Next-generation sequencing using a pre-designed gene panel for the molecular diagnosis of congenital disorders in pediatric patients

Eileen C. P. Lim¹, Maggie Brett¹, Angeline H. M. Lai²,³, Siew-Peng Lee¹, Ee-Shien Tan²,³, Saumya S. Jamuar²,³, Ivy S. L. Ng²,³ and Ene-Choo Tan¹,³*

Abstract

Background: Next-generation sequencing (NGS) has revolutionized genetic research and offers enormous potential for clinical application. Sequencing the exome has the advantage of casting the net wide for all known coding regions while targeted gene panel sequencing provides enhanced sequencing depths and can be designed to avoid incidental findings in adult-onset conditions. A HaloPlex panel consisting of 180 genes within commonly altered chromosomal regions is available for use on both the Ion Personal Genome Machine* (PGM™) and MiSeq platforms to screen for causative mutations in these genes.

Methods: We used this Haloplex ICCG panel for targeted sequencing of 15 patients with clinical presentations indicative of an abnormality in one of the 180 genes. Sequencing runs were done using the Ion 318 Chips on the Ion Torrent PGM. Variants were filtered for known polymorphisms and analysis was done to identify possible disease-causing variants before validation by Sanger sequencing. When possible, segregation of variants with phenotype in family members was performed to ascertain the pathogenicity of the variant.

Results: More than 97 % of the target bases were covered at >20x. There was an average of 96 novel variants per patient. Pathogenic mutations were identified in five genes for six patients, with two novel variants. There were another five likely pathogenic variants, some of which were unreported novel variants.

Conclusions: In a cohort of 15 patients, we were able to identify a likely genetic etiology in six patients (40 %). Another five patients had candidate variants for which further evaluation and segregation analysis are ongoing. Our results indicate that the HaloPlex ICCG panel is useful as a rapid, high-throughput and cost-effective screening tool for 170 of the 180 genes. There is low coverage for some regions in several genes which might have to be supplemented by Sanger sequencing. However, comparing the cost, ease of analysis, and shorter turnaround time, it is a good alternative to exome sequencing for patients whose features are suggestive of a genetic etiology involving one of the genes in the panel.

Keywords: Congenital disorders, Gene panel, ICCG, Mutation screening, Next-generation sequencing

* Correspondence: tanec@bigfoot.com
1KK Research Centre, KK Women’s and Children’s Hospital, 100 Bukit Timah Road, Singapore 229899, Singapore
2Paediatrics Academic Clinical Programme, SingHealth Duke-NUS Graduate Medical School, Singapore 169857, Singapore
Full list of author information is available at the end of the article

© 2015 Lim et al. Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background
Congenital disorders comprise conditions present at birth or those that developed during infancy or early childhood. Presentations include structural abnormalities, neuromuscular disorders, developmental delay, and intellectual disability which collectively affect more than 10 % of children. The European Surveillance of Congenital Anomalies (EUROCAT) reported the prevalence of major congenital anomalies to be about 2.4 % of live births [1], while the Center for Disease Control and Prevention (CDC) reported 3.3 % for birth defects [2]. The prevalence of developmental disabilities is reported to be 13.9 % in the USA [3].

Less than half of these disorders have an identifiable cause such as aneuploidy, metabolic disorder, maternal infection, parental exposure to teratogenic agents, or intrapartum events. The remaining cases are thought to have a genetic etiology such as submicroscopic chromosomal abnormalities or rare single/multiple nucleotide changes. The former can be detected by using chromosomal microarray analysis (CMA) which is now the recommended first-tier test for children with dysmorphism, multiple congenital anomalies, developmental delay/intellectual disability, and/or autism spectrum disorder [4]. Although CMA is more sensitive than conventional karyotyping, the diagnostic yield for this group of disorders is still only about 20 % in multiple studies [5–7]. Genetic causes for the rest are likely due to small deletions and insertions, balanced translocations involving gene disruptions, and point mutations which cannot be detected by commonly used CMA platforms.

