nature genetics

Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome

Sarah B Ng[1,7], Abigail W Bigham[2,7], Kati J Buckingham[2], Mark C Hannibal[2,3], Margaret J McMillin[2], Heidi I Gildersleeve[2], Anita E Beck[2,3], Holly K Tabor[2,3], Gregory M Cooper[1], Heather C Mefford[2], Choli Lee[1], Emily H Turner[1], Joshua D Smith[1], Mark J Rieder[1], Koh-ichiro Yoshiura[4], Naomichi Matsumoto[5], Tohru Ohta[6], Norio Niikawa[6], Deborah A Nickerson[1], Michael J Bamshad[1–3] & Jay Shendure[1]

We demonstrate the successful application of exome sequencing[1–3] to discover a gene for an autosomal dominant disorder, Kabuki syndrome (OMIM%147920). We subjected the exomes of ten unrelated probands to massively parallel sequencing. After filtering against existing SNP databases, there was no compelling candidate gene containing previously unknown variants in all affected individuals. Less stringent filtering criteria allowed for the presence of modest genetic heterogeneity or missing data but also identified multiple candidate genes. However, genotypic and phenotypic stratification highlighted *MLL2*, which encodes a Trithorax-group histone methyltransferase[4]: seven probands had newly identified nonsense or frameshift mutations in this gene. Follow-up Sanger sequencing detected *MLL2* mutations in two of the three remaining individuals with Kabuki syndrome (cases) and in 26 of 43 additional cases. In families where parental DNA was available, the mutation was confirmed to be *de novo* ($n = 12$) or transmitted ($n = 2$) in concordance with phenotype. Our results strongly suggest that mutations in *MLL2* are a major cause of Kabuki syndrome.

Kabuki syndrome is a rare, multiple malformation disorder characterized by a distinctive facial appearance (**Supplementary Fig. 1**), cardiac anomalies, skeletal abnormalities, immunological defects and mild to moderate mental retardation. Originally described in 1981 (refs. 5,6), Kabuki syndrome has an estimated incidence of 1 in 32,000 (ref. 7), and approximately 400 cases have been reported worldwide. The vast majority of reported cases have been sporadic, but parent-to-child transmission in more than a half dozen instances[8] suggests that Kabuki syndrome is an autosomal dominant disorder. The relatively low number of cases, the lack of multiplex families and the phenotypic variability of Kabuki syndrome have made the identification of the gene(s) underlying this disorder intractable to conventional approaches of gene discovery, despite aggressive efforts.

We sequenced the exomes of ten unrelated individuals with Kabuki syndrome: seven of European ancestry, two of Hispanic ancestry and one of mixed European and Haitian ancestry (**Supplementary Fig. 1** and **Supplementary Table 1**). Enrichment was performed by hybridization of shotgun fragment libraries to custom microarrays followed by massively parallel sequencing[1–3]. On average, 6.3 gigabases of sequence were generated per sample to achieve 40× coverage of the mappable, targeted exome (31 Mb). As with our previous studies, we focused our analyses here primarily on nonsynonymous variants, splice acceptor and donor site mutations and coding indels, anticipating that synonymous variants were far less likely to be pathogenic. We also predicted that variants underlying Kabuki syndrome are rare, and therefore likely to be previously unidentified. We defined variants as previously unidentified if they were absent from all datasets used for comparison, including dbSNP129, the 1000 Genomes Project, exome data from 16 individuals previously reported by us[2,3] and 10 exomes sequenced as part of the Environmental Genome Project (EGP).

Under a dominant model in which each case was required to have at least one previously unidentified nonsynonymous variant, splice acceptor and donor site mutation or coding indel variant in the same gene, only a single candidate gene (*MUC16*) was shared across all ten exomes (**Table 1** and **Supplementary Table 2**). However, we considered *MUC16* as a likely false positive due to its extremely large size (14,507 amino acids). Potential explanations for our failure to find a compelling candidate gene in which newly identified variants were seen in all affected individuals included: (i) Kabuki syndrome is genetically heterogeneous and therefore not all affected individuals will have mutations in the same gene; (ii) we failed to identify all mutations in the targeted exome; and (iii) some or all causative mutations were outside of the targeted exome, for example, in noncoding regions or unannotated genes. To allow for a modest degree of genetic heterogeneity and/or missing data, we conducted a less stringent analysis by looking for candidate genes shared among subsets of affected individuals. Specifically, we searched

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [2]Department of Pediatrics, University of Washington, Seattle, Washington, USA. [3]Seattle Children's Hospital, Seattle, Washington, USA. [4]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan. [5]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. [6]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan. [7]These authors contributed equally to this work. Correspondence should be addressed to J.S. (shendure@u.washington.edu) or M.J.B. (mbamshad@u.washington.edu).

Received 28 April; accepted 21 July; published online 15 August 2010; doi:10.1038/ng.646

Table 1 Number of genes common to any subset of *x* affected individuals.

Subset analysis (any *x* of 10)	1	2	3	4	5	6	7	8	9	10
NS/SS/I	12,042	8,722	7,084	6,049	5,289	4,581	3,940	3,244	2,486	1,459
Not in dbSNP129 or 1000 Genomes	7,419	2,697	1,057	488	288	192	128	88	60	34
Not in control exomes	7,827	2,865	1,025	399	184	90	50	22	7	2
Not in either	6,935	2,227	701	242	104	44	16	6	3	1
Is loss-of-function (nonsense or frameshift indel)	753	49	7	3	2	2	1	0	0	0

The number of genes with at least one nonsynonymous variant (NS), splice-site acceptor or donor variants (SS) or coding indel (I) are listed under various filters. Variants were filtered by presence in dbSNP or 1000 Genomes (not in dbSNP129 or 1000 genomes) and control exomes (not in control exomes) or both (not in either); control exomes refer to those from 8 Hapmap[3], 4 FSS[3], 4 Miller[2] and 10 EGP samples. The number of genes found using the union of the intersection of *x* individuals is given.

for subsets of *x* out of 10 exomes having ≥1 previously unidentified variant in the same gene, with $x = 1$ to $x = 10$. For $x = 9$, $x = 8$ and $x = 7$, previously unidentified variants were shared in 3 genes, 6 genes and 16 genes, respectively (**Table 1**). However, there was no obvious way to rank these candidate genes.

We speculated that genotypic and/or phenotypic stratification would facilitate the prioritization of candidate genes identified by subset analysis. Specifically, we assigned a categorical rank to each individual with Kabuki syndrome based on a subjective assessment of the presence of, or similarity to, the canonical facial characteristics of Kabuki syndrome (**Supplementary Fig. 1**) and the presence of developmental delay and/or major birth defects (**Supplementary Table 1**). The highest-ranked individual was one of a pair of monozygotic twins with Kabuki syndrome. We then categorized the functional impact (that is, nonsense versus nonsynonymous substitution, splice-site disruption and frameshift compared to in-frame indel) of each newly identified variant in candidate genes shared by each subset of two or more ranked cases. Manual review of these data highlighted distinct, previously unidentified nonsense variants in *MLL2* in each of the four highest-ranked cases. After sequential analysis of phenotype-ranked cases with a loss-of-function filter, *MLL2* was the only candidate gene remaining after addition of the second individual (**Table 2**). We found no such variant in *MLL2* in the individual with Kabuki syndrome ranked fifth; hence, the number of candidate genes dropped to zero after the individual ranked fourth in the set (**Table 2**). However, we found a 4-bp deletion in the individual ranked sixth, and we found nonsense variants in the individuals ranked seventh and ninth. Thus, exome sequencing identified a nonsense substitution or frameshift indel in *MLL2* in seven of the ten individuals with Kabuki syndrome analyzed here.

Retrospectively, we applied a loss-of-function filter to the subset analysis of exome data (**Table 1**), and at $x = 7$, found *MLL2* to be the only candidate gene. We also developed a *post hoc* ranking of candidate genes based on the functional impact of the variants present (variant score) and the rank of the cases in which each variant was observed (case score). When this was applied to the exome data as a combined metric, *MLL2* emerged as the top candidate gene (**Supplementary Fig. 2**).

In parallel with these analyses, we applied genomic evolutionary rate profiling (GERP)[9] to the exome data. GERP uses mammalian genome alignments to define a rejected substitution score for each variant regardless of functional class. We have previously shown that the quantitative ranking of candidate genes by the rejected substitution scores of their variants can facilitate the exome-based analysis of Mendelian disorders[10]. Following subset analysis with GERP-based ranking, *MLL2* remained on the candidate list up to $x = 8$, ranking third in a list of 11 candidate genes at this threshold (**Table 3** and **Supplementary Fig. 3**). Notably, the additional *MLL2* variant contributing to this analysis (such that *MLL2* was still considered at $x = 8$) was a synonymous substitution with a rejected substitution score of 0.368 in the individual ranked fifth.

We sought to confirm all newly identified variants in *MLL2*, particularly because loss-of-function variants identified through massively parallel sequencing have a high prior probability of being false positives. All seven loss-of-function variants in *MLL2* were validated by Sanger sequencing. We further analyzed the three cases in which we did not initially find a loss-of-function variant in *MLL2*, first by array comparative genomic hybridization (aCGH) to determine any gross structural changes and then by Sanger sequencing of all exons of *MLL2* in case of false negatives by exome sequencing. Because an average of 96% of the coding bases in *MLL2* were called at sufficient quality and coverage for single nucleotide variant detection, we anticipated that any missed variants were more likely to be indels because of the higher coverage required for confident indel detection in short-read sequence data. Indeed, although aCGH did not find any structural variants in the region, Sanger sequencing did identify frameshift indels in two of these three cases (specifically, the cases ranked eighth and tenth).

Ultimately, loss-of-function mutations in *MLL2* were identified in nine out of ten cases in the discovery cohort (**Fig. 1**), making this gene a compelling candidate for Kabuki syndrome. For validation, we screened all 54 exons of *MLL2* in 43 additional cases by Sanger sequencing. Previously unidentified nonsynonymous, nonsense or frameshift mutations in *MLL2* were found in 26 of these 43 cases (**Fig. 1** and **Supplementary Table 3**). In total, through either exome sequencing or targeted sequencing of *MLL2*, 33 distinct *MLL2* mutations were identified in 35 of 53 families (66%) with Kabuki syndrome (**Fig. 1** and **Supplementary Table 3**). In each of 12 cases for which DNA from both parents was available, the *MLL2* variant was found to have occurred *de novo*. Three mutations were found in two individuals each. One of these three mutations was confirmed to have arisen *de novo* in one of the cases, indicating that some mutations in individuals with Kabuki syndrome are recurrent. In addition, *MLL2* mutations (resulting in p.4527K>X and p.5464T>M) were also identified in each of two families in which Kabuki syndrome was transmitted from parent to child.

Table 2 Number of genes common in sequential analysis of phenotypically ranked individuals

Sequential analysis	1	+2	+3	+4	+5	+6	+7	+8	+9	+10
NS/SS/I	5,282	3,850	3,250	2,354	2,028	1,899	1,772	1,686	1,600	1,459
Not in dbSNP129 or 1000 Genomes	687	214	145	84	63	54	42	40	39	34
Not in control exomes	675	134	50	26	13	13	8	5	4	2
Not in either	467	89	34	18	9	8	4	4	3	1
Is loss-of-function (nonsense/frameshift indel)	25	1	1	1	0	0	0	0	0	0

Variants were filtered as in **Table 1**. Exomes were added sequentially to the analysis by ranked phenotype; for example, column "+3" shows the number of genes at the intersection of the three top ranked cases (**Supplementary Fig. 1**). The gene with at least one NS/SS/I in all individuals is *MUC16*, which is very likely to be a false positive due to its extreme length (14,507 amino acids).

Table 3 Analysis of exome variants using genomic evolutionary rate profiling

GERP score analysis (at least x of 10)	1	2	3	4	5	6	7	8	9	10
Variant RS score > 0	7,176	2,360	754	269	106	39	20	11	3	1
MLL2 rank	3,732	1,232	399	136	47	14	6	3	NA	NA

The number of genes with at least a single previously unidentified variant with a rejected substitution score[10] > 0 in at least x individuals is given. A gene rank is assigned based on the average GERP score[9] over all newly identified variants with rejected substitution score > 0 in all affected individuals.

None of the additional *MLL2* mutations was found in 190 control chromosomes from individuals of matched geographical ancestry.

