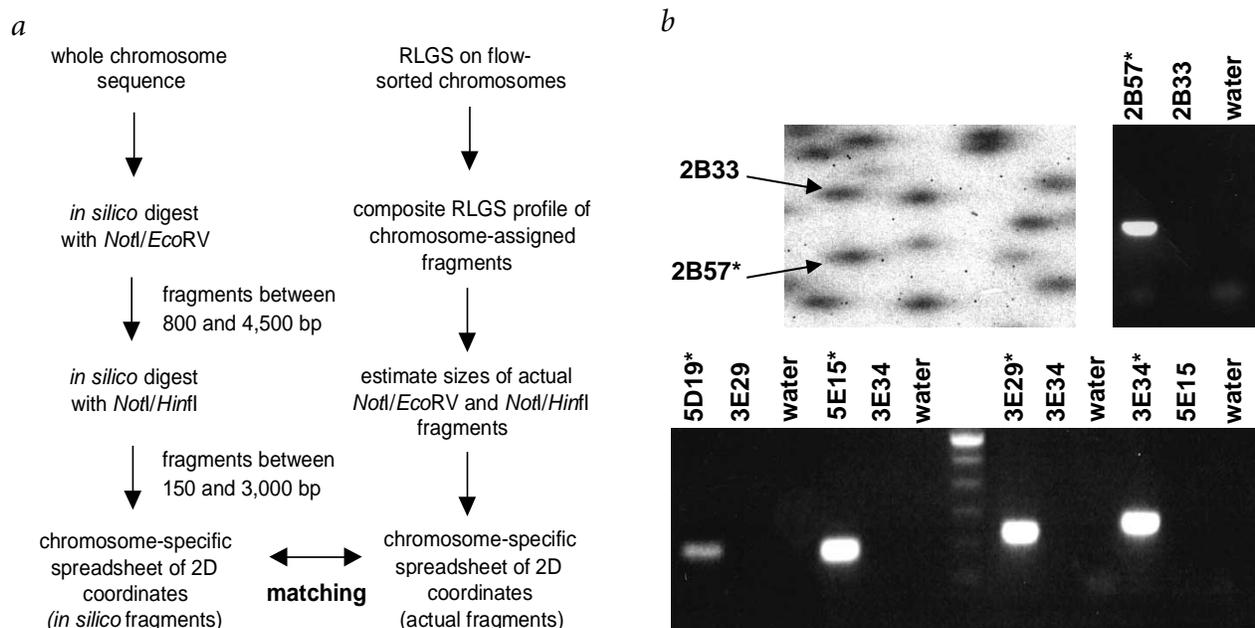
Fig. 1 Identification of the sequences and precise chromosomal localizations of RLGS fragments. *a*, Strategy for matching *in silico* *Not*I fragments generated from the human genome sequence (left side) to actual RLGS fragments (right side) on the basis of chromosome of origin and estimated two-dimensional size of each fragment. *b*, PCR confirmation of matches, using primers designed from the sequence of the predicted matched fragment. DNA templates for PCR were individual fragments excised from the RLGS gel (top, left) and included the putative matched fragment (marked with an asterisk) and unrelated negative control fragments or water only (top, right). Lower panel, PCR products confirming four additional predicted matched fragments.