With massively parallel sequencing, many regions and even the entire genome can be interrogated simultaneously to identify such mutations. Although the cost of whole genome sequencing has become progressively lower in the last few years, data analysis and interpretation remain challenging. Due to the large number of short-reads, the sequence data has to be mapped back to the reference genome and filtered through known databases to identify variants for each individual, leading to long turnaround time from clinic testing to reporting. There is also the issue of incidental findings unrelated to the indication for testing and the American College of Medical Genetics and Genomics (ACMG) have recommended the reporting of pathogenic variants for 56 genes [8]. Subsequently, the ACMG recommended that patients be given the choice of opting out of receiving such information [9]. For these reasons, many laboratories still use Sanger sequencing of single or a few genes when there are known causal genes for the suspected disorders.

Exome sequencing can partly overcome the issue of data throughput but not the possibility of incidental findings. Targeted gene panels can address both by focusing on a set of relevant candidate genes with known diagnostic yield, while providing cost-related advantage as well as easier data analysis without the need for specialized computing infrastructure and expertise. The American Society of Human Genetics (ASHG) also recommends that gene testing should be limited to single genes or targeted gene panels based on the clinical presentations of the patient [10]. Compared to Sanger sequencing of single genes, targeted gene panel sequencing has much higher throughput, but each design needs to be evaluated for coverage and sensitivity before being put to routine clinical diagnostic use.

Among several pre-designed catalog panels for pediatric congenital disorders, there is one comprising 180 genes located within chromosomal regions with a high frequency of cytogenetic abnormalities in constitutional disorders [11] according to publicly available data from the International Collaboration for Clinical Genomics (ICCG—previously known as International Standards for Cytogenomic Arrays or ISCA) [12, 13]. To assess the coverage and sensitivity of this ICCG gene panel for high-throughput next-generation sequencing in congenital disorders, we used the Ion Torrent PGM platform to perform mutation screening of 15 pediatric patients with suspected genetic disorders.

Materials and methods
Ethics statement
The patients were previously recruited under two separate projects (CIRB Ref: 2007/831/F and 2010/238/F). Approval to conduct this sequencing study was provided by the SingHealth Central Institutional Review Board (CIRB Ref: 2013/798/F). All the subjects were minors, and written informed consent had been obtained from the parents.

Study samples
The 15 patients were previously recruited from the hospital’s Genetics Clinics for testing of chromosomal imbalance using human 400 K CGH arrays (Agilent Technologies Inc., Santa Clara, USA). No significant pathogenic copy number changes were identified in all 15. Inclusion criteria include developmental delay/intellectual disability and multiple congenital anomalies. Each patient had been followed up and examined by a clinical geneticist. All of them have clinical features suggestive of a disorder associated with one of the 180 genes, although the features may not have been typical or completely fulfilled the clinical criteria of a specific syndrome at the time of recruitment.

DNA extraction
Genomic DNA was manually extracted from peripheral blood collected in EDTA tubes using the Gentra Puregene Blood Kit (Qiagen Inc., Valencia, USA) according to the manufacturer’s instructions. DNA quality and quantity were measured on a Nanodrop Spectrophotometer (Thermo Scientific, Wilmington, USA).
Library construction, sequencing, and data analysis

Genomic DNA (225 ng gDNA) was digested with 16 different restriction enzymes at 37 °C for 30 min to create a library of gDNA restriction fragments. Both ends of the targeted fragments were selectively hybridized to biotinylated probes from the HaloPlex ICCG panel (Agilent Technologies Inc., Santa Clara, CA, USA), which resulted in direct fragment circularization. During the 16-h hybridization process, HaloPlex ION Barcodes and Ion Torrent sequencing motifs were incorporated into the