Our results strongly suggest that mutations in *MLL2* are a major cause of Kabuki syndrome. *MLL2* encodes a large 5,262-residue protein that is part of the SET family of proteins, of which Trithorax, the *Drosophila* homolog of MLL, is the best characterized[11]. The SET domain of MLL2 confers strong histone 3 lysine 4 methyltransferase activity and is important in the epigenetic control of active chromatin states[12]. In mice, loss of *Mll2* on a mixed 129Sv/C57BL/6 background slows growth, increases apoptosis and retards development, leading to early embryonic lethality due in part to misregulation of homeobox gene expression[13]. However, no morphological defects have been reported in $Mll2^{+/-}$ mice[13].

Most of the *MLL2* variants identified in individuals with Kabuki syndrome are predicted to truncate the polypeptide chain before translation of the SET domain. Though it is not certain whether Kabuki syndrome results from haploinsufficiency or from a gain of function at *MLL2*, haploinsufficiency seems to be the more likely mechanism. Deletion of chromosome 12q12–q13.2, which encompasses *MLL2*, has been reported in a child with characteristics of Noonan syndrome[14]. However, we re-analyzed this case using oligo aCGH (including 21 probes that cover *MLL2*) and found the distal breakpoint to be located ~700 kb proximal to *MLL2* (data not shown). Also, all of the pathogenic missense variants identified here are located in regions of *MLL2* that encode C-terminal domains. This suggests that missense variants elsewhere in *MLL2* may be better tolerated or, alternatively, may be embryonically lethal.

For the 18 of 53 cases for which no previously unidentified protein-altering variant was found, it is possible that noncoding or other missed mutations in *MLL2* are responsible for this disorder. Alternatively, Kabuki syndrome could be genetically heterogeneous, and further analysis of these cases by exome sequencing may elucidate additional genes for Kabuki syndrome and potentially explain some of the phenotypic heterogeneity seen in this disorder. Notably, 9 of 10 individuals in the discovery cohort (90%), but only 26 of 43 individuals in the replication cohort (60%), were ultimately found to have mutations in *MLL2*. It is therefore possible that the careful selection of canonical Kabuki cases for the discovery cohort enriched for a shared genetic basis. This underscores the importance of access to deeply phenotyped and well-characterized cases.

In summary, we applied exome sequencing of a small number of unrelated individuals with Kabuki syndrome to discover that mutations in *MLL2* underlie this disorder. As predicted in previous analyses[2,3], allowing for even a small degree of genetic heterogeneity or missing data substantially confounds exome analysis by increasing the number of candidate genes consistent with the model of inheritance. To facilitate the prioritization of genes under such criteria, we stratified data by ranked phenotypes and found that *MLL2* was prominent in the higher ranked cases. However, nine of the ten individuals with Kabuki syndrome in the discovery cohort were ultimately found to have *MLL2* mutations, such that stratification by phenotype was of less importance than originally appeared to have been the case. Nonetheless, the sequential analysis of ranked cases may have reduced the probability of confounding due to genetic heterogeneity. All of the *MLL2* mutations found in the discovery set via exome sequencing were loss-of-function variants. As a result, *MLL2* ranked highly among candidate genes assessed by predicted functional impact. Such a pattern will likely occur for some, but not all, Mendelian phenotypes subjected to this approach. We anticipate that the further development of strategies to stratify data at both the genotypic and phenotypic level will be critical for exome and whole-genome sequencing to reach their full potential as tools for discovery of genes underlying Mendelian and complex diseases.

URLs. RefSeq 36.3, ftp://ftp.ncbi.nlm.nih.gov/genomes/MapView/Homo_sapiens/sequence/BUILD.36.3/updates/seq_gene.md.gz; Phaster, http://www.phrap.org; SeattleSeq Annotation, http://gvs.gs.washington.edu/SeattleSeqAnnotation/; 1000 Genomes Project, http://www.1000genomes.org/page.php/; dbGaP accession, http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000295.v1.p1.

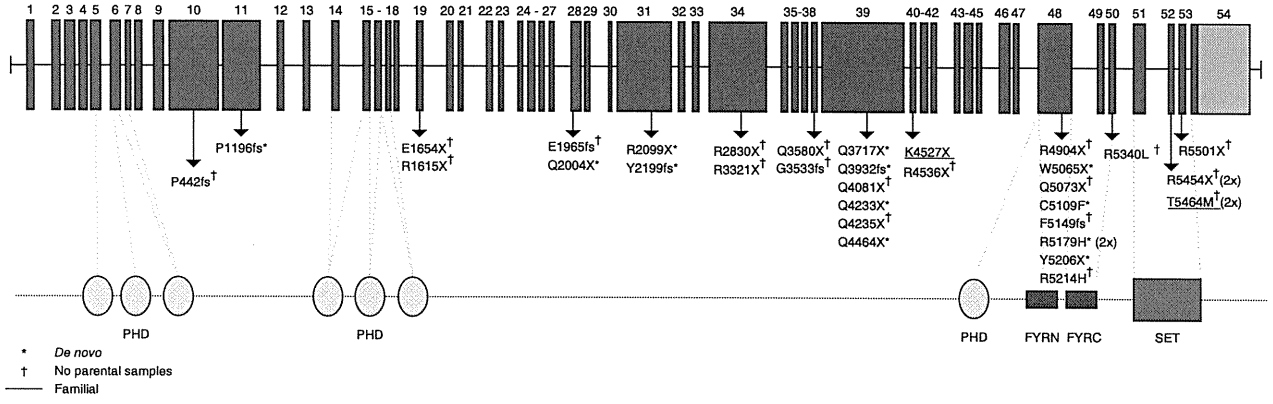

Figure 1 Genomic structure and allelic spectrum of *MLL2* mutations that cause Kabuki syndrome. *MLL2* is composed of 54 exons that encode untranslated regions (orange) and protein coding sequence (blue) including 7 PHD fingers (yellow), FYRN (green), FYRC (green) and a SET domain (red). Arrows indicate the locations of 32 different mutations found in 53 families with Kabuki syndrome including 20 nonsense mutations, 7 indels and 5 amino acid substitutions. Asterisks indicate mutations that were confirmed to be *de novo* and crosses indicate cases for which parental DNA was unavailable. The two underlined mutations were transmitted each within a family, from an affected parent to an affected child.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Exome data for the discovery cohort is available via the NCBI dbGaP repository under accession number phs000295.v1.p1.

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

ACKNOWLEDGMENTS

We thank the families for their participation and the Kabuki Syndrome Network for their support. We thank J. Allanson, J. Carey and M. Golabi for referral of cases and M. Emond for helpful discussion. We thank the 1000 Genomes Project for early data release that proved useful for filtering out common variants. Our work was supported in part by grants from the US National Institutes of Health (NIH)–National Heart, Lung, and Blood Institute (5R01HL094976 to D.A.N. and J.S.), the NIH–National Human Genome Research Institute (5R21HG004749 to J.S., 1RC2HG005608 to M.J.B., D.A.N. and J.S.; and 5RO1HG004316 to H.K.T.), NIH–National Institute of Environmental Health Sciences (HHSN273200800010C to D.N. and M.J.R.), Ministry of Health, Labour and Welfare (K.Y., N.M., T.O. and N.N.), Japan Science and Technology Agency (N.M.), Society for the Promotion of Science (N.M.), the Life Sciences Discovery Fund (2065508 and 0905001), the Washington Research Foundation and the NIH–National Institute of Child Health and Human Development (1R01HD048895 to M.J.B.). S.B.N. is supported by the Agency for Science, Technology and Research, Singapore. A.W.B. is supported by a training fellowship from the NIH–National Human Genome Research Institute (T32HG00035).

AUTHOR CONTRIBUTIONS

The project was conceived and the experiments were planned by M.J.B., D.A.N. and J.S. The review of phenotypes and the sample collection were performed by M.J.B., M.C.H., M.J.M., K.Y., N.M., T.O. and N.N. Experiments were performed by S.B.N., K.J.B., A.E.B., C.L., H.C.M., J.D.S., M.J.R., E.H.T. and H.I.G. Ethical consultation was provided by H.K.T. Data analysis was performed by A.W.B., M.J.B., K.J.B., G.M.C., S.B.N. and J.S. The manuscript was written by M.J.B., S.B.N. and J.S. All aspects of the study were supervised by M.J.B. and J.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101 (2009).
2. Ng, S.B. *et al.* Exome sequencing identifies the cause of a Mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
3. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
4. FitzGerald, K.T. & Diaz, M.O. MLL2: A new mammalian member of the *trx/MLL* family of genes. *Genomics* **59**, 187–192 (1999).
5. Niikawa, N., Matsuura, N., Fukushima, Y., Ohsawa, T. & Kajii, T. Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. *J. Pediatr.* **99**, 565–569 (1981).
6. Kuroki, Y., Suzuki, Y., Chyo, H., Hata, A. & Matsui, I. A new malformation syndrome of long palpebral fissures, large ears, depressed nasal tip, and skeletal anomalies associated with postnatal dwarfism and mental retardation. *J. Pediatr.* **99**, 570–573 (1981).
7. Niikawa, N. *et al.* Kabuki make-up (Niikawa-Kuroki) syndrome: a study of 62 patients. *Am. J. Med. Genet.* **31**, 565–589 (1988).
8. Courtens, W., Rassart, A., Stene, J.J. & Vamos, E. Further evidence for autosomal dominant inheritance and ectodermal abnormalities in Kabuki syndrome. *Am. J. Med. Genet.* **93**, 244–249 (2000).
9. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
10. Cooper, G.M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat. Methods* **7**, 250–251 (2010).
11. Prasad, R. *et al.* Structure and expression pattern of human *ALR*, a novel gene with strong homology to *ALL-1* involved in acute leukemia and to *Drosophila* trithorax. *Oncogene* **15**, 549–560 (1997).
12. Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* **27**, 1889–1903 (2007).
13. Glaser, S. *et al.* Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**, 1423–1432 (2006).
14. Tonoki, H., Saitoh, S. & Kobayashi, K. Patient with del(12)(q12q13.12) manifesting abnormalities compatible with Noonan syndrome. *Am. J. Med. Genet.* **75**, 416–418 (1998).

ONLINE METHODS

Cases and samples. For exome sequencing, we selected ten individuals of self-reported European, Hispanic or mixed European and Haitian ancestry with Kabuki syndrome from ten unrelated families. Phenotypic data were collected from review of medical records, phone interviews and photographs. All participants provided written consent, and the Institutional Review Boards of Seattle Children's Hospital and the University of Washington approved all studies. The clinical characteristics of the 43 individuals in the validation cohort who had been diagnosed with Kabuki syndrome have been reported previously[7]. Subjective assessment and ranking of the Kabuki phenotype was based on pictures of each subject (**Supplementary Fig. 1**) and clinical information (**Supplementary Table 1**). Informed consent was obtained for publication of each of the facial photos shown.

Exome definition, array design and target masking. We targeted all protein-coding regions as defined by RefSeq 36.3. Entries were filtered for the following: (i) CDS as the feature type, (ii) transcript name starting with "NM_" or "-", (iii) reference as the group_label, (iv) not being on an unplaced contig (for example, 17|NT_113931.1). Overlapping coordinates were collapsed for a total of 31,922,798 bases over 186,040 discontiguous regions. A single custom array (Agilent, 1M features, aCGH format) was designed to have probes over these coordinates as previously described[3], except here, the maximum melting temperature (T_m) was raised to 73 °C.

The mappable exome was also determined as previously described[3] using this RefSeq exome definition instead. After masking for 'unmappable' regions, 30,923,460 bases were left as the mappable target.

Targeted capture and massive parallel sequencing. Genomic DNA was extracted from peripheral blood lymphocytes using standard protocols. Five micrograms of DNA from each of ten individuals with Kabuki syndrome was used for construction of a shotgun sequencing library as described previously[3] using paired-end adaptors for sequencing on an Illumina Genome Analyzer II (GAII). Each shotgun library was hybridized to an array for target enrichment; this was then followed by washing, elution and additional amplification. Enriched libraries were then sequenced on a GAII to get either single-end or paired-end reads.

