

Mutations causing severe combined immunodeficiency: detection with a custom resequencing microarray

Tonya Lebet, BA¹, Richard Chiles, PhD², Amy P. Hsu, BA³, Elaine S. Mansfield, PhD⁴, Janet A. Warrington, PhD⁴, and Jennifer M. Puck, MD¹

Purpose: Mutation diagnosis of severe combined immunodeficiency is challenging because of the multiplicity of disease genes and large number of disease-causing mutations, including unique ones that continue to be found. A resequencing microarray could facilitate mutation detection, increasing the chance of diagnosing infants early for optimal rescue by hematopoietic stem cell transplantation. **Methods:** After analyzing cumulative mutations, we developed a custom Affymetrix GeneChip® microarray including probes representing exons and flanking regions of severe combined immunodeficiency disease genes. DNA samples were analyzed by array versus standard dideoxy genomic sequencing. We tested males and their mothers with X-linked *IL2RG* variants and patients and carriers with autosomal variants in *IL7R*, *JAK3*, and *DCLRE1C*. **Results:** New, unique severe combined immunodeficiency mutations are frequent. Resequencing array call rates of 95–98% exceeded GeneChip product specifications, and all of 47 point mutations in known samples were detected, as were the sites of 12 of 22 disease-causing insertions and deletions. Each gene had particular nucleotides that were often not called correctly and had to be excluded from analysis; exclusion rates ranged from 0.4% (hemizygous *IL2RG*) to 9.2% (heterozygous *JAK3*). **Conclusion:** Microarray resequencing is a promising technology for severe combined immunodeficiency mutation diagnosis that can detect both known and new mutations. Future customization of probe sequences and analysis algorithms could increase the number of accurately called nucleotides. *Genet Med* 2008;10(8):575–585.

Key Words: severe combined immunodeficiency, mutation detection, microarray, resequencing, sequencing, hot-spot, point mutation, indel, *IL2RG*, *IL7R*, *JAK3*, *DCLRE1C*

Severe combined immunodeficiency (SCID) describes a group of heritable diseases characterized by a profound lack of both T cell and B cell immunity.^{1–4} Infants with SCID develop severe infections and do not survive beyond their first years of life unless provided with functional immunity, such as by hematopoietic stem cell transplantation (HSCT) from a healthy donor. To date at least 12 distinct SCID disease genes have been identified.⁵ All patients with SCID have very low T cell numbers, but about two thirds have normal or even elevated numbers of B cells which, however, fail to produce specific antibody-

ies. These T[−]B⁺ SCID cases are almost all due to defects in one of three genes that encode lymphocyte receptors and mediators of cytokine signals. X-linked *IL2RG* is mutated in approximately half of all SCID cases and encodes the common γ chain, γ c, found in receptors for IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21. Between 5% and 15% of cases are caused by defects in each of *IL7R*, encoding the IL-7 receptor α chain, and *JAK3*, encoding an intracellular Janus activating kinase associated with the γ c cytoplasmic tail.

Early diagnosis of SCID, before the onset of severe infections, and prompt treatment in the first months of life bring about the best outcomes.⁶ However, most SCID infants lack a positive family history and appear normal at birth. Moreover, their early infections may not be distinguishable from those in otherwise healthy infants. Newborn screening for SCID is a desirable public health measure to identify affected infants at a preinfectious stage of their disease, but until this is instituted, physician awareness and ready access to definitive diagnosis are critical so that affected infants can be diagnosed in time for HSCT therapy.^{7,8}

Molecular diagnosis of SCID mutations is challenging because a large number of deleterious mutations can be found in each gene, and multiple genes can be mutated to give the same

From the ¹Department of Pediatrics, University of California San Francisco School of Medicine, San Francisco, California; ²Signal Demand, San Francisco, California; ³Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland; ⁴Affymetrix Inc., 3420 Central Expressway, Santa Clara, California.

Jennifer M. Puck, MD, University of California, San Francisco, Box 0519, 513 Parnassus Avenue, HSE 301A, San Francisco, CA 94143-0519. E-mail: puckj@peds.ucsf.edu.

Disclosure: Janet A. Warrington and Elaine S. Mansfield are employees of Affymetrix and own stock in Affymetrix, Inc. There are no other potential conflicts of interest.

Supplementary appendixes are available via the ArticlePlus feature at www.geneticsinmedicine.org.

Submitted for publication March 16, 2008.

Accepted for publication May 20, 2008.

DOI: 10.1097/GIM.0b013e31818063bc

phenotype. We investigated whether direct detection of known mutations versus resequencing would be appropriate for SCID mutation detection and developed an array-based resequencing approach.

MATERIALS AND METHODS

Patient samples

Patients with SCID and their parents, if available, were enrolled with informed consent in IRB-approved protocols to determine genotype and genotype/phenotype correlations. Some parents were enrolled for carrier determination whose SCID offspring were deceased with no stored tissue. Medical records were reviewed. Blood samples were subjected to enumeration of lymphocyte subsets, DNA extraction, and transformation with Epstein-Barr Virus (EBV) to create B cell lines,⁹ which in turn were used to prepare additional DNA and to measure STAT5 phosphorylation in response to IL-2 stimulation.¹⁰ Mutations found in 240 consecutive patients with *IL2RG* mutations enrolled between 1987 and 2006 were used to establish rates of new versus recurrent mutations.^{11–17} Samples from 59 SCID patients with known SCID mutations and 50 carrier parents^{10–19} (J. Puck, unpublished data) were used to assess array performance. An additional 10 samples from SCID patients of unknown genotype were tested by array prior to standard dideoxy sequence determination.

Resequencing microarray design

Our microarray covered all exons and 50 bp of flanking non-coding sequence from the SCID genes *IL2RG*, *IL7R*, *JAK3*, *ADA*, *RAG1*, *RAG2*, *DCLRE1C* (Artemis), and *LIG4* and additional human immunodeficiency genes (full list in Supplement, available online only). Successive 25-base oligonucleotides stepping through the sequence, each offset by one base, were synthesized on wafers as described.^{20–22} Both forward and reverse strands were included, and the 13th base in each oligo was represented by the reference sequence nucleotide and each of the other three nucleotides. GenBank reference mRNA sequences were aligned with genomic DNA using the Blat program (UC Santa Cruz Genome Browser, <http://genome.ucsc.edu>). Entries were subjected to Repeat Masker (Institute for Systems Biology, Seattle, WA) to identify internal duplications, ALU and LINE repetitive elements, and low-complexity repeats. With the exception of the terminal exons encoding polyadenylation regions of *IL2RG* and *IL7R*, noncoding repetitive elements were excluded from the array.

