

# Epigenetic differences arise during the lifetime of monozygotic twins

Mario F. Fraga\*, Esteban Ballestar\*, Maria F. Paz\*, Santiago Ropero\*, Fernando Setien\*, Maria L. Ballestar†, Damia Heine-Suñer‡, Juan C. Cigudosa§, Miguel Urioste¶, Javier Benitez¶, Manuel Boix-Chornet†, Abel Sanchez-Aguilera†, Charlotte Ling||, Emma Carlsson||, Pernille Poulsen\*\*, Allan Vaag\*\*, Zarko Stephan††, Tim D. Spector††, Yue-Zhong Wu\*\*, Christoph Plass\*\*, and Manel Esteller\*§§

\*Epigenetics, §Cytogenetics, and ¶Genetic Laboratories, Spanish National Cancer Centre (CNIO), Melchor Fernandez Almagro 3, 28029 Madrid, Spain; †Department of Behavioral Science, University of Valencia, 46010 Valencia, Spain; ‡Molecular Genetics Laboratory, Genetics Department, Son Dureta Hospital, 07014 Palma de Mallorca, Spain; §Department of Clinical Sciences, University Hospital Malmö, Lund University, S-205 02 Malmö, Sweden; \*\*Steno Diabetes Center, 2820 Gentofte, Denmark; ††Twin Research and Genetic Epidemiology Unit, St. Thomas' Hospital, London SE1 7EH, United Kingdom; and §§Human Cancer Genetics Program, Department of Molecular Virology, Immunology, and Medical Genetics, Ohio State University, Columbus, OH 43210

Edited by Stanley M. Gartler, University of Washington, Seattle, WA, and approved May 23, 2005 (received for review January 17, 2005)

**Monozygous twins share a common genotype. However, most monozygotic twin pairs are not identical; several types of phenotypic discordance may be observed, such as differences in susceptibilities to disease and a wide range of anthropomorphic features. There are several possible explanations for these observations, but one is the existence of epigenetic differences. To address this issue, we examined the global and locus-specific differences in DNA methylation and histone acetylation of a large cohort of monozygotic twins. We found that, although twins are epigenetically indistinguishable during the early years of life, older monozygous twins exhibited remarkable differences in their overall content and genomic distribution of 5-methylcytosine DNA and histone acetylation, affecting their gene-expression portrait. These findings indicate how an appreciation of epigenetics is missing from our understanding of how different phenotypes can be originated from the same genotype.**

DNA methylation | epigenetics | histones

Human monozygotic (MZ) twins account for 1 in 250 live births (1). The origin of MZ twins is attributed to two or more daughter cells of a single zygote undergoing independent mitotic divisions, leading to independent development and births. They are considered genetically identical, but significant phenotypic discordance between them may exist. This quality is particularly noticeable for psychiatric diseases, such as schizophrenia and bipolar disorder (2). MZ twins have been used to demonstrate the role of environmental factors in determining complex diseases and phenotypes, but the true nature of the phenotypic discordance nevertheless remains extremely poorly understood. In this context, differences in the placenta, amniotic sac, and vascularization of the separate cell masses or even mosaicism in genetic and cytogenetic markers in MZ may exist (3), although the published studies are very few in number. Thus, the real causes for MZ twin discordance for common diseases and traits remain to be established. Epigenetic differences may be an important part of the solution to this puzzle. Indeed, epigenetic profiles may represent the link between an environmental factor and phenotypic differences in MZ twins. Cloned animals provide another example of how epigenetics may explain phenotypic differences in beings that have identical genetic sequences. In this case, inefficient epigenetic reprogramming of the transplanted nucleus is associated with aberrations in imprinting, aberrant growth, and lethality beyond a threshold of faulty epigenetic control (4). MZ twins are another phenomenon in which epigenetics can “make the difference.” To address this possibility, we have profiled the epigenetic patterns related to global and locus-specific DNA methylation and histone H3 and H4 acetylation in the largest series of MZ twins for which molecular studies have been reported.

## Materials and Methods

**Subjects.** Eighty volunteer Caucasian twins from Spain were recruited in the study, including 30 male and 50 female subjects. Their mean ( $\pm$ SD) age was 30.6 ( $\pm$ 14.2) years (range, 3–74 years). Twins studied included monozygotic and dizygotic. All subjects, or in the case of children, the parents, gave their informed written consent to be included in the study. Lymphocyte cells were purified by standard procedures and stored at  $-80^{\circ}\text{C}$ . In eight cases, epithelial skin cells were obtained from buccal smears. Muscle biopsy tissues ( $n = 14$ ) from the vastus lateralis muscle and s.c. abdominal tissue ( $n = 4$ ) were obtained by needle suction under local anesthesia from volunteer MZ twins from Denmark and the United Kingdom, respectively. Homozygosity was determined by using highly polymorphic short tandem-repeat loci. With five markers, the probability that any twin pair was MZ if all markers were concordant was 99% (5).

**X-Inactivation Analysis.** Androgen receptor locus methylation analysis was performed based on PCR on genomic DNA digested with the methylation-sensitive restriction enzyme HpaII where only the androgen receptor gene residing on the inactivated X chromosome is amplified (6, 7).

