Genotyping Microarray (Disease Chip) for Leber Congenital Amaurosis: Detection of Modifier Alleles

Jana Zernant,[1,2] Maigi Külm,[2,3] Sharola Dharmaraj,[4] Anneke I. den Hollander,[5]
Isabelle Perrault,[6] Markus N. Preising,[7] Birgit Lorenz,[7] Josseline Kaplan,[6]
Frans P. M. Cremers,[5] Irene Maumenee,[8] Robert K. Koenekoop,[9] and Rando Allikmets[1,10]

PURPOSE. Leber congenital amaurosis (LCA) is an early-onset inherited disorder of childhood blindness characterized by visual impairment noted soon after birth. Variants in at least six genes (*AIPL1*, *CRB1*, *CRX*, *GUCY2D*, *RPE65*, and *RPGRIP1*) have been associated with a diagnosis consistent with LCA or early-onset retinitis pigmentosa (RP). Genetically heterogeneous inheritance complicates the analyses of LCA cases, especially in patients without a family history of the disorder, and conventional methods are of limited value.

METHODS. To overcome these limitations, arrayed primer extension (APEX) technology was used to design a genotyping microarray for early-onset, severe retinal degenerations that includes all of the >300 disease-associated variants currently described in eight genes (in addition to the six just listed, the early-onset RP genes *LRAT* and *MERTK* were added). The resultant LCA array allows simultaneous detection of all known disease-associated alleles in any patient with early-onset RP. The array was validated by screening 93 confirmed patients with LCA who had known mutations. Subsequently, 205 novel LCA cases were screened on the array, followed by segregation analyses in families, if applicable.

RESULTS. The microarray was >99% effective in determining the existing genetic variation and yielded at least one disease-associated allele in approximately one third of the novel pa-tients. More than two (expected) variants were discovered in a substantial fraction (22/300) of the patients, suggesting a modifier effect from more than one gene. In support of the latter hypothesis, the third allele segregated with a more severe disease phenotype in at least five families.

CONCLUSIONS. The LCA genotyping microarray is a robust and cost-effective screening tool, representing the prototype of a disease chip for genotyping patients with a genetically heterogeneous condition. Simultaneous screening for all known LCA-associated variants in large LCA cohorts allows systematic detection and analysis of genetic variation, facilitating prospective diagnosis and ultimately predicting disease progression. (*Invest Ophthalmol Vis Sci.* 2005;46:3052–3059) DOI: 10.1167/iovs.05-0111

L eber congenital amaurosis (LCA) defines a clinically and genetically heterogeneous group of congenitally blind patients. It is diagnosed as an early-onset, bilateral retinal dystrophy associated with a severely diminished or absent electroretinogram (ERG) before the age of 1 year.[1] Shortly after birth, patients usually present with poor fixation, nystagmus, photophobia, and amaurotic pupils. Later, in most patients, a large variety of retinal changes appear, including salt-and-pepper pigmentation, attenuated vessels, and atrophy of the retinal pigment epithelium (RPE). LCA is mostly inherited as an autosomal recessive, genetically heterogeneous (multigenic) disorder. Mutations in at least eight genes have been associated with the retinal pathology consistent with the diagnosis of LCA and/or early-onset RP. Since the identification of guanylate cyclase 2D (*GUCY2D*)[2] in 1996 as the first LCA gene, mutations in five more genes—*AIPL1*, *CRB1*, *CRX*, *RPE65*, and *RPGRIP1*—have been shown to cause LCA.[3,4] Variant alleles in some other genes, such as *LRAT*, *MERTK*, *PROML1*, and *TULP1* have been described in patients with early-onset RP. Altogether, our current knowledge explains <50% of the genetic causality of LCA.[5]

Clinical and genetic heterogeneity has substantially complicated the assessment of the genetic determinants in LCA cases, especially the sporadic ones. Disease-associated alleles in the six major genes have shown substantial heterogeneity. Currently, >300 disease-associated variants have been identified in all known LCA-associated genes; however, only a few of these alleles have been detected in more than 1 or 2 families. After tedious screening with single-strand conformational polymorphism (SSCP) analysis, denaturing high-performance liquid chromatography (DHPLC), and direct sequencing, the most comprehensive studies have been able to determine from ~28%[6,7] to ~45%[5] of all disease-causing alleles. Important factors that influence mutation analysis are the effort and the cost of screening all LCA-associated genes. Currently, mutation analysis involves screening at least 80 amplicons (exons) that encompass the entire open reading frames of the eight genes. Therefore, all available mutation detection techniques remain labor intensive, time consuming, and expensive.

To overcome these challenges and to generate a high volume, cost-effective, and efficient screening tool, we developed the LCA genotyping microarray. The LCA array contains all currently known (>300) disease-associated genetic variants and several common polymorphisms from eight genes involved in early-onset severe retinal dystrophies (Supplementary Table S1, http://www.iovs.org/cgi/content/full/46/9/3052/DC1). In this report, we describe the efficiency of the chip for the genetic analysis of LCA and discuss several applications made possible by the array.

METHODS

Study Subjects

LCA patient cohorts (298 unrelated individuals) were derived from two locations in North America: Wilmer Eye Institute (Johns Hopkins University, Baltimore, MD) and Montreal Children's Hospital (Montreal, Quebec, Canada); and from two centers in Europe: University Medical Centre (Nijmegen, The Netherlands); and Hôpital Des Enfants Malades (Paris, France). The Nijmegen Center cohort included patients from Germany, the Netherlands, and Belgium. Of the 298 patients, 93 had been screened for mutations in six LCA genes at the four centers. The remaining 205 patients, recruited at the U.S., Canadian, and Dutch centers, had not been screened. Collections of matched control samples (200 unrelated individuals) were ascertained at the same locations. Research procedures were in accordance with institutional guidelines and the Declaration of Helsinki. Informed consent was obtained at each Center from all patients after the nature of procedures to be performed was fully explained.

