Sultan Qaboos University Med J, February 2013, Vol. 13, Iss. 1, pp. 69–79, Epub. 27th Feb 13
Submitted 16th May 12
Revision Req. 7th Aug 12, Revision Recd. 6th Sep 12
Accepted 6th Oct 12

Array-based Identification of Copy Number Changes in a Diagnostic Setting

Simultaneous gene-focused and low resolution whole human genome analysis

Renate Marquis-Nicholson,[1] Elaine Doherty,[1] Jennifer M. Love,[1] Chuan-Ching Lan,[1] Alice M. George,[1] Anthony Thrush[2,] *Donald R. Love,[1,3]

استعراف مبني على منظومة تغييرات ارقام النسخ لضبط التشخيص

تحليل كامل للجينوم باستخدام طريقتي التركيز الجيني والتميز المنخفض في آن واحد

رينات مرقص نيكلسون، ايلين دوروتي، جنيفر م لوف، شون شينج لان، اليس م جورج، دونالد ر لوف،
انتوني ثراش

الملخص: الهدف: هدفت الدراسة إلى تطوير والتحقق من صحة منظومة تهجين جيني مقارن تسمح بتحليل كامل الجينوم بطريقتي الأستهداف لعدد من جينات المرض والتميز المنخفض في آن واحد. **طرق الدراسة:** تم تصميم منظومة بيسبوك روش نيمبلجين 12x135 ك للتهجين الجيني المقارن (شركة روش نيمبلجين، ماديسون، وينكونسينو الولايات المتحدة الأمريكية) لتمحيص المناطق الشفرية في 66 جين مع استخدام مسبارات إضافية متسعة القاعدة من أجل التغطية الكلية لكامل الجينوم. تم تحليل الحامض النووي ثنائي الريبوز (د ن ا) من عشرين مريضاً ذوي تغيرات أرقام نسخ سابقة التعريف و ثمانية مرضى لم يتم تحديد الجرعة الجينية لهم من قبل. **النتائج:** استطاعت منظومة روش نيمبلجين تهجين جيني مقارن أن تحدد تغيرات ارقام النسخ في كل المرضى العشرين. كما تم ايضا بنجاح التشخيص الجزيئي لأحد المرضى الأربعة الذين لم يتم تأكيد تشخيصهم السريري بواسطة تحليل المتوالية من قبل. تم أيضا بنجاح تحديد حمل صفة تغيرات أرقام النسخ للمرضى الأربعة الباقين. **الخلاصة:** منظومة روش نيمبلجين تهجين جيني مقارن الموصوفة في هذا البحث هي طريقة دقيقة، صامدة، وعالية المردود و مثالية للاستخدام في المختبرات التشخيصية الصغيرة. من الممكن استخدام هذة الطريقة كبديل للطرق الأخرى الشائعة الاستخدام مثل طريقة الربط المتعدد المعتمد علي تضخيم المسبار.

مفتاح الكلمات: منظومة تهجين جيني مقارن، جرعة جينية، متغيرات أرقام النسخ، ميكروواراي ال د ن ا، التشخيص الجزيئي.

ABSTRACT: ***Objectives:*** The aim of this study was to develop and validate a comparative genomic hybridisation (CGH) array that would allow simultaneous targeted analysis of a panel of disease genes and low resolution whole genome analysis. ***Methods:*** A bespoke Roche NimbleGen 12x135K CGH array (Roche NimbleGen Inc., Madison, Wisconsin, USA) was designed to interrogate the coding regions of 66 genes of interest, with additional widely-spaced backbone probes providing coverage across the whole genome. We analysed genomic deoxyribonucleic acid (DNA) from 20 patients with a range of previously characterised copy number changes and from 8 patients who had not previously undergone any form of dosage analysis. ***Results:*** The custom-designed Roche NimbleGen CGH array was able to detect known copy number changes in all 20 patients. A molecular diagnosis was also made for one of the additional 4 patients with a clinical diagnosis that had not been confirmed by sequence analysis, and carrier testing for familial copy number variants was successfully completed for the remaining four patients. ***Conclusion:*** The custom-designed CGH array described here is ideally suited for use in a small diagnostic laboratory. The method is robust, accurate, and cost-effective, and offers an ideal alternative to more conventional targeted assays such as multiplex ligation-dependent probe amplification.

Keywords: Array comparative genomic hybridization (aCGH); Gene dosage; Copy number variants (CNVs); DNA microarray; Molecular diagnosis.

Advances in knowledge

- *Customised comparative genomic hybridisation (CGH) arrays, such as the one described here, allow robust high density gene-targeted as well as low density whole genome analysis to be undertaken simultaneously in the diagnostic setting.*
- *Our data has shown that complicated gene rearrangements may underlie disease and that these rearrangements may be missed by more conventional diagnostic techniques.*

[1]Diagnostic Genetics, LabPLUS, Auckland City Hospital, Auckland, New Zealand; [2]Roche Diagnostics New Zealand Ltd., Auckland, New Zealand; [3]School of Biological Sciences, University of Auckland, Auckland, New Zealand

**Corresponding Author e-mail: donaldl@adhb.govt.nz*

Applications to Patient Care

- *The targeted CGH array with backbone format allows for diagnostic flexibility in a clinical laboratory setting.*
- *The added advantage of the approach described here is that it removes the need to batch the mutation screening of patients based on their clinical phenotype.*

The importance of gene deletion and duplication in the pathogenesis of disease has become increasingly evident over the last decade. These deletions/duplications range from intragenic changes that are too large to be detected by sequence analysis, to larger genomic rearrangements responsible for the microdeletion and microduplication syndromes, and finally to whole chromosome loss or gain as seen in the aneuploidies.