Read mapping and variant analysis. Reads were mapped and processed largely as previously described[3]. In brief, reads were quality recalibrated using Eland and then aligned to the reference human genome (hg18) using Maq. When reads with the same start site and orientation were filtered, paired-end reads were treated like separate single-end reads; this method is overly conservative and hence the actual coverage of the exomes is higher than reported here. Sequence calls were performed using Maq and these calls were filtered to coordinates with ≥8× coverage and consensus quality ≥ 20.

Indels affecting coding sequences were identified as previously described[3], but we used phaster instead of cross_match and Maq. Specifically, unmapped reads from Maq were aligned to the reference sequence using phaster (version 1.100122a) with the parameters -max_ins:21 -max_del:21 -gapextend_ins:-1 -gapextend_del:-1 -match_report_type:1. Reads were then filtered for those with at most two substitutions and one indel. Reads that mapped to the negative strand were reverse complemented and, together with the other filtered reads, were remapped using the same parameters to reduce ambiguity in the called indel positions. These reads were then filtered for (i) having a single indel more than 3 bp from the ends and (ii) having no other substitutions in the read. Putative indels were then called per individual if they were supported by at least two filtered reads that started from different positions. An 'indel reference' was generated as previously described[3], and all the reads from each individual were mapped back to this reference using phaster with default settings and -match_report_type:1. Indel genotypes were called as previously described[3].

To determine the novelty of the variants, sequence calls were compared against 16 individuals for whom we had previously reported exome data[2,3] and 10 EGP exomes. Annotations of variants were based on NCBI and UCSC databases using an in-house server (SeattleSeqAnnotation). Loss-of-function variants were defined as nonsense mutations (premature stop) or frame-shifting indels. For each variant, we also generated constraint scores as implemented in GERP[10].

***Post hoc* ranking of candidate genes.** Candidate genes were ranked by summation of a case score and variant score. The case score was calculated by counting the total number of Kabuki exomes in which a variant was identified at a given gene, weighted for case rank from 1 to 10. For example, the top ranked case was weighted by a factor of 10, whereas the case ranked tenth was weighted by a factor of 1. The variant score was calculated by first counting the total number of nonsense, nonsynonymous and synonymous variants across the ten Kabuki exomes and assigning a prior probability of the occurrence of each variant type per gene based upon the target of 18,918 genes. Next, for each candidate gene shared among two or more Kabuki exomes, the scores for each newly identified variant were summed across the gene. The case score and variant score were summed as the candidate gene score.

Mutation validation. Sanger sequencing of PCR amplicons from genomic DNA was used to confirm the presence and identity of variants in the candidate gene identified via exome sequencing and to screen the candidate gene in additional individuals with Kabuki syndrome.

Array comparative genomic hybridization (CGH). Samples were hybridized to commercially available whole-genome tiling arrays consisting of one million oligonucleotide probes with an average spacing of 2.6 kb throughout the genome (SurePrint G3 Human CGH Microarray 1x1M, Agilent Technologies). Twenty-one probes on this array covered *MLL2* specifically. Data were analyzed using Genomics Workbench software according to the manufacturer's instructions.

doi:10.1038/ng.646

RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART A

Two Missense Mutations of the IRF6 Gene in Two Japanese Families With Popliteal Pterygium Syndrome

Noriko Matsuzawa,[1,2]* Shinji Kondo,[3] Kazuo Shimozato,[2] Toru Nagao,[1,2] Motoi Nakano,[4] Masayoshi Tsuda,[4,5] Akiyoshi Hirano,[4] Norio Niikawa,[6,7] and Koh-ichiro Yoshiura[5,7]

[1]Department of Oral and Maxillofacial Surgery, Okazaki City Hospital, Okazaki, Japan

[2]Department of Maxillofacial Surgery, Aichi-Gakuin University School of Dentistry, Nagoya, Japan

[3]Department of Clinical Pharmacy, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

[4]Department of Plastic and Reconstructive Surgery, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

[5]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

[6]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Tobetsu, Japan

[7]Solution Oriented Research of Science and Technology (SORST), Japan Science and Technology Agency (JST), Tokyo, Japan

Received 27 February 2009; Accepted 3 January 2010

Mutations in the interferon regulatory factor 6 gene (*IRF6*) cause either popliteal pterygium syndrome (PPS) or Van der Woude syndrome (VWS), allelic autosomal dominant orofacial clefting conditions. To further investigate the *IRF6* mutation profile in PPS, we performed mutation analysis of patients from two unrelated Japanese families with PPS and identified mutations in *IRF6*: c.251G>T (R84L) and c.1271C>T (S424L). We also found R84L, which together with previous reports on R84 mutations, provided another line of evidence that both syndromes could result from the same mutation probably under an influence of a modifier gene(s). This supports the idea that the R84 residue in the DNA binding domain of *IRF6* is a mutational hot spot for PPS. A luciferase assay of the S424L protein in the other family demonstrated that the mutation decreased the *IRF6* transcriptional activity significantly to 6% of that of the wild-type. This finding suggests that the C-terminus region of IRF6 could have an important function in phosphorylation or protein interaction. To our knowledge, this is the first report of mutations observed in Japanese PPS patients. © 2010 Wiley-Liss, Inc.

Key words: *IRF6*; gene mutation; clinical spectrum; transcriptional activity

How to Cite this Article:
Matsuzawa N, Kondo S, Shimozato K, Nagao T, Nakano M, Tsuda M, Hirano A, Niikawa N, Yoshiura K-i. 2010. Two missense mutations of the IRF6 gene in two Japanese families with popliteal pterygium syndrome.
Am J Med Genet Part A 152A:2262–2267.

INTRODUCTION

Popliteal pterygium syndrome (PPS, OMIM #119500) and Van der Woude syndrome (VWS, OMIM #119300) are allelic, autosomal dominant orofacial cleft syndromes. Clinical manifestations of PPS are diverse, ranging from orofacial anomalies such as lower lip pits, cleft lip and/or palate and syngnathia to skin and genital abnormalities, including webbing of the lower limbs, syndactyly, hypoplasia of the labia majora and bifid or absent scrotum [Gorlin et al., 1968; Bixler et al., 1973]. VWS is characterized mainly by cleft lip and/or palate, and lower lip pits [Schinzel and Klausler, 1986]. The incidence of PPS is estimated to be 1/300,000 [Froster-Iskenius, 1990], while VWS is more frequent and its incidence in the Finnish population some 20 years ago was about 1/33,600 [Cheney et al., 1986]. Recently, Kondo et al. [2002] discovered mutations in the interferon regulatory factor 6 gene (*IRF6*) causing either PPS or

Grant sponsor: JST, Japan (SORST); Grant sponsor: Scientific Research (Priority Area "Applied Genomics"); Grant number: 17019055; Grant sponsor: Specially Promoted Research; Grant numbers: 17019056, 17790225; Grant sponsor: Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan; Grant sponsor: Scientific Research from the Ministry of Health, Labour and Welfare.

Noriko Matsuzawa and Shinji Kondo contributed equally to this work.

*Correspondence to:
Noriko Matsuzawa, DDD, PhD, Department of Oral and Maxillofacial Surgery, Okazaki City Hospital, Koryuji-cho 3-1, Goshoai, Okazaki, Aichi 444-8553, Japan. E-mail: mnoriko51@yahoo.co.jp

Published online 18 August 2010 in Wiley Online Library (wileyonlinelibrary.com)
DOI 10.1002/ajmg.a.33338

VWS. The gene product of *IRF6* belongs to the *IRF* family of transcriptional factors. All IRF proteins are characterized by two domains that consist of the highly conserved winged-helix DNA binding domain containing a penta-tryptophan motif and the less conserved protein binding domain termed SMAD/IRF (SMIR) or interferon association domain (IAD) [Eroshkin and Mushegian, 1999]. IRFs function not only to regulate interferon response after viral infection [Taniguchi et al., 2001], but also to regulate cell growth, differentiation, and apoptosis [Tanaka et al., 1994; Tamura et al., 1995; Holtschke et al., 1996; Harada et al., 1998; Lohoff et al., 2000; Bailey et al., 2008]. Among IRFs, the biological function of *IRF6* seems unique, but its involvement in viral infections during the early developmental stage, especially for PPS/VWS, remains unknown. Most *IRF6* mutations found in PPS are missense mutations located at R84 within the DNA binding domain or at surrounding amino acid residues [Kondo et al., 2002; Peyrard-Janvid et al., 2005], with one mutation, D430N, being reported in the SMIR domain. Mutations in VWS, however, occur across the entire coding region and consist of missense or truncation-type mutations [Kondo et al., 2002; Kayano et al., 2003; Wang et al., 2003; Matsuzawa et al., 2004, 2006; Peyrard-Janvid et al., 2005; Ye et al., 2005; Du et al., 2006; Tan et al., 2008; de Medeiros et al., 2008].

We studied two Japanese families, one of which had a patient with PPS and three other members with cleft lip with or without palate, and the other with a mother and son both affected with PPS. Here we report the results of mutation analysis in the two families and functional study of a mutated allele identified in one family. To our knowledge, this is the first report of mutations observed in Japanese PPS patients.

MATERIALS AND METHODS

Patients and Families

Family 1. Four PPS affected individuals appeared in three generations (Fig. 1A). The proband (III-2), a 1-year-old boy, had lip pits, bilateral cleft lip with palate, syngnathia, syndactyly as well as popliteal webbing, and atrophic testes (Fig. 2a–d). He had received a surgical treatment for syngnathia. His father had bilateral cleft lip with palate but no other signs of PPS. Both a paternal grandfather (I-1) and an aunt (II-2) had bifid uvula in addition to cleft palate but no lip pits. Variable expressivity was thus clinically evident in this family. Other individuals including his mother were phenotypically normal. There was no evidence that the mother had had a viral infection during early pregnancy with her son. The pedigree was constructed by physical examinations of the individuals concerned and by interviewing the parents of the proband.

Family 2. Two PPS affected members were present in two successive generations (Fig. 1B). The male proband (III-1) showed cleft palate, syngnathia, unclear scrotum, and syndactyly as well as mild popliteal webbing (Fig. 2e–g). His mother (II-1) was reported previously by Hirano et al. [1994], and had cleft lip and palate, syngnathia, syndactyly and hypoplasia of the labia majora (Fig. 2h). The proband had received an operation for the syngnathia. Other members of this family were phenotypically normal. The pedigree was constructed by physical examination of the individuals concerned and by interviewing a family member.

Blood samples were collected under informed consent from the proband and his parents in Family 1 and from the

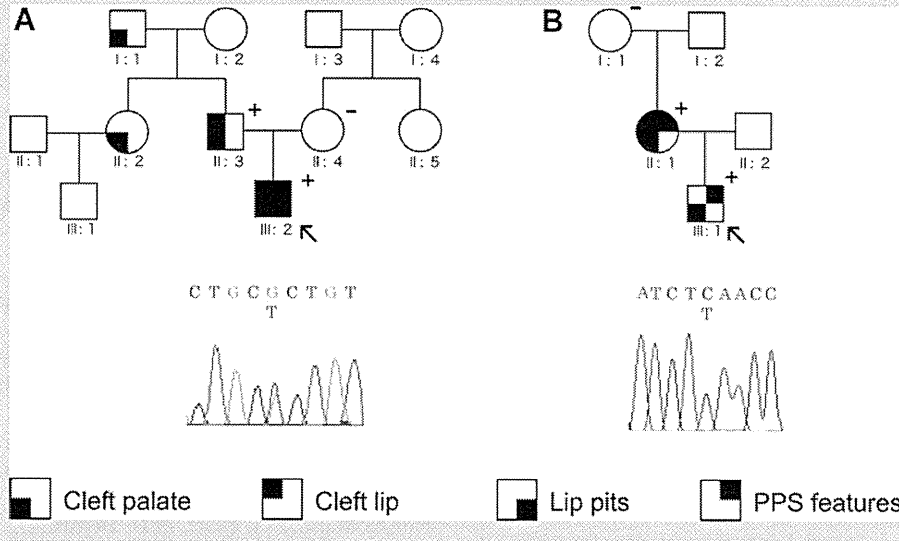

FIG. 1. Pedigrees, clinical findings and mutations of Families 1 (A) and 2 (B) with PPS. A: PPS Family 1, (B) PPS Family 2. Symbols "+" and "—" denote individuals with and without a mutation on DNA analysis, respectively. Electropherogram of DNA derived from each proband (arrow) is shown below the pedigree.