DNA extraction and PCR

DNA was extracted from blood samples or EBV cell lines using the Puregene Tissue Extraction kit (Gentra Systems, Minneapolis, MN). Long range polymerase chain reaction (PCR) to enrich targeted sequences with a minimum of reactions was performed using 500 ng DNA with a Takara LA PCR Kit (TaKaRa Bio Inc., Shiga, Japan). Primers were designed using the Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and WebPrimer (<http://seq.yeastgenome.org/cgi-bin/web-primer>) programs. PCR amplicons ranged from 0.41 to 9.8 kb with an average of 6.2 kb. A 7.5-kb plasmid (IQ-EX_7) included in the GeneChip® Resequencing Assay Kit (Affymetrix, Santa Clara, CA) was amplified as directed as a positive internal control for every hybridization.

Quantitation, pooling, fragmentation, and labeling of PCR products

PCR products were purified of residual primers and nucleotides using Clontech purification plates (Clontech, Mountain View, CA) and DNA concentration of each amplicon was measured by spectrophotometer (NanoDrop Technologies, Wilmington, DE). Equimolar amounts of PCR products were pooled and processed using the GeneChip Resequencing Assay Kit. After fragmentation with 0.1 U/μL DNase (0.06 U/μg DNA) at 37°C for 35 minutes and inactivation at 95°C for 15 minutes, DNA was biotinylated using terminal deoxynucleotidyl transferase as described.²²

Microarray hybridization and analysis

Hybridization was performed as described.²² Briefly, 120 μL of hybridization solution was placed into microarray chambers, and arrays were rotated at 60g at 45°C for 16 hours followed by washing and staining on an Affymetrix fluidics station using the Midi_DNAARRAY_450 protocol.²³ The arrays were then scanned on a GeneChip 3000 Scanner (Affymetrix), and the data were analyzed using Affymetrix GeneChip Resequencing Analysis Software (GSEQ® v4.0). False positive, false negative, and no-call rates for each gene and the control plasmid were evaluated. The performance of the array was judged correct upon identification of variant from reference sequence. In GSEQ analysis, one can set the genome model to either haploid or diploid mode. Haploid mode makes only homozygous calls (A, C, T, G), while diploid mode makes homozygous or heterozygous calls assuming two alleles are present (AC, AT, AG, etc.). Haploid mode was used to analyze the X-linked *IL2RG* gene for boys, while diploid mode was used for *IL2RG* analysis of female carrier samples and all autosomal gene analysis.

Chi-squared tests were used to determine whether observed SNP genotypes corresponded to Hardy-Weinberg expectations.

Replicate sample analysis

Sixteen genomic DNA samples from SCID patient EBV lines were assayed by different individuals in two locations, Affymetrix (Santa Clara, CA) and University of California, San Francisco. Reproducibility was defined by the ability of the array-based assay at each location to identify previously documented mutations.

Dideoxy sequencing

Genomic DNA amplified as described previously was sequenced using a Big Dye v3.1 Terminator DNA Sequencing kit (Applied Biosystems, Foster City, CA). Sequencing products were purified with Agencourt CleanSEQ beads (Beckman Coulter, Beverly, MA) and separated by capillary electrophore-

sis on an ABI 3100 DNA Analyzer (Applied Biosystems). Sequence analysis was performed with Sequencher (Version 4.7, GeneCodes Corporation, Ann Arbor, MI).

RESULTS

Proportion of unique versus recurrent mutations

For genetic disorders in which a known set of mutations underlie the vast majority of disease cases, such as cystic fibrosis,^{24,25} direct detection of mutant alleles can be an effective approach to clinical molecular diagnosis. However, primary immunodeficiency diseases, particularly those encoded on the X chromosome, are rare and have diverse mutations, including newly arising defects.^{14,16} For such genes, a resequencing strategy may be required to achieve an acceptable rate of mutation detection in a clinical setting.

We determined the relative contributions of previously known versus newly discovered mutations in 240 consecutive, unrelated cases of X-linked SCID studied in our laboratory in the past 20 years.^{10,12–17} The mutations in *IL2RG* in this cohort were listed by date of case referral and divided into 12 groups of 20 cases, as shown in Figure 1. The number of unique mutations as each successive group was added (orange boxes, left axis) continued to grow and did not reach a plateau, even though cases with recurrent mutations also rose (green circles, left axis). Of a total of 153 distinct mutations, 87 (64%) were found only once in this retrospective cohort. The cumulative percentage of unique mutations (Fig. 1, black triangles and broken line, right axis) decreased from 90% in the first group of 20 cases to less than 70% after 100 cases were analyzed (Fig. 1, Groups 1–5) but were still >60% after 240 cases (Fig. 1, Group 12).

Five hot-spot point mutations at CG dinucleotides, cDNA 684T, 690T, 691A, 868A, and 879T, were each found 10–13 times in this series, and mutation of the branch point adeno-

sine, exon3(–15)g, occurred 6 times; together these six hot-spots accounted for 26% of all mutations.

Of the 240 *IL2RG* cases in this series, 189 had point mutations, 31 were deletions, and 13 were insertions of one to 12 nucleotides. There were five complex microinsertion and -deletion events and two large deletions spanning four or more exons.

A custom resequencing microarray approach was chosen for its ability to detect new as well as previously known point mutations. Compared to *IL2RG* mutations, 5-fold fewer *JAK3* and 10-fold fewer *IL7R* mutations have been detected, either in our laboratory or from the literature.^{25,26} Lower numbers of cataloged gene mutations as well as documented newly arising autosomal *IL7R* mutations argued in favor of a resequencing strategy (Puck, unpublished data).