A more detailed description of clinical procedures has been published.[5,8] Briefly, the inclusion criteria for patients with LCA were the following: (1) onset of symptoms—that is, severe impairment of visual function in early infancy; (2) nondetectable or severely diminished ERG before the age of 1 year; and (3) an absence of systemic disease features. Collected detailed clinical data included the age of onset, history of the patient and family (compatible with autosomal recessive inheritance), visual acuity (if measurable), electroretinography, and fundus photography. Detailed interviews with parents (and patients, if applicable) were conducted, and pedigrees were established.

Molecular Methods

Sixty amplicons from eight genes were PCR amplified, as described previously.[9] Primer sequences are available on request. In the amplification mixture, 20% of the dTTP was substituted by dUTP.[10] The amplification products were concentrated and purified (GENErALL PCR kit; General Biosystems, Inc., Seoul, Korea). The fragmentation of amplification products was achieved by adding thermolabile uracil *N*-glycosylase (Epicenter Technologies, Madison, WI) and the following heat treatment.[10]

One-sixth of every amplification product was used in the primer extension reaction on the LCA array. Each arrayed primer extension (APEX) reaction consisted of a fragmented and denatured PCR product, 4 units of DNA Polymerase (ThermoSequenase; GE Healthcare, Piscataway, NJ), 1× reaction buffer, and 1.4 μM final concentration of each fluorescently labeled ddNTP: Texas red-ddATP, fluorescein-ddGTP (GE Healthcare), Cy3-ddCTP, and Cy5-ddUTP (NEN; Boston, MA). The reaction mixture was applied to a microarray slide for 15 minutes at 58°C. The reaction was stopped by washing the slide at 95°C in double-distilled water (Milli-Q; Millipore, Bedford, MA).[11] The slides were imaged (Genorama QuattroImager; Asper Biotech, Ltd., Tartu, Estonia) and the sequence variants were identified with a computer program (Genorama Genotyping Software; Asper Biotech, Ltd.).[9,12]

Array-identified variants were confirmed by direct sequencing with dye termination chemistry (*Taq* Dyedeoxy Terminator Cycle Sequencing kit; Applied Biosystems, Inc. [ABI], Foster, City, CA), according to the manufacturer's instructions. Sequencing reactions were resolved on an automated sequencer (model 377; ABI).

Electronic Databases

Databases used in the study included Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim (National Center for Biotechnology Information, National Institutes of health, Bethesda, MD); RetNet Retinal Information Network, http://www.sph.uth.tmc.edu/Retnet/home.htm (University of Texas Houston Health Science Center, Houston, TX); The Genome Database (GDB), http://gdbwww.gdb.org/gdb/ (RTI International, Research Triangle Park, NC); and the Retina International Scientific Newsletter Mutation Database: http://www.retina-international.org/sci-news/mutation.htm (maintained by Markus Preising, Molecular Genetics Laboratory, University of Regensburg, Regensburg, Germany).

RESULTS

Design of the LCA Microarray

Microarrays were designed and manufactured with APEX technology.[13,14] A detailed description of the methodology is available elsewhere (www.asperbio.com; Asper Biotech, Ltd.).[10,12] Briefly, 5'-modified, sequence-specific oligonucleotides are arrayed on a glass slide. In general, these oligonucleotides are designed with their 3' end immediately adjacent to the variable site. PCR-prepared and -fragmented target nucleic acids are annealed to oligonucleotides on the slide, followed by sequence-specific extension of the 3' ends of primers with dye-labeled nucleotide analogues (ddNTPs) by DNA polymerase. The APEX reaction is, in essence, a sequencing reaction on a solid support.

One advantage of APEX, compared with other common microarray-based techniques, such as hybridization with allele-specific oligonucleotide probes, is that all variable nucleotides are identified with optimal discrimination at the same reaction conditions, rendering this approach far more specific for multiplex mutation detection.[15] This method successfully detects single-nucleotide polymorphisms (SNPs), deletions, and insertions in heterozygous and homozygous patient samples.[11,12] APEX also permits at least one order of magnitude higher power of discrimination between genotypes, when compared with hybridization-based techniques.[15] Template-dependent extension reactions with four uniquely dye-labeled ddNTPs yield covalent bonds between an oligonucleotide and a dye terminator. This allows stringent washing of slides after APEX, yielding minimal background signal. The time required for complete APEX analysis, including sample preparation, is less than 4 hours. Once designed, these microarrays can be modified (upgraded) easily with new variants.

By systematic analysis of all published, reported, and communicated data, we compiled the most comprehensive database of LCA-associated variants (Supplementary Table S1). The number of sequence changes currently considered disease-associated exceeds 300. By design, we included on this chip all variants from the coding region and adjacent intronic sequences of the eight LCA/early-onset RP genes: *AIPL1*, *CRB1*, *CRX*, *GUCY2D*, *RPE65*, *RPGRIP1*, *MERTK*, and *LRAT*. Intronic sequences were included only in cases with predicted or documented involvement in splicing. Several common polymorphisms were also included, mainly from the coding regions (K192E and E1033Q in *RPGRIP1*, A52S and L782H in *GUCY2D*, D90H in *AIPL1*), to facilitate haplotype assignments. The resultant LCA microarray contains 307 variants (Supplementary Table S1, Table 1). Each sequence change was designed to be queried in duplicate from each sense and antisense strand by the software (Genorama Genotyping Software; Asper Biotech, Ltd.).

