ARTICLE

Differential decay of parent-of-origin-specific genomic sharing in cystic fibrosis-affected sib pairs maps a paternally imprinted locus to 7q34

Frauke Stanke*,[1], Colin Davenport[1], Silke Hedtfeld[1] and Burkhard Tümmler[1]

Cystic fibrosis (CF) is a monogenic disease characterized by a high variability of disease severity and outcome that points to the role of environmental factors and modulating genes that shape the course of this multiorgan disease. We genotyped families of cystic fibrosis sib pairs homozygous for F508del-*CFTR* who represent extreme clinical phenotypes at informative microsatellite markers spanning a 38 Mb region between *CFTR* and 7qtel. Recombination events on both parental chromosomes were compared between siblings with concordant clinical phenotypes and siblings with discordant clinical phenotypes. Monitoring parent-of-origin-specific decay of genomic sharing delineated a 2.9-Mb segment on 7q34 in which excess of recombination on paternal chromosomes in discordant pairs was observed compared with phenotypically concordant sibs. This 2.9-Mb core candidate region was enriched in imprinting-related elements such as predicted CCCTC-binding factor consensus sites and CpG islands dense in repetitive elements. Moreover, allele frequencies at a microsatellite marker within the core candidate region differed significantly comparing mildly and severely affected cystic fibrosis sib pairs. The identification of this paternally imprinted locus on 7q34 as a modulator of cystic fibrosis disease severity shows that imprinted elements can be identified by straightforward fine mapping of break points in sib pairs with informative contrasting phenotypes.
European Journal of Human Genetics (2010) **18**, 553–559; doi:10.1038/ejhg.2009.229; published online 6 January 2010

INTRODUCTION

Imprinted genes are defined by the functional nonequivalence of the maternal and paternal copy resulting in monoallelic expression in a parent-of-origin-dependent manner.[1–3] Imprinting is known to be affected in numerous human disorders ranging from cancer to psychiatric diseases such as autism and schizophrenia.[1–3] Imprinted elements have been identified in large chromosomal segments by parametric methods,[4–6] but none of these novel imprinted elements have been mapped to a highly restricted area as accomplished for the well-characterized genes responsible for the Beckwith–Wiedemann Syndrome (11p15.5) or the Prader–Willi and Angelman Syndrome (15q11–q13).[1–3]

Cystic fibrosis (CF) is a generalized exocrinopathy caused by two defective copies of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, whereby about 70% of CF-causing chromosomes in Caucasians carry the deletion F508del-*CFTR*.[7] The large clinical variability among F508del-*CFTR* homozygous CF patients[7,8] points to a major role of environment and genetic background in shaping the disease phenotype.[8] However, the majority of CF modifier studies deal with few or even one single-nucleotide polymorphism within one candidate gene,[8] in contrast to systematic mapping to dissect neighboring, but equally plausible, candidate genes, which allows the identification of novel CF modulators.[9]

Previously, we could detect a paternally imprinted CF modulator within a 40 cM chromosomal segment 3′ to *CFTR*.[10] In this study, we have used a highly refined customized microsatellite map to confine the paternally imprinted CF modulator to a 2.9-Mb locus on 7q34 in our study cohort, consisting of 29 clinically concordant and 19 clinically discordant F508del-*CFTR* homozygous sibling pairs and their parents. These sib pairs were selected by their extreme clinical phenotype and their *CFTR* mutation genotype from a set of 318 affected patient pairs,[9,10] accordingly representing the 15% most informative phenotypes of the European CF patient-pair population. CF nuclear families were genotyped at 11 microsatellite markers, chosen from 21 candidate marker sequences on the basis of their informativity and typability, which were selected from the genome database or developed *de novo* from raw genomic sequences spanning 7q31 to 7qtel.

PATIENTS AND METHODS

Informative F508del-*CFTR* homozygous patient pairs

As described in detail previously,[11] we recruited CF twin and sibling pairs and their parents from 158 CF clinics from central Europe. Briefly, we selected dizygous F508del-*CFTR* homozygous CF siblings showing extreme clinical phenotypes as judged by a ranking algorithm that relies on two clinical parameters most sensitive for the course and prognosis of CF, that is, weight as percentage of predicted weight for height, and CF population centiles for the forced expiratory volume in 1 s, expressed as percentage of predicted values.[11] We described concordant mildly affected patient pairs (CON+), concordant severely affected patient pairs (CON−), patient pairs who are concordant with an average disease severity (ND) and discordant sib pairs composed of one mildly and one severely affected sibling (DIS). Compared with the patient

[1]Department of Pediatrics, Hannover Medical School, Hannover, Germany
*Correspondence: Dr F Stanke, Department of Pediatrics OE6710, Hannover Medical School, Carl-Neuberg-Strasse 1, Hannover 30625, Germany.
Tel: +49 511 532 6722; Fax: +49 511 532 6723; E-mail: mekus.frauke@mh-hannover.de

cohort of 37 CON+, CON− and DIS families used to map CF modulators on 12p13 and 16p12,[9] we extended the cohort for the analysis of recombination break points on paternal chromosomes. For the analysis of parent-of-origin-specific decay of genomic sharing presented within this paper, all concordant phenotypes – CON+, CON− and ND – were pooled (CONC) and compared against the discordant patient pairs (DIS) of the European CF sibling population. As our analysis required the knowledge of the parental origin of the chromosomes, only families for whom we were able to recruit at least one parent or families with siblings who shared both their chromosomes along 7q could be enrolled. In total, we evaluated the genotypes of 48 families split into the following clinical phenotypes: 11 CON−, 13 CON+, 5 ND and 19 DIS.