Array design, production, and testing

Eight SCID disease genes and additional immunodeficiency genes were represented by tiled, overlapping 25-base oligonucleotides, representing both forward and reverse strands. In addition, 644 disease-associated insertion and deletion (indel) mutations from our own data and published reports were tiled as alternative oligonucleotides. These sequences were designed to test whether previously reported difficulties in detection of indel mutations could be improved with specific oligonucleotides matching the mutant sequences.²³ To pilot the array-based assay, we focused on X-linked *IL2RG* as well as the two next most commonly defective genes in T⁻B⁺ SCID, *IL7R*, and *JAK3*. Together these genes account for around two thirds of all SCID. Moreover, each of these genes presented features of interest as prototypes: we were able to compare hemizygous male patients with *IL2RG* defects to their heterozygous mothers; *IL7R* contains five frequent coding SNPs and as all autosomal recessive SCIDs requires identification of two defective alleles per patient, and *JAK3* is a larger autosomal gene that contains GC-rich regions that have proven difficult for dideoxy sequencing. In addition, we looked at a common homozygous founder mutation in *DCLRE1C* (also called Artemis).¹⁷

An early evaluation of reproducibility consisted of duplicate runs in two laboratories starting with genomic DNA aliquots with X-linked SCID point mutations previously verified by dideoxy sequencing. All 16 sequence variants were found in one laboratory and 15 of 16 in the other (data not shown).

Analysis of array data: effect of batch size

Array and GSEQ 4.0 software performance were assessed as a function of number of arrays analyzed in a batch. GSEQ contains learning algorithms derived from ABACUS, an adaptive background genotype calling scheme²⁰ to optimize data from multiple arrays analyzed together; larger batch sizes in an analysis are thus expected to have greater accuracy. The same arrays analyzed in batch sizes of 1, 3, 6, 12, and 24 DNA samples from males hemizygous for the X-linked gene *IL2RG* are shown in Table 1 (run in haploid mode). Only two false positive calls and seven or fewer false negative calls were observed with all batch sizes, giving an error rate (Table 1) below 2×10^{-4} ; thus, the limit of performance was almost entirely de-

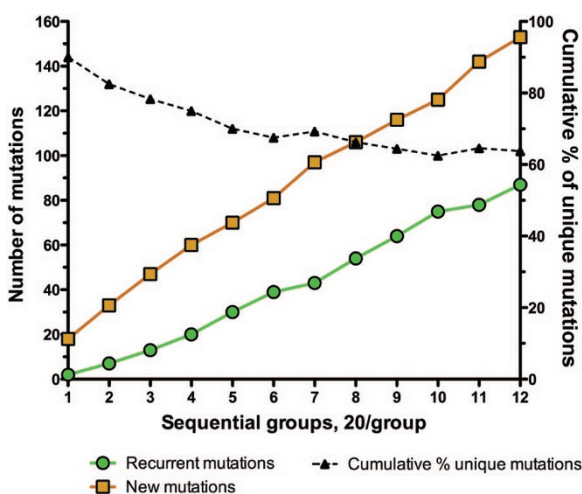


Fig. 1. New versus recurrent mutations in *IL2RG*, the disease gene for X-linked severe combined immunodeficiency (SCID). In 240 sequential mutations in groups of 20, those not observed before (orange boxes) and recurrent mutations (green circles) are shown (left axis). The percent of total mutations that were unique is also indicated (black triangles and dotted line, right axis).

Table 1

Effect of batch size on *IL2RG* call rates for hemizygous (haploid) male samples^a

No. files/batch	False positives ^b	False negatives ^c	No-calls ^d	No-call rate ^e	Error rate ^f ($\times 10^{-4}$)	False positive plus no call rate ^g
1	2	3	308	0.0061	0.99	0.0062
3	2	6	237	0.0047	1.6	0.0048
6	2	7	227	0.0045	1.8	0.0045
12	2	7	227	0.0045	1.8	0.0045
24	2	3	232	0.0046	0.99	0.0046

^a*IL2RG* total reference calls 50,496; GSEQ quality score set at 2; haploid mode.

^bCalled nonreference nucleotide (variant) that was not present on dideoxy sequence.

^cFailed to call a variant that was present on dideoxy sequence.

^dNo nucleotide called.

^eNo-calls/total reference calls.

^f(False positives plus false negatives)/total reference calls.

^g(False positives plus no-calls)/total reference calls.

pendent upon the rate of unassigned nucleotides (no-calls). Although no-calls decreased 25%, from 0.0061 to 0.0047, when batch size increased from one to three, there was little further improvement with up to 24 samples per batch.

In contrast to the hemizygous analysis, the diploid analysis with heterozygous female DNA samples sequenced for *IL2RG*, plus all samples sequenced for *IL7R* and *JAK3*, had many more false positive calls and no-calls (Table 2). Also, heterozygous call rates markedly improved when larger batches of arrays were analyzed together. At a batch size of 32, the diploid no-call rate was 5.5% for all three genes combined. As with the hemizygous X-linked gene, performance of the array for heterozygous mutation detection was largely limited by the no-call rate.

Analysis of array data: effect of quality score threshold

Utilization of a resequencing array for mutation screening in a clinical setting would require very low false negatives to avoid missing a diagnosis in an affected patient; false positives and no-calls would decrease specificity but were considered preferable in our selection of software settings. In GSEQ 4.0,

Table 2

Effect of batch size on call rates for diploid samples^a

No. files/batch	False positives	False negatives	No-calls	No-call rate	Error rate	False positive plus no-call rate
1	720	8	53733	0.159	0.0022	0.1612
4	780	10	31983	0.095	0.0023	0.0971
8	788	10	24755	0.073	0.0024	0.0757
16	762	9	20432	0.060	0.0023	0.0628
32	751	9	18575	0.055	0.0023	0.0572

^a*IL2RG* for females, *IL7R* and *JAK3* for both males and females; total reference calls 337,728; GSEQ quality score 3; diploid mode. Parameters as defined in Table 1.

Table 3

Effect of quality score on *IL2RG* call rates for hemizygous male samples^a

Quality score	False positives	False negatives	No-call rate	Error rate ($\times 10^{-5}$)	False positive plus no-call rate
2	2	3	0.0046	9.9	0.0046
3	2	3	0.0067	9.9	0.0068
5	1	3	0.0126	7.9	0.0126
10	0	3	0.0460	5.9	0.0460
15	0	3	0.1218	5.9	0.1218

^aTotal reference calls 50,496. Batch size 24, haploid mode. Parameters as defined in Table 1.

the quality score threshold can be manipulated to balance stringency for calling only correct bases versus not being able to make a definitive call. At high thresholds, GSEQ made few false calls, but many no-calls, while low thresholds maximized calls at the expense of accuracy. We analyzed our data at a variety of quality threshold scores to determine best overall performance. As shown in Table 3, X-chromosome hemizygous male samples had very few false calls in *IL2RG* at any threshold, and so optimal results were generated at our lowest quality score of 2, at which no-call rates were minimized. Thus, all further hemizygous analyses were done with batch sizes of 24 or more at quality score 2.