| Sample | Reads Total | Mapped | On target | Mean depth | Aligned ≥Q20 | On target Uniformity |
|--------|-------------|--------|-----------|------------|--------------|----------------------|
| 1      | 1,348,756   | 1,322,761 | 91.29 %   | 203.4      | 98.59 %      | 87.61 %              |
| 2      | 1,389,395   | 1,361,138 | 91.29 %   | 209.9      | 98.58 %      | 87.80 %              |
| 3      | 1,552,042   | 1,522,728 | 91.16 %   | 234.3      | 98.63 %      | 87.82 %              |
| 4      | 1,494,165   | 1,470,215 | 91.90 %   | 226.8      | 98.71 %      | 87.87 %              |
| 5      | 1,369,435   | 1,346,412 | 91.91 %   | 210.2      | 98.78 %      | 88.89 %              |
| 6      | 1,663,702   | 1,633,814 | 91.20 %   | 252.4      | 98.72 %      | 88.33 %              |
| 7      | 1,602,753   | 1,569,980 | 91.01 %   | 242.7      | 98.67 %      | 88.75 %              |
| 8      | 1,694,348   | 1,662,379 | 91.25 %   | 256.8      | 98.69 %      | 88.80 %              |
| 9      | 1,431,017   | 1,398,943 | 90.08 %   | 211.7      | 98.30 %      | 88.04 %              |
| 10     | 1,717,174   | 1,677,112 | 90.16 %   | 253.2      | 98.24 %      | 87.83 %              |
| 11     | 1,408,352   | 1,373,789 | 89.67 %   | 205.5      | 98.12 %      | 87.42 %              |
| 12     | 1,511,078   | 1,484,377 | 90.97 %   | 227.3      | 98.42 %      | 88.06 %              |
| 13     | 1,554,866   | 1,521,948 | 90.96 %   | 235.1      | 98.44 %      | 89.17 %              |
| 14     | 1,578,886   | 1,547,559 | 91.48 %   | 239.6      | 98.54 %      | 89.31 %              |
| 15     | 1,558,185   | 1,525,061 | 90.91 %   | 234.0      | 98.50 %      | 88.91 %              |

*aPercentage of target bases covered by at least 0.2× the average base read length

**Fig. 1** Percentage of bases at the different read depths
Table 2 Mean coverage with highest and lowest number of reads for target regions for each gene

| Gene  | Mean  | Lowest | Highest |
|-------|-------|--------|---------|
| 1     | ABCC8 | 338.07 | 81.92   | 786.09  |
| 2     | ABCD1 | 169.66 | 12.56   | 411.39  |
| 3     | ACSL4 | 164.77 | 21.30   | 492.11  |
| 4     | AFF2  | 214.46 | 36.76   | 580.05  |
| 5     | ALX4  | 222.07 | 84.73   | 558.92  |
| 6     | AP1S2 | 135.94 | 38.59   | 325.08  |
| 7     | AP2A  | 179.90 | 3.73    | 406.62  |
| 8     | AR    | 223.99 | 43.85   | 529.40  |
| 9     | ATP7A | 178.46 | 15.06   | 431.28  |
| 10    | ATRX  | 158.16 | 10.57   | 441.01  |
| 11    | AIP1R2| 212.24 | 91.24   | 401.30  |
| 12    | BMP4  | 277.08 | 184.26  | 355.73  |
| 13    | BMP1RA| 249.32 | 92.33   | 325.08  |
| 14    | BMP2  | 221.72 | 39.06   | 545.08  |
| 15    | BRCA4 | 226.44 | 69.62   | 659.97  |
| 16    | BRD3  | 158.01 | 1.94    | 403.65  |
| 17    | BSN   | 281.21 | 166.64  | 426.50  |
| 18    | BTR   | 248.63 | 71.65   | 522.36  |
| 19    | CACNA1C| 313.63 | 70.18   | 681.23  |
| 20    | CASK  | 174.65 | 3.07    | 469.39  |
| 21    | CDK1  | 61.17  | 21.98   | 111.66  |
| 22    | CFC1  | 0.00   | 0.00    | 0.06    |
| 23    | CHD7A | 238.95 | 6.00    | 491.12  |
| 24    | CHD8B | 241.14 | 3.17    | 571.36  |
| 25    | CHM   | 138.15 | 0.00    | 424.06  |
| 26    | CHRNA7| 133.17 | 0.00    | 649.69  |
| 27    | CLNKA | 207.37 | 41.54   | 632.50  |
| 28    | CLNKB | 227.42 | 19.10   | 558.00  |
| 29    | CNTN4 | 258.10 | 74.48   | 742.09  |
| 30    | COL2A1| 311.23 | 28.83   | 762.32  |
| 31    | COL4A5| 145.99 | 6.30    | 492.06  |
| 32    | CREEBP| 307.73 | 66.01   | 682.75  |
| 33    | CUL4B | 148.17 | 35.09   | 399.12  |
| 34    | CYP21A2| 42.13  | 0.00    | 317.76  |
| 35    | DOX   | 191.11 | 31.11   | 424.96  |
| 36    | DHCPR | 356.18 | 73.73   | 715.42  |
| 37    | DMRT1 | 317.71 | 99.58   | 526.08  |
| 38    | DYN   | 199.54 | 35.51   | 538.64  |
| 39    | DYSK1A| 238.22 | 56.46   | 539.50  |
| 40    | EDNRB | 244.22 | 108.94  | 440.60  |
| 41    | EHM1  | 322.86 | 0.00    | 914.42  |
| 42    | EMX2  | 191.49 | 89.92   | 367.85  |
| 43    | EXT1  | 255.03 | 122.82  | 531.24  |