TABLE 1. Genetic Variation in LCA Genes

LCA Gene	Exons	Amplicons	Known Mutations	Mutation Frequency in LCA (%)	LCA Array (%)
AIPL1	6	6	25	5.8[16]* 3.4[5]	7.8
CRB1	12	13	68	9.0–13.5[17,18] 10.0[5]	5.4
CRX	3	3	29	2.0[19] 2.8[20] 0.6[5]	1.5
GUCY2D	20	14	67	6.0[6,21] 20.3[22] 21.2[5]	11.7
RPE65	14	10	81	6.8[6] 8.2[23] 11.4[24] 15.6[25] 6.1[5]	2.4
RPGRIP1	24	18	32	5.3[26] 5.6[27] 4.5[5]	4.9
Total	79	64	302		33.7

* Mutation frequencies are shown with the reference sources.

Validation of the LCA Array

Validation of the array served a dual purpose: to estimate the efficiency of the array in detecting variants on the chip by screening samples with known variants and to gauge its efficiency in detecting disease-associated alleles in LCA cohorts, compared with other methods, such as SSCP and direct sequencing. The overall efficiency of the array was enhanced by designing primers with mismatched or modified bases for several variants where sequence presented additional challenges (e.g., hairpin loops, and repeats). Currently, all 307 variants are detected by the latest version of the chip; 294 positions are detected from both strands, whereas 13 (4.2%) variants are detected reliably from one strand. Since the same oligonucleotides detect both mutant and wild-type (wt) sequence, these data were obtained with positive controls for 102 variants and with wt (control) DNA for the rest.

The array, which originally contained 267 variants, was validated in a cohort of 93 confirmed LCA samples collected at the four centers. The initial screening by SSCP or DHPLC had resulted in the detection of 99 variants that were deemed disease-associated by confirmed clinical assessment and segregation analysis.[17,21,22] These comprised 145 alleles (78%) and 5 "modifier" variants (the third sequence change in a different gene from the causal) in 99 positions. Four variants, which had been described by SSCP/DHPLC gave a wt signal on the array;

these results were confirmed by direct sequencing. Most likely, these represent false-positive results due to an experimental error.

The microarray screening detected 20 additional alleles missed by other methods and found new changes in 15 of the 93 samples, bringing the total to 154 alleles and 16 "modifier" variants in 105 positions. The six variants missed (or not described) in studies employing SSCP/DHPLC methods included the V96I change in *AIPL1*, the R769Q and R1331H variants in *CRB1*, the A132T change in *RPE65*, the D1114G change in *RPGRIP1*, and the del V1020_L1022 variant in *GUCY2D*. Altogether, the LCA array identified 83% of all known disease-associated alleles in this preselected cohort. All variants detected by the array were confirmed by direct sequencing, and no conflicting sense/antisense calls or false positives were identified; i.e., all variants included on the array are detected with 100% accuracy.

Screening Test Populations on the LCA Microarray

To evaluate the LCA array further, we screened 205 previously unanalyzed patients with LCA of diverse ethnicity derived from three of the four participating centers. At each center, the best effort was made to avoid phenocopies and cases with questionable diagnoses. In each North American cohort, the ethnicity of subjects was mostly consistent with "European American"; however, approximately one-third of the samples originated from Asian ethnic groups (e.g., Indian, Pakistani, Iranian). The European cohort mainly contained patient samples from Germany, The Netherlands, and Belgium. The chip screening results for these cohorts are summarized in Table 2.

The screening efficiency of the LCA array was remarkably similar in all three cohorts, yielding from ~20.3% (the Canadian sample) to 23.8% (the U.S. sample) of all possible disease-associated LCA alleles (Table 2). The LCA array detected at least one disease-associated allele in approximately one third of patients. The remarkably similar mutation detection rate in all cohorts suggests a similar frequency of known disease-associated LCA alleles across all populations. However, the frequency of individual alleles showed substantial differences between the screened patient populations. For example, mutations were frequent in the *AIPL1* gene in the Dutch cohort (~20%), but very rare in the cohort from Baltimore. At the same time, *CRB1* alleles accounted for the LCA phenotype in 8.5% of cases from Baltimore, but they were not detected at all in the Nijmegen cohort included in this study. However, in an independent cohort from the same center, *CRB1* mutations were a frequent cause of LCA (Cremers F, unpublished data, July 2005). Disease-associated alleles from *GUCY2D* and *RPGRIP1* were relatively evenly distributed across all cohorts. Almost no mutations were found in the *CRX* and *RPE65* genes in the entire

TABLE 2. Screening Efficiency of the LCA Array in Three Patient Cohorts

Gene	Nijmegen, The Netherlands 41 Patients (82 Alleles)			Montreal, Canada 59 Patients (118 Alleles)			Baltimore, MD 105 Patients (210 Alleles)		
	Patients	Alleles	%	Patients	Alleles	%	Patients	Alleles	%
AIPL1	8	10	19.5	4	6	6.8	4	6	2.9
CRB1	0	0	0	2	2	3.4	9	14	8.6
CRX	0	0	0	1	1	1.7	2	2	1.9
GUCY2D	4	4	9.8	8	11	13.6	12	17	11.4
RPE65	0	0	0	1	1	1.7	4	5	3.8
RPGRIP1	2	3	4.9	2	3	3.4	6	6	5.7
Total	14 (34%)	17 (20.7%)	34	18 (30.5%)	24 (20.3%)	30	37 (35%)	50 (23.8%)	35

screened patient population, although the *RPE65* gene is represented on the array by the largest number (81) of alleles (Supplementary Table S1; Table 1). This observation suggests that allelic variation in the *RPE65* and *CRX* genes is more likely associated with the diagnosis of early-onset severe retinal dystrophies rather than with LCA, if characterized by strict criteria.[5]

Three variants were found in all three patient cohorts. The most frequent allele in the Dutch cohort, W278X in *AIPL1* (29% of all alleles found), was detected once in Canadian samples and twice in samples from Baltimore (both 4%). C948Y in *CRB1* and P701S in *GUCY2D* had similar frequencies in all the three sample cohorts, 6% to 8% and 11% to 12%, respectively, whereas P701S was also the most frequent allele in Canadian samples. The *CRB1* variant 611_617 delAAATAGG was the most frequent allele in samples from Baltimore (14% of alleles), but was completely absent in other cohorts. The segregation of this mutation with the disease in families of Puerto-Rican origin clearly suggests a founder effect. The Canadian cohort showed the highest heterogeneity —17 different variants accounting for 24 disease-associated alleles. In the Dutch samples, we found 17 alleles (8 different variants), whereas 50 alleles (23 individual variants) were detected in the Baltimore cohort. In summary, comprehensive microarray analysis currently allowed the identification of the causal gene in approximately one third of all LCA samples.