In contrast, the difficulty of making heterozygous calls in diploid samples caused unacceptably high false positives and more false negatives at score 2 than at higher quality thresholds (Table 4). However, no-call rates increased nearly 10-fold, from 0.038 to 0.36 as quality threshold was raised from 2 to 10. Failures to call the reference nucleotide were minimized at quality score threshold 3 for heterozygous analysis (Table 4, right column). All subsequent heterozygous analyses used batch sizes of 32 or greater and quality score threshold 3.

SNPs in *IL7R*

To evaluate proficiency of the array to make heterozygous calls, we examined five frequent coding SNPs in *IL7R* in 59 unrelated individuals (Table 5). GSEQ assigned the expected homozygous and heterozygous calls 93–98% of the time. The minority of DNA samples that had undergone *IL7R* dideoxy se-

Table 4

Effect of quality score on call rates for diploid samples^a

Quality score	False positives	False negatives	No call rate	Error rate	False positive and no call rate
2	1008	11	0.038	0.0030	0.0412
3	751	9	0.055	0.0023	0.0572
5	521	9	0.110	0.0016	0.1063
10	282	5	0.360	0.0009	0.3623
15	153	4	0.700	0.0005	0.7017

^a*IL2RG* for females, *IL7R* and *JAK3* for both males and females; total reference calls 337,728; batch size 32; diploid mode. Parameters as defined in Table 1.

Table 5
Detection of frequent SNPs in *IL7R*

SNP	<i>IL7R</i> cDNA ^a	Exon	% of DNA samples analyzed ^b	Array allele frequency	Array genotype calls			Hardy-Weinberg P
					Homozygote	Heterozygote	Homozygote	
rs1494558	197	2	97	0.39/0.61	8 T	28 CT	21 C	NS ^c
rs1494555	412	4	98	0.40/0.60	9 G	28 AG	21 G	NS
rs7737000	495	4	93	0.89/0.11	45 C	8 CT	2 T	NS
rs6897932	731	6	98	0.78/0.22	36 C	18 CT	4 T	NS
rs3194051	1066	8	98	0.77/0.23	36 A	17 AG	5 G	NS

^aRelative to transcription start site = 1.

^bOf 59 DNA samples from unrelated subjects those with no-calls and calls other than published SNP nucleotides were excluded.

^cNo significant difference between observed genotypes and Hardy-Weinberg expectations.

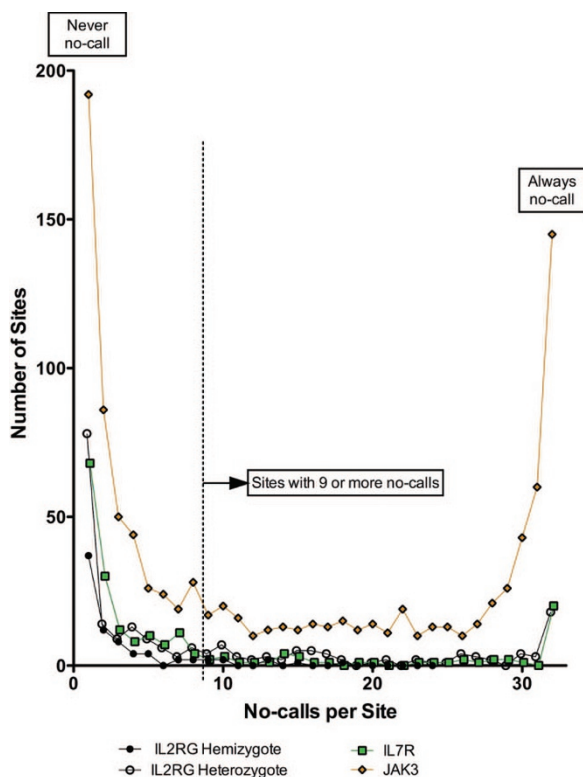


Fig. 2. Histograms derived from hemizygous samples analyzed in haploid mode for *IL2RG* ($n = 24$, filled circles), and heterozygous samples ($n = 32$) analyzed in diploid mode for *IL2RG* (open circles), *IL7R* (green squares) and *JAK3* (orange diamonds). Sites called most reliably are at the left side of the histograms, while those always generating “no-calls” are at the right. No-call rates were highest for *JAK3*, due to its larger size and high GC content.

quencing of these SNP regions showed concordance with allele calls by the array (not shown). While ethnic backgrounds were not recorded for all of our predominantly Caucasian subjects, our observed allele frequencies were consistent with frequencies in dbSNP (not shown). Moreover, the observed array calls were also in concordance with Hardy-Weinberg expectations (Table 5).

Sites often not called correctly

A total of 16,000 no-call designations were made at 337,000 positions queried, but several particular nucleotide sites ac-

counted for a disproportionate number of no-calls. Distributions of no-call sites are shown in Figure 2. Although the great majority of sites were called correctly in the entire test set, a number of sites had one no-call, as indicated at the left of the histograms. Progressively fewer sites received no-call assignments more often, and 681 difficult-to-call reference nucleotides each had at least 9 of our test samples yielding no-calls. For *IL2RG*, there were eight hemizygous (filled circles, Fig. 2) and 78 diploid (open circles) problematic sites out of 2104 nucleotides, or 0.4% and 3.7%, respectively. For *IL7R* there were 2% problematic sites (green squares), and for *JAK3*, 9.2% (orange diamonds). Analysis in heterozygous mode gave 100% no-calls at 145 sites in *JAK3* but only 20 and 18 in *IL2RG* and *IL7R*, respectively (Fig. 2, right side of graph). Hemizygous analysis of *IL2RG* showed no sites that were always no-calls. Positions with frequent bad calls included areas of high GC content, runs of the same nucleotide, directly repeated motifs, and other sites for which no reason for inaccuracy was obvious.