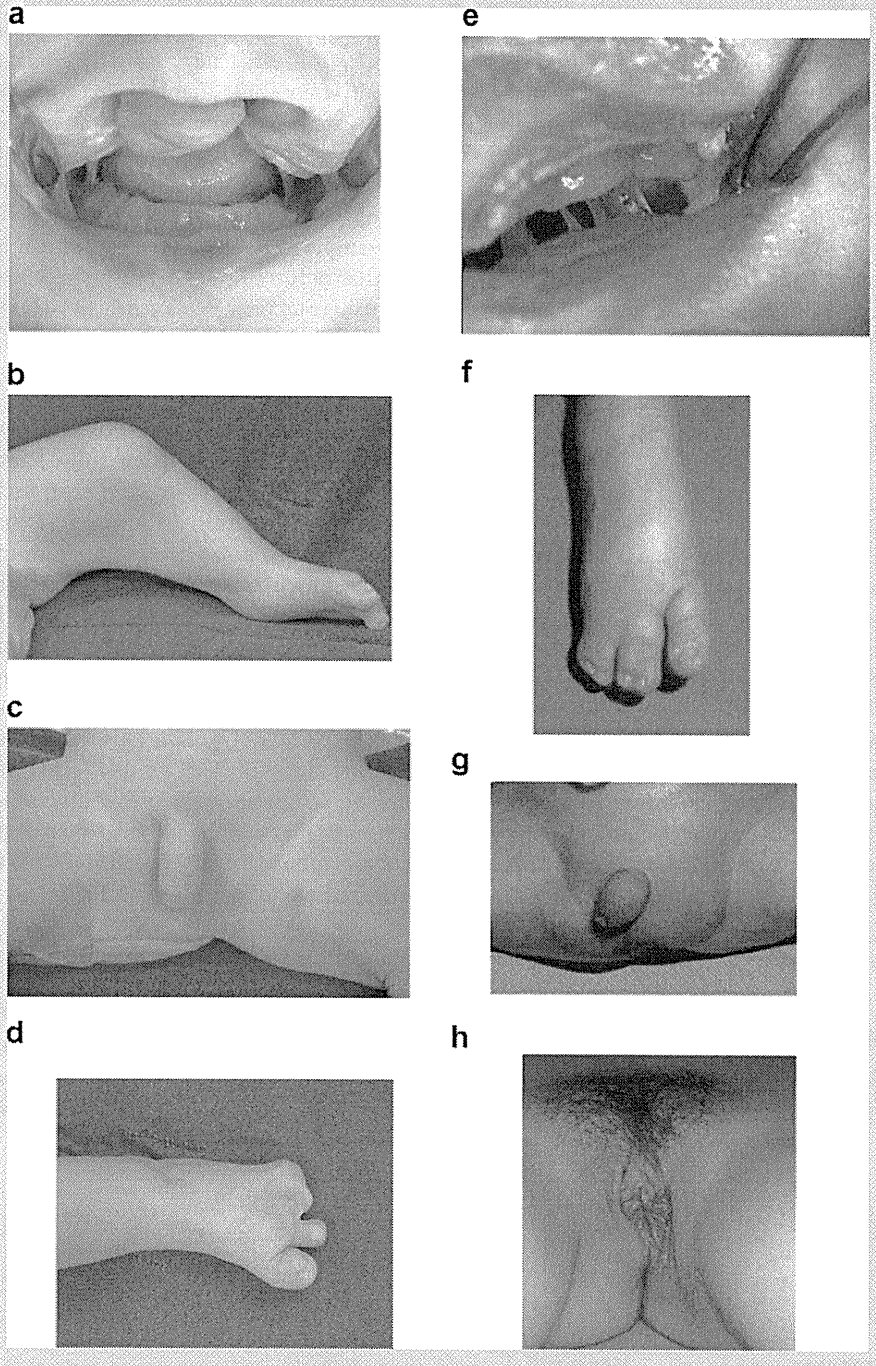

FIG. 2. Phenotypic appearance of the proband (a–d) from Family 1, and the proband (e–g) and his mother (h) from Family 2. a: Bilateral cleft lip and palate, lower lip pits, syngnathia; (b) popliteal web; (c) absent scrotum; (d) syndactyly; (e) syngnathia; (f) syndactyly; (g) unclear scrotum; (h) hypoplastic labia majora of mother.

proband, his parents and maternal grandmother in Family 2. All experimental procedures were approved by the Committees for the Ethical Issues on Human Genome and Gene Analysis in Aichi-Gakuin University and in Nagasaki University.

IRF6 Mutation Search

Screening for mutations in *IRF6* was performed in individuals from the two families. All exons and exon–intron boundaries of *IRF6* were amplified by PCR using primer pairs designed from the

genomic sequence. PCR was performed in a 10 μl mixture containing 5 ng of genomic DNA, 1 μM each primer, 200 μM each dNTPs, 0.3 U TaKaRa ExTaq HS version (TaKaRa, Kyoto, Japan), and 10× PCR buffer supplied by TaKaRa. PCR conditions were as follows; initial incubation at 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 30 sec, elongation at 72°C for 30 sec, and final elongation at 72°C for 7 min. PCR products were treated with ExoSAP-IT (GE Healthcare, Buckinghamshire, England) following the supplier's instruction manual, and then sequenced directly using BigDye Terminator ver.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Sequenced samples were purified with Sephadex G-50 (GE Healthcare) and run on an automated sequencer Model 3100 (Applied Biosystems). Sequence electropherograms were aligned using AutoAssembler software v2.1 (Applied Biosystems) to find a base alteration by visual inspection.

Construction of Plasmid Expressing *IRF6*

We generated plasmids expressing the wild-type (pS424) or mutant (pL424) *IRF6* by site-directed mutagenesis using the Genetailor Site-Directed Mutagenesis System (Invitrogen, Carlsbad, CA). To express a GAL4-IRF6 fusion protein, we fused the DNA sequence for IRF6 C-terminus portions (amino acid positions 114 to 467) to GAL4 DNA binding domain (DBD). We used pBIND-DEST expressing vector containing the GAL4-DNA binding domain in Gateway system (Invitrogen). Thus, each of the recombinant plasmids (pS424 and pL424) contained the sequence for the GAL4 -DBD and the IRF6 interaction domain with Ser or Leu at the 424th amino acid residue (Fig. 3a). All clones were verified for their integrity by sequencing, and their expressions were verified by western blot analysis with a polyclonal anti-IRF6 antibody (Active Motif, Carlsbad, CA) (data not shown).

Cell Culture, Transfection, and Luciferase Assay

We cultured 293T cells in Dulbecco's MEM (Sigma-Aldrich, St. Louis, MO) containing 10% FBS at 37°C in 5% CO_2. We transiently transfected expression plasmids (pS424 or pL424) to 293T cells with a luciferase-expression reporter plasmid driven by the GAL promoter, and a βgal-expression plasmid to normalize the transfection efficiency using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, soluble protein was extracted and luciferase activity was assayed by the Luciferase Assay System (Promega, Madison, WI) using Galacto-Light Plus (Applied Biosystems). Transcriptional activation by the GAL4DBD-IRF6 fusion protein was calculated as the ratio of luciferase activity in the sample to that of a positive control GAL4-DBD plasmid, after normalization to the βgal activity. Statistical analysis was carried out using StatView Version 5.0, and a *P*-value <0.05 was considered to be statistically significant.

RESULTS

The proband and his father from Family 1 showed a missense mutation, c.251G>T, in *IRF6*-exon 4 (Fig. 1A). This nucleotide substitution results in an amino acid change in the DNA binding domain from a polar charged basic arginine residue into a hydrophobic nonpolar leucine (R84L). No such mutation was present in either the mother or among 90 healthy Japanese individuals.

In the proband from Family 2, a missense mutation, c.1271C>T, was found in the last exon (exon 9) giving a predicted amino acid change from a polar uncharged serine to a hydrophobic nonpolar leucine (S424L) (Fig. 1B). This mutation was found in his affected mother, but neither in his unaffected maternal grandmother nor among 200 healthy Japanese individuals. The luciferase assay demonstrated that the S424L protein decreased the *IRF6* transcriptional promotion activity significantly to 6% of that of the wild-type S424 protein ($P < 0.05$) (Fig. 3b).

DISCUSSION

We have identified two mutations, R84L and S424L, in two Japanese families affected by PPS. Both mutations were hitherto undescribed, although R84H in two families and R84C in five families have previously been reported to be associated with PPS [Kondo

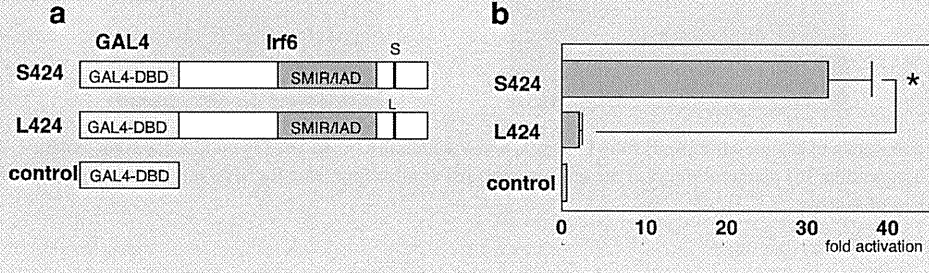

FIG. 3. Luciferase assay for transcriptional activity of mutated IRF6 protein by transient transfection of S424 into 293T cells. a: Illustration of mouse Irf6 protein (aa114–467) fused in frame to GAL4-DBD. The position of the residue 424 is indicated by S or L. GAL4-DBD, the GAL4 DNA binding domain; SMIR/IAD, SMAD/IRF or the interferon association domain. b: Luciferase value (fold activation) for each construct was measured as its ratio to a value measured for vector alone, after normalization to the co-tansfected βgal activity. The values represent the average of five experiments with standard deviation (error bars). *$P < 0.05$.

et al., 2002]. As R84 was assumed to be a critical site for DNA binding [Kondo et al., 2002], it is likely that R84L observed in Family 1 causes a change of DNA binding ability of *IRF6*, and the site R84 in the IRF6 protein is a mutational hot spot.

In Family 2, we identified another heterozygous base substitution, c.1271C>T, in the last exon of *IRF6*, and this mutation results in S424L at the C-terminus region of IRF6 protein. The S424L is a rare missense mutation found within the C-terminus domain, where only one mutation (D430N) in PPS [Kondo et al., 2002] and two mutations (P396S and R400W) in VWS [Kayano et al., 2003; Wang et al., 2003] have been reported. These findings suggest that the C-terminus region of IRF6 could have an important function, and the S424L identified in Family 2 could be associated with that function. A luciferase assay showed that the S424L mutation significantly reduces the transcriptional activity of IRF6, and two possible functions of the C-terminus are suggested. The region may act as an important domain for phosphorylation of *IRF6* via a kinase to allow correct localization as occurs in other IRFs. It has been suggested that S424 is a potential target residue of a kinase, although details are unknown [Bailey et al., 2005]. Studies on IRF3 and IRF7 have shown that phosphorylation at the C-terminus is required to unfold and dimerize them, and that the dimer is then translocated into the nucleus, where it binds with their targets and activates gene expression [Mamane et al., 1999; Sharma et al., 2003]. Alternatively, the C-terminus region could play a role in protein interaction or its regulation. This possibility is supported by a study showing an interaction of IRF6 to the mammary serine protease inhibitor (maspin) [Bailey et al., 2005]. We favor this second possibility, because a relationship between the phosphorylation state of IRF6 and subcellular localization was not supported either by immunofluorescent staining [Bailey et al., 2008] or by luciferase assay using GAL4-DBD [Little et al., 2009], and the present luciferase assay using GAL4-DBD suggests that the S424L mutation abolished transcriptional activation of mutant IRF6. However, why different mutations surrounding the SMIR domain result in either the PPS or the VWS phenotype still remains to be investigated.