(Continued)
| Gene     | Mean Coverage | Highest Number of Reads | Lowest Number of Reads |
|----------|---------------|-------------------------|------------------------|
| MECP2    | 116.40        | 21.60                   | 224.22                 |
| MJD2     | 188.60        | 44.28                   | 383.24                 |
| MITF     | 303.77        | 97.21                   | 559.04                 |
| MSX1     | 148.47        | 85.95                   | 232.61                 |
| MSX2     | 147.01        | 93.77                   | 230.81                 |
| MTM1     | 197.51        | 54.42                   | 517.75                 |
| MYCN     | 228.84        | 96.40                   | 407.08                 |
| NDF      | 237.17        | 92.91                   | 444.75                 |
| NDXFV1   | 299.67        | 104.88                  | 555.45                 |
| NF1      | 368.03        | 144.63                  | 794.05                 |
| NHS      | 189.40        | 24.18                   | 373.29                 |
| NIPBL    | 172.28        | 15.79                   | 382.14                 |
| NLG4A     | 251.27        | 119.13                  | 470.66                 |
| NOTCH2   | 281.99        | 0.00                    | 642.59                 |
| NRSAI1    | 222.01        | 116.84                  | 407.74                 |
| NRXN1    | 225.80        | 21.56                   | 577.97                 |
| NSD1     | 261.44        | 86.24                   | 500.38                 |
| OCA2     | 321.02        | 106.42                  | 685.92                 |
| OCR1      | 178.97        | 13.16                   | 440.97                 |
| OFD1     | 171.17        | 58.00                   | 394.13                 |
| OTC      | 190.78        | 1.46                    | 562.10                 |
| OTX2     | 311.41        | 198.51                  | 476.69                 |
| PAFAH1B1  | 234.58        | 15.99                   | 516.41                 |
| PAK3      | 181.03        | 47.33                   | 405.32                 |
| PXA3     | 235.81        | 80.62                   | 569.70                 |
| PXA6     | 242.37        | 25.79                   | 599.28                 |
| PXA9     | 258.47        | 7.00                    | 540.15                 |
| PGK1     | 252.23        | 95.17                   | 614.74                 |
| PHEX     | 200.00        | 48.82                   | 439.78                 |
| PHFS      | 165.22        | 63.07                   | 315.44                 |
| PIGB     | 205.86        | 19.26                   | 539.76                 |
| PITX2    | 283.28        | 122.96                  | 548.91                 |
| PKD1     | 99.18         | 0.00                    | 512.88                 |
| PKD2     | 225.02        | 47.45                   | 475.99                 |
| PLP1     | 247.34        | 9.76                    | 525.11                 |
| PREPL     | 215.09        | 36.12                   | 484.89                 |
| PRPS1    | 247.30        | 102.55                  | 418.52                 |
| PTC1     | 270.35        | 12.72                   | 733.37                 |
| PTEF      | 150.42        | 22.31                   | 371.28                 |
| PTTPN1    | 267.28        | 7.24                    | 610.72                 |
| RA1A      | 343.75        | 101.17                  | 648.66                 |
| RB1      | 131.02        | 14.38                   | 388.44                 |
| RET      | 271.31        | 71.71                   | 635.08                 |
| RPS19     | 321.73        | 108.91                  | 519.38                 |
targeted fragments. Circularized target DNA-HaloPlex probe hybrids containing biotin were then captured by HaloPlex Magnetic Beads on the Agencourt SRIPlate Super magnet magnetic plate. DNA ligase was added to close the nicks in the hybrids, and freshly-prepared NaOH was used to elute the captured target libraries. The target libraries were then amplified with 18 PCR cycles and purified using AMPure XP beads. Amplicons ranging from 150 to 550 bp were then quantified using an Agilent BioAnalyzer High Sensitivity DNA Assay kit on the 2100 Bioanalyzer to validate the enrichment of the libraries. Library preparation took approximately 1½ days.