In the discipline of cytogenetics, molecular karyotyping using high-density oligonucleotide arrays has recently become the recommended first-line diagnostic test for patients with developmental delay/intellectual disability, autistic spectrum disorder, or multiple congenital anomalies, replacing more conventional techniques such as G-banded karyotyping.[1,2] Large deletions and duplications have long been recognised as playing an important part in the pathogenesis of several disorders traditionally diagnosed using molecular techniques, such as Duchenne muscular dystrophy and Charcot-Marie-Tooth disease type 1A.[3,4] In addition to these classical deletion/duplication disorders, the role of partial or whole gene deletions in the aetiology of a wide variety of single-gene disorders is becoming more apparent. A 2008 review of the entries in the online Human Gene Mutation Database showed that large deletions and duplications comprise 10% of the listed mutations, compared to 6% in 2003.[5,6] This number is likely to increase further as more individuals are subjected to dosage analysis as part of routine molecular diagnostics.

A variety of dosage analysis methods are available to the diagnostic laboratory, including multiplex ligation-dependent probe amplification (MLPA), quantitative real-time polymerase chain reaction (qPCR), and customised fluorescence *in situ* hybridisation (FISH).[7–9] Each of these methods, however, is relatively expensive, principally as a result of the price of the probes, and in the case of MLPA and qPCR, is usually confined to a limited number of exons across a limited number of genes.[10,11] Finally, in the case of a small diagnostic laboratory, low sample throughput decreases cost-effectiveness, together with the attendant issue of maintaining staff proficiency in a range of dosage techniques.

In order to address the above difficulties, we designed a bespoke NimbleGen 12x135K comparative genomic hybridisation (CGH) array (Roche NimbleGen Inc., Madison, Wisconsin, USA). This array targets a panel of genes chosen to complement the sequencing assays offered in-house, as well as a number of other genes for which deletions and duplications are known to be implicated in a disease phenotype. In addition to this gene-focused coverage, the design of the array also involved low-density coverage of the entire human genome. Here, we report the use of this custom-designed array to analyse a series of 28 clinical samples in order to investigate the suitability of this approach for dosage analysis in the diagnostic environment.

Methods

A group of 20 individuals with a range of previously characterised copy number changes were selected for array comparative genomic hybridisation (aCGH) analysis. The patients, or parents in the case of neonates, provided informed consent for diagnostic testing; the New Zealand multi-region ethics committee has ruled that cases of patient management do not require formal ethics committee approval. The copy number changes included both cytogenetic and molecular abnormalities, and spanned a spectrum from aneuploidy to intragenic deletion with three cases of aneuploidy, two of unbalanced translocations, three microdeletions, two microduplications, seven intragenic deletions, and three intragenic duplications. These changes had been identified using a range of techniques, including conventional and molecular karyotyping, FISH and MLPA [Table 1]. aCGH was also completed for an additional 8 individuals without known copy number changes for whom dosage analysis was desirable either for

Table 1: Copy number changes used to validate the Roche NimbleGen custom-designed comparative genomic hybridisation array

Patient	Disorder/Copy Number Variants	Description	Previous testing method
1	Klinefelter syndrome	XXY	Karyotype
2	Down's syndrome	Trisomy 21	Karyotype
3	Edward's syndrome	Trisomy 18	Karyotype
4	Unbalanced translocation t (7;22)	46,XX,der(7)t(7;22)(q36.3;q13.1) mat.ish der(7)t(7;22)(q36.3;q13.1) (ARSA+)	Karyotype, FISH
5	Unbalanced translocation t (3;4)	46,XX,der(4)t(3;4)(q23;q35.1)pat	Karyotype
6	Microdeletion chr2	Del chr2:102176600-119933523	Illumina HumanCytoSNP 300K microarray
7	Autistic spectrum disorder	Del chr2: 44305631-44425668; dup chr15: 36188779-36655207; del chr16: 29522477-30107306	Affymetrix SNP 6.0 microarray
8	Prader-Willi syndrome	Deletion of *SNRPN* gene	Southern blot, FISH
9	Williams-Beuren duplication syndrome	Dup chr7: 71914639-73718403	Affymetrix SNP 6.0 microarray
10	Rett syndrome	Duplication *MECP2* gene	MLPA
11	LQTS	Del exons 6, 7,10,11,15 *KCNH2* gene	MLPA
12	LQTS	Dup exons 10,11, 15 *KCNH2* gene	MLPA
13	Familial adenomatous polyposis	Del exons 11-12 *APC* gene	MLPA
14	DMD	Del exon 45-52 *DMD* gene	MLPA and multiplex PCR
15	DMD, carrier	Dup exon 63 (heterozygous) *DMD* gene	MLPA
16	Familial breast cancer	Del exon 1, 2 *BRCA1* gene	MLPA
17	DMD	Del Ex3-44 *DMD* gene	MLPA
18	HNPCC	Del exon 6 *MLH1* gene	MLPA
19	Familial breast cancer	Dup exon13 *BRCA1* gene	MLPA
20	Familial breast cancer	Del exons 1,2 *BRCA2* gene	MLPA

FISH = fluorescence in situ hybridization; del chr = deleted chromosome; dup chr = duplicated chromosome; SNP = single nucleotide polymorphisms; LQTS = long QT syndrome; MLPA = multiplex ligation-dependent probe amplification; DMD = Duchenne muscular dystrophy; PCR = polymerase chain reaction; HNPCC = hereditary non-polyposis colorectal cancer.

diagnostic purposes or for completion of family studies.