Recognizing that certain sites could not be called correctly with the current set of oligomers on the array, we filtered out sites with nine or more no-calls (Fig. 2, right of broken line) and carried out additional analysis on the three genes after removing these sites as well as the five *IL7R* SNP frequent variant sites (Table 6). Not surprisingly, the hemizygous *IL2RG* assay remained the most accurate, with overall noncorrect calls of only 0.3%, followed by both diploid *IL2RG* and *IL7R*, with 0.5% noncorrect calls. After filtering out problematic bases, *JAK3* noncorrect calls improved to 0.7%. Thus, most, but not all, regions of our test genes could be successfully hybridized to the 25-nucleotide oligomers in standard CustomSeq® conditions and analyzed with GSEQ 4.0 software. The differences in overall performance between the genes reflected the fact that their primary sequences were not equally appropriate for the default oligomer design, hybridization, washing conditions, and analysis tools.

Mutation detection in *IL2RG*

IL2RG point mutations were recognized by the array in 100% of 25 male SCID patients, analyzed in haploid mode, and in all but 1 (95%) of 21 mothers of boys with X-linked SCID (Table 7, Fig. 3). Importantly, all six major hot-spots for recur-

Table 6
Array-based assay performance after filtering out frequent SNPs and problematic base positions

Gene	Remaining bases ^a	Fraction of bases that had to be removed	False positive calls (rate) ^b	False negative calls (rate) ^c	No-calls (rate) ^d
<i>IL2RG</i> hemizygous	2096	0.004	2 (0.004%)	2 (0.004%)	151 (0.3%)
<i>IL2RG</i> diploid	2026	0.037	2 (0.003%)	2 (0.003%)	335 (0.5%)
<i>IL7R</i> ^c	2408	0.020	80 (0.1%)	7 (0.009%)	393 (0.5%)
<i>JAK3</i> ^c	5439	0.092	232 (0.13%)	0	1321 (0.7%)

^aReference bases minus SNP sites minus sites miscalled in 9 or more of 24 hemizygous or 32 diploid samples.

^bNon-correct calls per filtered reference base per DNA sample.

^cFailure to call a documented variant per filtered reference base per DNA sample.

^dNo-calls per filtered reference base per DNA sample.

rent mutation (Table 7, bold type) were detected in hemizygous males. Figure 3, A illustrates precise GSEQ calling of male point mutations at cDNA 678 and hot-spots 684 and 691 in *IL2RG* exon 5. Even when the precise mutation was not specified (cDNA 709, red circle, sample line 8), a single no-call at the position of the mutation prompted examination of signal intensities at that site. A two-fold or greater intensity of one nucleotide over the others diagnosed the exact cDNA 709 mutation (Fig. 3, A, right).

Figure 3, B, illustrates how frequent no-calls in multiple samples at *IL2RG* exon 1 cDNA 75 to 92 would have precluded mutation identification had one existed in this region. Examination of the individual nucleotide signal intensities, such as the forward strand at cDNA 92 of sample line 9 (Fig. 3, B, right), showed no predominant nucleotide. In contrast, a true splice mutation in the 4th sample row in this illustration (green highlight) was readily diagnosed at cDNA 129(+1)g>c in a region accurately called in all samples.

Detection of indel and complex mutations in *IL2RG* was less successful than detection of point mutations, in part because the available software had no algorithm designed for this application. Nonetheless, the locations of mutated nucleotides were correctly pinpointed in 10 of 17 boys with X-linked SCID (59%) and 5 of 14 mutation-carrying mothers (36%; Table 8). Although certain insertions and deletions were completely missed, as with the nine base tandem duplication at cDNA 725 (Table 8), others were suggested by calls of nonreference nucleotides or single or multiple no-calls in a region otherwise generally called well (not shown).

Because alternative tilings spanning known insertion and deletion mutations had been placed on our array, manual examination of these specific oligonucleotides confirmed the correct mutation in three hemizygous male samples, two with insertions at cDNA 373, a known frequent insertion/deletion site following seven A residues,¹⁵ and one at cDNA 750 adjacent to 5 G residues.

Mutation detection in autosomal recessive SCID

Point mutation detection in SCID patients and carrier parents with previously proven autosomal recessive *IL7R* and *JAK3* defects is illustrated in Table 9. All of 22 DNA samples with mutations at 11 sites in *IL7R* and in *JAK3* were identified

precisely by GSEQ. In addition, the SCID founder mutation in Athabaskan Indian tribes in *DCLRE1C* was detected in both homozygous patient and heterozygous carrier DNA (not shown).

Three of five indel mutations in *IL7R* were identified in the correct nucleotide position on the array, with confirmation upon examination of the mutation specific alternate tilings as with *IL2RG* indels discussed above (not shown).

Mutations in newly diagnosed patients with SCID

Our laboratory has continued to receive blood samples from newly diagnosed patients with SCID, affording an opportunity to test the resequencing array as an initial molecular diagnostic tool. In 10 of 11 new SCID cases mutations were found in *IL2RG* or *IL7R* (Table 10), all subsequently confirmed by dideoxy sequencing. One of the 11 cases had no mutation found by either the array or dideoxy sequencing (not shown). Four of the cases (Table 10, bold type) had *IL2RG* mutations not previously reported in the literature.

In one case with the hot-spot mutation at *IL2RG* cDNA 690, an initial dideoxy sequence in a clinical laboratory had been erroneously reported as normal. In another case, we were able to demonstrate a newly arising *IL2RG* missense mutation causing sporadic SCID. The array (Fig. 4, A) indicated that the patient, but not his mother, had a C to T transition at cDNA 684, and patient and maternal genotypes were confirmed by subsequent dideoxy sequencing (Fig. 4, B). As discussed previously, the cDNA 684T mutation is also one of the most frequent hot-spots associated with X-linked SCID.