Equimolar amounts of four multiplexed bar-coded libraries were pooled and clonally amplified by emulsion PCR, using the Ion PGM Template OT2 200 Kit 9 (Life Technologies, Carlsbad, CA, USA). The template-positive Ion Sphere Particles (ISPs) were then enriched with the Ion OneTouch™ ES and loaded on an Ion 318™ Chip v1. Four separate runs were performed for the 15 samples, with one sample sequenced twice on two different chips. Sequencing was carried out in the Ion PGM™ System using the Ion PGM™ Sequencing 200 Kit v2 according to the manufacturer’s instructions with 500 flow runs.

The data from the sequencing runs were analyzed using the Torrent Suite v4.0.2 analysis pipeline, which includes raw sequencing data processing (DAT processing), splitting of the reads according to the barcode for the individual sample output sequence, classification, signal processing, base calling, read filtering, adapter trimming, and alignment QC. Single-nucleotide polymorphisms (SNP), multi-nucleotide polymorphisms (MNP), insertions, and deletions were identified across all samples. More than 88 % of called bases had a quality score of Q20 (Table 1). At the gene level, 137 of the 180 genes had mean coverage of at least 20×, of which 99 had a mean of >50× and 40 had a mean of >100× (Table 2). Despite the high target region coverage, amplification failed for at least 26 exons across the 180 genes. Thirteen genes (CFC1, CHRNA7, CYP21A2, EHMT1, F8, HBA1, HBA2, IKBKG, NOTCH2, PKD1, SGC, SRY, TSC2) had at least one region that was not amplified and therefore not sequenced (lowest number of reads “0” in Table 2). The sequencing coverage of CFC1, IKBKG, HBA1, and HBA2 previously classified as pathogenic were evaluated for coverage depth and also visually inspected using the Integrative Genomics Viewer before validation by dideoxy sequencing using standard protocol for BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, CA, USA). Segregation analysis was performed when DNA from family members was available. Sequencing was carried out on the Applied Biosystems® 3130 Genetic Analyzer (Life Technologies, Carlsbad, CA, USA). In addition, SIFT (sift.bii.a-star.edu.sg) and Polyphen2 (genetics.bwh.harvard.edu/pph2) were used to check the likely functional significance of missense variants for clinical interpretation.

**Table 2** Mean coverage with highest and lowest number of reads for target regions for each gene (Continued)

| Gene    | Highest coverage | Lowest coverage | Mean coverage |
|---------|------------------|-----------------|--------------|
| ZFM2    | 211.09           | 0.00            | 393.25       |
| ZC1     | 228.52           | 127.92          | 369.74       |
| ZC2     | 128.92           | 29.81           | 320.70       |
| ZC3     | 202.01           | 120.08          | 320.90       |
| ZC4     | 291.25           | 127.92          | 616.00       |

*Target regions do not include non-coding first exons*
was low with >50% of these genes sequenced at >20× (Table 3). The gene with the highest mean coverage was SALL1 (358×). The poorest coverage was for CFC1. Mean read depth for individual exons for three different genes were shown in Figs. 2, 3, and 4.

Overall, 2326 single-nucleotide variants (SNVs) and 25 indels were identified in the 15 patients. These variants identified from the Ion Reporter had an average coverage of 595× and an average Qscore of 38. Variant annotation indicated that 2203 were common variants present in dbSNP and 1000 Genome Project databases. The number of variants ranged from 154 to 175 per patient, with an average of 9.6 novel variants each. Synonymous variants were the most common.