Peripheral blood ethylenediaminetetraacetic acid (EDTA) samples from each of these 28 individuals were submitted to the Diagnostic Genetics Department at LabPLUS, Auckland City Hospital, New Zealand, for either molecular or cytogenetic analysis, as clinically indicated.

Genomic diribonucleic acid (gDNA) was extracted from peripheral blood leucocytes using the Gentra Puregene DNA Extraction Kit (QIAGEN, Germantown, Maryland, USA). In those samples referred for conventional karyotype or FISH analysis, classical phenol/chloroform extraction with ethanol precipitation was used to isolate DNA from cultured leucocytes, in order to provide a source of gDNA for molecular testing.

A primer design protocol was used to design primers flanking the region spanning exons 11–14 of the *KCNH2* gene.[12,13] In brief, the messenger RNA (mRNA) sequence of interest was identified using the University of California Santa Cruz (UCSC) genome browser.[14] All primers were checked for single nucleotide polymorphisms using the software tool available from the National Genetic Reference Laboratory, Manchester, UK.[15] The primers were tailed with M13 sequences and were synthesised by Invitrogen Ltd., Renfrewshire, UK (primer sequences are available on request).

Table 2: Human disease genes selected for inclusion on the Roche NimbleGen custom-designed comparative genomic hybridisation array

Disorder	Gene	Accession number (Transcript)	Accession number (Protein)	Uniprot number	OMIM
LQT	*KCNQ1*	NM_000218.2	NP_000209	P51787	607542
	KCNH2	NM_000238.2	NP_000229	Q12809	152427
	SCN5A	NM_198056.2	NP_932173	Q14524	600163
	GPD1L	NM_015141.2	NP_055956	Q8N335	611778
	SCN1B	NM_001037.4	NP_001028	Q07699	611778
		NM_199037.3	NP_950238	Q6TN97	600235
	SCN3B	NM_018400.3	NP_060870	Q9NY72	608214
	CACNB2	NM_201596.2	NP_963890	Q08289	600003
	KCNE3	NM_005472.4	NP_005463	Q9Y6H6	604433
	ANK2	NM_001148.3	NP_001139	Q01484	106410
	KCNE1	NM_000219.3	NP_000210	P15382	176261
	KCNE2	NM_172201.1	NP_751951	Q9Y6J6	603796
	KCNJ2	NM_000891.2	NP_000882	P63252	600681
	CACNA1c	NM_001129827.1	NP_001123299	Q13936	114205
	CAV3	NM_033337.1	NP_203123	P56539	601253
	SCN4B	NM_174934.3	NP_777594	Q8IWT1	608256
	AKAP9	NM_005751.4	NP_005742	Q8IWT1	604001
HCM	*MYH7*	NM_000257.2	NP_000248	P12883	160760
	MYBPC3	NM_000256.3	NP_000247	Q14896	600958
	TNNT2	NM_000364.2	NP_000355	P45379	191045
	TNNI3	NM_000363.4	NP_000354	P19429	191044
	TPM1	NM_001018020.1	NP_001018020	O15513	191010
	ACTC1	NM_005159.4	NP_005150	P68032	102540
	MYL2	NM_000432.3	NP_000423	P10916	160781
	MYL3	NM_000258.2	NP_000249	P08590	160790
	LAMP2	NM_001122606.1	NP_001116078	Q6Q3G8	309060
	PRKAG2	NM_016203.3	NP_057287	Q9UGJ0	602743
	GLA	NM_000169.2	NP_000160	P06280	301500
CPVT	*RYR2*	NM_001035.2	NP_001026	Q92736	180902
	CASQ2	NM_001232.2	NP_001223	O14958	114251
ARVC	*DSP*	NM_004415.2	NP_004406	P15924	125647
	PKP2	NM_001005242.2	NP_001005242	A0AV37	602861
	DSG2	NM_001943.3	NP_001934	Q14126	125671
	DSC2	NM_024422.3	NP_077740	Q02487	125645
	JUP	NM_002230.2	NP_002221	P14923	173325
	TGFB3	NM_003239.2	NP_003230	P10600	190230
	TMEM43	NM_024334.2	NP_077310	Q9BTV4	612048