The molecular mechanisms resulting in PPS and VWS and the genotype–phenotype correlations between them remain unclear. A possible explanation why mutations in the same gene cause two different syndromes is the nature of the mutations. One idea is that haploinsufficiency of *IRF6* causes VWS, while PPS develops by a dominant negative mechanism or by a situation of less than half insufficiency [Kondo et al., 2002]. According to this hypothesis, mutated IRF6 at R84 found in PPS would completely abolish the transcriptional activator function without any disturbance of localization or dimerization regulation. The R84 mutations found in Family 1 and in seven unrelated PPS families [Kondo et al., 2002] support this hypothesis. Likewise, mutations in the C-terminus region resulting in PPS, such as S424L seen in Family 2, would abolish completely the transcriptional activity. A quantitative assay reflecting intrinsic *IRF6* transcriptional activity should provide an answer. Alternatively, as the two disorders cannot always be differentiated clearly [Khan et al., 1986], they may fall into a wide clinical spectrum, resulting in little genotype–phenotype correlation.

Variable expressivity within a family is also evident, as seen in Family 1, where only the proband had typical clinical signs for PPS while other affected members had only cleft palate with or without cleft lip. Clinical severity is very different between typical PPS and VWS patients. Although these findings can be explained by an assumption that the two conditions represent each end of the spectrum, it is also plausible that various mutated IRF6 proteins would show consecutive intrinsic transcriptional activity under an influence of a modifier gene(s) and the PPS/VWS phenotypes would be distinctively categorized through thresholds according to the level of activity. All five mutations (R9W, R45Q, R84C, E349V, and P396S) reported in Japanese VWS patients [Kayano et al., 2003; Matsuzawa et al., 2004] are located at either the DNA binding domain or the SMIR domain, and there is no evidence for the existence of a Japanese-specific mutation spectrum. However, as we previously found R84C to be quite common [Matsuzawa et al., 2004], some genetic factors modifying IRF6 function might exist more frequently in the Japanese.

In conclusion, two mutations associated with PPS have been identified, c.251G>T (R84L) and c.1271C>T (S424L), one each in the two families studied. The nature of the mutations suggests R84 and S424 are biologically important sites for DNA binding ability and for transcriptional activation of IRF6, respectively. To our knowledge, this is the first report of mutations observed in Japanese PPS patients.

ACKNOWLEDGMENTS

The participation of the patients in this study is highly appreciated. We thank Ms. Y. Noguchi, K. Miyazaki, C. Hayashida, and M. Ohga for their technical assistance. SORST from JST, Japan for N. N.; Grants-in-Aid for Scientific Research (Priority Area "Applied Genomics," No. 17019055) for N. N., and for Specially Promoted Research (Nos. 17019056 and 17790225) for S. K. from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. K.Y. was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare.

REFERENCES

Bailey CM, Khalkhali-Ellis Z, Kondo S, Margaryan NV, Seftor REB, Wheaton WM, Amir S, Pins MR, Schutte BC, Hendrix MJC. 2005. Mammary serine protease inhibitor (maspin) binds directly to interferon regulatory factor 6. J Biol Chem 280:34210–34217.

Bailey CM, Abbott DE, Margaryan NV, Khalkhali-Ellis Z, Hendrix MJC. 2008. Interferon regulatory factor 6 promotes cell cycle arrest and is regulated by the proteasome in a cell cycle-dependent manner. Mol Cell Biol 28:2235–2243.

Bixler D, Poland C, Nance WE. 1973. Phenotypic variation in the popliteal pterygium syndrome. Clin Genet 4:220–228.

Cheney ML, Cheny WR, Lejeune FE. 1986. Familial incidence of labial pits. Am J Otolaryngol 7:311–313.

de Medeiros F, Hansen L, Mawlad E, Eiberg H, Asklund C, Tommerup N, Jakobsen LP. 2008. A novel mutation in IRF6 resulting in VWS-PPS spectrum disorder with renal aplasia. Am J Med Genet Part A 146A: 1605–1608.

Du X, Tang W, Tian W, Li S, Li X, Liu L, Zheng X, Chen X, Lin Y, Tang Y. 2006. Novel IRF6 mutations in Chinese patients with Van der Woude syndrome. J Dent Res 85:937–940.

Eroshkin A, Mushegian A. 1999. Conserved transactivation domain shared by interferon regulatory factors and Smad morphogens. J Mol Med 77:403–405.

Froster-Iskenius UG. 1990. Popliteal pterygium syndrome. J Med Genet 27:320–326.

Gorlin RJ, Sedano HO, Cervenka J. 1968. Popliteal pterygium syndrome. A syndrome comprising cleft lip-palate, popliteal and intercrural pterygia, digital and genital anomalies. Pediatrics 41:503–509.

Harada H, Taniguchi T, Tanaka N. 1998. The role of interferon regulatory factors in the interferon system and cell growth control. Biochimie 80:641–650.

Hirano A, Iio Y, Murakami R, Fujii T. 1994. Recurrent trismus: Twenty-year follow-up result. Cleft Palate Craniofac J 31:309–312.

Holtschke T, Lohler J, Kanno Y, Fehr T, Giese N, Rosenbauer F, Lou J, Knobeloch KP, Gabriele L, Waring JF, Bachmann MF, Zinkernagel RM, Morse HC III, Ozato K, Horak I. 1996. Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. Cell 87:307–317.

Kayano S, Kure S, Suzuki Y, Kanno K, Aoki Y, Kondo S, Schutte BC, Murray JC, Yamada A, Matsubara Y. 2003. Novel IRF6 mutations in Japanese patients with Van der Woude syndrome: Two missense mutations (R45Q and P396S) and a 17-kb deletion. J Hum Genet 48:622–628.

Khan SN, Hufnagle KG, Pool R. 1986. Intrafamilial variability of popliteal pterygium syndrome: A family description. Cleft Palate 23:233–236.

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC. 2002. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32:285–289.

Little HJ, Rorick NK, Su LI, Baldock C, Malhotra S, Jowitt T, Gakhar L, Subramanian R, Schutte BC, Dixon MJ, Shore P. 2009. Missense mutation that cause Van der Woude syndrome and popliteal pterygium syndrome affect the DNA-binding and transcriptional activation functions of IRF6. Hum Mol Genet 18:535–545.

Lohoff M, Duncan GS, Ferrick D, Mittrücker HW, Bischof S, Prechtl S, Röllinghoff M, Schmitt E, Pahl A, Mak TW. 2000. Deficiency in the transcription factor interferon regulatory factor (IRF)-2 leads to severely compromised development of natural killer and T helper type 1 cells. J Exp Med 192:325–336.

Mamane Y, Heylbroeck C, Genin P, Algarté M, Servant MJ, LePage C, DeLuca C, Kwon H, Lin R, Hiscott J. 1999. Interferon regulatory factors: The next generation. Gene 237:1–14.

Matsuzawa N, Yoshiura K, Machida J, Nakamura T, Niimi T, Furukawa H, Toyoda T, Natsume N, Shimozato K, Niikawa N. 2004. Two missense mutations in the IRF6 gene in two Japanese families with Van der Woude syndrome. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 98: 414–417.

Matsuzawa N, Shimozato K, Natsume N, Niikawa N, Yoshiura K. 2006. A novel missense mutation in Van der Woude syndrome: Usefulness of fingernail DNA for genetic analysis. J Dent Res 85:1143–1146.

Peyrard-Janvid M, Pegelow M, Koillinen H, Larsson C, Fransson I, Rautio J, Hukki J, Larson O, Karsten AL, Kere J. 2005. Novel and de novo mutations of the IRF6 gene detected in patients with Van der Woude or popliteal pterygium syndrome. Eur J Hum Gene 13:1261–1267.

Schinzel A, Klausler M. 1986. The Van der Woude syndrome (dominantly inherited lip pits and cleft). J Med Genet 23:291–294.

Sharma S, tenOever BR, Grandvaux N, Zhou GP, Lin R, Hiscott J. 2003. Triggering the interferon antiviral response through an IKK-related pathway. Science 300:1148–1151.

Tamura T, Ishihara M, Lamphier MS, Tanaka N, Oishi I, Aizawa S, Matsuyama T, Mak TW, Taki S, Taniguchi T. 1995. An IRF-1-dependent pathway of DNA damage-induced apoptosis in mitogen-activated T lymphocytes. Nature 376:596–599.

Tan EC, Lim EC, Yap SH, Lee ST, Cheng J, Por YC, Yeow V. 2008. Identification of IRF6 gene variants in three families with Van der Woude syndrome. Int J Mol Med 21:747–751.

Tanaka N, Ishihara M, Kitagawa M, Harada H, Kimura T, Matsuyama T, Lamphier MS, Aizawa S, Mak TW, Taniguchi T. 1994. Cellular commitment to oncogene-induced transformation or apoptosis is dependent on the transcription factor IRF-1. Cell 77:829–839.

Taniguchi T, Ogasawara K, Takaoka A, Tanaka N. 2001. IRF family of transcription factors as regulators of host defense. Annu Rev Immunol 19:623–655.

Wang X, Liu J, Zhang H, Xiao M, Li J, Yang C, Lin X, Wu Z, Hu L, Kong X. 2003. Novel mutations in the IRF6 gene for Van der Woude syndrome. Hum Genet 113:382–386.

Ye XO, Jin HX, Shi LS, Fan MW, Song GT, Fan HL, Bian Z. 2005. Identification of novel mutations of IRF6 gene in Chinese families with Van der Woude syndrome. Int J Mol Med 16:851–856.

PRENATAL DIAGNOSIS
Prenat Diagn 2010; **30**: 849–861.
Published online 27 July 2010 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/pd.2570

The possibility of microarray-based analysis using cell-free placental mRNA in maternal plasma

Kiyonori Miura[1]*, Shoko Miura[1], Kentaro Yamasaki[1], Takako Shimada[1], Akira Kinoshita[2], Norio Niikawa[3], Koh-ichiro Yoshiura[2] and Hideaki Masuzaki[1]

[1]*Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan*
[2]*Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan*
[3]*Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan*

Objective The purpose of this study is to investigate a possibility of overall assessment of cell-free (CF) placental mRNAs in maternal plasma.

Methods First, placenta-predominantly expressed transcripts were selected by the analysis of GeneChip using three sets of placental tissues and corresponding maternal blood cells. Subsequently, a custom cDNA array panel of placenta-predominantly expressed transcripts was designed and used to compare the RNA profiles of maternal plasma collected from 12 preeclamptic and 12 uncomplicated pregnancies. Scatter plots for the signal intensities of the comparative cDNA hybridization revealed either unchanged or aberrant patterns.

Results We selected top 50 placenta-predominantly expressed transcripts that were >2500 times higher in placental tissues than in corresponding whole blood samples. A custom cDNA array analysis detected the aberrant pattern in five preeclamptic women with severe hypertension but not in seven preeclamptic women with mild hypertension ($P < 0.05$, Fisher's direct method). The aberrant pattern of above RNA transcripts in maternal plasma was validated by quantitative real-time reverse transcription-polymerase chain reaction. The mean (range) value of coefficient of variations in this custom array quantification was 9.4% (3.0–16.2%).

Conclusion Our custom cDNA array is expected to be useful for overall assessment of CF placental mRNAs in maternal plasma in a single experiment. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS: microarray-based analysis; cell-free placental mRNA; maternal plasma; placental status; preeclampsia; noninvasive diagnosis

INTRODUCTION

Assessment of the fetal/placental status during pregnancy is currently carried out by fetal cardiotocography, ultrasonography and/or tests using biological marker molecules, but they often produce false-positive results (Bobby, 2003). To obtain direct information regarding the fetal/placental status, conventional prenatal diagnostic procedures such as amniocentesis, cordocentesis or fetal scalp blood sampling may remain useful. However, these invasive procedures always involve the risk of serious complications, for example fetal loss, rupture of the membrane and infections. Therefore, noninvasive detection of fetus or placenta-derived molecular markers in pregnant women is desirable for accurate monitoring of the fetal/placental status. The recent discovery of cell-free (CF) placental mRNA in maternal plasma has provided possibilities for exploring placental dysfunction (Lo and Chiu, 2007; Maron and Bianchi, 2007). CF placental transcripts can be detected in maternal plasma by week 4 of gestation and have a median half-life of 14 min (Chiu *et al.*, 2006). The pregnancy specificity of CF placental mRNA has been demonstrated by its rapid clearance from the maternal plasma after delivery. Therefore, its detection in the maternal circulation appears to be a promising approach for the development of sex- and polymorphism-independent fetal/placental molecular markers for prenatal gene expression profiling. CF placental mRNA in maternal plasma is measurable by quantitative real-time reverse transcription-polymerase chain reaction (RT-PCR) (Miura *et al.*, 2008), and placental transcripts are more readily detectable in plasma than in whole blood (Heung *et al.*, 2009).