[1]Comprehensive Cancer Center, [2]The Brain Tumor Research Center and Department of Neurological Surgery, [3]Department of Laboratory Medicine, [4]Cancer Research Institute and [5]Department of Pathology, University of California, San Francisco, San Francisco, California 94115, USA. Correspondence should be addressed to J.C. (e-mail: jcostello@cc.ucsf.edu).
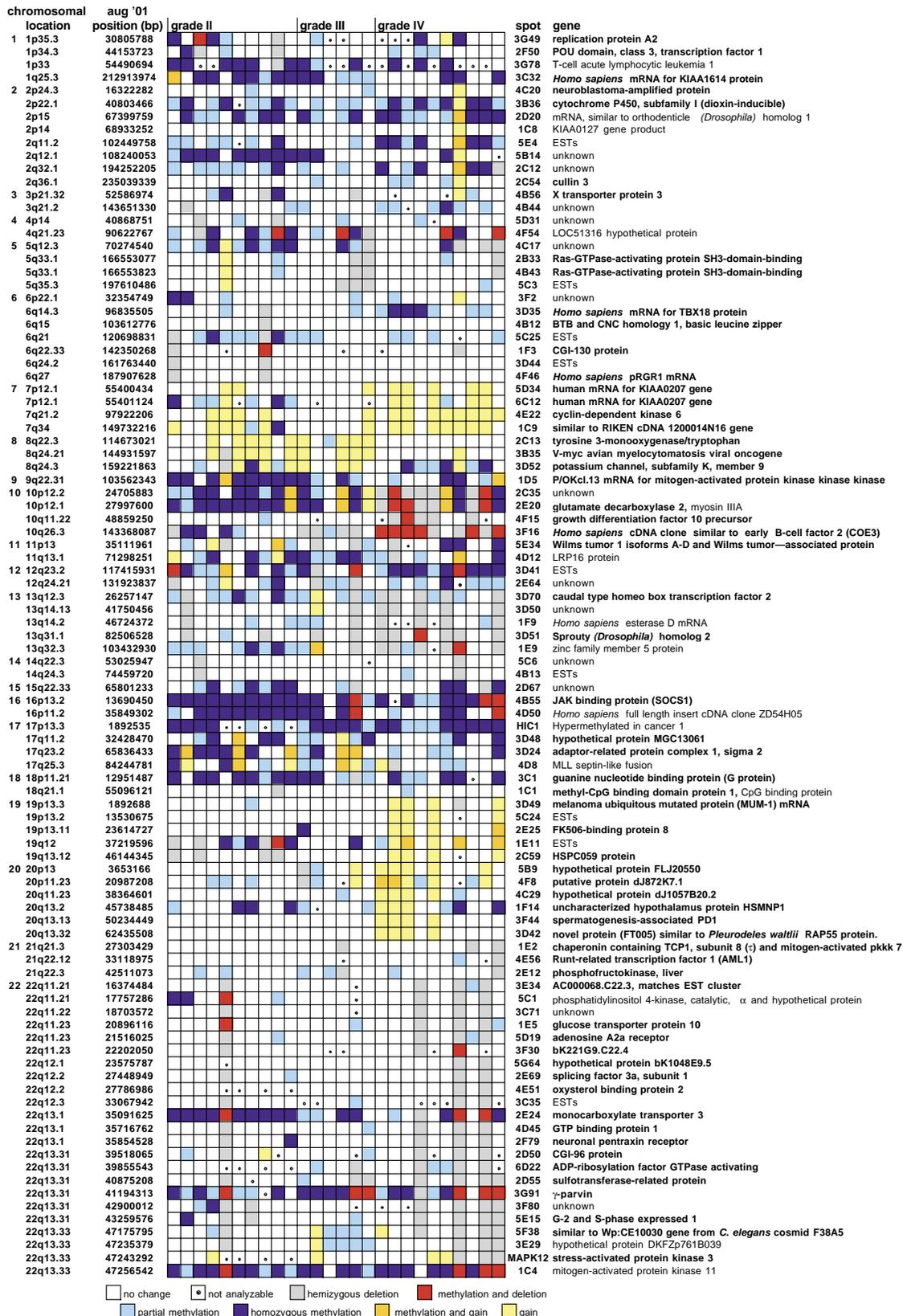
Fig. 2 data — Aberrant methylation and gain or loss of copy number in 26 human gliomas. (The grade II / grade III / grade IV sample grid is a colored data matrix; the text columns are transcribed below.)