**High-Performance Capillary Electrophoresis Quantification of Global Histone H3 and H4 Acetylation.** Global histone H4 acetylation (AcH4) and histone H3 acetylation (AcH3) were quantified as described in ref. 8. In brief, individual histone H3 and histone H4 fractions were prepared from cell nuclei and further purified by reversed-phase high-performance liquid chromatography (HPLC) on a Jupiter C18 column (Phenomenex, Torrance, CA). Histones were eluted with an acetonitrile gradient (20–60%) in 0.3% trifluoroacetic acid using an HPLC gradient system (Beckman Coulter). Nonacetylated, monoacetylated, and diacetylated histone H3 and H4 derivatives were resolved by high-performance capillary electrophoresis. All samples were analyzed in duplicate, and three measurements were made per replicate.

**Quantification of Total DNA Methylation.** 5-Methylcytosine (5mC) content was determined by high-performance capillary electrophoresis, as described in ref. 8. Briefly, genomic DNA samples were

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: AIMS, amplification of intermethylated sites; ESD, Euclidean squared distance; 5mC, 5-methylcytosine; MZ, monozygotic; RLGs, restriction landmark genomic scanning.

See Commentary on page 10413.

§§To whom correspondence should be addressed. E-mail: mesteller@cnio.es.

© 2005 by The National Academy of Sciences of the USA

boiled, treated with nuclease P1 for 16 h at 37°C, and treated with alkaline phosphatase (Sigma) for an additional 2 h at 37°C. After hydrolysis, total cytosine and 5mC content was measured by capillary electrophoresis using a P/ACE MDQ system (Beckman Coulter). Relative 5mC content was expressed as a percentage with respect to the total cytosine content.

**Amplification of Intermethylated Sites (AIMS).** AIMS was developed as described in ref. 9. The nonmethylated sites were cut in an initial digestion using the methylation-sensitive endonuclease SmaI (Amersham Pharmacia Biotech, Buckinghamshire, U.K.), which leaves blunt ends. A second digestion was performed using the isoschizomer PspAI (Stratagene), which leaves a CCGG overhang. DNA fragments flanked by two ligated adaptors were amplified by PCR using specific primers that hybridize to the adaptor sequence and the restriction site and one or more additional, arbitrarily chosen nucleotides. The PCR products were run on polyacrylamide urea sequencing gels. Bands appearing differentially methylated were excised from dried polyacrylamide gels and cloned into plasmid vectors. Automatic sequencing of multiple colonies was performed to ascertain the unique identity of the isolated band.

**Analysis of Sequence-Specific DNA Methylation.** The methylation status of specific genomic DNA sequences was established by bisulfite genomic sequencing (9). Automatic sequencing of 12 colonies for each sequence was performed to obtain data on the methylation status of every single CpG dinucleotide for statistical analysis. Primer sequences are available in Table 2, which is published as supporting information on the PNAS web site.

**Real-Time Quantitative RT-PCR Expression Analyses.** Quantitative RT-PCRs were performed in 96-well optical plates, in a volume of 20  $\mu$ l per well. Each reaction mixture contained 1  $\mu$ l of the appropriate dilution of the experimental cDNA (1:10 to 1:100), 5 pmol of each primer, and 10  $\mu$ l of 2 $\times$  SYBRGreen PCR Master Mix (Applied Biosystems). All measurements were performed in triplicate. Expression values were normalized against the expression of GAPDH as an endogenous control. PCR reactions were run and analyzed by using the Prism 7700 Sequence Detection System (Applied Biosystems).

**Competitive Hybridization of AIMS Products to Metaphase Chromosomes.** To study the distribution of the AIMS-amplified DNA within the chromosomes, we hybridized equimolecular quantities from both members of the twin pair. The AIMS product from one of the twins was labeled with Spectrum Red dUTP using a comparative genomic hybridization nick-translation kit (Vysis), and the AIMS products of the other twin was labeled with Spectrum Green. The mixture was then hybridized onto normal metaphase chromosome spreads, following the comparative genomic hybridization protocol described in ref. 9. All experiments were carried out twice.

**Restriction Landmark Genomic Scanning (RLGS).** RLGS of genomic DNA was performed as described in ref. 10. DNA samples were analyzed by using the restriction enzyme combination of NotI–EcoRV–HinfI. Briefly, high-molecular-weight DNA was digested with the methylation-sensitive restriction enzyme NotI, and the restriction sites were filled in with [ $\alpha$ -<sup>32</sup>P]dCTP and [ $\alpha$ -<sup>32</sup>P]dGTP. Methylation of the NotI restriction site makes it resistant to digestion, thus preventing the site from becoming radioactively labeled. Next, the DNA was digested with EcoRV, electrophoresed overnight in a 0.8% agarose gel, digested *in situ* with HinfI, and electrophoresed overnight in a 5% acrylamide gel. Gels were dried and exposed to x-ray film for 7 days. Southern blot analysis was carried out as described in ref. 10.

**Affymetrix GeneChip.** Biotinylated target RNA was prepared from 5  $\mu$ g of total RNA by following the Affymetrix protocol and was hybridized on the Human U133 Plus 2.0 array GeneChips (Affymetrix, Santa Clara, CA). The hybridization reactions were processed and scanned according to the standard Affymetrix protocols. To select genes that were differently expressed in twin siblings, ANOVA *P* values were adjusted by using multiple comparison procedures. The expression profiles of twins were compared in scatter plots.