Genotyping
All microsatellite markers were typed using PCR amplification with one biotinylated primer and high-resolution direct blotting electrophoresis with subsequent chemoluminescence detection as described elsewhere (Supplementary Table 1).[9]

Data evaluation
We compared the observed number of recombined and unrecombined chromosomes between concordant and discordant F508del-*CFTR* homozygous CF sib pairs. Statistical evaluation was performed using the program CLUMP by Sham and Curtis,[12] relying on the hypothesis-free permutation analysis by Monte Carlo simulation. Results obtained with the CLUMP program were corrected for multiple testing by Bonferroni taking the seven independent genomic fragments observed on discordant paternal chromosomes (Figure 1) into account. All further case–control and family-based analyses were executed using the FAMHAP software package,[13] which allows family-based analysis[13] and accepts data evaluation in association studies on unrelated individuals, as well as on affected sib pairs.[13] Nuclear families were analyzed using the Monte Carlo simulation-based association test,[14] which can be viewed as an extension of the transmission–disequilibrium test[15] to both nuclear families with more than one affected child and to multimarker haplotypes.

Identification of DNA motifs
Motifs associated with imprinted genes[16] or CTCF-binding sites[17] were identified on NT_079596 with the aligner Mummer.[18] Alignments to shorter consensus sequences of 19–26 bp[16] were rejected if they covered less than 80%, and alignments of longer consensus sequences of 35–48 bp[16] were rejected if they covered less than 65% of the consensus motif.[16] Alignments to the 14 bp degenerate consensus sequence of CTCF[17] were only accepted if the similarities covered 11 consecutive bases without gaps or mismatches.

Bioinformatics: repeats in CpG islands
Global GC content of the entire contig was calculated using the program OligoWords[19] with a window size of 300 bp and step size of 150 bp. Repetitive elements were surveyed using eTandem[20] on the Galaxy server[21] with default settings except for a maximum repeat size of 50 bp. CpG islands were localized using CpGPlot[20] with a minimum island size of 300 bp but otherwise default settings. These two data sets were combined using the Galaxy intersect program. Graphs were plotted using the statistical language R.[22]

WWW resources
The URLs for internet resources and databases presented herein are as follows: Genomic sequences were retrieved from NCBI database resources (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore&itool=toolbar).[23] microRNA sequences and targets were predicted using miRbase at http://microrna.sanger.ac.uk/sequences/.[24,25] CpG islands and repetitive elements were located using CpGPlot and eTandem from the EMBOSS package[20] available on the Galaxy server at http://main.g2.bx.psu.edu/.[21] The Imprinting Gene Catalogue was accessed at http://igc.otago.ac.nz/home.html.[26]
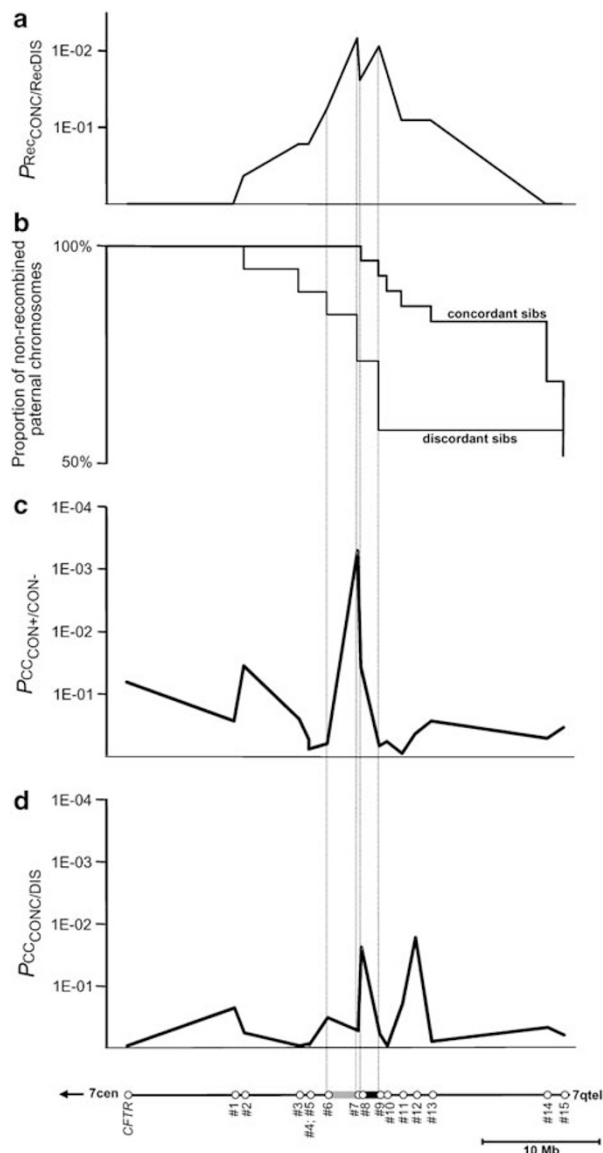

Figure 1 Analysis of parent-of-origin-specific decay of sharing and allelic association at *CFTR* to 7qtel among concordant and discordant sibs. Recombination at markers between *CFTR* and 7qtel (**a** and **b**), as well as allelic association to disease severity (**c**) and intra-pair discordance (**d**), is shown. The physical map is displayed below panel (**d**), whereby the 4.43-Mb candidate region, flanked by markers no. 6 and no. 9, and the 2.87-Mb core candidate region, flanked by markers no. 7 and no. 9, are visualized as gray and black thick lines on the physical map below. Markers no. 6 to no. 9 are indicated by vertical lines throughout this figure. (**a**) Comparison of distributions of recombined and nonrecombined paternal CF chromosomes between concordant and discordant CF patient pairs (uncorrected single locus *P*-values). Bonferroni's correction for multiple testing for observed independent genomic fragments yields $P_{corr}=0.047$ for marker no. 7 and $P_{corr}=0.061$ for marker no. 9. (**b**) Decay of sharing among concordant and discordant F508del-*CFTR* homozygous sib pairs visualized as proportion of nonrecombined paternal chromosomes among concordant and discordant sib pairs. The candidate region and the core candidate region are defined on the basis of this analysis (see text for details). (**c**) case–control (CC) comparison of allele distributions between concordant mildly affected sib pairs (CON+) and concordant severely affected sib pairs (CON−). The minimal $pCC_{CON+/CON−}$ is observed at marker no. 7 (D7Sat3), localized centrally in the candidate region (*P*=0.0005). (**d**) case–control (CC) comparison of allele distributions between concordant (CONC) and discordant (DIS) sib pairs.