#	chromosomal location	aug '01 position (bp)	spot	gene
1	1p35.3	30805788	3G49	replication protein A2
	1p34.3	44153723	2F50	POU domain, class 3, transcription factor 1
	1p33	54490694	3G78	T-cell acute lymphocytic leukemia 1
	1q25.3	212913974	3C32	*Homo sapiens* mRNA for KIAA1614 protein
2	2p24.3	16322282	4C20	neuroblastoma-amplified protein
	2p22.1	40803466	3B36	cytochrome P450, subfamily I (dioxin-inducible)
	2p15	67399759	2D20	mRNA, similar to orthodenticle *(Drosophila)* homolog 1
	2p14	68933252	1C8	KIAA0127 gene product
	2q11.2	102449758	5E4	ESTs
	2q12.1	108240053	5B14	unknown
	2q32.1	194252205	2C12	unknown
	2q36.1	235039339	2C54	cullin 3
3	3p21.32	52586974	4B56	X transporter protein 3
	3q21.2	143651330	4B44	unknown
4	4p14	40868751	5D31	unknown
	4q21.23	90622767	4F54	LOC51316 hypothetical protein
5	5q12.3	70274540	4C17	unknown
	5q33.1	166553077	2B33	Ras-GTPase-activating protein SH3-domain-binding
	5q33.1	166553823	4B43	Ras-GTPase-activating protein SH3-domain-binding
	5q35.3	197610486	5C3	ESTs
6	6p22.1	32354749	3F2	unknown
	6q14.3	96835505	3D35	*Homo sapiens* mRNA for TBX18 protein
	6q15	103612776	4B12	BTB and CNC homology 1, basic leucine zipper
	6q21	120698831	5C25	ESTs
	6q22.33	142350268	1F3	CGI-130 protein
	6q24.2	161763440	3D44	ESTs
	6q27	187907628	4F46	*Homo sapiens* pRGR1 mRNA
7	7p12.1	55400434	5D34	human mRNA for KIAA0207 gene
	7p12.1	55401124	6C12	human mRNA for KIAA0207 gene
	7q21.2	97922206	4E22	cyclin-dependent kinase 6
	7q34	149732216	1C9	similar to RIKEN cDNA 1200014N16 gene
8	8q22.3	114673021	2C13	tyrosine 3-monooxygenase/tryptophan
	8q24.21	144931597	3B35	V-myc avian myelocytomatosis viral oncogene
	8q24.3	159221863	3D52	potassium channel, subfamily K, member 9
9	9q22.31	103562343	1D5	P/OKcl.13 mRNA for mitogen-activated protein kinase kinase kinase
10	10p12.2	24705883	2C35	unknown
	10p12.1	27997600	2E20	glutamate decarboxylase 2, myosin IIIA
	10q11.22	48859250	4F15	growth differentiation factor 10 precursor
	10q26.3	143368087	3F16	*Homo sapiens* cDNA clone similar to early B-cell factor 2 (COE3)
11	11p13	35111961	5E34	Wilms tumor 1 isoforms A-D and Wilms tumor—associated protein
	11q13.1	71298251	4D12	LRP16 protein
12	12q23.2	117415931	3D41	ESTs
	12q24.21	131923837	2E64	unknown
13	13q12.3	26257147	3D70	caudal type homeo box transcription factor 2
	13q14.13	41750456	3D50	unknown
	13q14.2	46724372	1F9	*Homo sapiens* esterase D mRNA
	13q31.1	82506528	3D51	Sprouty *(Drosophila)* homolog 2
	13q32.3	103432930	1E9	zinc family member 5 protein
14	14q22.3	53025947	5C6	unknown
	14q24.3	74459720	4B13	ESTs
15	15q22.33	65801233	2D67	unknown
16	16p13.2	13690450	4B55	JAK binding protein (SOCS1)
	16p11.2	35849302	4D50	*Homo sapiens* full length insert cDNA clone ZD54H05
17	17p13.3	1892535	HIC1	Hypermethylated in cancer 1
	17q11.2	32428470	3D48	hypothetical protein MGC13061
	17q23.2	65836433	3D24	adaptor-related protein complex 1, sigma 2
	17q25.3	84244781	4D8	MLL septin-like fusion
18	18p11.21	12951487	3C1	guanine nucleotide binding protein (G protein)
	18q21.1	55096121	1C1	methyl-CpG binding domain protein 1, CpG binding protein
19	19p13.3	1892688	3D49	melanoma ubiquitous mutated protein (MUM-1) mRNA
	19p13.2	13530675	5C24	ESTs
	19p13.11	23614727	2E25	FK506-binding protein 8
	19q12	37219596	1E11	ESTs
	19q13.12	46144345	2C59	HSPC059 protein
20	20p13	3653166	5B9	hypothetical protein FLJ20550
	20p11.23	20987208	4F8	putative protein dJ872K7.1
	20q11.23	38364601	4C29	hypothetical protein dJ1057B20.2
	20q13.2	45738485	1F14	uncharacterized hypothalamus protein HSMNP1
	20q13.13	50234449	3F44	spermatogenesis-associated PD1
	20q13.32	62435508	3D42	novel protein (FT005) similar to *Pleurodeles waltlii* RAP55 protein.
21	21q21.3	27303429	1E2	chaperonin containing TCP1, subunit 8 (τ) and mitogen-activated pkkk 7
	21q22.12	33118975	4E56	Runt-related transcription factor 1 (AML1)
	21q22.3	42511073	2E12	phosphofructokinase, liver
22	22q11.21	16374484	3E34	AC000068.C22.3, matches EST cluster
	22q11.21	17757286	5C1	phosphatidylinositol 4-kinase, catalytic, α and hypothetical protein
	22q11.22	18703572	3C71	unknown
	22q11.23	20896116	1E5	glucose transporter protein 10
	22q11.23	21516025	5D19	adenosine A2a receptor
	22q11.23	22202050	3F30	bK221G9.C22.4
	22q12.1	23575787	5G64	hypothetical protein bK1048E9.5
	22q12.2	27448949	2E69	splicing factor 3a, subunit 1
	22q12.2	27786986	4E51	oxysterol binding protein 2
	22q12.3	33067942	3C35	ESTs
	22q13.1	35091625	2E24	monocarboxylate transporter 3
	22q13.1	35716762	4D45	GTP binding protein 1
	22q13.1	35854528	2F79	neuronal pentraxin receptor
	22q13.31	39518065	2D50	CGI-96 protein
	22q13.31	39855543	6D22	ADP-ribosylation factor GTPase activating
	22q13.31	40875208	2D55	sulfotransferase-related protein
	22q13.31	41194313	3G91	γ-parvin
	22q13.31	42900012	3F80	unknown
	22q13.31	43259576	5E15	G-2 and S-phase expressed 1
	22q13.33	47175795	5F38	similar to Wp:CE10030 gene from *C. elegans* cosmid F38A5
	22q13.33	47235379	3E29	hypothetical protein DKFZp761B039
	22q13.33	47243292	MAPK12	stress-activated protein kinase 3
	22q13.33	47256542	1C4	mitogen-activated protein kinase 11

Legend:
- no change
- not analyzable
- hemizygous deletion
- methylation and deletion
- partial methylation
- homozygous methylation
- methylation and gain
- gain

Fig. 2 Aberrant methylation and gain or loss of copy number in 26 human gliomas. Gene names in bold indicate that the *Not*I fragment (spot) is in the 5′ region. The two genes with the highest incidence of convergent methylation and deletion were 5′ CpG islands 3G91 and 3F16. The total number of methylation-and-deletion (MD) events was similar between putative 5′ and non-5′ sites, with 60% (27 of 45) of MD events in 5′ CpG islands and 40% (18 of 45 events) in non-5′ CpG islands. RLGS spots were given three-variable coordinates for identification purposes (see Web Fig. A online). Fragments that were potentially imprinted or commonly polymorphic at the restriction sites (fragments that vary between 'diploid intensity', 50% reduction and complete absence of a spot) in normal tissue were excluded from the analysis. The samples, from left to right, were: grade II; 1A, 2A, #6, #7, #8, #20, #22, #23, #24, #25, grade III: #1, #2, #4, #7, #8, #10, grade IV; #1–9, #15.