DMD	*DMD*	NM_004006.2	NP_003997	P11532	300377
ALD	*ABCD1*	NM_000033.2	NP_000024.2	P33897	300371
FAP	*APC*	NM_000038.3	NP_000029.2	P25054	611731
Type 1 citrullinaemia	*ASS1*	NM_000050.4	NP_000041.2	P00966	603470
Type II citrullinaemia	*SLC25A13*	NM_014251.2	NP_001153682.1	Q9UJS0	603859
Thyroid carcinoma/melanoma	*BRAF1*	NM_004333.4	NP_004324.2	P15056	164757
Familial breast and ovarian cancer	*BRCA1*	NM_007294.2	NP_009225.1	P38398	113705
	BRCA2	NM_000059.3	NP_000050.2	P51587	600185
X-linked congential stationary night blindness type 2	*CACNA1F*	NM_005183.2	NP_005174.2	O60840	300110
E-cadherin related stomach cancer	*CDH1*	NM_004360.2	NP_004351.1	P12830	192090
Larsen syndrome	*FLNB*	NM_001457.2	NP_001157789.1	O75369	603381
NKH	*GLDC*	NM_000170.2	NP_000161.2	P23378	238300
Holocarboxylase synthetase deficiency	*HLCS*	NM_000411.4	NP_000402.3	P50747	609018
MODY	*GCK*	NM_000162.3	NP_000153.1	P35557	138079
	HNF1a	NM_000545.4	NP_000536.5	P20823	142410
	HNF1b	NM_000458.2	NP_000449.1	P35680	189907
	HNF4a	NM_000457.3	NP_000448.3	P41235	600281
Familial hypercholesterolemia	*LDLR*	NM_000527.3	NP_000518.1	P01130	606945
Rett syndrome	*MECP2*	NM_004992.3	NP_001104262.1	P51608	300005
HNPCC	*MLH1*	NM_000249.2	NP_000240.1	P40692	120436
	MSH2	NM_000251.1	NP_000242.1	P43246	609309
	PMS1	NM_000534.4	NP_000525.1	P54277	600258
	PMS2	NM_000535.5	NP_000526.1	P54278	600259
MEN2A	*RET*	NM_020630.4	NP_065681.1	P07949	64761
Familial phaeochromocytoma/ paraganglioma	*SDHAF2*	NM_017841.1	NP_060311.1	Q9NX18	613019
	SDHB	NM_003000.2	NP_002991.2	P21912	185470
	SDHC	NM_003001.3	NP_001030588.1	Q99643	602413
	SDHD	NM_003002.1	NP_002993.1	O14521	602690
DYT11	*SGCE*	NM_003919.2	NP_001092870.1	O43556	604149
VHL	*VHL*	NM_000551.2	NP_000542.1	P40337	608537

LQT = long QT syndrome; HCM = hypertrophic cardiomyopathy; CPVT = catecholaminergic polymorphic ventricular tachycardia; ARVC = arrhythmogenic right ventricular cardiomyopathy; DMD = Duchenne muscular dystrophy; ALD = adrenoleukodystrophy; FAP = familial adenomatous polyposis; NKH = nonketotic hyperglycinemia; MODY = maturity onset diabetes of the young: HNPCC = hereditary non-polyposis colorectal cancer; MEN2A = multiple endocrine neoplasia type 2A; DYT11 = Myoclonus dystonia; VHL = Von-Hippel Lindau syndrome.

Polymerase chain reaction (PCR) amplification was performed in a total volume of 25 µL, containing 50 ng of genomic deoxyribonucleic acid (DNA), 0.20 µM of each primer, 1 mM of each dNTP, and 1.75 U of expand long template enzyme mix in buffer 2 (F. Hoffmann-La Roche Ltd., Basel, Switzerland). After an initial denaturation for 2 minutes at 94° C, the PCR amplification included 10 cycles of 94° C for 10 seconds, 60° C for 30 seconds, and 68° C for 2 minutes, followed by 20 cycles of 94° C for 15 seconds, 60° C for 30 seconds, 68° C for 4 minutes, and a final extension at 68° C for 10 minutes. PCR products were separated by a 2% agarose gel and the lower band, corresponding to the allele carrying the

deletion, was excised and purified using the Roche High Pure PCR Cleanup Micro Kit (Roche Applied Sciences, Roche Diagnostics, Penzberg, Germany).

Bidirectional DNA sequencing was performed using M13 forward and reverse primers and Big-Dye Terminator, Version 3.0 (Applied Biosystems Ltd., Carlsbad, California, USA). Using an automated Clean-Seq procedure (Agencourt Bioscience Corp., Beverly, Massachusetts, USA), 20 µL of sequenced product was purified with the aid of an epMOTION 5075 liquid handling robot (Eppendorf, Hamburg, Germany). Using the Applied Biosystems model 3130xl genetic analyser (Applied Biosystems, Inc., Foster City, California USA), 15 µL of purified product was then subjected to capillary electrophoresis.

Genes of interest, including those already sequenced in-house and those pertaining to common disorders known to frequently involve deletions/duplications (such as Duchenne muscular dystophy), were selected and the appropriate NM accession numbers identified using the UCSC genome browser. The final gene list comprising 66 genes was forwarded to NimbleGen and formed the basis of their design for a 12-plex 135K oligonucleotide array (see Table 2 for gene list). Each probe was 60–85 bp in length and possessed similar isothermal characteristics. Exonic probes were designed to overlap by 25 bp in order to provide high resolution detection of deletions or duplications within the coding regions of the genes of interest. Intronic probes were spaced on average every 175 bp. To minimise the occurrence of false positive results due to a one-off failure of hybridisation to a particular probe, each gene-focused probe was spotted in duplicate. In addition to the targeted probes tiled over the genes of interest, approximately 75,000 'backbone' probes were also included. These probes were spaced across the entire genome (with a mean probe interval of 46 kbp) to provide low-density whole genome interrogation, as well as increase the accuracy of data normalisation during the analysis procedure. Following completion of the design process, the array was manufactured by NimbleGen, Inc.

A total of 250 nanograms of genomic deoxyribonucleic acid (gDNA) were processed according to the NimbleGen Array User's Guide: CGH and CNV Arrays, Version 6.0. In brief, extracted gDNA from samples and Promega controls was denatured in the presence of a Cy3- for the test group or Cy5- for the control group, labelled random primers and incubated with the Klenow fragment of DNA polymerase, together with deoxyribonucleotide triphosphates (dNTPs) (5 mM of each dNTP), at 37° C for 2 hours. The reaction was terminated by the addition of 0.5 M EDTA (21.5 µL), prior to isopropanol precipitation and ethanol washing. Following quantification, the test and sex-matched control samples were combined in equimolar amounts and applied to one of the twelve arrays on the microarray slide. Hybridisation was carried out in a NimbleGen Hybridisation Chamber for a period of 48 hours. Slides were washed and scanned using a NimbleGen MS 200 microarray scanner. Array image files (.tif) produced by the MS 200 Data Collection Software were imported into NimbleScan Version 2.6 for analysis. Each genomic region exhibiting a copy number change within one of the genes of interest was examined using the UCSC genome browser to determine the location and significance of the change. Data was filtered using the default $\log^2$ ratio thresholds recommended in the NimbleGen Array User's Guide of less than -0.2 for a deletion and greater than 0.2 for duplication.