RESULTS

Monitoring parent-of-origin-specific decay of genomic sharing on 7q31-7qtel

We genotyped 29 clinically concordant and 19 clinically discordant F508del-*CFTR* homozygous sibling pairs[9,10] at 11 loci spanning a 38-Mb genomic area from the *CFTR* gene on 7q32 to 7qtel. CF is an autosomal recessive trait, and hence all CF siblings obligatorily share both alleles at the *CFTR* locus. Starting with IVS17bTA, a highly informative microsatellite marker within the *CFTR* gene, maternal and paternal chromosomes of both sibs were reconstructed along 7q. Using the information at neighboring markers to interpret loci with a noninformative phase, we implemented four previously typed markers[10] into an integrated 15-marker map (Supplementary Table 1). The resulting microsatellite genotyping data set, consisting of approximately 2100 individual genotypes, was evaluated to compare the parent-of-origin-specific decay of intra-pair genomic sharing along 7q between concordant and discordant sib pairs (Figure 1a and b).

The frequency of recombined and non-recombined maternal chromosomes was similar comparing concordant and discordant pairs ($P > 0.26$ at all loci, data not shown). In contrast, the distribution of nonrecombined and recombined paternal chromosomes was significantly different between concordant and discordant sib pairs at markers no. 6 ($P = 0.057$), no. 7 ($P = 0.007$), no. 8 ($P = 0.025$), no. 9 ($P = 0.009$) and no. 10 ($P = 0.018$) (Figure 1a). After Bonferroni's correction for multiple testing of observed independent genomic fragments, significance was retained at marker no. 7 ($P_{corr} = 0.047$). Recombination events on discordant paternal chromosomes were detected 10 Mb earlier toward 7cen compared with concordant paternal chromosomes (Figure 1b).

Owing to the more proximal onset of recombination on paternal chromosomes of discordant pairs, an elevated frequency of recombined chromosomes on paternal but not on maternal chromosomes was found to be associated with intrapair discordance in the clinical phenotype of F508del-*CFTR* homozygous siblings at several markers on 7q. This association can be explained by the presence of a paternally imprinted CF modifier in the analyzed genomic area, which determines intrapair discordance among CF siblings. To delineate the region that contains the putative imprinted gene(s) and/or their control elements, we described a candidate region, defined by the marker for which the significant onset of decay of sharing was observed among discordant pairs, and a core-candidate region, starting with the marker for which the maximum divergence of sharing was observed comparing the two contrasting phenotypes. Using these definitions, divergence in sharing of the paternal genomic sequence

comparing concordant and discordant sibs points toward a 4.4 Mb candidate region, flanked by markers no. 6 and no. 9, which encompasses an even smaller 2.9 Mb core-candidate region, flanked by markers no. 7 and no. 9 (Figure 1).

Association with CF disease severity is observed within the 2.9-Mb core candidate region

We evaluated the allele distributions between concordant mildly affected, concordant severely affected and discordant patient pairs at all 7q marker loci (Figure 1c and d). Allele frequencies were significantly different at
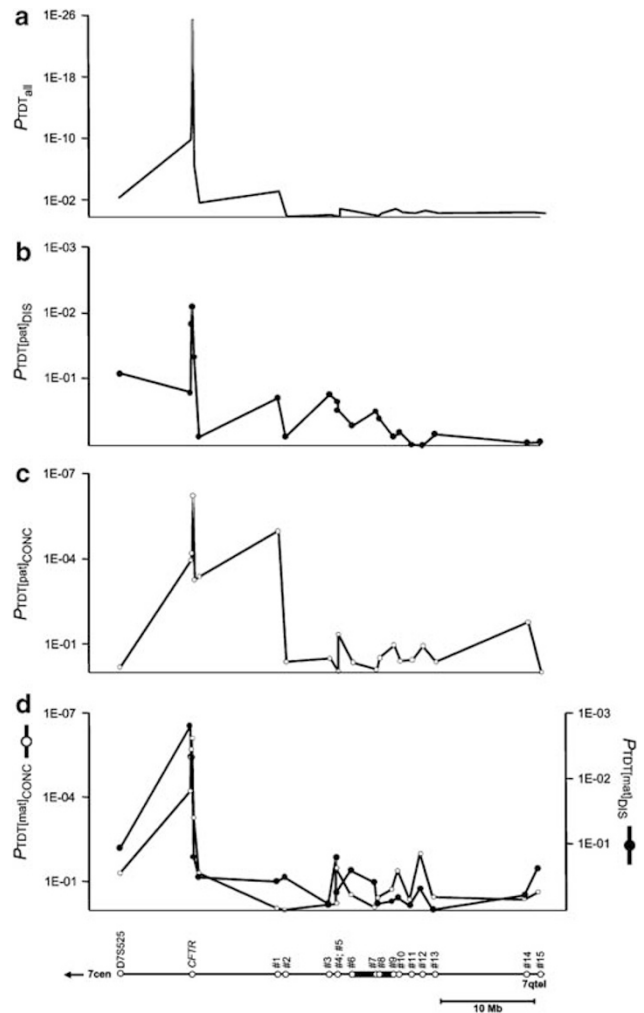