An integrated approach for mapping genomic and epigenomic alterations across tumor genomes is crucial for identifying candidate tumor-suppressor genes. Identifying such genes by either deletion mapping or methylation analysis alone has been problematic, and most work has focused on single regions identified by positional cloning and evaluation of candidate genes. These studies have established that methylation and deletion do converge on individual tumor-suppressor genes[5,6]. To develop a genome-wide approach for mapping genomic and epigenomic alterations, we combined RLGS, a two-dimensional gel-based method of analyzing the methylation status of thousands of genes containing CpG islands[1], with high-resolution analysis of DNA copy number, array CGH (refs 3,7) and quantitative microsatellite analysis (QuMA; ref. 8). In RLGS, DNA is cleaved with the methylation-sensitive enzyme *Not*I and with *Eco*RV and radioactively labeled at the cleaved *Not*I sites. After size separation in the first dimension, DNA is further digested with *Hinf*I and separated by electrophoresis in a second dimension to generate a reproducible profile of more than 1,000 spots (see Web Fig. A online). The intensity of each spot reflects the methylation status and copy number of the fragment[9–13]. Because the identities of most of the RLGS loci are unknown, precluding their alignment with copy-number maps of chromosomes, we developed a custom PERL program to generate RLGS digestion fragments from the human genome sequence, similar to an informatics tool termed Virtual Genome Scan[14]. We matched *in silico* restriction fragments with actual RLGS fragments on the basis of their two-dimensional fragment size and chromosome of origin[2] and confirmed each match by PCR (Fig. 1). Here, we applied this RLGS map, comprised of all RLGS fragments from chromosome 22 and 2–8 RLGS loci from all other somatic chromosomes, to the analysis of brain tumors (Fig. 2).

To measure the interaction of methylation and deletion at the 96 loci, we analyzed DNA from 26 human brain tumors by RLGS, array CGH and QuMA (see Web Table A online for QuMA results). We attributed RLGS fragment loss (decreased intensity) to (i) partial methylation, if the decrease in *Not*I fragment intensity was at least 30% but less than 70%, relative to normal brain, and deletions were not observed; (ii) homozygous methylation, if the decrease in *Not*I intensity exceeded 70% and no deletions were observed; (iii) deletion, if the *Not*I fragment intensity was decreased by approximately 50%, and deletion was observed; or (iv) methylation and deletion, if the decrease in *Not*I fragment

intensity was more than 70%, and hemizygous deletion was indicated. Using these criteria, we found that aberrant CpG island methylation rarely occurred together with deletion (45 events in 2,406 sites, 1.8%, or, excluding sites with no alteration, 45 events in 989 sites, 4.6%; Fig. 2). For ten loci, however, we detected recurrent convergence of deletion and methylation. If these patterns of biallelic alteration were random, it is unlikely that the events would have a functional impact on tumorigenesis. But the pattern of convergence is non-random ($P < 0.001$; Fig. 3a), suggesting that some of these events arise through growth selection, increased locus-specific susceptibility to deletion and methylation, or both.

We also examined the pattern of gain-and-methylation events in these tumors (Fig. 3b). The analysis indicated that a subset of CpG islands were preferentially altered by such events ($P < 0.001$). The data suggest that some aberrant CpG island methylation may mitigate the effect of copy-number gain in tumors, although this remains to be tested.

We found that the RLGS fragment 4B55, corresponding to *CISH1* (also known as *SOCS1*) at 16p13.2 (refs 15,16), was aberrantly methylated on both alleles in 18 of 26 tumors (Fig. 2). In a further three tumors, including one grade IV tumor with a relatively small deletion (approximately 3.16 Mb), the locus was both methylated and deleted. Our genomic and epigenomic analyses indicating that the locus is affected in almost all brain tumors in this small set of samples, together with other observations identifying methylation of *CISH1* in hepatocellular carcinoma and functional assays implicating CISH1 in growth control[4], implicates *CISH1* as a candidate tumor-suppressor gene.

We found a second locus, 3F16 at 10q26.3 (a region frequently deleted in grade IV tumors[17–20]), to be aberrantly methylated and deleted in our grade IV tumors and homozygously methylated in grade II tumors (Fig. 4a). This fragment maps to the 5′ end of the putative *COE3*, a member of the COE family of transcription factors that regulate neurogenesis and differentiation[21]. *COE3* is expressed in normal adult brain but repressed in four glioma cell lines (Fig. 4b). We were able to reactivate *COE3* expression by treating cells with the demethylating agent 5-aza-2-deoxycytidine, suggesting that aberrant methylation caused silencing of *COE3*. Consistent with this, a more in-depth bisulfite-sequencing analysis showed that *COE3* was extensively methylated in cells with little or no *COE3* expression and was unmethylated in normal brain (Fig. 4c).
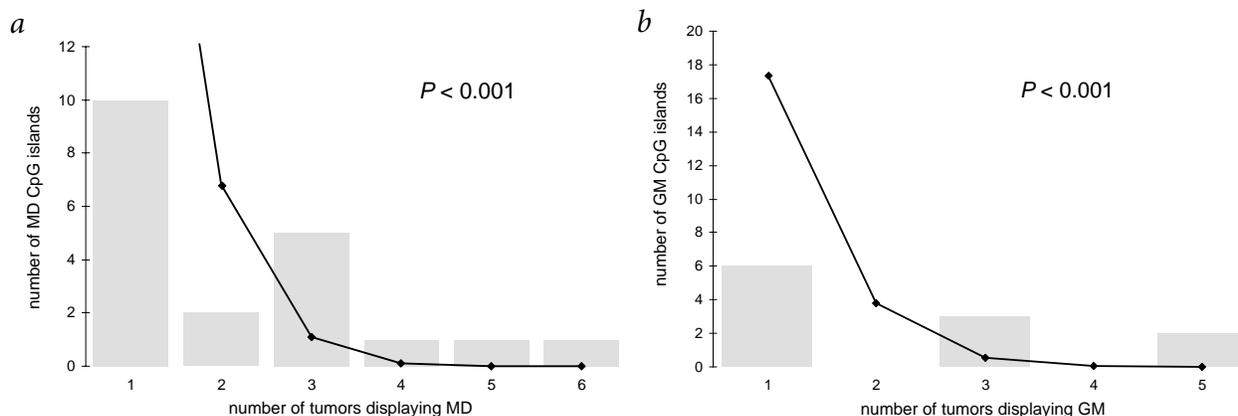
a