The placental functions during pregnancy are regulated by a large variety of genes. Preeclampsia is a serious complication of pregnancy, and its pathogenesis is known to be associated with multiple factors, such as abnormal placentation, reduced placental perfusion, endothelial cell dysfunction and systemic vasospasm. The plasma concentrations of circulating CF placental mRNA including transcripts for corticotrophin-releasing hormone, placenta-specific gene 1 and selectin P were reported to be higher for pregnant women with preeclampsia than for normal pregnant women (Lo and Chiu, 2007; Maron and Bianchi, 2007; Purwosunu *et al.*, 2007). These findings suggest that certain different kinds of CF placental mRNA in maternal plasma can be used as markers for preeclampsia. However, overall assessment of CF placental mRNA by a single quantitative

*Correspondence to: Dr Kiyonori Miura, Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, 1-7-1 Sakamoto, Nagasaki 852-8501, Japan.
E-mail: kiyonori@nagasaki-u.ac.jp

Received: 23 February 2010
Revised: 13 April 2010
Accepted: 10 May 2010
Published online: 27 July 2010

real-time RT-PCR analysis is difficult because of the limitation of the blood sampling volume. To resolve this technical issue, the development of a high-throughput microarray-based approach is promising. In addition, microarray comparative genomic hybridization analysis of CF fetal DNA in amniotic fluid already provides rapid screening for chromosome abnormalities with copy-number changes (Larrabee *et al.*, 2004; Miura *et al.*, 2006), though amniotic fluid cannot be obtained noninvasively. The application of a microarray-based method using CF placental mRNA in maternal plasma is feasible and may allow a noninvasive overall assessment of the placental status in a single experiment.

In this study, we initially identified placenta-predominantly expressed transcripts in maternal blood by comparisons between their expression levels in the placenta and maternal blood (Tsui *et al.*, 2004). Subsequently, we generated a custom cDNA microarray panel composed of the cDNAs of these genes. Finally, we adopted a custom array-based comparative cDNA hybridization using CF mRNA in maternal plasma to establish an overall assessment of CF placental mRNA levels in pregnant women.

MATERIALS AND METHODS

Subjects and sample collection

The study subjects comprised 12 pregnant women with preeclampsia and 12 normal pregnant women whose gestational weeks were matched with those of the preeclamptic women. CF plasma samples (12 mL) were prepared from maternal blood by a double centrifugation method. All these women attended the Department of Obstetrics and Gynecology at the Nagasaki University Hospital. Preeclampsia was defined as gestational hypertension (systolic pressure $\geq$140 mm Hg and/or diastolic blood pressure $\geq$90 mm Hg on at least two occasions during measurements after 20 weeks of gestation) with proteinuria ($\geq$0.3 g/day). Severe preeclampsia was defined by the presence of one or more of the following findings: severe gestational hypertension (systolic pressure $\geq$160 mm Hg and/or diastolic blood pressure $\geq$110 mm Hg on two occasions at least 6 h apart while the patient was on bed rest); severe proteinuria ($\geq$5 g of protein in a 24-h urine specimen, or 3 or greater in two random urine samples collected at least 4 h apart). Women in whom the onset of preeclampsia was before 32 weeks of gestation were defined as having the early-onset type, while those with onset after 32 weeks of gestation were defined as having the late-onset type.

Placental tissue samples were obtained from three normal pregnant women immediately after termination of pregnancy during the first (7–12 weeks of pregnancy), second (18–21 weeks) or third (37–39 weeks) trimesters (Figure 1), immediately placed in RNAlater™ (Ambion, Austin, TX) and stored at −80 °C until RNA extraction. Blood samples (6 mL) were collected from the women into PAXgene™ blood RNA tubes (PreAnalytiX, Hombrechtikon, Switzerland) before the termination of pregnancy at each trimester. All the samples

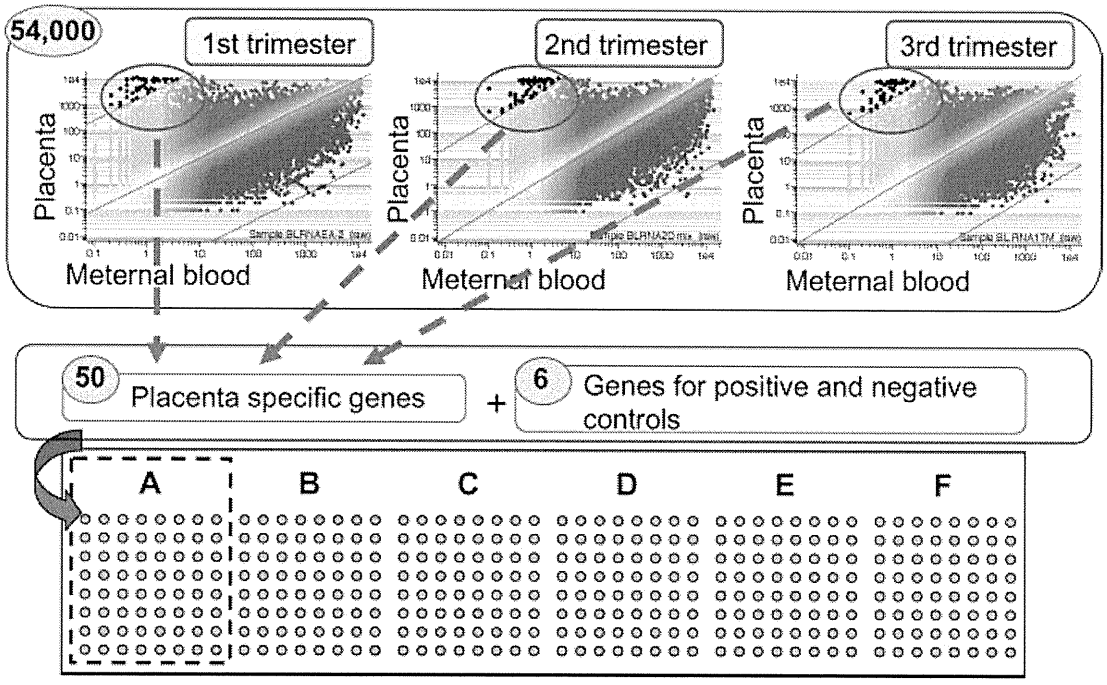

Figure 1—Outline of the strategy used for the development of a custom microarray panel of placenta-specific genes in maternal blood. Upper panel: gene expression profiles of three paired samples of placenta tissues and corresponding whole maternal blood (at the first, second and third trimesters). Black dots indicate placenta-specific transcripts exhibiting signal intensities that are >2500 times higher in the placenta in every trimester compared with the corresponding whole maternal blood samples. Middle panel: collection of 50 placenta-specific genes plus 6 positive and negative control genes. Lower panel: a custom cDNA microarray panel where 6 blocks containing a total of 56 gene-specific sequences were robotically spotted onto a single slide. The 50 placenta-specific genes were spotted in the order of their high expression levels in the placenta

DOI: 10.1002/pd

were obtained after receiving written informed consent, and the study protocol was approved by the Institutional Review Board for Ethical, Legal and Social Issues at the Nagasaki University.

Identification of placenta-predominantly expressed transcripts in the maternal circulation

Total RNA from the placental tissue samples and the peripheral blood were extracted as described, respectively, by Tsui *et al.* (2004) and Okazaki *et al.* (2007). The same amount of RNA from the placental or corresponding maternal blood sample was subjected to GeneChip® analysis (Affymetrix, Santa Clara, CA). Because the expression pattern of placental transcripts is changing during pregnancy, three sets (placental tissue and corresponding maternal blood sample) at each trimester were analyzed. For each sample, the extracted RNA (1 μg) was labeled by the one-cycle target labeling method, and biotinylated cDNA (15 μg) was hybridized to a GeneChip® Human Genome U133 Plus 2.0 (Affymetrix) according to the manufacturer's instructions. After the hybridization, the arrays were washed and stained in a GeneChip Fluidics Station 450 (Affymetrix). The chips were scanned with a GeneChip Scanner 3000 (Affymetrix) and analyzed using GeneChip Microarray Suite 5.0 (Affymetrix). The raw intensity data were analyzed with the GeneChip Operating Software (Affymetrix), and data mining was performed with the GeneSpring software (Agilent Technologies, Palo Alto, CA). The corrected signal intensities for the placental tissues and blood samples were expressed as scatter plots (Figure 1). Three sets of placental tissue and maternal blood cell samples across the three trimesters were used to look for the placenta-predominantly expressed genes that were detected in one or more out of the three profiles. The top 50 placenta-predominantly expressed transcripts exhibiting raw signal intensities that were >2500 times higher in the placental tissues than in the corresponding whole blood samples were selected according to a previously reported strategy for systematic identification of pregnancy-specific placental mRNA markers expressed in the maternal circulation (Tsui *et al.*, 2004).

Development of a custom microarray panel of placenta-predominantly expressed genes

Preparation of probe cDNAs

The 50 placenta-predominantly expressed transcripts (Nos. 1–50) and six control transcripts (Nos. 51–56) were selected as probe cDNAs to be spotted onto a custom microarray panel (Table 1, Figure 1). The control transcripts included five placenta-nonspecific genes as positive controls (Nos. 51–55) and the plasmid sequence of the pBluescriptII SK(−) vector as a negative control (No. 56). The 50 primer sets for the placenta-specific genes and 5 primer sets for the placenta-nonspecific genes (TaqMan® Gene Expression Assays) were purchased from Applied Biosystems (Foster City, CA, USA) (each assay ID is listed in Table 1). RT-PCR amplifications of the placenta mRNA samples were carried out, and the PCR products were sequenced to confirm the accuracy of each gene-specific primer set. As the primer sets for *CSH1* and *HERV-FRD* failed to amplify their exact cDNA sequences, custom primers were made for these genes as described previously (Ng *et al.*, 2003b; Okahara *et al.*, 2004). Each placental cDNA aliquot (2 μL) was subjected to 30 cycles of PCR in a total volume of 20 μL. The PCR product was cloned into the TOPO II vector (Invitrogen (Carlsbad, CA, USA)), and the plasmid DNA was extracted. Several clones for each transcript were sequenced using an ABI3100 to confirm the sequence integrity, and a total of 55 clones, 1 for each from the placenta-specific and nonspecific transcripts, were selected.

As probe DNA resources for the microarray panel, the sequences integrated into the TOPO II vector were amplified with primers that were modified to have sequences for an amino residue ($-NH_2$) at their 5′ ends. The primer sequences were as follows: topo-forward: 5′-agtgtgctggaattcgccct-3′; topo-reverse: 5′-gatatctgcagaattcgccct-3′. After purification using a Microcon YM-10/30/100 (Millipore Corporation, Billerica, MA), the concentration of the collected DNA was 138–580 ng/μL and each volume was 50–100 μL. The PCR amplifications involved 30 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s in a 100-μL mixture containing 1 ng of cDNA, 10 pM of primers (topo-forward and topo-reverse), 250 μM dNTP, 0.5 U of Ex Taq polymerase (TaKaRa Bio Inc., Tokyo, Japan) and 10× PCR buffer (TaKaRa Bio Inc.). PCR amplification was performed four times for each cDNA clone.

DNA spotting

The PCR products were dissolved in distilled water, and an equal volume of spotting solution DSP0050 (Matsunami, Osaka, Japan) was added. The final concentration of the PCR products was 0.15 μg/μL. The resulting DNA samples (0.15 μg/μL at 200 pL/spot) were robotically spotted using an inkjet printing technique (NGK, Nagoya, Japan) in six blocks onto CodeLink™ activated slides (Amersham Biosciences, Piscataway, NJ). The design of the custom cDNA microarray is shown in Figure 1.