For MLPA, the SALSA MLPA P114 LQT kit (lot 0805) was purchased from MRC-Holland (Amsterdam, Netherlands). This mix contains probes for 17 exons of the *KCNQ1* gene, 9 probes for the *KCNH2* gene, 4 probes for the *SCN5A* gene, as well as 4 and 3 probes for *KCNE1* and *KCNE2*, respectively. This kit also contains four control probes mapping to other autosomes. MLPA analysis was carried out according to the MRC Holland protocol. Briefly, 125 ng of genomic DNA from each sample was diluted in 5 µl TE buffer and denatured at 98° C for 5 minutes. MLPA buffer and probe mix (1.5 µl of each) were then added to allow the probes to anneal to their target sequences by heating at 95° C for one minute and incubating for 16 hours at 60° C. A buffer/ligase mixture (32 µl) was added to each sample and incubated at 54° C for 15 minutes followed by heating to 98° C for 5 minutes. Ten microlitres of the ligation reaction were used for multiplex PCR amplification using a single universal primer pair suitable for all the probes in the kit. The SALSA polymerase was added at 60° C, followed by 36 cycles of 95° C for 30 seconds, 60° C for 30 seconds, 72° C for one minute, and a final extension

Table 3: Custom- designed CGH array results for all samples

Patient	Previous result	Custom array raw result	Significance of result
1	XXY	arr Xp22.33q28(6,329-154,894,377)x3	XXY
2	Trisomy 21	arr 21q11.2q22.3(9,931,865-46,914,745)x3	Trisomy 21
3	Trisomy 18	arr 18p11.32q23(102,328-76,093,443)x3	Trisomy 18
4	46,XX,der(7)t(7;22)(q36.3;q13.1)mat.ish der(7)t(7;22)(q36.3;q13.1)(ARSA+)	arr 7q36.3(156,973,768-158,816,034)x1,22 q13.1q13.33(37,139,349-49,522,598)x3	t(7;22), coordinates consistent with previous result
5	46,XX,der(4)t(3;4)(q23;q35.1)pat	arr 4q34.3q35.2(182,454,628-191,220,565) x1,3q23q29(144,114,087-199,377,478)x3	t(3;4), coordinates consistent with previous result
6	Del chr2:102176600-119933523	arr 2q12.1q14.2(102,195,252-119,812,387) x1	Del chr2, coordinates consistent with previous result
7	Del chr2: 44305631-44425668; dup chr15: 36188779-36655207; del chr16: 29522477-30107306	arr 2p21(44,325,958-44,373,442) x1,15q14(36,244,896-36,615,176)x3 16p11.2(29,653,824-30,100,122)x1	Multiple CNVs, coordinates consistent with previous result
8	Deletion of *SNRPN* gene	arr 15q11.2q13.1(21,450,428-26,192,737) x1	Del entire *SNRPN* gene
9	Dup chr7: 71914639-73718403	arr 7q11.23(71,964,201-73,874,826)x3	Dup chr7, coordinates consistent with previous result
10	Dup *MECP2* gene	arr Xq28(152,900,329-153,202,330)x3	Dup entire *MECP2* gene
11	Del exons 6, 7,10,11,15 *KCNH2* gene	arr 7q36.1(150,250,593-150,283,627)x1	Del exons 6-15 (inclusive)
12	Dup exons 10,11, 15 *KCNH2* gene	arr 7q36.1(150,250,593-150,275,172x3,150,275,345-150,276,020x1,150,276,456-150,279,665x3)	Dup exons 7,8,9,10,11; del exons 12,13; dup exons 14,15
13	Del exons 11-12 *APC* gene	arr 5q22.2(112,190,700-112,191,901)x1	Del exons 11,12
14	Del exon 45-52 *DMD* gene	arr Xp21.1(31,625,116-31,904,144)x0	Del exons 45-52 (inclusive)
15	Dup exon 63 (heterozygous) *DMD* gene	arr Xp21.2(31,155,081-31,194,353)x3	Dup exon 63 (heterozygous)
16	Del exon 1, 2 *BRCA1* gene	arr 17q21.31(38,525,107-38,531,019)x1	Del exons 1,2
17	Del Ex3-44 *DMD* gene	arr Xp21.2p21.1(31,048,707-32,916,496)x0	Del exons 3-44 (inclusive)
18	Del exon 6 *MLH1* gene	arr 3p22.2(37,025,008-37,027,636)x1	Del exon 6
19	Dup exon13 *BRCA1* gene	arr 17q21.31(38,484,216-38,488,483)x3	Dup exon 13
20	Del exons 1,2 *BRCA2* gene	arr 13q13.1(31,787,734-31,788,803)x1	Del exons 1,2
Individuals with no known copy number change			
	Referral reason	**Custom array raw result**	**Significance of result**
21	Mother of patient 9	No CNV detected	*De novo* dup chr7 in patient 9
22	Father of patient 9	No CNV detected	*De novo* dup chr7 in patient 9
23	Mother of patient 7	arr 15q14(36,188,779-36,655,207)x3	Carrier of chr15 dup; *de novo* deletion chr16 in patient 7
24	Father of patient 7	arr 2p21(44,325,958-44,373,442)x1	Carrier of chr2 del; *de novo* deletion chr16 in patient 7
25	LQTS	No CNV detected	Pathogenic mutation not detected
26	LQTS	No CNV detected	Pathogenic mutation not detected
27	MODY	No CNV detected	Pathogenic mutation not detected
28	HNPCC	arr 2p21(47,486,274-47,559,311)x1	Del exons 2-14 *MSH2* gene