b

Fig. 3 Subsets of CpG islands were preferentially altered by converging genomic and epigenomic mechanisms. The histograms display the number of tumors (of 26 total) in which the particular CpG islands were either (***a***) methylated and deleted (MD events) or (***b***) gained and methylated (GM events). The line plot in each panel represents the expected distribution assuming that all CpG islands have an equal frequency of alteration. For example, there are five CpG islands that exhibit MD in three of the tumors, whereas from a random pattern only one such CpG island would be expected. *P* values indicate that the observed and expected frequencies of alteration were significantly different from one another.

a

alteration	frequency			
	II	**III**	**IV**	**all**
	n = 10	*n* = 6	*n* = 10	*n* = 26
O	5	1	1	7
MD	0	0	6	6
MM	3	0	0	3
M	1	1	0	2
D	1	4	3	8

b

d

c

Fig. 4 Aberrant methylation and deletion at putative *COE3* (fragment 3F16). *a*, RLGS profiles highlighting the *COE3* *Not*I fragment (→) along with the *COE3* locus (↓) on the copy-number plot from the same tumor. Lower panel, summary of *COE3* alterations, grades II–IV. O, no alteration; M, partial methylation; MM, homozygous methylation; D, hemizygous deletion; MD, methylation and deletion. *b*, RT–PCR analysis of *COE3* and *ACTB* expression in four human glioma cell lines. un, untreated cells; A1, A2, independent treatments of cells with 1 μM 5-aza-2-deoxycytidine for 3 d. The *COE3* status prior to demethylation treatment is listed. *c*, Analysis of the methylation status of *COE3* in normal brain (NB3, NB7, NB8) and brain tumor cell lines determined by the bisulfite-sequencing method. Each row of circles represents sequence analysis of an individual clone from the PCR product (open circle, unmethylated CpG; filled circle, methylated CpG). CpG 41 and 42 are within the original *Not*I site. *d*, RT–PCR analysis of *COE3* and *ACTB* expression in normal brain from two different individuals (N#4, NB8) and in primary brain tumors with and without *COE3* alteration. The graphs display the statistically significant difference ($P = 0.013$) in average and standard deviation of the amount of RT–PCR product (ratio of *COE3* to *ACTB*) in tumors with alteration (MM/MD) relative to that in normal brain and tumors without alteration (O). The lower panel shows an independent analysis of the samples without *COE3* alteration along with an additional grade IV brain tumor (J1) exhibiting methylation and deletion of *COE3*. We expected that all samples would exhibit some *COE3* expression, as each contains a fraction of non-tumor cells, particularly blood cells that express *COE3* at a significant level (see Web Fig. C online).

COE3 expression was also significantly lower ($P = 0.013$) in primary brain tumors with biallelic alteration (Fig. 4*d*). Thus, *COE3* is a novel target of biallelic inactivation in brain tumors and has been missed by traditional approaches. This raises the possibility that there may be additional undetected genes with a similar pattern of biallelic inactivation in tumors.

In summary, our integrated approach addresses a large gap in the understanding of tumor genomes. Mapping RLGS fragments *in silico* is widely applicable to all chromosomes and enzyme combinations. These maps will facilitate a more comprehensive assessment of DNA copy number and methylation and provide a new capability for genome-wide analyses of methylation and gene expression. In contrast with candidate-gene studies, our global approach suggests that aberrant methylation and deletion primarily affect separate sets of genes in brain tumors.

Methods

Tumor samples. We obtained 26 glioma samples and 11 non-tumor brain samples from the Neurosurgery Tumor Bank at the University of California, San Francisco. The samples included ten grade II gliomas (as graded by the World Health Organization (WHO)), six grade III anaplastic astrocytomas and ten grade IV glioblastomas (eight primary and two recurrent; Fig. 2). We obtained 11 non-tumor brain samples from different individuals, including two samples of brain adjacent to tumor (BAT4, BAT6), one autopsy sample (N#4) and eight surgical samples from individuals with epilepsy (NB1–8). Samples were obtained with informed consent and their use was approved by the Committee on Human Research at University of California, San Francisco.