Microarray-based comparative cDNA hybridization between samples from preeclamptic women and control pregnant women

cDNA preparation from the women and T7-based CF mRNA amplification

CF mRNA was extracted as described previously (Zhong *et al.*, 2008) and amplified by a T7-based RNA amplification method (Abe *et al.*, 2003), because the concentration of CF mRNA in maternal plasma was too low for direct use in a microarray analysis (Figure 2). A total of 56 specific primers [RNA amplification primers (RAPs)]

Table 1—Fifty genes highly expressed in the placenta compared with the maternal blood

	Gene symbol (name)	GenBank accession number	Assay ID	Chromosome localization	Gene expression[a]	Gene function	CV value[b] (custom array) (%)	CV value (real-time RT-PCR) (%)
1	LUM (lumican)	NM_002345.3	Hs00158940_m1	12q21.3-q22	P	A member of the small leucine-rich proteoglycan family	12.0	1.7
2	RAI14 (retinoic acid induced 14)	NM_015577.1	Hs00210238_m1	5p13.3-p13.2	P	Retinoic acid regulated gene	14.9	3.3
3	CDH1 [cadherin 1, type 1, E-cadherin (epithelial)]	NM_004360.2	Hs00170423_m1	16q22.1	P	Calcium ion-dependent cell adhesion molecule	7.8	5.8
4	CSRP2 (cysteine and glycine-rich protein 2)	NM_001321.1	Hs00426717_m1	12q21.1	P	Member of the CRP (cysteine- and glycine-rich protein) family of LIM (Lin/Isl/Mec) domain proteins	13.2	4.8
5	ERVWE1 [endogenous retroviral family W, env(C7), member 1, syncytin 1]	NM_014590.3	Hs00205893_m1	7q21-q22	P	Part of an HERV provirus, which is expressed in the placental syncytiotrophoblast and is involved in fusion of the cytotrophoblast cells to form the syncytial layer of the placenta	10.2	3.1
6	INHBA (inhibin, beta A)	NM_002192.2	Hs00170103_m1	7p15-p13	P	Growth/differentiation hormone	17.2	4.7
7	PSG5 (pregnancy-specific beta-1-glycoprotein 5)	NM_002781.2	Hs00818332_m1	19q13.2	P	A group of molecules that are mainly produced by the placental syncytiotrophoblasts during pregnancy	12.1	10.6
8	TFPI (tissue factor pathway inhibitor)	NM_001032281.2	Hs00196731_m1	2q31-q32.1	P	A protease inhibitor that regulates the tissue factor-dependent pathway of blood coagulation	6.9	6.0
9	INSL4 (insulin-like 4)	NM_002195.1	Hs00171411_m1	9q24	P	A member of the insulin superfamily	9.1	4.6
10	LEP (leptin)	NM_000230.1	Hs00174877_m1	7q31.3	P	Secreted by adipocytes and by placenta	13.0	9.1
11	TFPI2 (tissue factor pathway inhibitor 2)	NM_006528.2	Hs00197918_m1	7q22	P	A placental glycoprotein that inhibits plasmin, trypsin and thrombin	7.3	8.9
12	GH1 (growth hormone 1, transcript variant 1)	NM_000515.3	Hs00236859_m1	17q24.2	P	A member of the somatotrophin/prolactin family of hormones which play an important role in growth control	13.1	2.6
13	ADAM12 (a disintegrin and metalloproteinase domain 12)	NM_003474	Hs01106104_m1	10q26.3	P	Candidate regulator of trophoblast fusion	12.7	5.2
14	ANGPT2 (angiopoietin 2)	NM_001147.1	Hs00169867_m1	8p23.1	P	An antagonist of angiopoietin 1 (ANGPT1) and endothelial TEK (endothelial-specific receptor tyrosine kinase) tyrosine kinase (TIE-2 (tunica internal endothelial cell kinase 2), TEK)	10.9	6.1
15	GH1 (growth hormone 1), transcript variant 4	NM_022561.2	Hs00737955_m1	17q24.2	P	A member of the somatotrophin/prolactin family of hormones that play an important role in growth control	11.4	7.6
16	PSG9 (pregnancy-specific beta-1-glycoprotein 9)	NM_002784.2	Hs00358192_m1	19q13.2	P	A group of molecules that are mainly produced by the placental syncytiotrophoblasts during pregnancy	16.2	7.7
17	CGA (glycoprotein hormones, alpha polypeptide)	NM_000735.2	Hs00174938_m1	6q12-q21	P	The alpha subunit and belongs to the glycoprotein hormones alpha chain family	12.0	5.2
18	KISS1 (KiSS-1 metastasis suppressor)	NM_002256.2	Hs00158486_m1	1q32	P	A putative role in the regulation of events downstream of cell-matrix adhesion	8.9	2.8

Table 1—(*Continued*)

	Gene symbol (name)	GenBank accession number	Assay ID	Chromosome localization	Gene expression[a]	Gene function	CV value[b] (custom array) (%)	CV value (real-time RT-PCR) (%)
19	CAPN6 (calpain 6)	NM_014289.2	Hs00560073_m1	Xq23	P	Ubiquitous, well-conserved family of calcium-dependent, cysteine proteases	12.5	6.1
20	PSG6 (pregnancy-specific beta-1-glycoprotein 6)	NM_001031850.1	Hs00747417_m1	19q13.2	P	A group of molecules that are mainly produced by the placental syncytiotrophoblasts during pregnancy	9.9	6.3
21	TIMP3 (TIMP metallopeptidase inhibitor 3)	NM_000362.4	Hs00165949_m1	22q12.1-q13.2	P	Inhibitors of the matrix metalloproteinases	5.8	3.3
22	FBLN1 (fibulin 1)	NM_001996.2	Hs00242546_m1	22q13.3	P	A secreted glycoprotein that becomes incorporated into a fibrillar extracellular matrix	13.6	9.4
23	PRG2 (plasticity-related gene 2)	NM_002728.4	Hs00794928_m1	19p13.3	P	Lipid phosphate phosphatase family	8.4	3.0
24	CYP19A1 (cytochrome P450, family 19, subfamily A, polypeptide 1)	NM_031226.1	Hs00240671_m1	15q21.1	P	A member of the cytochrome P450 superfamily of enzymes	8.4	4.9
25	PSG3 (pregnancy-specific beta-1-glycoprotein 3)	NM_021016.3	Hs00360732_m1	19q13.2	P	A group of molecules that are mainly produced by the placental syncytiotrophoblasts during pregnancy	6.0	2.8
26	PPAP2B (phosphatidic acid phosphatase type 2B)	NM_177414.1	Hs00170359_m1	1pter-p22.1	P	A member of the phosphatidic acid phosphatase family	12.6	0.6
27	P11 (26 serine protease, placental protein 11)	NM_006025.2	Hs00195731_m1	12q13.1	P	A serine protease specifically expressed in the syncytiotrophoblast	6.9	4.5
28	PAGE4 (P antigen family, member 4)	NM_007003.2	Hs00199655_m1	Xp11.23	P	Expressed in a variety of tumors and in some fetal and reproductive tissues including placenta, a member of the GAGE (G antigen) family	9.6	9.3
29	SMARCA1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1)	NM_003069.2	Hs00161922_m1	Xq25	P	A member of the SWI/SNF (switch/sucrose nonfermentation) family of proteins	4.7	4.9
30	COL1A2 (collagen, type I, alpha 2)	NM_000089.3	Hs00164099_m1	7q22.1	P	The pro-alpha 2 chain of type I collagen whose triple helix comprises two alpha 1 chains and one alpha 2 chain	8.6	0.9
31	GULP1 (GULP, engulfment adaptor PTB (phosphotyrosine binding) domain containing 1)	NM_016315.2	Hs00169604_m1	2q32.3-q33	P	An evolutionarily conserved adaptor protein required for efficient engulfment of apoptotic cells by phagocytes	8.4	6.2
32	PLEKHC1 (pleckstrin homology domain containing, family C member 1)	NM_006832.1	Hs00235033_m1	14q22.1	P	A component of ECM (cell-extracellular matrix) structures in mammalian cells	6.9	1.8

Table 1—(*Continued*)

	Gene symbol (name)	GenBank accession number	Assay ID	Chromosome localization	Gene expression[a]	Gene function	CV value[b] (custom array) (%)	CV value (real-time RT-PCR) (%)
33	PKIB [protein kinase (cAMP-dependent, catalytic) inhibitor beta]	NM_032471.4	Hs00261162_m1	6q22.31	P	A member of the cAMP-dependent protein kinase inhibitor family	6.6	4.8
34	CXCL14 (chemokine (C-X-C motif) ligand 14)	NM_004887.3	Hs00171135_m1	5q31	P	The cytokine gene family which encode secreted proteins involved in immunoregulatory and inflammatory processes	6.9	4.6
35	PEG3 (paternally expressed 3)	NM_006210.1	Hs00377844_m1	19q13.4	P	A Kruppel-type ZNF (zinc finger protein) protein, high levels of PEG3 in the human placenta and localized the signal to the layer of villous cytotrophoblast cells	5.6	3.7
36	ESRRG (estrogen-related receptor gamma)	NM_206594.1	Hs00155006_m1	1q41	P	A member of the steroid/thyroid/retinoid receptor superfamily	10.8	5.9
37	EBI3 (Epstein-Barr virus induced 3)	NM_005755.2	Hs00194957_m1	19p13.3	P	A secreted glycoprotein belonging to the hematopoietin receptor family	7.0	3.3
38	HSD3B1 (hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1)	NM_000862.2	Hs00426435_m1	1p13.1	P	Type I enzyme that is expressed mainly in the placenta and peripheral tissues	6.3	4.7
39	PAPPA (pregnancy-associated plasma protein A)	NM_002581.3	Hs00361692_m1	9q33.2	P	A large zinc glycoprotein of placental origin	10.4	6.7
40	PLAP (alkaline phosphatase, placental)	NM_001632.3	Hs01654626_s1	2q37	P	A membrane-bound glycosylated enzyme, which appears in the serum during pregnancy	9.4	6.1
41	CSH1 (chorionic somatomammotropin hormone 1 (placental lactogen))	NM_022640.2	Custom primer[c]	17q24.2	P	A member of the somatotrophin/prolactin family of hormones and plays an important role in growth control	3.5	1.0
42	SLC7A2 (solute carrier family 7 (cationic amino acid transporter, y+ system), member 2)	NM_003046.3	Hs00161809_m1	8p22-p21.3	P	Member of the APC (amino acid-polyamine-organocation) family of transporters	9.3	5.6
43	EFEMP1 (EGF-containing fibulin-like extracellular matrix protein 1)	NM_001039348.1	Hs00244575_m1	2p16	P	Extracellular matrix protein, which is expressed in many tissues but it is not expressed in brain and lymphocytes	11.1	3.7
44	CGB (chorionic gonadotrophin, beta polypeptide)	NM_000737.2	Hs00361224_gH	19q13.32	P	A member of the glycoprotein hormone beta chain family	10.5	10.4
45	COL3A1 (collagen, type III, alpha 1)	NM_000090.3	Hs00164103_m1	2q31	P	The pro-alpha 1 chains of type III collagen	13.0	6.5
46	LIFR (leukemia inhibitory factor receptor alpha)	NM_002310.3	Hs01123581_m1	5p13-p12	P	A protein that belongs to the type I cytokine receptor family	6.9	8.0

Table 1—(*Continued*)

	Gene symbol (name)	GenBank accession number	Assay ID	Chromosome localization	Gene expression[a]	Gene function	CV value[b] (custom array) (%)	CV value (real-time RT-PCR) (%)
47	SERPINE1 [serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1, plasminogen activator inhibitor-1 (PAI1)]	NM_000602.1	Hs00167155_m1	7q21.3-q22	P	This inhibitor acts as 'bait' for tissue plasminogen activator, urokinase and protein C	4.4	5.3
48	HERV-FRD (human endogenous retrovirus FRD envelope protein, syncytin 2)	NM_207582.1	Hs01652148_m1	6p24.1	P	HERV-FRD envelop protein, which is expressed predominantly in placenta, has a potential role in placenta formation	5.7	3.8
49	PSG2 (pregnancy-specific beta-1-glycoprotein 2)	NM_031246.1	Hs01652779_m1	19q13.1-q13.2	P	A group of molecules that are mainly produced by the placental syncytiotrophoblasts during pregnancy	8.5	10.4
50	HERV-FRD (human endogenous retrovirus FRD envelope protein, syncytin 2)	NM_207582.1	Custom primer[c]	6p24.1	P	HERV-FRD env protein, which is expressed predominantly in placenta, has a potential role in placenta formation	3.0	1.9
51	GAPDH (glyceraldehyde-3-phosphate dehydrogenase)	NM_002046.3	Hs99999905_m1	12p13	P&B	A kinase involved in the glycolysis-dependent endogenous phosphorylation of the alpha 1 subunit of the GABA-A receptor	7.6	8.4
52	MRPS16 (mitochondrial ribosomal protein S16)	NM_016065.3	Hs00831691_s1	10q22.1	P&B	Protein synthesis within the mitochondrion	12.6	10.0
53	EEF1A2 (eukaryotic translation elongation factor 1 alpha 2)	NM_001402.5	Hs00265885_g1	20q13.3	P&B	Isoform of the alpha subunit of the elongation factor 1 complex	8.7	1.5
54	IL8RA (interleukin 8 receptor, alpha)	NM_000634.2	Hs00174146_m1	2q35	P&B	A member of the G protein-coupled receptor family	13.8	9.4
55	P2RY13 (purinergic receptor P2Y, G protein coupled, 13)	NM_023914.2	Hs03043902_s1	3q24	P&B	The family of G protein-coupled receptors	7.5	5.1
56	Blue script plasmid DNA	—	—	—	—	No DNA sequence in human	—	—