DISCUSSION

Our analysis of accumulated mutations causing human SCID argues that a resequencing approach, rather than direct detection of previously known mutations, will be required for the foreseeable future due to the substantial proportion of unique mutations that continue to be found, exceeding 60% even after 240 patient genotypes were accumulated. The large number of SCID genes and many mutations in each gene suggest that arrays can become cost effective compared to dideoxy sequencing, particularly with long PCR enrichment of target areas to minimize the number of PCR reactions required. In-

Table 7
Detection of point mutations in *IL2RG*

Point mutations queried			Hemizygous SCID males		Diploid mothers	
Exon	cDNA position ^a	Reference nucleotide	Dideoxy sequence	Array sequence	Dideoxy sequence	Array sequence
1	17	G	A	A	G/A	G/A
	129 (+1)	G	C	C	<i>na</i> ^b	<i>na</i>
2	171	G	A	A	<i>na</i>	<i>na</i>
	229	G	A	A	<i>na</i>	<i>na</i>
3	284 (-15)^c	A	G	G	A/G	A/G
	284 (-15)	A	G	G	<i>na</i>	<i>na</i>
	342	G	T	T	G/T	Single no-call ^d
	355	G	<i>na</i>	<i>na</i>	G/A	G/A
	405	C	T	T	C/T	C/T
4	536	G	A	Single no-call	G/A	G/A
5	678	C	T	T	C/T	C/T
	684	C	T	T	C/T	C/T
	684	C	T	<i>na</i>	C (WT) ^e	C (WT) ^e
	690	C	T	T	C/T	C/T
	691	G	A	A	G/A	G/A
	691	G	A	A	G/A	G/A
	709	G	T	Single no-call, T ^{d,f}	<i>na</i>	<i>na</i>
	717	C	T	T	C/T	C/T
	717	C	T	T	<i>na</i>	<i>na</i>
	717	C	<i>na</i>	<i>na</i>	C/T	C/T
6	816	G	C	C	G/C	G/C
	868	G	A	A	G/A	G/A
	868	G	A	A	<i>na</i>	<i>na</i>
	879	C	T	T	C/T	Many no-calls ^g
7	937	C	A	<i>na</i>	C (WT) ^e	C (WT) ^e
	938	G	C	C	<i>na</i>	<i>na</i>
	938	G	C	Single no-call, C ^{d,f}	G/C	G/C
	938 (+1)	G	A	A	<i>na</i>	<i>na</i>
	938 (+1)	G	C	C	G/C	G/C

^aRelative to transcription start site = 1 (GenBank sequence NM_000206).

^b*na*, not available.

^cBold type, 6 mutation hot-spots, accounting for 26% of all X-linked SCID patient mutations.

^dPrecise mutant call not made by GSEQ 4.0, but isolated no-call in a series of perfectly called bases suggested an abnormality at the true mutant site (see Fig. 3).

^eMother with wild type sequence in blood DNA; son's new mutation documented elsewhere.

^fCorrect mutation identified by examination of signal intensities on array (see Fig. 3).

^gMutation missed, could not distinguish true mutation from other no-calls nearby.

deed, in our research laboratory, the cost of sequence analysis of seven SCID genes by the dideoxy method was nearly twice that of the array method (\$1970 for dideoxy, \$1040 for array; for details see Supplement, available online only).

Although its first application was in SNP discovery and genotyping, recent reports have evaluated resequencing arrays

as a means of finding human disease-causing mutations in hypogonadotropic hypogonadism,²² congenital myasthenia,²² retinitis pigmentosa,²⁷ lung cancer,²⁸ and intrahepatic cholestasis.²⁹ All of these studies achieved overall call rates exceeding 90% using methods based on adaptive background genotype calling scheme algorithms of Cutler et al.,²⁰ which

Table 8
Detection of insertion and deletion mutations in *IL2RG*

Insertion/deletion mutations queried		Hemizygous SCID males		Diploid mothers	
Exon	cDNA position	Dideoxy sequence	Array sequence	Dideoxy sequence	Array sequence
1	32	<i>na</i> ^a	<i>na</i>	insA/WT ^b	Single no-call ^c
	88	delC	Single no-call	<i>na</i>	<i>na</i>
	88	del5	7 no-calls ^d	<i>na</i>	<i>na</i>
2	143	delC	<u>WT</u> ^e	delC/WT	<u>WT</u>
3	351	delT	<u>WT</u>	<i>na</i>	<i>na</i>
	373	insT	G	insT/WT	A/T
	373	insA	A ^f	<i>na</i>	<i>na</i>
	375	<i>na</i>	<i>na</i>	delG/WT	<u>Many no-calls</u> ^e
	375	<i>na</i>	<i>na</i>	insA/WT	<u>WT</u>
	375	<i>na</i>	<i>na</i>	delG/WT	Single no-call
	383	delCCTC insGCTTT	17 no-calls ^d	<i>na</i>	<i>na</i>
	447	delA	Single no-call	delA/WT	Single no-call
	477	insA	<u>WT</u>	insA/WT	<u>WT</u>
	4	557	delCT	GA	delCT/WT
5	617	insATCA	<u>WT</u>	<i>Na</i>	<i>na</i>
	725	ins9, tandem dup	<u>WT</u>	<i>na</i>	<i>na</i>
	750	insT	G	insT/WT	Single no-call
	807	del23	<u>WT</u>	del23/WT	<u>WT</u>
6	829	delT	<u>WT</u>	delT/WT	<u>WT</u>
	868 (+5)	delGA	2 no-calls ^d	delGA/WT	<u>WT</u>
8	1000	del8	19 no-calls ^d	del8/WT	<u>WT</u>

^a*na*, not available.

^bWT, wild type (present on one allele of all heterozygous female carriers).

^cPrecise mutant call not made, but isolated no-call suggested an abnormality at the true mutant site.

^dPrecise mutant call not made, but multiple no-calls found at the true mutant site.

^eMutation missed, could not distinguish true mutation from other no-calls nearby; underlining signifies variants missed by array.

^fCorrect mutation missed by automated analysis, but identified by manual examination of alternative tiling on array.

tilings to make more accurate consensus calls and ideally to automatically identify insertions and deletions.

Different SCID-associated genes had different proportions of sequence that proved difficult-to-call as was also the case for genes on an array designed to analyze hypogonadotropic hypogonadism.²² Although some problematic regions of our genes were predictable, such as GC-rich portions of *JAK3* or direct repeats,^{20,23} other regions had no apparent reason for this behavior. Further analysis might reveal why these regions are difficult.³⁰ A future customized array design to address difficult-to-read regions might include redundant representation of specific sequences; refined analysis algorithms could improve the accuracy of calls and increase the proportion of readable nucleotides. In addition to multiplicity of representation of problematic regions, alternative approaches to custom tiling could be evaluated. Where the thirteenth position on current 25-mer oligos is currently the only varied nucleotide, a differ-

ent varying position could generate a more diverse set of sequence tilings with improved resultant sensitivity as observed by Cronin et al.³¹ in the original report of photolithographic arrays for mutation detection. Another approach also presented in this study was to alter the length of oligos in regions with extreme GC content.³¹

Importantly, new molecular diagnoses in previously undiagnosed patients were made with the SCID gene array, including four mutations never reported previously (Table 10). These mutations extend and confirm our observations that the genes are not yet saturated for mutations and that many more new mutations will be found as more patients are diagnosed with SCID and their DNA is sequenced. Thus, resequencing approaches will continue to be more appropriate for SCID than alternative methods that identify only known mutations.