RLGS. We did RLGS according to published protocols[1]. Briefly, non-specific sheared ends of 1–3 μg of genomic DNA were blocked in a 10-μl reaction by the addition of nucleotide analogs (αS-dGTP, αS-dCTP, ddATP, ddTTP) with 2 U DNA polymerase I (37 °C, 20 min) followed by enzyme inactivation (65 °C, 30 min). We adjusted the buffer and digested DNA (37 °C, 2 h) with 20 U *Not*I (Promega), which is sensitive to methylation. We then used Sequenase (version 2.O, U.S.B.) to fill in the *Not*I ends with [α-^{32}P]dGTP and [α-^{32}P]dCTP (Amersham) for 30 min at 37 °C. We digested the labeled DNA (37 °C, 1 h) with 20 U *Eco*RV (Promega) and separated a portion by electrophoresis through a 60-cm long, 0.8% agarose tube gel (first-dimension separation). The agarose gel was next equilibrated in *Hin*fI digestion buffer, and the DNA was digested in the gel with 700 U *Hin*fI (Promega) at 37 °C for 2 h. We then placed the agarose gel horizontally (rotated 90° relative to the first direction of electrophoresis) across the top of a non-denaturing 5% polyacrylamide gel, connected the two gels with molten agarose and separated the DNA by electrophoresis in the second dimension. We dried the gels and exposed them to X-ray film in the presence of intensifying screens (Quanta III, DuPont) for 2–10 d. We carried out RLGS at least twice for each tumor, except for three of the grade III samples.

RLGS coordinate system. Each radioactively labeled *Not*I fragment displayed on the typical *Not*I/*Eco*RV (first-dimension) and *Not*I/*Hin*fI (second-dimension) profile was assigned a three-variable designation consisting of its *y* coordinate, *x* coordinate and fragment number. For example, 3C21 represents *y* coordinate 3, *x* coordinate C and fragment number 21 (see Web Fig. A online).

RLGS profile analysis. We compared profiles of tumor and normal brain DNA (from three unrelated individuals, BAT4, BAT6, N#4) by visual inspection of overlaid autoradiographs and corroborated them using a two-dimensional gel analysis program (Non-linear Dynamics). To diminish the contribution of common restriction-site polymorphism to apparent spot loss, we selected only those spots that were consistently present in profiles from normal brain (three individuals), peripheral blood lymphocyte DNA (three individuals) and normal muscle (four individuals). This technique also distinguished methylation due to normal imprinting from abnormal methylation in the tumors. Our two-dimensional gel analysis program automatically detected, matched and quantified spots, and used 'mode of non-spot' for background correction. To normalize differences in DNA loading and autoradiography times, the program calculated the ratio of the

volume of each spot to the sum of the volumes of at least 900 single-copy spots in the normal and tumor profiles, and then calculated the ratio of these ratios. The program output graphically the results of the quantification (see Web Fig. B online). We used these software-derived values and visual inspection to determine the alteration status for each data point (as described earlier), with visual inspection overriding the software in numerous cases. The actual average spot intensity for the categories of alteration in the tumors were as follows: relative to normal brain (100%), the intensity for partial methylation was 50.7%, for methylation and deletion, 10.8%, for homozygous methylation, 16.7% and for hemizygous deletion, 65%. The high average for deletion events is a result of the conservative criteria required for an event to be considered methylation and deletion. Although some alterations were near the threshold values, and thus could be considered either deletion or methylation and deletion, the average values and distribution of intensities indicate that the vast majority were clearly classified as one type or the other. Criteria for the category of methylation and deletion were particularly stringent, requiring evidence of deletion by both array CGH and RLGS and also that approximately 40% of the remaining allele (20% of remaining 50%, or 70% spot loss in total) be methylated. For approximately 20% of spots, there was disagreement between spot intensity assessed visually and quantified with the software. This reflects the limitation of the software in accurately quantifying spots that are very close to other spots, diffuse spots or spots with a high local background. Visual assessment was more reliable in these cases. In no case was the software quantification used without visual assessment.

In silico digests. We used the August 2001 freeze of the University of California, Santa Cruz version of the draft assembly of human genome for the *in silico* restriction analysis. Briefly, we downloaded the whole-chromosome assemblies and analyzed them using a custom PERL program on a Sun Ultra 80 workstation. The program accepted as arguments the name of a FASTA sequence file to be analyzed and the names and recognition sites of the restriction endonucleases that were used for generation of experimental RLGS profiles. The results were output to a FASTA file with predicted fragment sequences, the two-dimensional size for each fragment and the nucleotide position along the chromosome. This tab-delimited file could be directly imported into commercial spreadsheet analysis programs such as Microsoft Excel. The program can be adjusted for any restriction enzyme combination, and can process the human genome in less than one hour, facilitating consecutive analysis of each version of the genome assembly as it becomes available. We generated similar spreadsheets of actual fragments by determining the whole chromosome of origin of each fragment from RLGS analysis of flow-sorted human chromosomes[2] and estimating the two-dimensional size of each fragment based on the exact known sizes of previously cloned fragments. We then matched the *in silico* fragments with the actual fragments based on two-dimensional size and chromosome of origin.