[a] P indicates dominantly expressed in placental tissues and P&B indicates transcripts equally expressed in both placenta and blood.
[b] Coefficient of variations that is the ratio of the standard deviation to the mean value.
[c] TaqMan probe and gene-specific primers were described as previously (Ng *et al.*, 2003b; Okahara *et al.*, 2004).

were designed. These RAPs had a 45-bp T7 RNA polymerase consensus sequence at their 5′ end and also contained approximately 20-bp sequences of the genes corresponding to the cDNAs in the spots (their sequences are available on request). An antisense sequence was also designed for each spotted probe containing about 50 bp at its 3′ end. RNA amplification was performed as follows. CF mRNA was reverse transcribed using the RAPs and converted into double-stranded cDNA. To obtain a large amount of each target RNA, the double-stranded cDNA was used as a template for transcription with T7 RNA polymerase, and the amplified RNA was dissolved in 30 μL of diethylpyrocarbonate-treated (DEPC) H_2O. The specificity of the cDNA amplification in maternal plasma was confirmed by sequencing analysis.

Labeling and hybridization of CF placental mRNA

Comparative cDNA hybridization of CF placental mRNA from the preeclamptic women and normal control women was performed using the cDNA microarray panel (Figure 2). Amplified test cDNA (2 μg) from preeclamptic women and reference cDNA from normal control women were labeled with the fluorescent dyes Cy-5 and Cy-3, respectively. The two types of labeled cDNAs (final volume, 40 μL) were hybridized competitively with the custom cDNA microarray panel in a hybridization buffer composed of 5× saline-sodium citrate (SSC) and 0.5% sodium dodecyl sulfate (SDS) at 42 °C for 16 h. After the comparative hybridization, each panel was washed three times with buffer solutions composed of 2× SSC and 0.1% SDS, 2× SSC, 1× SSC and 0.5× SSC, respectively. The panels were scanned with a G2565AA Microarray Scanner System (Agilent Technologies) and analyzed using GenePix Pro Ver. 4.0.1.17 (Axon Instruments, Foster City, CA). The data for the comparative cDNA hybridization were corrected by the difference between the signal intensities of the test and reference cDNAs. The mean value of the signal intensities for each spot was calculated from the data of the six blocks on each slide (Figure 1). The corrected signal intensities for preeclamptsia and control pregnancies were expressed as scatter plots. Plasma samples from the only control pregnancies were used for the computation of the coefficient of variations (CV) values, which is the ratio of the standard deviation to the mean value, in this custom array quantification.

Quantitative real-time RT-PCR

One-step quantitative real-time RT-PCR assay was performed using ABI 7900T Sequence Detector (Perkin-Elmer, Foster City, CA) as described previously (Ng *et al.*, 2003b). The PCR products of 56 genes were cloned into the TOPO II vector (Invitrogen), respectively. Each of the extracted plasmid DNA was used

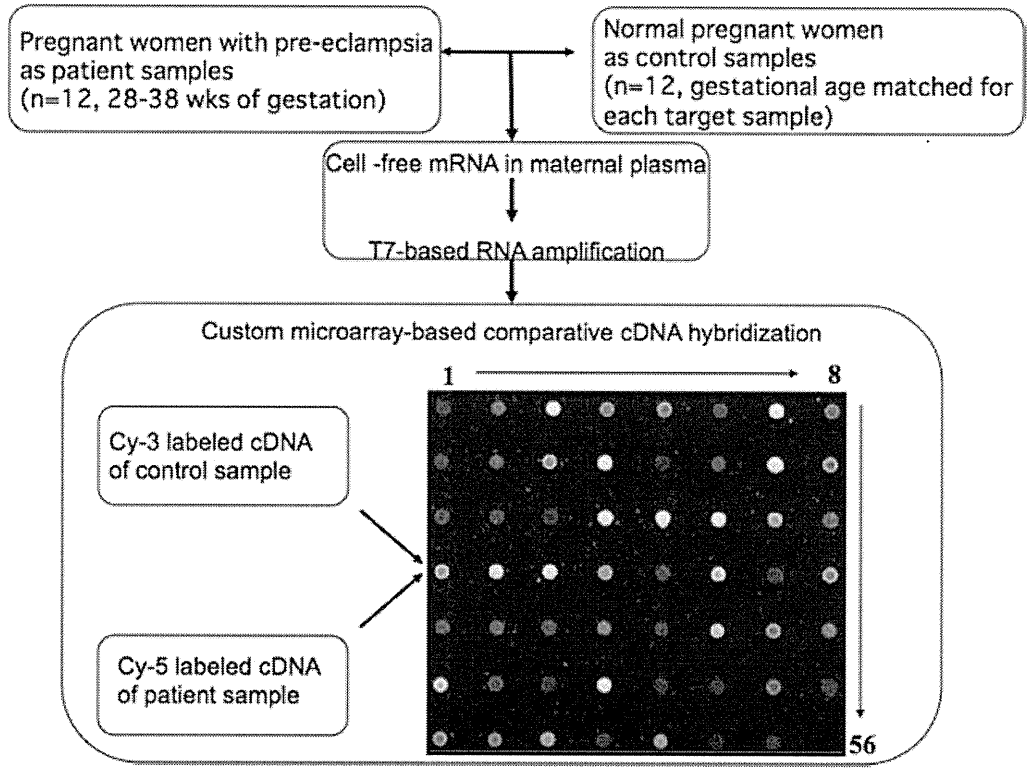

Figure 2—Outline of the custom microarray-based comparative cDNA hybridization. CF mRNA was extracted from 1.2 mL of individual plasma samples from 12 pregnant women with preeclampsia (27–38 weeks of gestation) and 12 normal pregnant women with matched gestational ages. The CF mRNA amplified by a T7-based RNA amplification method was subjected to comparative cDNA hybridization, in which amplified cDNA (2 μg) from preeclamptic women and control amplified cDNA from normal pregnant women were labeled with the fluorescent dyes Cy-5 and Cy-3, respectively. The same volumes of the labeled products were comparatively hybridized with the custom cDNA microarray panel. The mean values of six signal intensities for each sample were expressed as a scatter plot

DOI: 10.1002/pd

for a calibration curve of each gene. RT-PCR validation of the 12 normal control pregnancies was done for all 56 genes.

RESULTS

Identification of placenta-predominantly expressed genes in maternal plasma by microarray analysis, and development of a microarray panel from their cDNAs

We identified 50 placenta-specific transcripts that showed >2500 times higher expression in the placental tissues than in the corresponding whole blood samples in the respective trimesters of pregnancy (Table 1). Using the cDNAs of these transcripts, we created a custom cDNA array panel (Figure 1). To confirm whether a single experiment using the panel led to an accurate result, equal volumes of the same placental cDNA sample labeled with either Cy-5 or Cy-3 were hybridized in a comparative manner. Approximately equal signal intensities from Cy-5 and Cy-3 for the placental cDNA were successfully detected on all 55 spots, comprising the 50 placental and 5 positive control transcripts (*GAPDH, MRPS16, EEF1A2, IL8RA* and *P2RY13*), while no signal was obtained for the 56th spot as a negative control (the plasmid sequence from pBluescriptII SK(−) vector) (Figure 2).

The accuracy of the total plasma CF mRNA amplification using the T7-based RNA amplification method was confirmed by quantitative real-time PCR measurements (Abe *et al.*, 2003). Specifically, the average cycle threshold (Ct) values of the nonamplified positive control transcripts CF *GAPDH* mRNA, CF *MRPS16* mRNA, CF *EEF1A2* mRNA, CF *IL8RA* mRNA and CF *P2RY13* mRNA were 32.292, 32.211, 31.981, 33.017 and 32.143, while the Ct values for the same mRNAs after amplification were 24.927, 24.833, 23.163, 25.933 and 24.812, respectively. Similar results between amplified and nonamplified CF mRNA were also confirmed on a panel of placental RNA transcripts in maternal plasma. These data indicated that the total CF mRNA was reproducibly amplified using the T7-based RNA amplification method, which generated gene expression profiles that were comparable with the nonamplified total RNA as previously reported (Abe *et al.*, 2003). Valid signals for the amplified CF mRNA samples on the 55 spots were also detected, but not for the 56th spot comprising the negative control.

Identification of 17 different transcripts with increased or decreased levels of maternal plasma CF mRNA in severely hypertensive preeclamptic women

We created scatter plots for the signal intensities of the comparative hybridization between the target cDNA from the CF mRNA of preeclamptic women and the same amount of control cDNA from the CF mRNA of control pregnant women. The plots showed either an unchanged or an aberrant pattern (Figure 3). The aberrant pattern was defined when the signal intensity of the target cDNA was increased or decreased by twofold or more compared with the intensity of the gestational age-matched control cDNA. The aberrant pattern was specifically observed in five preeclamptic women with severe hypertension (cases 1–5) and not in the seven preeclamptic women with mild hypertension (cases 6–12) (Table 2) ($P < 0.05$, Fisher's direct method). Therefore, the aberrant pattern was associated with preeclampsia with severe hypertension. There were no correlations between the aberrant pattern and the proteinuria severity or disease onset.

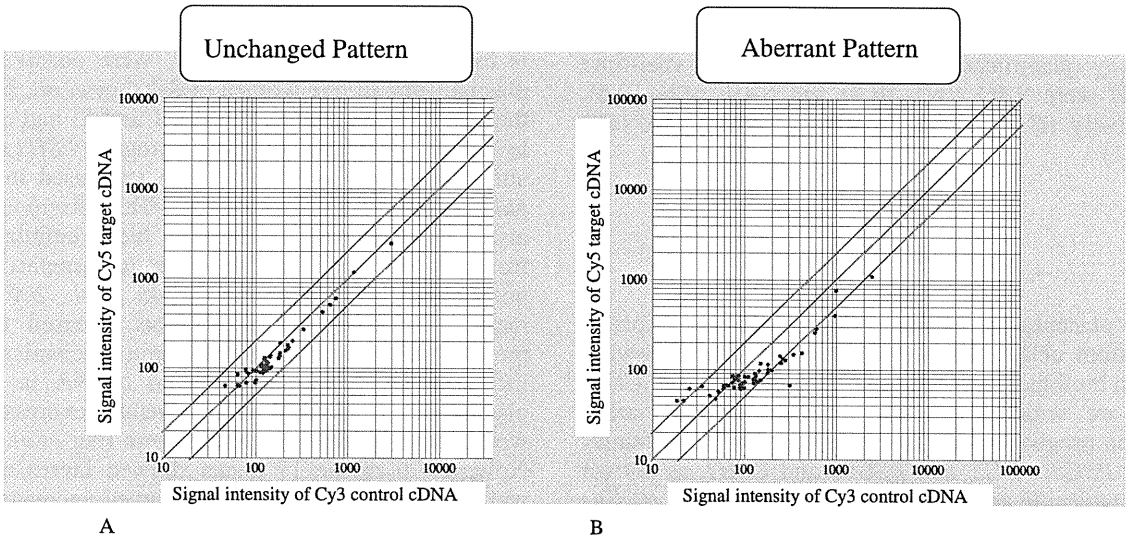

Figure 3—Scatter plots of the signal intensities obtained by comparative hybridization on the custom microarray between target cDNA from CF mRNA in plasma samples from preeclamptic women and control cDNA from normal control pregnant women. (a) Unchanged pattern showing that all the plots are within ±twofold from their equal intensity value. (c) Aberrant pattern showing that some plots are above or below the twofold line