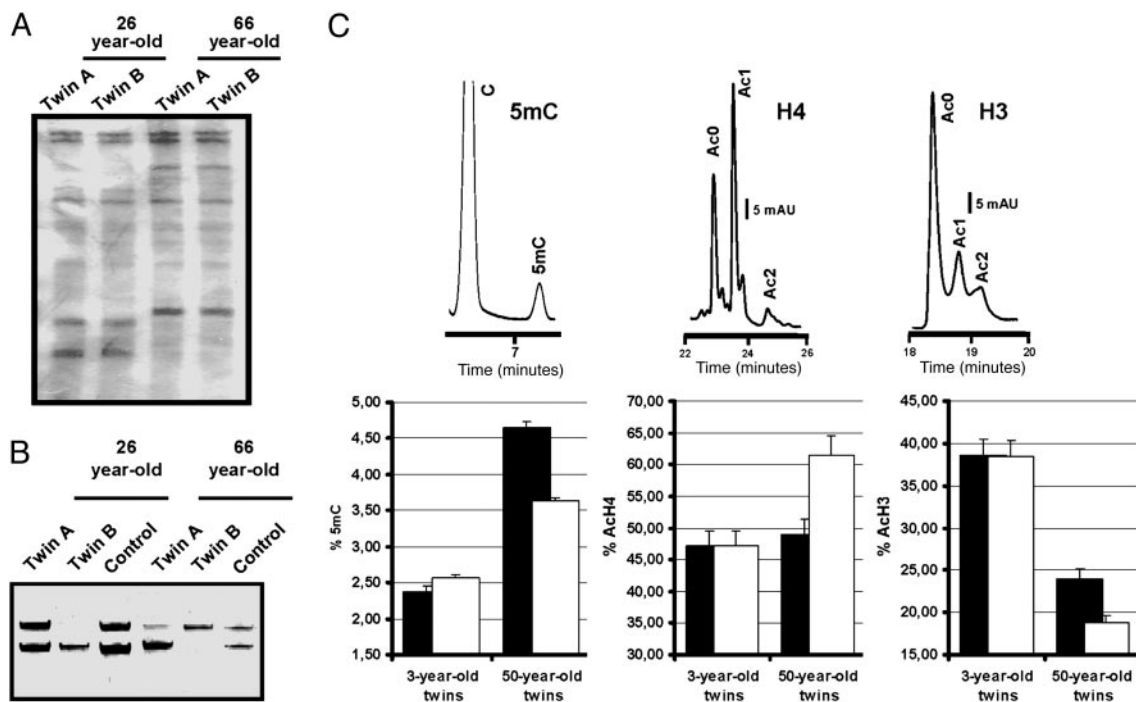
**Questionnaires.** All subjects were interviewed by properly trained personnel to complete the questionnaire about their health, nutritional habits, physical activities, pharmacological treatments, and tobacco, alcohol, and drug consumption. Weight and height were measured, and a family tree of genetic history was drawn up by the interviewer. The data collected in the questionnaires were entered into a database as nominal (natural health history), ordinal (percentage of lifetime shared, nutritional habits, physical activity, and consumption of folates, alcohol, tobacco, and drugs), and numerical (age, weight, and height) variables, which were subsequently used to estimate the phenotypic/environmental distance between twin pairs.

**Statistics.** First, a descriptive value was obtained for each individual corresponding to each epigenetic variable (5mC, AcH4, and AcH3). The similarity between twin pairs for each variable was estimated as the Euclidean squared distance (ESD) by subtracting their respective descriptive values. By using these distances, the relationships between the phenotypic-environmental data and the epigenetic variables were examined by categorical principal component analysis, using SPSS software (Version 11.5; SPSS, Chicago). Categorical principal component analysis reduced all of the original variables from the questionnaire to two new uncorrelated components or variables (“aging” and “health”). The aging principal component included the variables of age, weight, and height and is an indicator of the ontogenic development of the twins. The health principal component grouped the variables of diseases and pharmacological treatments. To identify mixed-effects models and to estimate the contribution of each random effect to the variance of the dependent variable, the variance component procedure was assessed by ANOVA. The relationship between all of the single questionnaire variables and the new phenotypic-environmental variables (aging and health) vs. the epigenetic data was evaluated by the Pearson test.

## Results

**Inclusion of Cases and Zygosity Typing.** The starting biological material studied was DNA and RNA from peripheral lymphocytes. In all cases, the true MZ twin status was confirmed by microsatellite analysis (Fig. 1A) (7).