PCR confirmation of sequence assignments. We excised individual *Not*I fragments from an RLGS gel containing human genomic DNA. DNA was eluted in 0.5 M ammonium acetate, pH 8, 1 mM EDTA (150 μl) at 37 °C overnight and precipitated with 2.5 volumes of cold ethanol at –70 °C overnight using 1 μg glycogen as carrier. We resuspended DNA in 10 μl of 10 mM Tris–HCl, pH 8.0, 1 mM EDTA buffer. We amplified 1 μl of each sample by PCR (35 cycles) using primers derived from the sequences predicted for each *Not*I/*Hin*fI fragment. We tested the specificity of the PCR reaction using water and unrelated spot DNA.

Array CGH. To determine the copy number across all chromosomes, we did comparative genomic hybridizations on whole-genome arrays of approximately 2,400 chromosomally mapped BAC clones (HumArray1.14) following previously described methods (ref. 3; see Web Table A online). Briefly, we hybridized arrays simultaneously with 600 ng each of tumor DNA labeled with Cy3–dCTP by random priming and Cy5–labeled reference DNA from normal brain tissue. We counterstained the spotted BAC DNA with 4′,6′-diamidino-2-phenylindole hydrochloride (DAPI) and collected and processed the images of the three fluorochromes using custom software[7] that calculates the raw ratios and mean log_2 ratios of triplicates of tumor to reference DNA hybridization. After normalization, we plotted mean log_2 ratios and analyzed the resultant graphs for deletions and gains along each chromosome.

QuMA. To corroborate the array CGH results, we also assessed copy number with 17 microsatellite markers from chromosome 22 using a previous-

ly described real-time PCR-based method[8]. Briefly, we amplified triplicates of each DNA sample (5 ng) simultaneously with primers specific for each microsatellite marker and with pooled reference primers in separate wells. We measured the accumulation of PCR product using a dual-labeled fluorogenic oligonucleotide TaqMan probe, 5′-(FAM)-TGTGTGTGTGTGTGTG TGTGTGT-(TAMRA)-3′, which detects the CA repeats in the PCR products. The reporter (FAM) fluorescence is liberated from the quencher (TAMRA) as a PCR product is generated, and serves as a direct measure of the amount of PCR product. To measure the relative copy number of the tumor sample compared with normal DNA, we did control reactions on normal brain tissue from nine different individuals.

The names and positions of the markers (based on the August 2001 freeze of the draft human genome from University of California Santa Cruz) are shown in Web Table A online. The genomic positions of three microsatellite markers (*D22S539*, *D22S1150*, *D22S426*) overlapped exactly with three array CGH BAC clones (at 18,956 kb, 26,197 kb and 33,638 kb). We selected seven additional markers used in a reference PCR pool from genomic positions that are typically unaffected in gliomas (*D8S1715* at 8p21.3, *D2S391* at 2p21, *D1S210* at 1q23-24, *D3S1554* at 3p21-23, *D5S643* at 5q22-32, *D5S478* at 5p14-15.2 and *D21S1904* at 21p13). We obtained primer sequences for most of the microsatellite markers from the Whitehead Institute for Genome Research and designed those for U20569 based on published sequences of component AP000346 in the GenBank database. There was concordance between deletions detected by QuMA and by array CGH at both the overlapping and adjacent sites (see Web Table A online).

Statistical analysis of the patterns of methylation-and-deletion and gain-and-methylation events. To formally test whether methylation-and-deletion events and gain-and-methylation events were randomly or preferentially distributed among the CpG islands, we first estimated the frequency of methylation and deletion for each tumor. Assuming no preference among sites, the estimated probability of methylation and deletion for any CpG island in a particular tumor is the number of methylation-and-deletion events divided by the number of islands evaluated. The estimated expected frequency of methylation and deletion at a given island for the total of the 26 tumors is the sum of these estimates. Assuming that there is a Poisson distribution of methylation-and-deletion events, and that the occurrence at one site is independent of the occurrence at another site, we calculated the expected number of islands having methylation and deletion on a certain number of tumors and compared that to the observed number (Fig. 3a) to create a χ^2 goodness-of-fit statistic. We calculated the *P* value using a Monte Carlo estimate of the exact permutation test *P* value (StatXact).

Drug treatment, reverse transcription, PCR and northern blotting. We plated four glioma cell lines at low density, incubated them with 1 μM 5-aza-2-deoxycytidine for 3 d with fresh drug and medium every 24 h and then isolated total cellular RNA using Trizol (Gibco-BRL). We carried out reverse transcription on RNAs using oligo-dT and random primers. We processed total RNA isolated from brain tissue and primary tumors in a similar fashion.

We did PCR in 25 μl reactions with one-fortieth of each reverse transcription reaction product as template. We amplified the *ACTB* mRNA (encoding β-actin) by incubating the reaction product at 95 °C for 4 min followed by 27 cycles (exponential phase of *ACTB* PCR amplification, data not shown) of 95 °C for 30 s, 68 °C for 30 s and 72 °C for 2 min. We amplified the *COE3* mRNA by incubating the reaction product at 95 °C for 5 min followed by 35 cycles of 95 °C for 1 min, 65 °C for 1 min and 72 °C for 1 min. All primers are available upon request. We used water and genomic DNA as negative controls and detected no PCR products (data not shown). We carried out independent replications of drug treatment, reverse transcription and PCR reactions with similar results.