Variants were prioritized for Sanger confirmation based on the individual's clinical presentations. Pathogenic variants were confirmed in six patients. The identified CHD7 (two patients), SHH, TCF4, TSC2, and MECP2 variants and the clinical features of these six patients are listed in Table 4. Another five patients had candidate variants

| Table 3 Percentage of coverage for each gene at 20× |
|-----------------------------------------------------|
| ABCG8  100.00 % | DIMRT1  100.00 % | HNF1B  100.00 % | OTX2  100.00 % | SLC16A2  100.00 % |
| ABCD1  100.00 % | DYM  100.00 % | HOXD13  100.00 % | PAFAH1B1  100.00 % | SLC3A1  100.00 % |
| ACSL4  100.00 % | DYNK1A  100.00 % | HPRT1  100.00 % | PAX3  100.00 % | SLC6A8  94.71 % |
| AFF2  100.00 % | EDNRB  100.00 % | IDS  89.40 % | PAX3  100.00 % | SLC9A6  100.00 % |
| ALX4  100.00 % | EHM1  99.47 % | IKBKG  99.47 % | PAX6  100.00 % | SMAD4  100.00 % |
| AP1S2  100.00 % | EMX2  100.00 % | IRF6  100.00 % | PAX9  99.61 % | SOX2  100.00 % |
| APC  98.72 % | ETTY  100.00 % | JAG1  100.00 % | PGK1  100.00 % | SPINK1  100.00 % |
| AR  100.00 % | EXT2  100.00 % | KAL1  100.00 % | PHEX  100.00 % | SRY  100.00 % |
| ATP7A  100.00 % | EYA1  100.00 % | KCNJ1  100.00 % | PHF6  100.00 % | SYN1  100.00 % |
| ATRX  99.29 % | F8  99.66 % | KCHQ1  100.00 % | PIGB  100.00 % | synaptin1  100.00 % |
| AVPR2  100.00 % | F9  100.00 % | L1CAM  100.00 % | PIIX2  100.00 % | TBCE  100.00 % |
| BMP4  100.00 % | FANC  100.00 % | LAMP2  100.00 % | PKD1  86.06 % | TBI1  100.00 % |
| BMPR1A  100.00 % | FANC  100.00 % | LEMD3  100.00 % | PKD2  100.00 % | TBI1  100.00 % |
| BMPR2  100.00 % | FBN1  100.00 % | LH1X  100.00 % | PLP1  79.74 % | TB5  100.00 % |
| BRC2  100.00 % | FGD1  100.00 % | LMX1B  100.00 % | PREPL  100.00 % | TCF4  100.00 % |
| BRWD3  99.43 % | GFR1  100.00 % | MECP2  100.00 % | PRPS1  100.00 % | TCF1  100.00 % |
| BSN  100.00 % | FLNA  100.00 % | MID1  100.00 % | PDC1  97.80 % | TGFB1  93.58 % |
| BTK  100.00 % | FMR1  100.00 % | MITF  100.00 % | PTEN  100.00 % | TGFB2  100.00 % |
| CACNA1C  100.00 % | FOXC1  100.00 % | MSX1  100.00 % | PTPN11  89.17 % | TGF1  100.00 % |
| CAS5  94.17 % | FOXG1  100.00 % | MSX2  100.00 % | RAI1  100.00 % | TMMPA  100.00 % |
| CDKN1C  100.00 % | FOXL2  100.00 % | MTM1  100.00 % | RB1  100.00 % | TRPS1  100.00 % |
| CFC1  100.00 % | FZD4  100.00 % | MYCN  100.00 % | RET  100.00 % | TSC1  100.00 % |
| CHD7  100.00 % | GATA3  100.00 % | NBP  100.00 % | RPS19  100.00 % | TSC2  98.30 % |
| CHD8  99.11 % | GATA4  100.00 % | NDUFA1  100.00 % | R51  100.00 % | TWIST1  100.00 % |
| CHM  95.10 % | GDF5  100.00 % | NF2  100.00 % | RUNX2  100.00 % | URF18  100.00 % |
| CHRNA7  84.46 % | GJB2  100.00 % | NHS  100.00 % | SALL1  100.00 % | USHC1  100.00 % |
| CLCN1A  100.00 % | GLA  100.00 % | NIF1  100.00 % | SALL4  100.00 % | VHL  100.00 % |
| CLCN1B  100.00 % | GL2  100.00 % | NINO4X1  100.00 % | SAB2  100.00 % | WT1  100.00 % |
| CNT1A  100.00 % | GLB3  100.00 % | NOTCH2  95.39 % | SCN1A  100.00 % | XIAP  100.00 % |
| COL2A1  100.00 % | GPC3  100.00 % | NRSN1  100.00 % | SGCE  94.86 % | ZDHHC9  100.00 % |
| COL4A5  98.76 % | GPC6  100.00 % | NRXN1  100.00 % | SH2D1A  100.00 % | ZEB2  100.00 % |
| CREBBP  100.00 % | GPR56  100.00 % | NOS1  100.00 % | SHANK3  96.32 % | ZFPM2  98.84 % |
| CUL4B  100.00 % | GRA1  100.00 % | OCA2  100.00 % | SHH  100.00 % | ZIC1  100.00 % |
| CYP21A2  67.67 % | HBA1  30.07 % | ORC1  100.00 % | SICX  100.00 % | ZIC2  100.00 % |
| DDX1  100.00 % | HBA2  30.07 % | ODF1  100.00 % | SLC12A1  100.00 % | ZIC3  100.00 % |
| DHCR7  100.00 % | HCCS  100.00 % | OTC  91.74 % | SLC12A3  100.00 % | ZIC4  100.00 % |
for which further evaluation and segregation analysis are ongoing.