del chr = deleted chromosome; dup chr = duplicated chromosome; LQTS = long QT syndrome; MODY = maturity onset diabetes of the young; HNPCC = hereditary non-polyposis colorectal cancer.

step of 72° C for 20 minutes. One microlitre of each PCR product was mixed with 0.5 µl GeneScan 600 Liz size standard (Applied Biosystems, Ltd.) and 8.5 µl of deionized formamide and 1µl was injected into a 36 cm capillary (Applied Biosystems model 3130XL)) at 60° C. The electropherogram was analysed using GeneMapper software (Applied Biosystems Ltd.). For each sample, the relative peak area (RPA) was calculated and compared to 5 healthy controls using custom-designed software. The software calculates RPAs for each probe within the same test and compares each RPA to those obtained from the 5 controls.

Results

We developed a custom-designed NimbleGen 12x135K aCGH that combines targeted high-density coverage of 66 genes of interest with genome-wide coverage to produce a low-resolution molecular karyotype. For the validation of this array we analysed 20 patients with known copy number abnormalities. The custom designed NimbleGen CGH array was able to accurately identify these copy number changes in all 20 patients [Table 3].

The array results for patient 12 revealed an additional alteration that had not been recognised previously. Patient 12 is a member of a large pedigree with multiple members suffering from long QT syndrome (LQTS). Analysis using the MRC-Holland SALSA P114 LQT MLPA kit, which interrogates a limited number of exons of the *KCNH2* gene (exons 1-4,6,7,10,11,15), had identified a duplication of exons 10, 11, and 15 in all affected individuals [Figure 1, panel A].[16] This duplication had therefore been the focus of predictive testing using MLPA for additional at-risk members of the family. The aCGH results clarified the extent of the duplication, not only showing that it involved a breakpoint within exon 7 and encompassed the whole of exons 8, 9, 10, 11, 14 and 15, but also that the genotype was more complex than previously thought. A critical micro-deletion encompassing exons 12 and 13 was

Figure 1: Graphic representation of copy number changes in the *KCNH2* gene in patient 12. **(A)** Dosage changes were detected using a multiplex ligation-dependent probe amplification (MLPA) approach. The graphic representation shows the increased dosage detected by probes that lie in exons 10, 11, and 15 of the *KCNH2* gene. **(B)** Dosage changes were detected in the *KCNH2* gene with a copy number gain (X3 copy number) defined by the chromosome 7 coordinates (NCBI36/hg18 assembly) 150,276,456-150,279,665bp (within exon 7 to within exon 11, log2 ratio 0.45) and 150,250,593-150,275,172 (encompassing exons 14 and 15, $\log^2$ratio 0.5), and an apparent 676bp deletion (X1 copy number, $\log^2$ratio -0.53) located at 150,275,345-150,276,020bp (encompassing exons 12 and 13). **(C)** Transcripts expressed from the *KCNH2* gene are shown, together with the distal exons of transcript 1 of the *KCNH2* gene (RefSeq accession number NM_000238.3).