Figure 2 Decay of linkage disequilibrium at *CFTR* to 7qtel in concordant and discordant F508del-*CFTR* families. The transmission disequilibrium test (TDT) was carried out for all families and both parental chromosomes (**a**), restricted to paternal transmissions among discordant (**b**) and concordant (**c**) families and restricted to maternal transmissions among discordant (filled circles) and concordant (open circles) families (**d**) TDT was carried out using the software package FAMHAP with the setting 'haptdt'.[13] The map was extended to encompass the area 3′ to *CFTR* by incorporating D7S525[10] and four SNPs in the 5′ (XV2c, KM19) and 3′ (HUG16RS,[42] J3.11) region of the *CFTR* gene.[10] The finding of a highly significant TDT at the polymorphic microsatellite marker IVS17bTA within the *CFTR* gene of $P = 4.26 \times 10^{-26}$ is trivial, as this represents the imbalance between transmitted F508del-*CFTR* alleles and nontransmitted wt-*CFTR* alleles for the autosomal recessively inherited disease CF. However, the transmission disequilibrium on paternal, but not on maternal chromosomes of concordant pairs is still detectable at D7S514 (Marker1), located at a distance of 10 Mb 3′ to *CFTR* (see also Table 2).

Table 1 Allele frequencies for concordant mildly (CON+) and concordant severely (CON−) affected F508del-*CFTR* homozygous patient pairs at marker no. 7 (D7Sat3)

Allele	Allele Freq.		$PCC_{CON+/CON-}$
	CON+	CON−	
D7Sat3-9	0.060	0.069	
D7Sat3-10	0.158	0.023	
D7Sat3-11	0.003	0.226	0.0005
D7Sat3-12	0.582	0.661	
D7Sat3-13	0.197	0.021	

Alleles at the trinucleotide repeat D7Sat3 are scored by arbitrary repeat units with D7Sat3-9 being the smallest allele observed among the analyzed CF families. Alleles D7Sat3-10, D7Sat3-11 etc refer to a size increase of the main PCR signal by 3 bp per arbitrary repeat unit. Case–control analysis (CC) was carried out using the software package FAMHAP with the settings 'sibvssib' and 'hapcc'[13] ($PCC_{CON+/CON-}$).
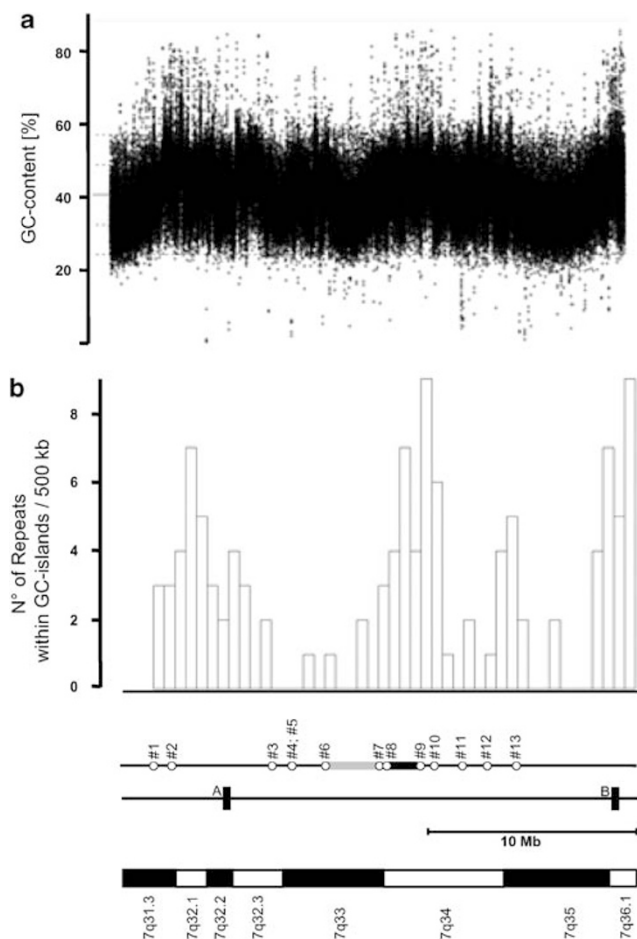
Figure 3 Density of GC-rich repetitive elements on 7q31.3–q36.1. The overall GC content (**a**) and the density of repeats within GC islands (**b**) are shown for a 24-Mb genomic segment on 7q. The 4.43-Mb candidate region is located centrally within this genomic segment (see physical map below (**b**)). (**a**) GC content was averaged for overlapping segments of 300 bp with a step size of 150 bp. (**b**) To display the density of repeats within GC islands, 24 Mb was divided into 48 segments of 500 000 bp each. GC islands were defined as a stretch of 300 bp with an average GC content exceeding 50%. Repetitive elements within GC islands were localized by comparing both data sets. Among the forty-eight 500 kb segments, 18 did not contain any repeats within GC islands. The remaining thirty 500 kb segments contained between 1 and 9 repeats within GC islands. A list of these GC-rich motifs located within the candidate region is displayed in Supplementary Table 3. The position of the 7q32 cluster of imprinted genes[26] *CPA4*, *PEG1/MEST* and its natural antisense transcript *MESTIT*, *COPG2* and its natural antisense transcript *COPG2IT*, and *KLF14* is indicated below the physical map of markers no. 1 to no. 13 by the black box labeled A. The black box labeled B denotes the position of *CNTNAP2* on 7q35, reported as an imprinted gene involved in autism.[26,43] Chromosome bands along 7q are visualized in the cytogenetic map at the bottom of the figure.