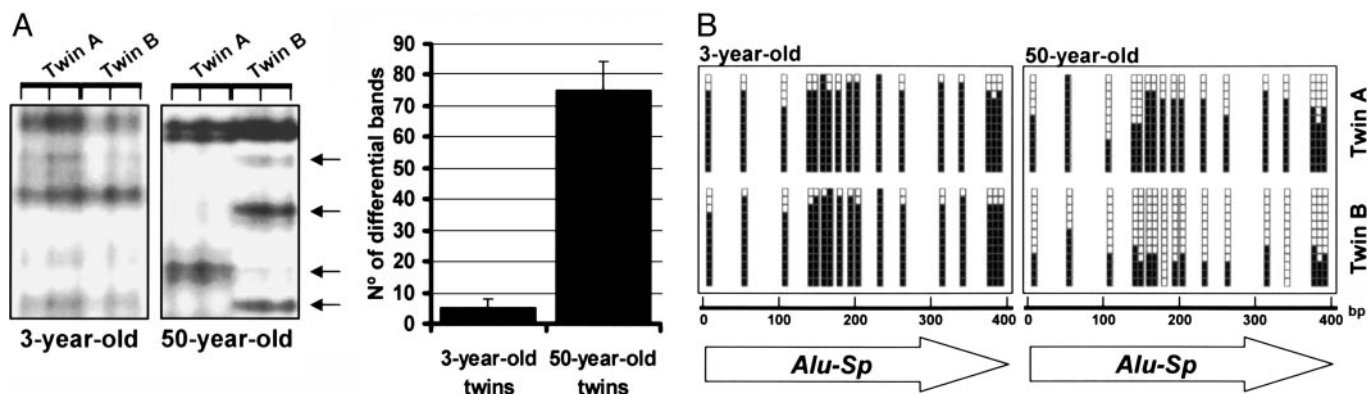
**Comparison of X Chromosome Inactivation Patterns in MZ Twins.** One of the most obvious epigenetic layers where MZ twins may differ is X chromosome inactivation in females, where discordant skewing of X chromosome inactivation has been observed for specific X-linked diseases (11). We found that among MZ female twins informative for our PCR approach, 81% (13 of 16) had the same X chromosome methylation pattern, and of these, 85% (11 of 13) had an unskewed random distribution of each methylated allele. Although the remaining 15% (2 of 13) displayed a skewed X chromosome pattern, this pattern was always concordant between siblings. Among all female MZ twins, only 19% (3 of 16) had a skewed X chromosome methylation pattern that differed between corresponding siblings (Fig. 1B). The presence of a differential X chromosome inactivation pattern in the latter cases was not associated with any epidemiological and clinicopathological feature but raises the possibility that epigenetic discordance can arise early on in the development of some MZ twins.



**Fig. 1.** Epigenetic differences arise in MZ twins. (A) Two representative examples of the determination of monozygosity using microsatellite markers. (B) Quantification of X chromosome inactivation by PCR amplification of the androgen receptor locus after digestion with the DNA methylation-sensitive and -insensitive restriction enzymes HpaII and Msp I, respectively. Two examples of a different pattern of X inactivation between MZ twin pairs are shown. (C Upper) Quantification of global 5mC DNA content (Left), histone H4 acetylation (Center), and histone H3 acetylation (Right) by HPLC and high-performance capillary electrophoresis. (C Lower) Comparison of epigenetic values between the siblings of each 3- and 50-year-old twin pair. Results are expressed as mean  $\pm$  SD.

**Comparison of 5mC DNA Content and Histone H3 and H4 Acetylation Levels in MZ Twin Pairs.** The epigenetic circumstances of the MZ twins proved to be more remarkable when we took a more global approach to the comparison of the DNA methylation and histone acetylation content of the MZ twins. For each pair, we determined the 5mC genomic content and the acetylation levels of histones H3 and H4 by HPLC and high-performance capillary electrophoresis (8) (see Table 3, which is published as supporting information on the PNAS web site). Replicate samples at time-points 0, 2, 4, 6, and 12 weeks for eight MZ twin pairs were assessed to determine short-time fluctuations in the described epigenetic parameters, and no significant changes were observed (Pearson test,  $P > 0.05$ ) (see Fig. 5, which is published as supporting information on the PNAS

web site). Illustrative examples of the three different analyses are shown in Fig. 1C. We found that 65% (26 of 40) of MZ twins had almost identical 5mC genomic content and overall acetylation levels of histones H3 and H4 in each sibling pair. However, in 35% (14 of 40) of the twin pairs all three different epigenetic characters were significantly different in each twin (Pearson test,  $P < 0.05$ ) (Fig. 1). Most importantly, we found a direct association between the remarkable epigenetic differences observed and the age of the MZ twins: the youngest pairs were epigenetically similar, whereas the oldest pairs were clearly distinct (Pearson test,  $P < 0.05$ ) (Fig. 1C). To discard the possibility that the observed age-dependent epigenetic differences in the MZ twin pairs were simply due to a general increase in the epigenetic variability in an older population,



**Fig. 2.** Mapping sequences with differential DNA methylation between MZ twins using AIMS. (A Left) Example of an AIMS analysis in 3- and 50-year-old twin pairs. Differential anonymous bands corresponding to sibling-specific changes of DNA methylation are indicated with arrows. (A Right) Number of differential bands obtained by AIMS in 3- and 50-year-old twin pairs. (B) Bisulfite genomic sequencing of 12 clones of the repetitive DNA sequence Alu-SP obtained by AIMS in 3- and 50-year-old twin pairs. Schematic representations of the methylation status of each CpG dinucleotide. Black and white dots indicate methylated and unmethylated CpGs, respectively.



**Table 1. Summary of DNA sequences with methylation changes identified by AIMS in MZ twin pairs**

Locus	Accession	Function	Chr. loc.	No. CpGs	CpG methylation, * %		
					Twin A	Twin B	Change
<i>Alu-Sx</i>	AC006271	—	19p13.2	30	72.1	55.3	16.8
<i>Alu-Sp</i>	U14572	—	19p13+3	17	68.1	29.9	38.2
<i>Alu-E2F3</i>	AF547386	—	6p22	2	81.8	100	18.2
<i>PKD1P2</i>	NG_002795	Polycystic kidney disease 1	16p13	7	90.9	83.5	7.4
<i>Alu-Sc</i>	U14571	—	8p11+1	4	66.6	63.3	3.3
<i>GUK1</i>	BC006249	Guanylate kinase 1	1q42	18	2.8	9.7	6.9

Chr. loc., chromosome location; —, unknown.