The pathogenicity of alleles on the chip was inferred from all available published data and further analyses of specific alleles in this study. All alleles, except for a small selection of clearly defined common polymorphisms, had been defined as "disease-associated" by a combination of criteria applied to all genetic studies, including: (1) the cosegregation of an allele with the disease phenotype in a family; (2) the absence of the allele in a substantial number (>200 chromosomes) of subjects defined as "general population"; and (3) the determined or deduced effect of a variant on the protein function. However, even including all these, several alleles cannot be called highly penetrant and/or 100% disease associated. These fall into two major categories: The first group includes those variants detected only once in a patient with LCA, which should be called "rare variants" by most strict criteria, unless definite information on a functional consequence of a sequence change is available. Analyses by mutagenesis to assess the effect of various alleles on the protein function have been performed for *GUCY2D*,[28–31] *AIPL1*,[32,33] and *CRX*,[34,35] but these studies have examined a limited number of alleles in tests that are often not directly coupled to the (proposed) protein function.

The second group includes a few variants, with higher than expected frequencies in the general population for a highly penetrant allele. Some variants from this category include the D1114G variant in *RPGRIP1* and the P701S missense change in *GUCY2D*. For example, the allele frequency of the P701S variant in the general population of white origin is ~2%. If this variant were disease-associated with complete penetrance, the disease prevalence for *GUCY2D*-associated LCA alone would be 1 in 2500, at least 20 times more than observed. We detected this variant in >5% of patients with LCA, including three homozygous probands derived from extended families of Iranian origin. Segregation analyses confirmed that the P701S variant cosegregated with the disease in all three extended pedigrees, one of which (LCA-108) is shown in Figure 1. Moreover, patient IV-3 from this pedigree was heterozygous for P701S. Direct sequencing of the *GUCY2D* gene in this patient revealed another, previously unknown, missense allele H945R on the other chromosome, further supporting the conclusion that *GUCY2D* is the disease gene in this family. Another explanation for the P701S allele's being disease associated in Iranians is the founder effect—that is, the possibility of another
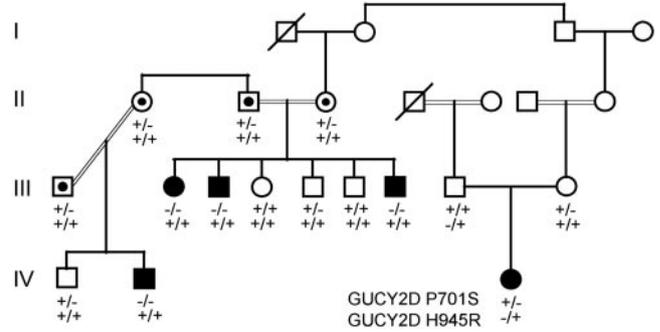

FIGURE 1. Pedigree of a family segregating LCA with the GUCY2D P701S variant.

variant's being in *cis* with P701S in Iranians, but not in other populations. However, we did not find any other variants in *GUCY2D* on the same chromosome with P701S. In conclusion, our current data— complete segregation of this variant with the disease in three extended pedigrees, lack of nonsymptomatic homozygotes, and compound heterozygosity in one patient—support P701S as being associated with LCA, at least in some populations.

Multiple Alleles in Patients with LCA, Suggesting Modifiers of the LCA Phenotype

Screening on the LCA chip revealed multiple cases in which possibly disease-associated alleles were detected in more than one gene in one patient. These included 12 cases in which heterozygous alleles from two genes were found, and 10 cases in which homozygous or compound heterozygous variants from one gene were detected, together with a heterozygous allele from a different LCA gene—a total of 22 of 300 (7.3%).

The odds of detecting a third allele by chance can be calculated as follows. The incidence of LCA is estimated to be between 1 in 50,000 and 1 in 100,000 persons (i.e., approximately 1/75,000). Assuming that six LCA genes account for approximately 50% of the cases, we could expect a total of 12 LCA genes, resulting in 1 in 900,000 individuals harboring a specific genetic form of LCA. This means that 1 in 225,000 marriages is at risk, and that 1 in 474 individuals carry a specific LCA-associated genetic defect. Therefore, it is expected that 12 of 474, or 1 in 40, individuals are carriers of a disease-causing LCA mutation by chance. Consequently, in this study we expected to find 300/40/2, or 3 patients, with LCA who carry a disease-causing mutation by chance alone. The same result is obtained by using the Hardy-Weinberg equilibrium and basic binomial statistics. However, we detected a modifier allele in 22 cases, a statistically significant difference ($P < 0.0001$; Fisher exact test). An inclusion of a more prevalent allele in the analysis, such as the *GUCY2D* P701S variant described earlier, would increase (double) the number of expected modifier alleles. The data shown in the previous section strongly suggest at least a modifier, if not a causal, status for P701S, which justifies the inclusion of this allele in the statistical analyses. Moreover, it would not change the conclusion that the frequency of detection of the modifier alleles by the array is significantly different from that expected by chance alone.