Table 2 Transmission disequilibrium at *CFTR* (IVS17bTA) and marker no. 1 (D7S514) among families of concordant and discordant F508del-*CFTR* homozygous CF sib pairs

| | | | PTDT | |
			Concordant	Discordant
116.85 Mb	IVS17bTA	Paternal	6×10^{-7}	7.72×10^{-3}
		Maternal	8×10^{-7}	4.78×10^{-3}
126.04 Mb	D7S514	Paternal	1×10^{-5}	0.189
	(marker no. 1)	Maternal	0.891	0.371

The transmission disequilibrium test (TDT) was carried out using the software package FAMHAP with the setting 'haptdt'[13] (*PTDT*).

disease severity and the paternally imprinted CF modifier predicted by the parent-of-origin-specific decay of genomic sharing might indicate that the underlying genes are the same.

Linkage disequilibrium on 7q differs by phenotype and parental origin

We asked for the decay of linkage disequilibrium on CF chromosomes using the transmission–disequilibrium test (TDT) on all markers along 7q. For maternal transmissions, concordant and discordant families behaved similarly in the TDT, whereas differences between these two phenotypes were observed for paternal transmissions (Figure 2). At marker no. 1, localized at a 10-Mb distance 3′ to *CFTR*, no imbalance between transmitted and nontransmitted chromosomes was observed for paternal transmissions among discordant pairs and maternal transmissions of both phenotypes. In contrast, TDT was still significant at marker no. 1 for paternal transmissions of concordant pairs (Table 2). Although our previous analysis of decay of genomic sharing monitors only the direct effect of the most recent recombination event within the last generation on the patient's phenotype, the analysis by TDT reflects recombination events that have occurred on ancestral chromosomes. The correlation between the concordant phenotype and the TDT signal that is still detectable at a 10-Mb distance to *CFTR* on paternal chromosomes reflects an enrichment of those *CFTR* haplotypes in this highly selected sub-population on which *CFTR* and marker no. 1 are in LD in the present-day CF population. In summary, concordance among CF sibs is observed if the paternal, but not the maternal, genomic region 3′ to *CFTR* is shared among siblings.

Exploratory analysis of the candidate region for imprinted genes

Although paternal imprinting has been reported for the 7q32 genes *MEST*[27] (alias *PEG1*) and *COPG2*,[28] this known imprinted cluster does not colocalize with the candidate region and hence is unlikely to be responsible for the clinical discordance among CF siblings with recombined paternal chromosomes. We have analyzed 7 Mb of genomic sequence of NT_079596 encompassing markers no. 6–no. 10 for elements that have been described to be associated with imprinted genes.

Wang *et al*[16] described 16 consensus motifs of 19–48 bp near or within imprinted genes. Overall, 26 similarities covering at least 80% of the short 19–26 bp consensus sequences or at least 65% of the long 35–48 bp consensus sequences were observed in the analyzed 7-Mb region. Shorter alignments were rejected. In total, matches for 8 of the 16 imprinting gene-associated motifs were recognized on the 7q34 genomic sequence. As Wang *et al*[16] differentiated between upstream-, downstream- and intronic consensus motifs, we have used these annotations to distinguish anchored and unanchored similarities.

marker no. 7 (*P*=0.0005) comparing mildly and severely affected patient pairs (Figure 1c). Two alleles at marker no. 7, D7Sat3-10 and D7Sat3-13, accounted for 35% of CON+ chromosomes, whereas these alleles were observed on less than 5% of CON− chromosomes (Table 1). Less-prominent allelic associations were observed comparing concordant and discordant patient pairs (Figure 1d) at markers no. 8 (*P*=0.024) and no. 12 (*P*=0.017). As marker no. 7 defines the onset of the 2.9 Mb core candidate region, the colocalization of the association signal with CF

Half of the extragenic motifs could not be anchored to any nearby gene (seven anchored and eight unanchored similarities to upstream consensus motifs; two anchored and two unanchored similarities to downstream consensus motifs), whereas the majority of intronic motifs were found within genes (five anchored and two unanchored intronic consensus motifs). In total, 11 genes localized on the 7-Mb fragment were associated with imprinting-related motifs as described by Wang et al[16] and/or predicted to be imprinted by Luedi et al,[29,30] who applied a genome-wide prediction algorithm to identify murine imprinted genes. The growth factor pleiotropin (PTN) and an aldo-keto reductase (AKR1D1), located immediately before and at the start of the core candidate region, were the only paternally imprinted genes within the 7q33-qtel syntenic region of the murine genome,[29] suggesting that, besides the MEST/COPG2-area, an additional chromosomal segment is regulated by paternal imprinting on 7q. Pollard et al[31] provided experimental evidence for imprinting KIAA1466 and HSPC049, both located in the candidate region near marker no. 6. In summary, 13 genes within the 4.4-Mb candidate region have been associated with imprinting by either in silico or experimental evidence (Table 3), although none of these genes seem to be directly connected to the pathophysiology of CF. In other words, on the basis of their annotation, none of these 13 genes is a straightforward candidate for a modifier gene for CF disease severity.