\*Twelve clones per locus.

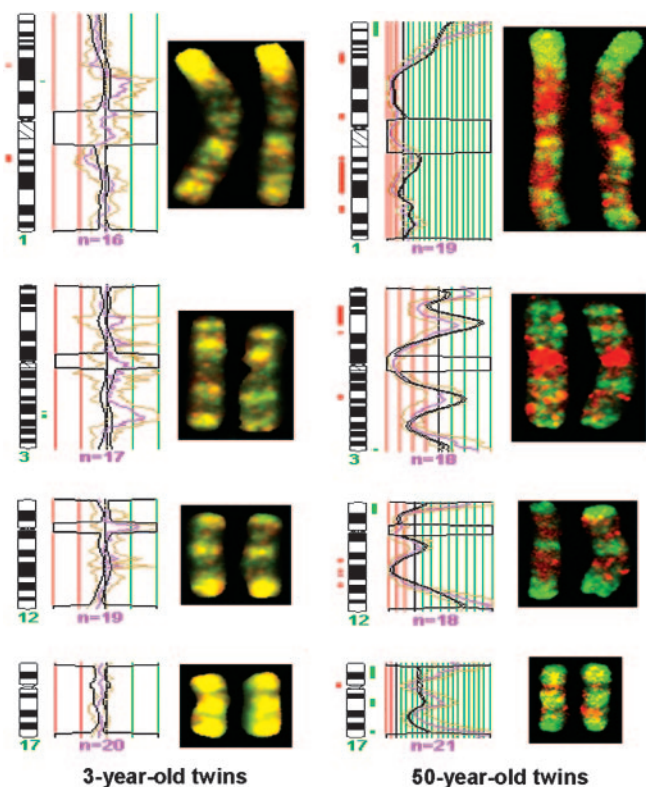
we performed ANOVA (see Table 4, which is published as supporting information on the PNAS web site). ANOVA was used to compare two parameters: individual epigenetic values, to compare epigenetic variability in the entire population, and ESD between twins, as a measure of variability between twin pairs. First, we analyzed the descriptive epigenetic values for each individual and compared it with the rest of the population, organized into young (<28 years old) vs. old (>28 years old) age-groups. Statistical analysis of individual descriptive parameters provides information about variability within the whole population and in the two age groups. Second, we used the ESD to test whether older vs. younger MZ twins are significantly more different within their pairs. We found that the epigenetic variability among individuals is high and similar (see Table 5, which is published as supporting information on the PNAS web site), regardless of the age group to which individuals belong (Pearson test,  $P > 0.05$ ). In contrast, the variance corresponding to the ESD in the older MZ twin group is significantly higher than that obtained for the younger group (Pearson test,  $P < 0.05$ ). These results suggest that older twins are epigenetically more different between pairs than younger twins, and this difference is not associated with an increased variance in the descriptive epigenetic parameters of the older population.

Finally, we also found that those twin pairs who, according to the questionnaire, had spent less of their lifetime together and/or had a more different natural health–medical history were those who also showed the greatest differences in levels of 5mC DNA and acetylation of histones H3 and H4 levels (Pearson test,  $P < 0.05$ ).