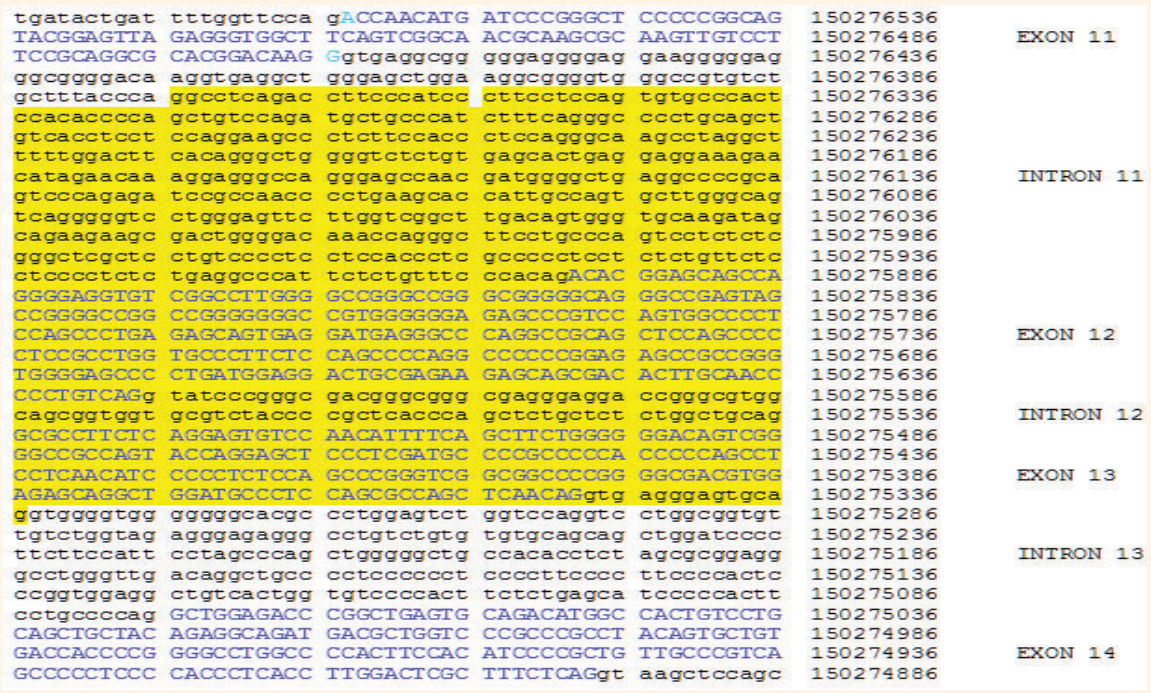

Figure 2: Location and extent of the *KCNH2* gene deletion in patient 12. A partial sequence of the *KCNH2* gene is shown that encompasses exons 11 to 13, inclusive (in blue). The sequence-confirmed location and extent of the 1041bp deletion detected in the genome of patient 12 is highlighted in yellow (chromosome 7: 150,276,375-150,275,335bp; NCBI36/hg18 assembly).

detected [Figure 1, panels B and C]. PCR and DNA sequencing determined the exact breakpoints of the 1041 bp deletion, the length of which compares favourably to the 676 bp copy number change detected by the array [Figure 2].

Of the 8 patients who had not yet undergone any form of copy number analysis, 4 had a clinical diagnosis that had not been confirmed by sequence analysis of the implicated genes: two had a diagnosis of long QT syndrome, one of hereditary non-polyposis colorectal cancer (HNPCC), and one of maturity onset diabetes of the young (MODY). No copy number changes were identified in the panel of long QT syndrome genes in either of the long QT patients, nor within the MODY genes in the MODY patient. However, a large deletion involving exons 2–14 inclusive of the *MSH2* gene was detected in the individual with a clinical diagnosis of hereditary non-polyposis colorectal cancer (HNPCC). Mutations in the mismatch repair gene *MSH2* are known to be responsible for 40% of cases of HNPCC; 20% of these mutations involve exonic or full gene deletions.[17]

The referral reason for aCGH analysis for the remaining 4 individuals without a known copy number change was to provide additional information for genetic counselling and family planning. Individuals 21 and 22 are the parents of patient 9, an eight-year-old girl with mild dysmorphic features and speech delay, who had been found to have a duplication involving the Williams-Beuren syndrome (WBS) critical region at 7q11[23] using an Affymetrix single nucleotide polymorphisms (SNP) 6.0 array (Affymetrix, Santa Clara, California, USA). While a microdeletion of the WBS critical region results in a well-characterised pattern of facial dysmorphism, supravalvular aortic stenosis, connective tissue abnormalities, hypercalcaemia, and a recognisable behavioural phenotype, duplication of the same region results in a much less distinctive set of characteristics.[18] Foremost among these, as was seen in our patient, are mildly dysmorphic facial features and prominent speech delay. Parental transmission of the 7q11.[23] duplication is relatively frequent in the WBS duplication syndrome, but reduced penetrance and variable expression mean that determination of carrier status based on phenotype alone is not simple. An approximately 1.5 Mb duplication of the WBS critical region was readily indentified in the affected girl by our custom-designed NimbleGen CGH array, which agreed with the earlier Affymetrix SNP 6.0 array data, but was not detected in either of her parents.

The conclusion is that the genomic copy number change detected in patient 9 is a *de novo* event and that future pregnancies are not at high risk of this mutation event.

Individuals 23 and 24 are the parents of patient 7, a six-year-old boy who was referred for investigation of developmental delay and features consistent with autistic spectrum disorder. High-density Affymetrix SNP 6.0 microarray analysis had revealed several copy number changes in the child, including a deletion at chromosome 2p21, a duplication at chromosome 15q14, and a deletion at chromosome 16p11.2 (see Table 3 for full coordinates). Each of these changes was also identified by our NimbleGen custom CGH array, with only minor differences in breakpoint location, despite the difference in probe density [Table 3]. The 16p11.2 deletion is consistent with the phenotypic features in this case, as dosage changes at 16p11.2 have been described in association with autistic spectrum disorder.[19] The aCGH results confirmed that the chromosome 16p11.2 deletion is *de novo* and that each of the other two copy number changes are most likely to be benign, as each is inherited from one of his parents.

Discussion

The purpose of the work described above was to design and validate a CGH array that could be used as an alternative to MLPA, quantitative PCR, and customised FISH in the diagnostic genetics laboratory. Although there have been several reports in the recent literature of custom-designed CGH arrays being used to screen for either exonic dosage changes in a large set of disease-specific genes, or for one of a panel of known genomic disorders, this is the first report, to our knowledge, of a custom-designed CGH array that provides both high-resolution coverage of a comprehensive set of genes and low-resolution whole genome coverage.[20-25]

The array design we report here is ideally suited to a small diagnostic laboratory. It enables the simultaneous interrogation of a large number of genes using a process that eliminates the risk of false negatives inherent in PCR-based techniques due to the possibility of polymorphisms lying under primer binding sites. Twelve patient samples are able to be tested at once, reducing the overall cost of the assay. The overlapping probes tile the exons at a high density and allow changes involving the coding regions of the gene(s) of interest, including single exon changes, to be readily and reliably detected. This design feature is in contrast to some previously reported designs which could not reliably detect single exon changes due to insufficient probe coverage over affected regions.[23] The intron probes enable clarification of breakpoints, which is not possible with MLPA or qPCR, and the backbone probes facilitate the identification of larger genomic rearrangements, either as confirmation following high-density molecular karyotyping, or for carrier testing and family studies.