CCCTC-binding factor consensus sites and CpG islands enriched with repetitive elements cluster in the core candidate region
Ishihara et al[17] have described a 14-bp evolutionarily conserved consensus sequence for CCCTC factor (CTCF) binding sites from sequences near H19/IGF2. Even though this consensus binding site is derived from one genomic region only, CTCF is known to act as a universal chromatin insulator, involved in the regulation of mono- and biallelically expressed genes.[32] Similarities to this consensus sequence covering at least 11 bp, whereby no gap or mismatch of the aligned sequence was tolerated, were observed at eight positions in the analyzed 7-Mb genomic sequence of NT_079596. Of these eight motifs, seven are located distal to marker no. 7, defining the onset of the 2.9-Mb core candidate region. In other words, the core candidate region, but not the preceding 3 Mb genomic sequence, is enriched in CTCF consensus sites described for an evolutionarily conserved insulator element of the imprinted H19/IGF2 domain (Supplementary Table 2).

An enrichment of CpG islands with repetitive elements near imprinted genes has been observed.[33–35] As the GC content and hence the density of CpG islands vary considerably throughout the human genome, we have decided to monitor a larger 24-Mb region rather than merely the 7 Mb encompassing the candidate region and flanking sequences. We compared the position of GC islands and the position of tandem repeats, thus being able to interrogate the entire genomic region for uncommon GC-rich repetitive elements (Supplementary Table 3). Three GC-rich repeat regions are observed at 7q32.1–32.2, 7q34 and 7q35–36.1 (Figure 3). The cluster of imprinted genes PEG1/COPG2 is located near the GC-rich tandem repeat enriched region at 7q32.1–32.2. CTNAP2, reported as an imprinted gene involved in autism, is located near the GC-rich tandem repeat enriched region at 7q35–36.1. The third GC-rich tandem repeat enriched region at 7q34 colocalizes with the 2.9-Mb core candidate region for which parent-of-origin-specific decay of genomic sharing predicts a paternally imprinted CF modifier. It is tempting to speculate that these CG-rich repetitive elements are most vulnerable to structural alterations on germline methylation of cytosines, which is the characteristic of all imprinting control elements identified to date.[1–3] Interestingly, genes that show imprinting in

Table 3 Suggestions for imprinted genes on 7q33–7q34

Gene	Annotation	Evidence for imprinting
Marker no. 6 (D7S500)	—	Defines onset of 4.4 Mb candidate region (see Figure 1)
KIAA1466	No annotation	Shows experimental evidence for imprinting at rs292539, rs1254445, rs292540[31]
HSPC049	No annotation	Shows experimental evidence for imprinting at rs292544[31]
LOC346673	Homolog of murine STRA8 (stimulated by retinoic acid gene 8)	Contains intronic consensus motif[16]
LOC641853	Similar to peptidase 26S subunit	Upstream consensus motif[16] 20 K to LOC641853
PTN	Pleiotropin	Upstream consensus motif 126K to PTN paternal imprinting predicted[29]
DGKI	Diacylglycerol kinase	Contains two intronic consensus motifs[16] contains natural antisense transcript LOC646483
Marker no. 7 (D7Sat3)	—	Defines onset of 2.9-Mb core candidate region (see Figure 1)
CREB3L2	cAMP-responsive element binding protein 3 – like 2	Upstream consensus motif[16] 34K to CREB3L2
AKRD1	Aldo-keto-reductase	Paternal imprinting predicted[29]
Marker #8 (D7S495)	—	Within 2.9-Mb core candidate region (see Figure 1)
TRIM24	Transcriptional intermediary factor 1 (alias TIF1)	Upstream consensus motif[16] 44 K to TRIM24
LOC155006	Hypothetical protein	Two upstream consensus motifs[16] 17 and 55 K to LOC155006
HIPK2	Homeodomain-interacting protein kinase 2	Upstream consensus motif[16] 145 K to HIPK2 maternal imprinting predicted[29]
Marker no. 9 (D7Sat12)	—	Defines end of candidate region (see Figure 1)
JHDM1D	Homolog to yeast jumonji C domain containing histone demethylase 1	Intronic consensus motif[16]
BRAF	v-raf murine sarcoma viral oncogene homolog B1	Intronic consensus motif[16]

Loci are listed according to genomic position towards 7qtel. Position of markers no. 6 to no. 9 are shown to assist comparison with the physical map in Figures 1–3. Evidence for imprinting was defined by sequence alignment of a motif associated with imprinted genes described by Wang et al[16] or a prediction by Luedi et al[29] or the observation that the gene is associated with differential allele expression by Pollard et al[31] JHDM1D is a predicted target for hsa-miR-490,[24,25] a micro-RNA encoded in the gene preceding PTN (CHRM2; cholinergic receptor. muscarinic 2).

human, mouse and cattle have repetitive elements that are conserved between the species,[36] supporting the hypothesis that these repeats provide a structural signal that results from the primary sequence of the DNA. Such control elements might provide a robust and easily recognizable signal, whereby the structure is altered on parent-of-origin-specific methylation, in the DNA structure for downstream processes such as binding of proteins and alignment of noncoding RNAs,[37] which establish the permanent silencing of the nonexpressed chromosomal copy of an imprinted gene.[1–3,38]

DISCUSSION

We monitored the parent-of-origin-dependent decay of genomic sharing among CF sib pairs between *CFTR* and 7qtel. Concordant,
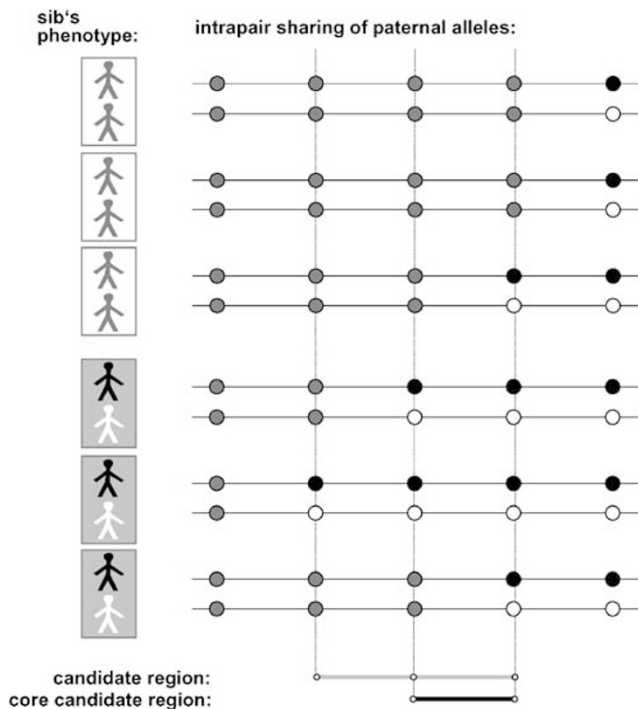