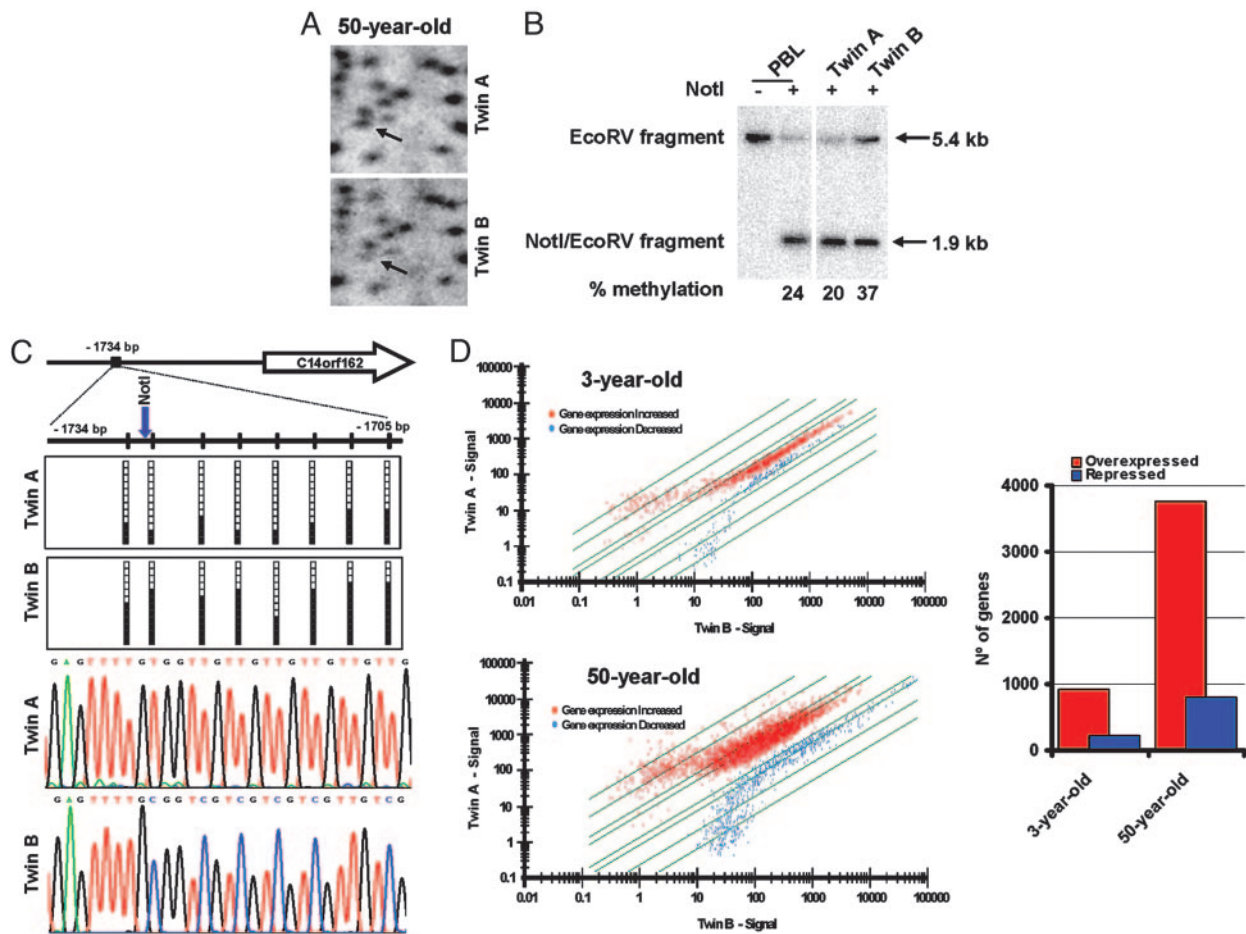
**Genomic Screening and Loci Identification of DNA Methylation Differences in MZ Twin Pairs.** We next examined where in the MZ twin genomes these epigenetic differences arose by using a global methylation DNA fingerprinting technique, AIMS (9). The AIMS approach provides a methylation fingerprint constituted by multiple anonymous bands or tags, representing DNA sequences flanked by two methylated sites, which can be isolated. An illustrative AIMS result in a MZ twin pair is shown in Fig. 2A. Approximately 600 AIMS bands were resolved in the gels, and we found that between 0.5% and 35% of the bands were different (on the basis of their presence or absence) within MZ twin pairs. Those MZ twins with the most differential AIMS bands were those with the greatest differences in 5mC DNA levels and acetylation levels of histones H3 and H4 (Pearson test,  $P < 0.05$ ). Most importantly, the twin pairs with the most differential AIMS bands corresponded to MZ twins who were older, had spent less of their lifetimes together, or had different natural health–medical histories (Pearson test,  $P < 0.05$ ).

We then selected a subset of 53 AIMS bands differentially present in MZ twin pairs for further characterization. These bands were cloned, sequenced, and blasted against multiple sequence databases. We found that 43% (23 of 53) of the clones matched Alu sequences (mainly from the Alu 6, Alu 7, and Alu 8 subfamilies), 9% (5 of 53) matched other repetitive sequences (2 LINES, 2 MER, and 1 MIR), 34% (18 of 53) matched ESTs deposited in databases, and 13% (7 of 53) of the clones corresponded to identified

single-copy genes. To confirm that these sequences featured DNA methylation differences in the twins, we carried out bisulfite genomic sequencing of multiple clones. The analysis of four different Alu sequences and two single-copy genes demonstrated that in those MZ twin pairs from which they were isolated, one sibling had dense CpG hypermethylation, whereas the other was predominantly hypomethylated at that particular sequence (Fig. 2B). Most importantly, for both Alus and single-copy genes, differential methylation was associated with a different expression of that particular sequence in the MZ twin pair, the presence of DNA methylation being associated with silencing or reduced expression. Table 1 summarizes the results.



**Fig. 3.** Mapping chromosomal regions with differential DNA methylation in MZ twins by using comparative genomic hybridization for methylated DNA. Competitive hybridization onto normal metaphase chromosomes of the AIMS products generated from 3- and 50-year-old twin pairs. Examples of the hybridization of chromosomes 1, 3, 12, and 17 are displayed. The 50-year-old twin pair shows abundant changes in the pattern of DNA methylation observed by the presence of green and red signals that indicate hypermethylation and hypomethylation events, whereas the 3-year-old twins have a very similar distribution of DNA methylation indicated by the presence of the yellow color obtained by equal amounts of the green and red dyes. Significant DNA methylation changes are indicated as thick red and green blocks in the ideograms.



**Fig. 4.** Differences in promoter CpG island DNA methylation and expression profiles in MZ twins are associated with aging. (A) Example of an RLGS profile from a 50-year-old twin pair showing the presence of an RLGS fragment C14orf162 in Twin A, indicated by the arrow, in whose sister (Twin B) it is significantly reduced. (B) Southern blot analysis confirms the CpG island methylation difference in C14orf162 between twins. All lines contain DNA digested with EcoRV. The DNA was also digested with NotI, as indicated by +. (C) Bisulfite genomic sequencing of 12 clones of the C14orf162 promoter CpG island in the 50-year-old twin pair. Black and white dots indicate methylated and unmethylated CpGs, respectively. Below the sequences are two representative electropherograms showing sequence-specific DNA methylation changes at the C14orf162 promoter in the 50-year-old twin pair. (D) The older MZ twins display the greatest differences in gene expression profiles between siblings. (Left) Scatter plots showing expression profiles of 3- and 50-year-old twin pairs. (Right) Number of overexpressed and repressed genes obtained by DNA array analyses in the 3-year-old compared with the 50-year-old twin pair.