Unfortunately, nearly all of these cases were sporadic or simplex, and no other siblings were available for segregation analysis and/or phenotype comparison. Five families with more than two LCA-associated alleles and multiple affected individuals were available for detailed clinical and genetic analysis (Fig. 2).
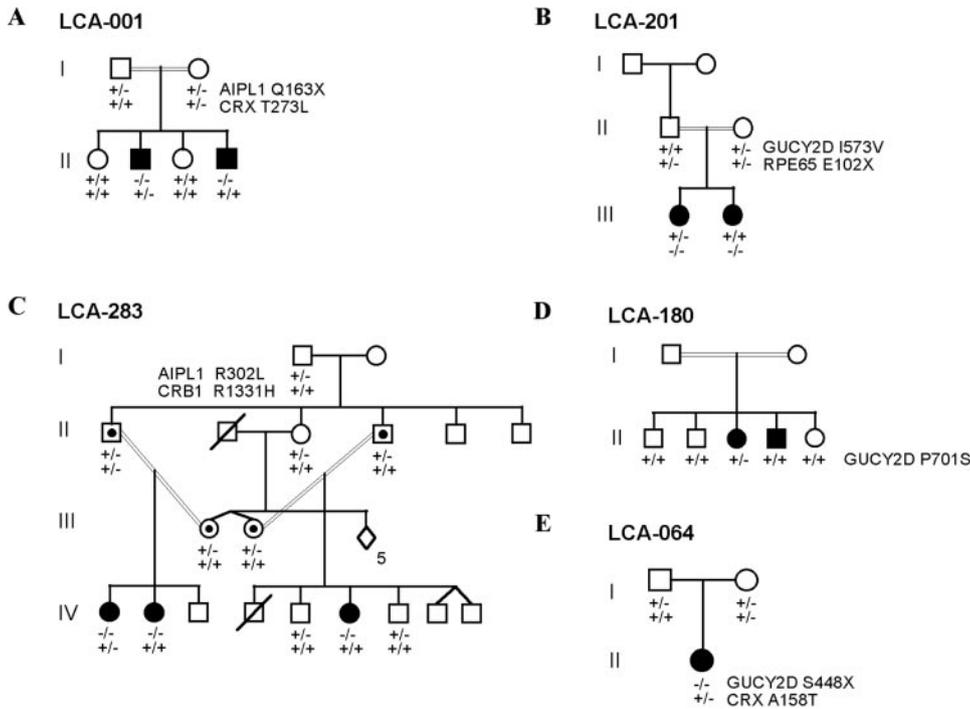
FIGURE 2. LCA pedigrees harboring mutant alleles from more than one LCA gene.

In a family of Palestinian origin, LCA-001 (Fig. 2A), the homozygous Q163X allele of the *AIPL1* gene segregated with the disease. An additional T273L missense allele from *CRX* was detected in patient II-2. LCA was diagnosed in the affected proband (II-2) in early infancy with a severely diminished ERG. The presence of keratoconus and cataracts was noted by the end of the first decade of life. There was a history of night blindness, and the visual acuity was light perception (LP) in both eyes. Retinal examination showed pale optic discs, arteriolar attenuation, and marked intraretinal pigment migration. Atrophic changes were seen in the macula, and the foveal reflex was ill defined. Individual II-4, who did not harbor the *CRX* T273L allele, presented with a similar phenotype, although he was clearly less severely affected than II-2 at a comparable age.

In a consanguineous Greek family, LCA-201, the disease phenotype was caused by a homozygous E102X *RPE65* mutation (Fig. 2B). The older sibling (III-1) had inherited an additional *GUCY2D* I573V variant from her mother. Because the mother is asymptomatic, we can exclude multiallelic inheritance; however, the older sister with three alleles clearly had a more severe visual dysfunction than did her younger sibling (III-2) at a comparable age. Specifically, at age 23, the older sibling (III-1) had best corrected visual acuity of 20/1300 (0.015) in both eyes, very limited visual fields by confrontation, night blindness, sluggish pupils, and roving nystagmus. The fundus examination revealed pale optic discs, dull macular reflexes with diffuse yellowish stippling, significant arteriolar attenuation, and considerable peripheral pigmentary changes. At a similar age, the younger sibling (patient III-2) had a visual acuity of 20/400 (0.05) in both eyes, with less-pronounced nystagmus. Ophthalmoscopic examination revealed mild arteriolar attenuation with areas of pigment clumping and atrophy in the macula, in addition to generalized mottling of the retina. However, unlike her older sibling, she was able to read normal-sized newsprint with the aid of a magnifying device and to navigate independently.[36]

In a consanguineous family from India, LCA-283, a homozygous R302L variant of the *AIPL1* gene cosegregated with the disease (Fig. 2C). Of the three affected members, one (IV-1)

harbored an additional *CRB1* R1331H allele. This patient's phenotype was advanced compared with her sister's (IV-2). Specifically, her retinal features included chorioretinal atrophy, extensive pigmentary mottling, narrowing of retinal arterioles, and stippling of the macula, with (extensive) areas of atrophy. In addition, she had severe keratoconus and cortical cataracts. Her younger sister, examined at a comparable age, had only a few peripheral bone-spicule–like pigmentary deposits and tiny areas of pigmentary atrophy and stippling in the macular region, moderate keratoconus, and clear lenses. Their cousin (IV-5), who also did not have the third allele, was younger at the time of clinical examination, excluding a direct phenotypic comparison.

In an extended pedigree belonging to the Old Order River Brethren, a religious isolate of Swiss descent residing in eastern Pennsylvania (LCA-180; Fig. 2D), the disease-causing gene had been linked to a new locus on 6q14.[37] Of the two affected siblings, one had inherited an additional *GUCY2D* P701S allele.[21] When compared at the same age, the sibling with three alleles presented with a more severe ocular phenotype and experienced greater difficulty with decreased visual function. Retinal examination revealed higher hyperopia and more extensive peripheral pigmentary mottling.