One new mutation we found was at a splice site in *IL2RG* and three others were missense changes that required func-

Table 9
Detection of point mutations in *IL7R* and *JAK3*

Gene	Exon	cDNA	Reference sequence	Sample description	Dideoxy sequence	Array sequence
<i>IL7R</i>	1	2	T	Homozygous SCID	G	G
		2	T	Parent	T/G	T/G
		2	T	Heterozygous SCID	T/G ^a	T/G ^a
		2	T	Parent	T ^a	T ^a
	2	104	T	Heterozygous SCID	T/A	T/A
		104	T	Parent	T/A	T/A
	3	353	G	Heterozygous SCID	G/A	G/A
		353	G	Parent	G/A	G/A
	5	616	C	Parent	C/T	C/T
		644	G	Heterozygous SCID	G/T ^a	G/T ^a
		644	G	Parent	G ^a	G ^a
		662	G	Heterozygous SCID	G/T	G/T
		662	G	Homozygous SCID	T	T
		662	G	Homozygous SCID	T	T
<i>JAK3</i>	5	507	C	Heterozygous SCID	C/A	C/A
	9	1208	G	Heterozygous SCID	G/A	G/A
	11	1569	G	Heterozygous SCID	G/A	A ^b
	11	1569	G	Parent	G/A	A ^b
	17	2350	G	Parent	G/A	G/A
	20	2712	C	Heterozygous SCID	C/A	C/A
	20	2712	C	Parent	C/A	C/A

^aNovel mutation in patient, parent had wild type reference sequence.

^bMutation site identified but called homozygous A instead of heterozygous G/A.

Table 10

Mutations discovered by array assay in newly diagnosed patients with SCID

Gene	Exon	cDNA	Reference sequence	Array sequence ^a	Defect
<i>IL2RG</i>	2	223^b	T	G	M70R
		280	A	G	Y89C
	4	565^b	A	G	E184G
		574^b	T	G	V187G
	5	678	C	T	R222C
		684	C	T	R224W
6	690 ^c	C	T	R226C	
	868 (+4)^b	A	G	Destroys splice site	
<i>IL7R</i>	5	644 ^d	G	T	G215V
		662 ^d	G	T	S221I

^aAll confirmed by subsequent dideoxy sequence.

^bBold type, mutation not previously reported.

^cPrevious dideoxy sequence by clinical laboratory initially reported as no mutation found.

^dHomozygous defect.

Ten new patients analyzed and 4 novel mutations discovered.

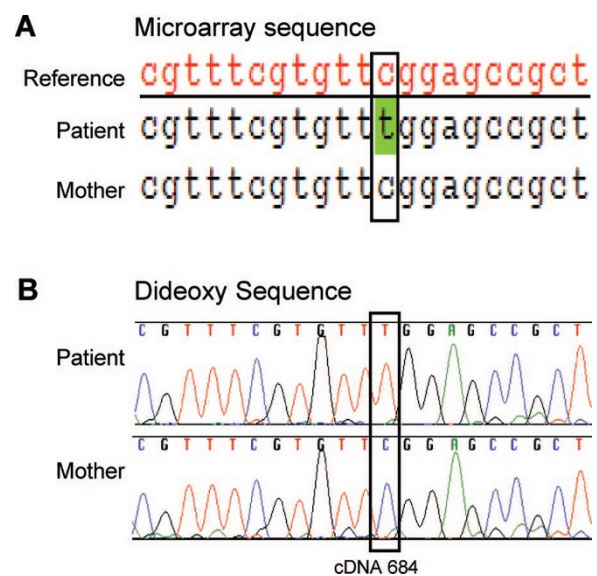


Fig. 4. Discovery of *IL2RG* exon 5 missense 684C>T; R224W variant. A, Variant T first identified by resequencing with array, highlighted in green. B, Variant, confirmed by dideoxy sequencing. This was a new mutation in the patient; DNA derived from maternal blood had only reference sequence.

tional assays to be performed to know whether they were deleterious. Functional characterization of *IL2RG* defects can be determined by IL-2 stimulation of EBV-transformed cells from controls and patients, after which phosphorylation of the downstream signaling molecule STAT5 is assayed by intracellular staining and flow cytometry.¹⁰ In each of these mutations tested, phospho-STAT5 was not detected.

Rapid, inexpensive and accurate genotyping is needed for genetic and medical management of patients with SCID and their relatives. Genetic heterogeneity makes molecular diagnosis essential for determining a family's recurrence risk and for conducting prenatal diagnosis. Increasingly HSCT treatment protocols are being tailored to patients' specific genotypes; for example, Athabascan SCID patients are highly sensitive to alkylating agents and should not receive standard doses of pre-transplant chemotherapy.¹⁹ Moreover, the proven benefit of early diagnosis of SCID in affording better outcomes should increase awareness and use of diagnostic testing. Array diagnosis could even become part of newborn screening programs for SCID. With our current array's demonstrated effectiveness in diagnosing all the common hot-spot mutations and other point mutations in the most common SCID genes, we estimate that it could identify the molecular basis of most T⁻B⁺ SCID cases. Further refinements in array design, analysis algorithms, or both could make this tool even more valuable.

ACKNOWLEDGMENTS

A portion of the research referred to in this publication was carried out under a research program initiated, developed, and funded in part by the Jeffrey Model Foundation (www.info4PI.org). Additional funding provided by Affymetrix, Inc. is gratefully acknowledged. JMP is supported in part by an NIAID, NIH, award to USIDNET, AI-30070. The authors thank Charles Haines for *IL7R* analysis, the immunologists (especially Mort Cowan and Zheng Xiao) and geneticists who referred cases, and the families with SCID who participated in our research.