**Identification of Chromosomal Regions with DNA Methylation Differences in MZ Twin Pairs.** We also investigated the “geographical” distribution of DNA methylation differences within the chromosomes of MZ twin pairs by competitively hybridizing the AIMS PCR products to metaphases to determine the differentially methylated DNA regions (9). The AIMS products from each component of one individual of an MZ twin pair were competitively hybridized with those of his or her sibling. The data obtained demonstrated that the twin pairs with similar 5mC DNA content, acetylation levels of H3 and H4, and AIMS DNA methylation patterns shared a common distribution of DNA methylation signal throughout all chromosomal regions (Fig. 3). However, those twins with different global epigenetic patterns also showed a distinct profile of DNA methylation signal in their metaphase chromosomes. In particular, these differences were located in almost all telomeres and a few selected gene-rich regions (such as 1p36, 1q21, 3p21, 8q21, 9q21, and 22q11.2). Furthermore, our DNA methylation-competitive genomic hybridization results again showed that MZ twins who were younger, had similar lifestyles, and had spent more of their lifetimes together displayed minimal DNA methylation changes in all chromosomes, whereas those who were older, had different lifestyles, and had spent less of their lives together had unevenly

distributed DNA methylation events (hypermethylation and hypomethylation) throughout the chromosomes described above.

**Global CpG Island Methylation Differences in MZ Twin Pairs Characterized by RLGS.** Twins’ differentially methylated sequences obtained by the AIMS approach also included CpG islands located in the promoter regions of single-copy genes. We implemented a specific strategy for further characterizing global genomic methylation differences in these important DNA regions. We subjected the two opposite sites of the MZ twin spectrum, those from the youngest 3- and 22-year-old vs. those completely epigenetically different from the 50-year-old, to methylation-oriented RLGS (10). This approach allows global determination of the methylation status of  $\approx 1,800$  unselected CpG islands. We observed that the older twin pairs presented 2.5 times as many DNA methylation differences in the CpG islands of single-copy genes as did the younger twin-pairs (Fig. 4A). These DNA methylation differences observed in the older twins also were demonstrated for these particular gene-associated CpG islands by methylation-sensitive Southern blot (Fig. 4B) and bisulfite genomic sequencing of multiple clones (Fig. 4C).



**Portraits of Gene Expression in MZ Twin Pairs.** Finally, we addressed the ultimate goal of all major epigenetic modification: global change in gene expression. We examined whether older twin pairs who differed most with respect to DNA methylation and histone acetylation levels in general and at specific loci also displayed the most different gene expression profiles. To this end, we extracted RNA from the two most distinct pairs of twins (the 3- and 50-year-old pairs) and performed gene expression microarray analysis. The results obtained when each twin expression portrait was confronted with its own sibling demonstrated that although the expression patterns of the 3-year-olds were almost identical, the 50-year-old twins had extremely different expression profiles (Fishers exact test,  $P = 0.029$ ) (Fig. 4D). From a quantitative standpoint, there were four times as many differentially expressed genes in the older twin pair as in the younger twin pair. Furthermore, the older twin sibling with the most severe DNA hypomethylation and histone H3 and H4 hyperacetylation (the epigenetic changes usually associated with transcriptional activation) was that with the highest number of overexpressed genes (Fig. 4D). Most importantly, and internally corroborating our data, all of the aforementioned single-copy genes identified by AIMS or RLGS as being differentially methylated in the twin pairs also showed different gene expression in our microarray analysis.

**Epigenetic Differences in MZ Twin Pairs Occur Across Different Tissue Types.** Although our study had focused on the DNA methylation and histone acetylation changes observed in the lymphocytes of MZ twins, we thought that it would be extremely interesting to analyze the existence of epigenetic differences among other cell types, especially because the described discordances between MZ twins are also present across different organs and cellular functions. To this end, we studied the 5mC DNA content, the acetylation status of histone H3 and H4, and the patterns of loci-specific DNA methylation by the AIMS approach, methyl-specific comparative genomic hybridization, and bisulfite genomic sequencing in MZ twins couples by using epithelial mouth cells, intraabdominal fat, and skeletal muscle biopsies. We found that in these three new tissues, in the same manner that we had observed in lymphocytes, marked epigenetic differences were present in older MZ twins with different lifestyles and that had spent less of their lives together (Pearson test,  $P < 0.05$ ) (see Fig. 6, which is published as supporting information on the PNAS web site). Thus, distinct profiles of DNA methylation and histone acetylation patterns among many different tissues arise during the lifetime of MZ twins that may contribute to the explanation of some of their phenotypic discordances and underlie their differential frequency/onset of common diseases.

## Discussion

Although genomic information is uniform among the different cells of a complex organism, the epigenome varies from tissue to tissue, controlling the differential expression of genes and providing specific identity to each cell type. DNA methylation and histone modifications store epigenetic information that mainly controls heritable states of gene expression (12, 13), and it is now well established that both epigenetic layers are mechanistically linked (12, 13).