We also hybridized multiple-human-tissue northern blots (Clontech) sequentially with a *COE3* probe and an *ACTB* probe (see Web Fig. C online). The results confirmed that *COE3* was expressed in normal human brain and most other tissues tested. Because the clinical tumor samples contained variable amounts of blood, we expected that RNA from the tumor samples would be partly derived from blood cells expressing *COE3*.

Bisulfite sequencing method. Briefly, we digested 2.5 μg of DNA from normal tissue and tumor cell line samples with 5 U *Eco*RV and then denatured it in 0.3 M NaOH for 20 min at 42 °C. We added freshly prepared sodium bisulfite (pH 5.0, 208 μl, 3.6 M) and hydroquinone (12 μl, 10 mM) to the

denatured DNA in a final volume of 240 μl and then incubated it for 18 h at 55 °C. We then purified DNA using the Promega Wizard DNA Clean-Up System and eluted it in 100 μl of sterilized water. We added freshly prepared NaOH to a final concentration of 0.3 M and incubated samples for 15 min at 37 °C or 20 min at 42 °C. After neutralizing the samples with 3 M ammonium acetate (pH 7), we precipitated and resuspended DNA in 10 μl of sterilized water and stored it at −20 °C until use. We carried out nested PCR on 1 μl of each DNA sample. PCR products were gel purified and cloned into the TOPO TA Cloning/pCR2.1 TOPO kit (Invitrogen). We subjected individual bacterial colonies to PCR using vector-specific primers (sequences available upon request), and the products were sequenced.

URLs. Our custom PERL program used to generate RLGS digestion fragments from the human genome sequence is available at http://shark. ucsf.edu/~stas/Costello/rlgs.html. The August 2001 freeze of the University of California, Santa Cruz version of the draft assembly of human genome is available at http://genome.ucsc.edu. The website for the sequence-tagged site–based map of the Whitehead Institute for Genome Research is http:// www-genome.wi.mit.edu/cgi-bin/contig/phys_map.

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

1. Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H. & Mukai, T. A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. Natl Acad. Sci. USA* **88**, 9523–9527 (1991).
2. Yoshikawa, H. *et al.* Chromosomal assignment of human genomic *Not*I restriction fragments in a two-dimensional electrophoresis profile. *Genomics* **31**, 28–35 (1996).
3. Snijders, A.M. *et al.* Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* **29**, 263–264 (2001).
4. Yoshikawa, H. *et al.* SOCS-1, a negative regulator of the JAK/STAT pathway, is silenced by methylation in human hepatocellular carcinoma and shows growth-suppression activity. *Nat. Genet.* **28**, 29–35 (2001).
5. Makos, M. *et al.* Distinct hypermethylation patterns occur at altered chromosome loci in human lung and colon cancer. *Proc. Natl Acad. Sci. USA* **89**, 1929–1933 (1992).
6. Herman, J.G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl Acad. Sci. USA* **91**, 9700–9704 (1994).
7. Jain, A.J. *et al.* Fully automatic quantification of microarray image data. *Genome Res.* **12**, 325–332 (2002).
8. Ginzinger, D.G. *et al.* Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. *Cancer Res.* **60**, 5405–5409 (2000).
9. Costello, J.F. *et al.* Aberrant CpG-island methylation has non-random and tumor-type-specific patterns. *Nat. Genet.* **24**, 132–138 (2000).
10. Rush, L.J. *et al.* Novel methylation targets in *de novo* acute myeloid leukemia with prevalence of chromosome 11 loci. *Blood* **97**, 3226–3233 (2001).
11. Smiraglia, D.J. *et al.* Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10**, 1413–1419 (2001).
12. Fruhwald, M.C. *et al.* Aberrant promoter methylation of previously unidentified target genes is a common abnormality in medulloblastomas—implications for tumor biology and potential clinical utility. *Oncogene* **20**, 5033–5042 (2001).
13. Dai, Z.Y. *et al.* Global methylation profiling of lung cancer identifies novel methylated genes. *Neoplasia* **3**, 314–323 (2001).
14. Rouillard, J.M. *et al.* Virtual genome scan: a tool for restriction landmark-based scanning of the human genome. *Genome Res.* **11**, 1453–1459 (2001).
15. Starr, R. *et al.* A family of cytokine-inducible inhibitors of signalling. *Nature* **387**, 917–921 (1997).
16. Endo, T.A. *et al.* A new protein containing an SH2 domain that inhibits JAK kinases. *Nature* **387**, 921–924 (1997).
17. Mohapatra, G. *et al.* Genetic analysis of glioblastoma multiforme provides evidence for subgroups within the grade. *Genes Chromosomes Cancer* **21**, 195–206 (1998).
18. Fults, D. & Pedone, C. Deletion mapping of the long arm of chromosome 10 in glioblastoma multiforme. *Genes Chromosomes Cancer* **7**, 173–177 (1993).
19. Karlbom, A.E. *et al.* Loss of heterozygosity in malignant gliomas involves at least three distinct regions on chromosome 10. *Hum. Genet.* **92**, 169–174 (1993).
20. von Deimling, A. *et al.* Comprehensive allelotype and genetic analysis of 466 human nervous system tumors. *J. Neuropathol. Exp. Neurol.* **59**, 544–558 (2000).
21. Dubois, L. & Vincent, A. The COE—Collier/Olf1/EBF—transcription factors: structural conservation and diversity of developmental functions. *Mech. Dev.* **108**, 3–12 (2001).