Conclusion

We have shown that our custom-designed NimbleGen CGH array can be used to accurately identify exonic deletions and duplications in a gene set of interest as well as offer a low resolution whole genome screen for larger genomic rearrangements. The technique is robust and cost-effective and allows for comprehensive analysis. This approach overcomes the problems associated with the use of expensive kits in the context of low sample throughput, and allows for consolidation of dosage analysis assays to a single validated technique.

References

1. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 2010; 86:749–64.
2. Manning M, Hudgins L. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. Genet Med 2010; 12:742–5.
3. Kunkel LM, Hejtmancik JF, Caskey CT, Speer A, Monaco AP, Middlesworth W, et al. Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. Nature 1986; 322:73–7.
4. Roa BB, Garcia CA, Lupski JR. Charcot-Marie-Tooth disease type 1A: Molecular mechanisms of gene dosage and point mutation underlying a common inherited peripheral neuropathy. Int J Neurol 1991-1992; 25–26:97–107.
5. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, et al. The Human Gene Mutation

Database: 2008 update. Genome Med 2009; 1:13.

6. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, et al. Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat 2003; 21:577–81.
7. Eijk-Van Os PG, Schouten JP. Multiplex ligation-dependent probe amplification (MLPA) for the detection of copy number variation in genomic sequences. Methods Mol Biol 2011; 688:97–126.
8. Sieber OM, Lamlum H, Crabtree MD, Rowan AJ, Barclay E, Lipton L, et al. Whole gene APC deletions cause classical familial adenomatous polyposis, but not attenuated polyposis or "multiple" colorectal adenomas. Proc Natl Acad Sci USA 2002; 99:2954–8.
9. Bendavid C, Kleta R, Long R. FISH diagnosis of the common 57-kb deletion in CTNS causing cystinosis. Hum Genet 2004; 115:510–14.
10. Gouas L, Goumy C, Veronese L, Tchirkov A, Vago P. Gene dosage methods as diagnostic tools for the identification of chromosome abnormalities. Pathol Biol (Paris) 2008; 56:345–53.
11. Armour JAL, Barton DE, Cockbuen DJ, Taylor GR. The detection of large deletions or duplications in Genomic DNA. Hum Mutat 2002; 20:325–37.
12. Doherty E, Marquis-Nicholson R, Brookes C, Love JM, Prosser D, Love DR. From primer design to sequence analysis: A pipeline tool for use in a diagnostic genetics laboratory. In: Ivanov O. Applications and Experiences of Quality Control. Pittsburgh, PA: InTech Publishers, 2011. Pp. 257–72.
13. Lai D, Love DR. Automation of a primer design and evaluation pipeline for subsequent sequencing of the coding regions of all human Refseq genes. Bioinformation 2012; 8:365–8.
14. Genome Browser. University of California Santa Cruz (UCSC) From: http://genome.ucsc.edu Accessed: Apr 2012.
15. National Genetic Reference Laboratory, Manchester, UK, Software tool. From: http://ngrl.man.ac.uk/SNPCheck.html Accessed: Apr 2012.
16. Eddy C-A, MacCormick J, Crawford JR, Chung S-K, Crawford JR, Love DR, et al. Identification of large gene deletions and duplications in KCNQ1 and KCNH2 in patients with long QT syndrome. Heart Rhythm 2008; 5:1275–81.
17. Kohlmann W, Gruber SB. Lynch Syndrome. In: Pagon RA, Bird TD, Dolan CR, Stephens K. GeneReviews [Internet]. Seattle: University of Washington; 1993–2004.
18. Merla G, Brunetti-Pierri LM, Fusco C. Copy number variants at Williams–Beuren syndrome 7q11.23 region. Hum Genet 2010; 128:3–26.
19. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 2008; 358:667–75.
20. Landsverk ML, Wang J, Schmitt ES, Pursley AN, Wong LJ. Utilization of targeted array comparative genomic hybridization, MitoMet®, in prenatal diagnosis of metabolic disorders. Mol Genet Metab 2011; 103:148–52.
21. Wong LJ, Dimmock D, Geraghty MT, Quan R, Lichter-Konecki U, Wang J, et al. Utility of oligonucleotide array-based comparative genomic hybridization for detection of target gene deletions. Clin Chem 2008; 54:1141–8.
22. Saillour Y, Cossée M, Leturcq F, Vasson A, Beugnet C, Poirier K, et al. Detection of exonic copy-number changes using a highly efficient oligonucleotide-based comparative genomic hybridization-array method. Hum Mutat 2008; 29:1083–90.
23. del Gaudio D, Yang Y, Boggs BA, Schmitt ES, Lee JA, Sahoo T, et al. Molecular diagnosis of Duchenne/Becker muscular dystrophy: Enhanced detection of dystrophin gene rearrangements by oligonucleotide array-comparative genomic hybridization. Hum Mutat 2008; 29:1100–7.
24. Piluso G, Dionisi M, Del Vecchio Blanco F, Torella A, Aurino S, Savarese M, et al. Motor chip: A comparative genomic hybridization microarray for copy-number mutations in 245 neuromuscular disorders. Clin Chem 2011; 57:1584–96.
25. Cheung SW, Shaw CA, Yu W, Li J, Ou Z, Patel A, et al. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. Genet Med 2005; 7:422–32.