Finally, in a Puerto-Rican family (LCA-064, Fig. 2E), the LCA phenotype was caused by a homozygous S448X nonsense mutation in *GUCY2D* in a single affected child. Although the patient had no affected siblings, the phenotype was not consistent with the usual one in *GUCY2D*-associated cases.[5,38] It appears to have been modified by an additional *CRX* allele, A158T, which resulted in the development of a prominent macular coloboma in both eyes. Most recently,[38] we reviewed the clinical phenotype of 28 patients with LCA from three independent studies, who had *GUCY2D* mutations and found that none of them had documented maculopathy, whereas we found that most of the patients with LCA with *CRX* mutations have maculopathy or macular coloboma.

In summary, these cases, although limited to five families, offer the first clues to how alleles from other than the primary disease-causing LCA gene can modify the disease presentation. The fact that heterozygous alleles from LCA genes can have a

modifying effect on the disease expression is further supported by observations that heterozygous carriers of LCA-associated genes, such as parents of patients with LCA with *CRB1,* *GUCY2D, AIPL1,* and *RPGRIP1* mutations, who can have mild, but clearly detectable by ERG, defects of visual processes[38,39] (Koenekoop RK, unpublished data, July 2005).

DISCUSSION

We designed and validated a comprehensive mutation-detection system for Leber congenital amaurosis, the "disease chip" which contains all known disease-associated genetic variations in eight genes. The DNA from >300 patients, who had phenotypes consistent with LCA, was screened with the array. Validation included determining the efficiency of detecting known mutations and the overall efficiency of screening new cohorts with LCA or early-onset autosomal recessive (ar)RP. The chip was 100% effective in detecting known variants; 293 positions were reliably detected from both strands and 13 (4.2%) positions from one strand (Supplementary Table S1). During routine screening, an average ~1% of positions on the chip remained unreliably detected in any given experiment. This resulted from weak or undetectable signals in some positions, either due to a weaker PCR product, inefficient hybridization-extension on a particular oligonucleotide, or from a combination of both, in some experiments. Therefore, in an average experiment, the LCA chip's efficiency can be estimated to be ~99%.

The chip alone determined possible disease-associated alleles in approximately one third of patients with LCA from novel, previously unscreened cohorts. These results suggest that the LCA array is an efficient screening tool for known variants and that its efficiency for screening patient populations with LCA is comparable to other methods, but substantially (>10 times) faster, much cheaper, and less labor intensive. The LCA array supplies several major applications, including (1) quick and affordable (pre-) screening of all patients with a clinical diagnosis of LCA, allowing the pinpointing of the disease-causing gene in >30% of patients with LCA; (2) confirmation of the clinical diagnosis and selection of applicable patients for emerging clinical trials (e.g., for those with *RPE65* mutations); and (3) simultaneous screening for all known LCA-associated variants in large LCA cohorts, for systematic detection and analysis of genetic variation, including modifier alleles, therefore facilitating prospective diagnosis and predicting disease progression.

Efficiency of the LCA Array

Because by design the LCA array detects only known variants, the expected efficiency of the chip can be calculated based on several assumptions: (1) the number of patients with LCA screened for new mutations from each ethnic group (i.e., whether the array is representative of any given population); (2) the average fraction of alleles that are expected to be found by direct sequencing; and (3) the estimated number of yet unknown alleles for any ethnic group. For example, recent studies screening sizable cohorts of patients with LCA have detected LCA-associated alleles in the six major genes from 28%[7] to 44% to 45%[5] of all screened patients with LCA.

Assuming that, in the six genes, an average of 40% of the mutations remain to be discovered across all populations of European descent and that no more than 45% of all possible alleles can be detected even by direct sequencing of patients with LCA,[5] the LCA array is currently expected to detect ~27% (100 × 0.6 × 0.45) of disease-associated alleles in populations of European origin. The number could be higher in some populations in which more cases have been screened (Dutch

and French) and lower in the others, in which studies with limited numbers have been conducted (English and Spanish). These calculations also assume that all major alleles have been found in European populations, such as the W278X allele in *AIPL1* and the C948Y allele in *CRB1.*

The experimental data correlated with theoretical calculations. The array detected ~22% of all possible LCA-associated alleles in ~33% of patients across all cohorts (Table 2). As expected, in most of the patients, the array detected both disease-associated alleles, which is the result of the enrichment of individual alleles in specific ethnic groups due to founder effects. More robust and closer-to-predicted results are expected in smaller, carefully characterized cohorts derived from a single clinical source. In addition, after the completion of the screening described in this study, we updated the array with the most recently discovered new gene, *RDH12*[40,41] and 68 new mutations, bringing the total number of alleles on the LCA array to 370. Therefore, it is reasonable to suggest that the latest version of the LCA array is even more efficient than the one described herein.

Comparison to Other Studies and/or Methods

Two classes of mutation-detection methods have been applied to screening the six genes in LCA cohorts: those involving separation of PCR products based on conformation changes (e.g., SSCP, DHPLC) and those directly sequencing the open reading frames of all six genes. Although direct comparison of the LCA array's efficiency in detecting genetic variation in patients with LCA to these methods is tempting, one should consider the basic difference between mutation detection (SSCP, direct sequencing) and genotyping (LCA array) methods. The former are supposed to detect all variation in analyzed sequences—the latter, only known variants. However, even with these restrictions, the efficiency of the LCA array stands out.

The array routinely detected LCA-associated alleles in approximately one third (~34%) of patients, which is better than SSCP-based studies,[7] and approximately 10% lower than in studies in which direct sequencing of well-characterized cohorts was used and in which the average detection rate was 44%. Available data suggest that direct sequencing of patients with sporadic cases of LCA would currently detect ~10% alleles in addition to those found with the LCA array. However, the cost and effort involved (~160 sequencing reactions per sample) is at least one order of magnitude higher than with the chip. Therefore, although direct sequencing of large cohorts of patients with LCA remains expensive, one could consider selective sequencing of the specific gene in the fraction of all patients in whom the LCA array currently identifies only one allele.