Figure 4 Mapping imprinted genes by parent-of-origin-specific decay of genomic sharing. The principle of the assay is illustrated by three concordant sib pairs (top panel, sibs are shown as gray pictograms) and three discordant sib pairs (bottom panel, sibs are shown as pairs of black and white pictograms). Only one parental chromosome – the paternal chromosome for mapping of a paternally imprinted gene – is shown next to the sibs. Alleles at five adjacent marker loci are visualized as circles, whereby gray color for both sibs of a pair denotes a shared chromosomal segment, and loci depicted in black for one and white for the other sib of a pair denote an unshared genomic segment. Intrapair sharing of the first locus in all sib pairs is obligatory to monitor decay of sharing. For this study on CF sibs at 7q, this locus is represented by the *CFTR* gene, and, as all siblings are F508del-*CFTR* homozygous, this locus is also shared between the pairs, although interpair sharing is not required if the method is applied to other diseases. In other words, application of this method to other regions and diseases than CF and 7q31-qtel requires that a subset of affected pairs be selected from the complete cohort for whom genomic intrapair sharing is ascertained through observed identity by descent of alleles at a marker locus, which is expected for a quarter of any sib pair sample according to Mendelian law. The candidate region, starting with the marker for which significant decay of sharing is observed among discordant pairs, and the core candidate region, starting with the marker for which maximum divergence of sharing between the two contrasting phenotypes is observed, enclose the imprinted gene(s) and/or regulatory elements that control imprinting.

but not discordant, CF sib pairs coinherited a segment of the paternal F508del-*CFTR* chromosome, indicative of a paternally imprinted gene that modifies the course of CF disease. Two independent observations in our study substantiate the description of a paternally imprinted CF modulator on 7q34, which is most likely localized within or controlled by elements mapping to the 2.9-Mb core candidate region: first, the more proximal onset of recombination on paternal chromosomes is associated with a discordant clinical phenotype among CF siblings (Figure 1b), whereas the onset of recombination on maternal chromosomes is similar for CONC and DIS pairs. Second, imprinting-associated elements such as CTCF consensus sites and repeat-rich CpG islands are enriched in the 2.9-Mb core candidate region on 7q34 (Supplementary Table 2, Figure 3). Interestingly, allele distribution at marker D7Sat3, also defining the onset of the core candidate region, differs significantly between mildly and severely affected CF patient pairs (Figure 1, Table 1), which points to the presence of at least one modifier in the 7q34 region. It is tempting to speculate that underlying genetic entities that cause the association signal at D7Sat3 and the parent-of-origin effect that mapped to the core candidate region are the same, although no further evidence than positional overlap can be provided for this hypothesis so far. Besides the research reported in this paper, the larger 7q region has been implicated to contain a paternally imprinted gene responsible for growth retardation by maternal isodisomy 7 and/or Silver Russel Syndrome,[39] and two QTLs for stature have been mapped in the vicinity of the paternally imprinted modulator on 7q34 described here.[40,41] Taken together, these findings indicate that the paternally imprinted 7q34 CF modifier acts on growth and stature.

We propose that the method applied here, that is, monitoring the decay of sharing in phenotypically informative sib pairs using a customized microsatellite map, can aid in the physical fine mapping of imprinted modulators of any disease that is shaped by parent-of-origin effects (Figure 4). As a quarter of sib pairs can be expected to share any genomic locus of interest according to Mendelian law, many large studies on affected patient pairs for diseases such as autism, schizophrenia and bipolar disorder will allow *a priori* classification of clinically concordant and discordant patient pairs and subsequent sampling of subgroups defined by intrapair sharing of parental haplotypes near a locus of interest. Paternal and maternal chromosomes of these selected subsamples of affected patient pairs can be followed up on a dense marker map as accomplished in this study for the paternally imprinted element on 7q34 to delineate novel imprinted genes in the human genome.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Reinhart B, Caillet JR: Genomic imprinting: cis-acting sequences and regional control. *Int Rev Cytol* 2005; **243**: 173–213.
2 Wood AJ, Oakey RJ: Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet* 2006; **2**: e147.
3 Robertson KD: DNA methylation and human disease. *Nat Rev Genet* 2005; **6**: 597–610.
4 Knapp M, Strauch K: Affected-sib-pair test for linkage based on constraints for identical-by-descent distributions corresponding to disease models with imprinting. *Genet Epidemiol* 2005; **26**: 273–285.

5 Wu CC, Shete S, Amos CI: Linkage analysis of affected sib pairs allowing for parent-of-origin effects. *Ann Hum Genet* 2005; **69**: 113–126.

6 Shete S, Amos CI: Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting. *Am J Hum Genet* 2002; **70**: 751–757.

7 Bush A, Alton EWFW, Davies JC, Griesenbach U, Jaffe A: *Cystic Fibrosis in the 21st Century*. Basel: Karger, 2006.

8 Cutting GR: Modifier genetics: cystic fibrosis. *Annu Rev Genomics Hum Genet* 2005; **6**: 237–260.

9 Stanke F, Becker T, Cuppens H *et al*: The TNFalpha receptor *TNFRSF1A* and genes encoding the amiloride-sensitive sodium channel ENaC as modulators in cystic fibrosis. *Hum Genet* 2006; **119**: 331–343.