Our study reveals that the patterns of epigenetic modifications in MZ twin pairs diverge as they become older. Differences in epigenetic patterns in genetically identical individuals could be explained by the influence of both external and internal factors. Smoking habits, physical activity, or diet, among others, are external factors that have been proposed to have a long-term influence on epigenetic modifications (12, 13). However, it is possible that small defects in transmitting epigenetic information through successive cell divisions, or maintaining it in differentiated cells, accumulate in a process that could be considered as an “epigenetic drift” associated with the aging process (14, 15). Identification of proteins that mediate these effects has provided an insightful view into this complex process and in the diseases that occur when it is perturbed (12, 13). Accumulation of epigenetic defects would probably occur at a faster rate than that corresponding to genetic mutations because their consequences in survival are probably less dramatic and cells have not developed a comparable amount of mechanisms to correct them.

MZ twins constitute an excellent example of how genetically identical individuals can exhibit differences and therefore provide a unique model to study the contribution/role of epigenetic modifications in the establishment of the phenotype (2, 16–18). What does make MZ twins differ? By using whole-genome and locus-specific approaches, we found that approximately one-third of MZ twins harbored epigenetic differences in DNA methylation and histone modification. These differential markers between twins are distributed throughout their genomes, affecting repeat DNA sequences and single-copy genes, and have an important impact on gene expression. We also established that these epigenetic markers were more distinct in MZ twins who were older, had different lifestyles, and had spent less of their lives together, underlining the significant role of environmental factors in translating a common genotype into a different phenotype. Our findings also support the role of epigenetic differences in the discordant frequency/onset of diseases in MZ twins (2, 16–18).

Other evidence indicates that relatively small differences in epigenetic patterns can have a large impact in phenotype, for instance in cloned animals (4), with MZ twins representing natural human clones. Another powerful example is provided by the *agouti* mouse (19). In this model, diet affects the methylation status of an inserted intracisternal A particle element that changes the animal's coat color: an environmental factor interacting with a single genotype, mediated by an epigenetic change, to produce a different phenotype. In humans, the investigation of how assisted reproductive technology that uses media with undisclosed concentrations of methyl-donors associates with epigenetic errors such as imprinting defects and cancer has been proposed (20). Our comparison of MZ twins suggests that external and/or internal factors can have an impact in the phenotype by altering the pattern of epigenetic modifications and thus modulating the genetic information. Future studies should now address the specific mechanisms responsible for the observed epigenetic drift of MZ twins.

We thank all our volunteer twins and Sara Casado, Lidia Lopez-Serra, Miguel Alaminos, and Alicia Barroso from the Spanish National Cancer Centre. M.F.F. is funded by the Foundation of the Spanish Association Against Cancer (AECC).

- Hall, J. G. & Lopez-Rangel, E. (1966) in *Twins and Twinning*, eds. Emery, A. E. H. & Rimoin, D. L. (Churchill Livingstone, New York), pp. 395–404.
- Cardno, A. G., Rijdsdijk, F. V., Sham, P. C., Murray, R. M. & McGuffin, P. (2002) *Am. J. Psychiatry* **159**, 539–545.
- Machin, G. A. (1996) *Am. J. Med. Genet.* **61**, 216–228.
- Rideout, W. M., III, Eggen, K. & Jaenisch, R. (2001) *Science* **293**, 1093–1098.
- Becker, A., Busjahn, A., Faulhaber, H. D., Bähring, S., Robertson, J., Schuster H. & Luft, F. C. (1997) *J. Reprod. Med.* **42**, 260–266.
- Kubota, T., Nonoyama, S., Tonoki, H., Masuno, M., Imaizumi, K., Kojima, M., Wakui, K., Shimadzu, M. & Fukushima, Y. (1999) *Hum. Genet.* **104**, 49–55.
- Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M. & Belmont, J. W. (1992) *Am. J. Hum. Genet.* **51**, 1229–1239.
- Fraga, M. F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., et al. (2005) *Nat. Genet.* **37**, 391–400.
- Paz, M. F., Wei, S., Cigudosa, J. C., Rodriguez-Perales, S., Peinado, M. A., Huang, T. H. & Esteller, M. (2003) *Hum. Mol. Genet.* **12**, 2209–2219.
- Costello, J. F., Smiraglia, D. J. & Plass, C. (2002) *Methods* **27**, 144–149.
- Tiberio, G. (1994) *Acta Genet. Med. Gemellol.* **43**, 207–214.
- Jaenisch, R. & Bird, A. (2003) *Nat. Genet.* **33**, Suppl., 245–254.
- Bjornsson, H. T., Fallin, M. D. & Feinberg, A. P. (2004) *Trends Genet.* **20**, 350–358.
- Cooney, C. A. (1993) *Growth. Dev. Aging* **57**, 261–273.
- Bennett-Baker, P. E., Wilkowski, J. & Burke, D. T. (2003) *Genetics* **165**, 2055–2062.
- Weksberg, R., Shuman, C., Caluseriu, O., Smith, A. C., Fei, Y. L., Nishikawa, J., Stockley, T. L., Best, L., Chitayat, D., Olney, A., et al. (2002) *Hum. Mol. Genet.* **11**, 1317–1325.
- Petronis, A., Gottesman, I. I., Kan, P., Kennedy, J. L., Basile, V. S., Paterson, A. D. & Popendikyte, V. (2003) *Schizophr. Bull.* **29**, 169–178.
- Wong, A. H., Gottesman I. I. & Petronis, A. (2005) *Hum. Mol. Genet.* **14**, Suppl. 1, R11–R18.
- Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. (1999) *Nat. Genet.* **23**, 314–318.
- Nimemitz, E. L. & Feinberg, A. P. (2004) *Am. J. Hum. Genet.* **74**, 599–609.