When comparing chip screening results with those obtained from various studies in which different screening methods (mostly SSCP) were used, several interesting observations emerge (Table 1). First, the fraction of the disease load was consistent in only three of six genes (*AIPL1, CRX,* and *RPGRIP1*). Of these, *CRX* seems to be associated with the lowest percentage (~2%) of LCA cases. The estimated role of the other three genes, *CRB1, GUCY2D,* and *RPE65,* varied widely from 2.5% to >20% (Table 1). Some variation of a causal gene load estimated from the analysis in separate cohorts is expected. The objective reasons include, for example, substantial differences in the ethnicity of patients with LCA (e.g., the high incidence, >20%, of *GUCY2D* alleles in populations of Mediterranean origin[5]) and in screening methods. The LCA array identifies known alleles, whereas the other methods (e.g., SSCP) also detect new variants. However, subjective reasons, which tend to influence the results even more, include differ-

ences in clinical diagnostic criteria and in interpreting results (i.e., determining whether a variant is disease associated, as discussed earlier). For example, mutations in *RPE65* have been associated with the disease in 2.4% to 15.6% of cases (Table 1). It has been suggested[24] that this reflects differences in defining the disease phenotype (early-onset RP versus LCA) rather than the actual variation in the disease phenotype. Another likely explanation has to do with the screening of relatively small cohorts in which even a small number of detected alleles would have a significant impact on the disease load. It is clear, however, that if a cohort has been ascertained by strict criteria for LCA and is of sufficient size, the fraction of *RPE65*-associated alleles will remain relatively small. Furthermore, these observations strongly suggest that the molecular screening of patients with similar diagnoses (e.g., the entire spectrum of arRP) has to be performed for *all* RP-associated genes, to achieve a conclusive diagnosis in each case.

Modifying Effects of Third Alleles

In 22 of 300 patients with LCA, we found variants in more than one gene. In 10 of these, we found a homozygous defect in one and a heterozygous defect in a second gene. In five of these latter cases, we were able to obtain enough clinical information and compare the severity and type of phenotypes between age-matched affected siblings, one with the homozygous defect alone, another with the same homozygous defect plus a heterozygous defect in another gene. We assume that the homozygous defect is causal, and we provide evidence that the heterozygous defect modifies the LCA phenotype by an additive effect. In four cases, the phenotype was clearly more severe in the affected individual with the three alleles. In one case, the phenotype was strikingly different in the affected patient with three alleles from the reported phenotypes associated with this gene (*GUCY2D*).

Summary

In conclusion, although the assessment of disease-associated variation in LCA still represents a difficult task for ophthalmologists and geneticists, the LCA array offers a prototype diagnostic tool to advance our knowledge substantially, specifically in LCA-associated disease and in ophthalmic genetics in general. It also represents the first instance in which modifier alleles can be detected in a single screening reaction, therefore improving molecular diagnosis and providing more accurate visual prognosis and suggestions of treatment options. The LCA disease chip is available to the entire scientific community for effective, high-throughput, accurate, and affordable screening of all populations of interest.

Acknowledgments

References

1. Franceschetti A, Dieterle P. Importance diagnostique et prognostique de l'electroretinogramme (ERG) dans les degenerescences tapeto-retiniennes avec retrecissement du champ visuel et hemeralopie. *Confin Neurol.* 1954;14:184–186.

2. Perrault I, Rozet JM, Calvas P, et al. Retinal-specific guanylate cyclase gene mutations in Leber's congenital amaurosis. *Nat Genet.* 1996;14:461–464.

3. Cremers FP, van den Hurk JA, den Hollander AI. Molecular genetics of Leber congenital amaurosis. *Hum Mol Genet* 2002;11:1169–1176.

4. Allikmets R. Leber congenital amaurosis: a genetic paradigm. *Ophthalmic Genet.* 2004;25:67–79.

5. Hanein S, Perrault I, Gerber S, et al. Leber congenital amaurosis: comprehensive survey of the genetic heterogeneity, refinement of the clinical definition, and genotype-phenotype correlations as a strategy for molecular diagnosis. *Hum Mutat.* 2004;23:306–317.

6. Lotery AJ, Namperumalsamy P, Jacobson SG, et al. Mutation analysis of 3 genes in patients with Leber congenital amaurosis. *Arch Ophthalmol* 2000;118:538–543.

7. Stone EM. Finding and interpreting genetic variations that are important to ophthalmologists. *Trans Am Ophthalmol Soc.* 2003;101:437–484.

8. Koenekoop RK. An overview of Leber congenital amaurosis: a model to understand human retinal development. *Surv Ophthalmol.* 2004;49:379–398.

9. Jaakson K, Zernant J, Kulm M, et al. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. *Hum Mutat.* 2003;22:395–403.

10. Kurg A, Tonisson N, Georgiou I, et al. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genetic Testing.* 2000;4:1-7.

11. Tonisson N, Zernant J, Kurg A, et al. Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. *Proc Natl Acad Sci USA.* 2002;99:5503–5508.

12. Tonisson N, Kurg A, Lohmussaar E, Metspalu A. Arrayed primer extension on the DNA chip—method & applications. In: Schena M, ed. *Microarray Biochip Technology.* Natick, MA: Eaton Publishing; 2000:247–263.

13. Pastinen T, Partanen J, Syvanen AC. Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. *Clin Chem.* 1996;42:1391–1397.

14. Shumaker JM, Metspalu A, Caskey CT. Mutation detection by solid phase primer extension. *Hum Mutat.* 1996;7:346–354.

15. Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* 1997;7:606–614.