10 Mekus F, Laabs U, Veeze H, Tümmler B: Genes in the vicinity of *CFTR* modulate the cystic fibrosis phenotype in highly concordant or discordant F508del homozygous sib pairs. *Hum Genet* 2003; **112**: 1–11.

11 Mekus F, Ballmann M, Bronsveld I, Bijman J, Veeze H, Tümmler B: Categories of deltaF508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics. *Twin Res* 2000; **3**: 277–293.

12 Sham PC, Curtis D: Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 1995; **59**: 97–105.

13 Herold C, Becker T: Genetic association analysis with FAMHAP: a major program update. *Bioinformatics* 2009; **25**: 134–136.

14 Knapp M, Becker T: Family-based association analysis with tightly linked markers. *Hum Hered* 2003; **56**: 2–9.

15 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus IDDM. *Am J Hum Genet* 1993; **52**: 506–516.

16 Wang Z, Fan H, Yang HH *et al*: Comparative sequence analysis of imprinted genes between human and mouse to reveal imprinting signatures. *Genomics* 2004; **83**: 395–401.

17 Ishihara K, Sasaki H: An evolutionarily conserved putative insulator element near the 3′ boundary of the imprinted Igf2/H19 domain. *Hum Mol Genet* 2002; **11**: 1627–1636.

18 Kurtz S, Phillippy A, Delcher AL *et al*: Versatile and open software for comparing large genomes. *Genome Biol* 2004; **5**: R1.

19 Reva ON, Tümmler B: Differentiation of regions with atypical oligonucleotide composition in bacterial genomes. *BMC Bioinformatics* 2005; **6**: 251.

20 Rice P, Longden I, Bleasby A: EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000; **16**: 276–277.

21 Giardine B, Riemer C, Hardison RC *et al*: Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 2005; **15**: 1451–1455.

22 R Development Core Team: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2008.

23 Scherer SW, Cheung J, MacDonald JR *et al*: Human chromosome 7: DNA sequence and biology. *Science* 2003; **300**: 767–772.

24 Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: miRbase: microRNAs sequences, targets and gene nomenclature. *Nucl Acids Res* 2006; **34**: D140–D144.

25 Griffiths-Jones S: The microRNA Registry. *Nucl Acids Res* 2004; **32**: D109–D111.

26 Morison IM, Paton CJ, Cleverley SD: The imprinted gene and parent-of-origin effect database. *Nucl Acids Res* 2001; **29**: 275–276.

27 Kobayashi S, Kohda T, Miyoshi N *et al*: Human *PEG1/MEST*, an imprinted gene on chromosome 7. *Hum Molec Genet* 1997; **6**: 781–786.

28 Blagitko N, Schulz U, Schinzel AA, Ropers HH, Kalscheuer VM: Gamma-2-COP, a novel imprinted gene on chromosome 7q32, defines a new imprinting cluster in the human genome. *Hum Molec Genet* 1999; **8**: 2387–2396.

29 Luedi PP, Hartemink AJ, Jirtle RL: Genome-wide prediction of imprinted murine genes. *Genome Res* 2005; **15**: 875–884.

30 Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ: Computational and experimental identification of novel human imprinted genes. *Genome Res* 2007; **17**: 1723–1730.

31 Pollard KS, Serre D, Wang X *et al*: A genome-wide approach to identifying novel-imprinted genes. *Hum Genet* 2008; **122**: 625–634.

32 Hore TA, Deakin JE, Marshall Graves JA: The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. *PLoS Genet* 2008; **4**: e1000169.

33 Hutter B, Helms V, Paulsen M: Tandem repeats in the CpG islands of imprinted genes. *Genomics* 2006; **88**: 323–332.

34 Walter J, Hutter B, Khare T, Paulsen M: Repetitive elements in imprinted genes. *Cytogenet Genome Res* 2006; **113**: 109–115.

35 Daura-Oller E, Cabré M, Montero MA, Paternáin JL, Romeu A: A first-stage approximation to identify new imprinted genes through sequence analysis of its coding regions. *Comp Funct Genomics* 2009; 549387, e-pub ahead of print 8 April 2009. PMID: 19360135.

36 Khatib H, Zaitoun I, Kim ES: Comparative analysis of sequence characteristics of imprinted genes in human, mouse, and cattle. *Mamm Genome* 2007; **18**: 538–547.

37 Mattick JS: The genetic signatures of noncoding RNAs. *PLoS Genet* 2009; **5**: e1000459.

38 Royo H, Cavaillé J: Non-coding RNAs in imprinted gene clusters. *Biol Cell* 2008; **100**: 149–166.

39 Hannuila K, Lipsanen-Nymawn M, Kontiokari T, Kere J: A narrow segment of maternal uniparental disomy of chromosome 7q31-qter in Silver-Russel Syndrome delimits a candidate gene region. *Am J Hum Genet* 2001; **68**: 247–253.

40 Hirschhorn JN, Lindgren CM, Daly MJ *et al*: Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. *Am J Hum Genet* 2001; **69**: 106–116.

41 Perola M, Ohman M, Hiekkalinna T *et al*: Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. *Am J Hum Genet* 2001; **69**: 117–123.

42 Rowntree R, Harris A: DNA polymorphisms in potential regulatory elements of the CFTR gene alter transcription factor binding. *Hum Genet* 2002; **111**: 66–74.

43 Arking DE, Cutler DJ, Brune CW *et al*: A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am J Hum Genet* 2008; **82**: 160–164.