References

- Puck JM. X-linked severe combined immunodeficiency. In: Ochs H, Smith CIE, Puck JM, editors. Primary immunodeficiency diseases: a molecular and genetic approach, 2nd ed. NY: Oxford University Press, 2007:123–136.
- Candotti F, Notarangelo L. Autosomal recessive severe combined immunodeficiency due to defects in cytokine signaling pathways. In: Ochs H, Smith CIE, Puck JM, editors. Primary immunodeficiency diseases: a molecular and genetic approach, 2nd ed. NY: Oxford University Press, 2007:136–152.
- De Villartay J-P, Schwarz K, Villa A. V(D)J recombination defects. In: Ochs H, Smith CIE, Puck JM, editors. Primary immunodeficiency diseases: a molecular and genetic approach, 2nd ed. NY: Oxford University Press, 2007:153–168.
- Buckley RH. Molecular defects in human severe combined immunodeficiency and approaches to immune reconstitution. *Annu Rev Immunol* 2004;22:625–655.
- Geha RS, Notarangelo LD, Casanova J-L, et al. Primary immunodeficiency diseases: an update, by the International Union of Immunological Societies (IUIS) Primary Immunodeficiency Diseases (PID) Classification Committee. *J Allergy Clin Immunol* 2007;120:776–794.
- Myers LA, Patel DD, Puck JM, Buckley RH. Hematopoietic stem cell transplantation for SCID in the neonatal period leads to superior thymic output and improved survival. *Blood* 2002;99:872–878.
- Chan K, Puck JM. Development of population-based newborn screening for severe combined immunodeficiency. *J Allergy Clin Immunol* 2005;115:391–398.
- Puck JM on behalf of The SCID Newborn Screening Working Group. Population-based newborn screening for severe combined immunodeficiency: steps toward implementation. *J Allergy Clin Immunol* 2007;120:760–768.
- Miller G. Immortalization of human lymphocytes by Epstein-Barr virus. *Yale J Biol Med* 1982;55:305–310.
- Kellermayer R, Hsu AP, Stankovics J, et al. Genetic Aspects of Severe Combined Immunodeficiency. *J Hum Genet* 2006;51:495–497.
- Puck JM, Nussbaum RL, Conley ME. Carrier detection in X-linked severe combined immunodeficiency based on patterns of X chromosome inactivation. *J Clin Invest* 1987;79:1395–1400.
- Puck JM, Deschènes SM, Porter JC, et al. The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-linked severe combined immunodeficiency, SCIDX1. *Hum Molecular Genet* 1993;2:1099–1104.
- Tassara C, Pepper AE, Puck JM. Intron point mutation in IL-2 receptor γ chain causing X-linked severe combined immunodeficiency. *Hum Mol Genet* 1995;4:1693–1695.
- Pepper AE, Buckley RH, Small TN, Puck JM. Two CpG mutational hot spots in the interleukin-2 receptor γ chain gene causing human X-linked severe combined immunodeficiency. *Am J Hum Genet* 1995;57:564–571.
- Puck JM, de Saint Basile G, Schwarz K, et al. IL2RGbase: a database of γ c-chain defects causing human X-SCID. *Immunol Today* 1996;17:507–511.
- Puck JM, Pepper AE, Henthorn PS, et al. Mutation analysis of IL2RG in human X-linked severe combined immunodeficiency. *Blood* 1997;89:1968–1977.
- Niemela JE, Puck JM, Fischer R, Fleisher TA, Hsu AP. Efficient detection of thirty-seven new *IL2RG* mutations in human X-linked severe combined immunodeficiency. *Clin Immunol* 2000;95:33–38.
- Butte MJ, Haines C, Bonilla FA, Puck JM. IL-7 receptor deficient SCID with a unique intronic mutation and post-transplant autoimmunity. *Clin Immunol* 2007;125:159–164.
- Li L, Moshous D, Zhou Y, et al. A founder mutation in Artemis, an SNM1-like protein, causes SCID in Athabascan-speaking Native Americans. *J Immunol* 2002;168:6323–6329.
- Cutler DJ, Zwick ME, Carrasquillo MM, et al. High throughput variation detection and genotyping using microarrays. *Genome Res* 2001;11:1913–1925.
- Warrington JA, Shah NA, Chen X, et al. New developments in high throughput resequencing and variation detection using high density microarrays. *Hum Mutation* 2002;19:402–409.
- Xu N, Podolsky RH, Chudgar P, et al. Screening candidate genes for mutations in patients with hypogonadotropic hypogonadism using custom genome resequencing microarrays. *Am J Obstet Gynecol* 2005;192:1274–1284.
- Denning L, Anderson JA, Davis R, Gregg JP, Kuzdenyi J, Maselli RA. High throughput genetic analysis of congenital myasthenic syndromes using resequencing microarrays. *PLoS ONE* 2007;2:e918.
- Cheng E, Cutting GR, Gibson R, et al. CFTR-related disorders. GeneReviews online publication of genetic disease information. Available at: <http://www.genetests.org/servelet/access?id=8888891&key=m7KRYmZ4YULO&fnc=y&fw=ke-I&filename=/about/content/reviews.html>. Accessed January, 2008.
- Moskowitz SM, Gibson RL, Stern DL, et al. CFTR-related disorders (Updated 2005). In: GeneReviews at GeneTests: medical genetics information resource (database online). Copyright, University of Washington, Seattle, 1997–2008. Available at: <http://www.genetests.org>.
- Stenson PD, Ball EV, Mort M, et al. Human gene mutation database (HGMD): 2003 update. *Hum Mutat* 2003;21:577–581. Available at: <http://www.hgmd.cf.ac.uk/ac/index.php>. Accessed January, 2008.
- Mandal MN, Heckenlively JR, Burch T, et al. Sequencing arrays for screening multiple genes associated with early-onset human retinal degenerations on a high-throughput platform. *Invest Ophthalmol Vis Sci* 2005;46:3344–3363.
- Tengs T, Lee JC, Guillermo Paez J, et al. A transforming MET mutation discovered in non-small cell lung cancer using microarray-based resequencing. *Cancer Lett* 2005;229:227–233.
- Liu C, Aronow BJ, Jegga AG, et al. Novel resequencing chip customized to diagnose mutations in patients with inherited syndromes of intrahepatic cholestasis. *Gastroenterol* 2007;132:119–126.
- Hacia JG. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet* 1999;21(suppl):42–47.
- Cronin MT, Fucini RV, Kim SM, Masino RS, Wespi RM, Miyada CG. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum Mutat* 1996;7:244–255.