16. Sohocki MM, Perrault I, Leroy BP, et al. Prevalence of AIPL1 mutations in inherited retinal degenerative disease. *Mol Genet Metab.* 2000;70:142–150.

17. den Hollander AI, Heckenlively JR, van den Born LI, et al. Leber congenital amaurosis and retinitis pigmentosa with Coats-like exudative vasculopathy are associated with mutations in the crumbs homologue 1 (CRB1) gene. *Am J Hum Genet.* 2001;69:198–203.

18. den Hollander AI, ten Brink JB, de Kok YJ, et al. Mutations in a human homologue of Drosophila crumbs cause retinitis pigmentosa (RP12). *Nat Genet.* 1999;23:217–221.

19. Freund CL, Gregory-Evans CY, Furukawa T, et al. Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell.* 1997;91:543-553.

20. Swain PK, Chen S, Wang QL, et al. Mutations in the cone-rod homeobox gene are associated with the cone-rod dystrophy photoreceptor degeneration. *Neuron.* 1997;19:1329–1336.

21. Dharmaraj SR, Silva ER, Pina AL, et al. Mutational analysis and clinical correlation in Leber congenital amaurosis. *Ophthalmic Genet.* 2000;21:135–150.

22. Perrault I, Rozet JM, Gerber S, et al. Spectrum of retGC1 mutations in Leber's congenital amaurosis. *Eur J Hum Genet* 2000;8:578–582.

23. Simovich MJ, Miller B, Ezzeldin H, et al. Four novel mutations in the RPE65 gene in patients with Leber congenital amaurosis. *Hum Mutat.* 2001;18:164.

24. Thompson DA, Gyurus P, Fleischer LL, et al. Genetics and phenotypes of RPE65 mutations in inherited retinal degeneration. *Invest Ophthalmol Vis Sci.* 2000;41:4293–4299.

25. Morimura H, Fishman GA, Grover SA, et al. Mutations in the RPE65 gene in patients with autosomal recessive retinitis pigmentosa or

leber congenital amaurosis. *Proc Natl Acad Sci USA.* 1998;95: 3088–3093.

26. Dryja TP, Adams SM, Grimsby JL, et al. Null RPGRIP1 alleles in patients with Leber congenital amaurosis. *Am J Hum Genet.* 2001; 68:1295–1298.

27. Gerber S, Perrault I, Hanein S, et al. Complete exon-intron structure of the RPGR-interacting protein (RPGRIP1) gene allows the identification of mutations underlying Leber congenital amaurosis. *Eur J Hum Genet.* 2001;9:561–571.

28. Wilkie SE, Newbold RJ, Deery E, et al. Functional characterization of missense mutations at codon 838 in retinal guanylate cyclase correlates with disease severity in patients with autosomal dominant cone-rod dystrophy. *Hum Mol Genet.* 2000;9:3065–3073.

29. Rozet JM, Perrault I, Gerber S, et al. Complete abolition of the retinal-specific guanylyl cyclase (retGC-1) catalytic ability consistently leads to Leber congenital amaurosis (LCA). *Invest Ophthalmol Vis Sci.* 2001;42:1190–1192.

30. Ramamurthy V, Tucker C, Wilkie SE, et al. Interactions within the coiled-coil domain of RetGC-1 guanylyl cyclase are optimized for regulation rather than for high affinity. *J Biol Chem.* 2001;276: 26218–26229.

31. Tucker CL, Ramamurthy V, Pina AL, et al. Functional analyses of mutant recessive GUCY2D alleles identified in Leber congenital amaurosis patients: protein domain comparisons and dominant negative effects. *Mol Vis.* 2004;10:297–303.

32. Akey DT, Zhu X, Dyer M, et al. The inherited blindness associated protein AIPL1 interacts with the cell cycle regulator protein NUB1. *Hum Mol Genet.* 2002;11:2723–2733.

33. Ramamurthy V, Roberts M, van den Akker F, et al. AIPL1, a protein implicated in Leber's congenital amaurosis, interacts with and aids in processing of farnesylated proteins. *Proc Natl Acad Sci USA.* 2003;100:12630–12635.

34. Chau KY, Chen S, Zack DJ, Ono SJ. Functional domains of the cone-rod homeobox (CRX) transcription factor. *J Biol Chem.* 2000;275:37264–37270.

35. Swaroop A, Wang QL, Wu W, et al. Leber congenital amaurosis caused by a homozygous mutation (R90W) in the homeodomain of the retinal transcription factor CRX: direct evidence for the involvement of CRX in the development of photoreceptor function. *Hum Mol Genet.* 1999;8:299–305.

36. Silva E, Dharmaraj S, Li YY, et al. A missense mutation in GUCY2D acts as a genetic modifier in RPE65-related Leber congenital amaurosis. *Ophthalmic Genet.* 2004;25:205–217.

37. Dharmaraj S, Li Y, Robitaille JM, et al. A novel locus for Leber congenital amaurosis maps to chromosome 6q. *Am J Hum Genet.* 2000;66:319–326.

38. Dharmaraj S, Leroy BP, Sohocki MM, et al. The phenotype of Leber congenital amaurosis in patients with AIPL1 mutations. *Arch Ophthalmol.* 2004;122:1029–1037.

39. Koenekoop RK, Fishman GA, Iannaccone A, et al. Electroretinographic abnormalities in parents of patients with Leber congenital amaurosis who have heterozygous GUCY2D mutations. *Arch Ophthalmol.* 2002;120:1325–1330.

40. Perrault I, Hanein S, Gerber S, et al. Retinal dehydrogenase 12 (RDH12) mutations in leber congenital amaurosis. *Am J Hum Genet.* 2004;75:639–646.

41. Janecke AR, Thompson DA, Utermann G, et al. Mutations in RDH12 encoding a photoreceptor cell retinol dehydrogenase cause childhood-onset severe retinal dystrophy. *Nat Genet.* 2004; 36:850–854.