**Discussion**

The HaloPlex ICCG panel is a pre-designed made-to-order panel targeting 180 genes. It follows the ICCG recommendations for design and resolution and is available through SureDesign from Agilent Technologies. The targeted panel includes genes in the most commonly altered chromosomal regions according to the ISCA/ICCG database. The 180 genes are covered by 2509 target regions which range in size from 2 to 6575 nucleotides. Depending on its size, a region is covered by between 1 and 547 amplicons.

The recommended minimum read depth for clinical diagnostic sequencing is 20× [15, 16], which was achieved for over 90% of the target for 170 genes. For *CHD7*, even the exon with the poorest coverage had a mean of 36 (Fig. 2). Of the remaining ten, four genes had 80–90% coverage, and the other six (*CFC1, CYP21A, HBA1, HBA2, IKBKG, NOTCH2, PLP1*) had <80%. More than half of the targets in these individual genes are within GC-rich regions. Less efficient PCR for these templates might have resulted in sequencing failure during library preparation, or insufficient sequence data were produced [17]. In addition, the HaloPlex protocol uses restriction enzymes which are sequence-dependent and nonrandom, this method might have contributed further to uneven coverage and also gaps in coverage [18]. For *IKBKG*, the presence of a pseudogene might have caused non-specific alignment and contributed to the low capture of target sequences [19]. Nijman et al. have almost no mapped reads in *IKBKG* in their targeted sequencing, and generally poor coverage of *CFC1* and *IKBKG* had been reported in multiple studies [20–22]. For the gene with the poorest coverage *CFC1*, all six exons had no reads across all 15 samples. This gene is associated with the generation of left-right asymmetry via the TGF pathway. There were 23 mutations in HGMD, 13 of which were found in patients with congenital

![Fig. 2 Average target base read depth for exons 2–38 of CHD7](image)

![Fig. 3 Average target base read depth for exons 1–4 of MECP2](image)
This panel would not be useful for patients with clinical suspicion of \textit{CFC1} gene mutations. The first exon of 64 genes was not included in the design (indicated with “*” in Table 2). All the 64 genes have one or more non-coding exon. The entire exon 1 of these genes (and additional exons for some others) contains only untranslated regions. In general, amplification of exon 1 of some genes was problematic because of the generally higher GC content and sequence complexity [24–26]. Our results showed that \textit{MECP2} had an average target base read depth of 118×. The coverage for exon 1 is the lowest among all, but it is still two times that of the minimum of 20× recommended for clinical diagnostics (Fig. 3). SATB2 had an average target base read depth of 300×, but exon 1 was not covered in the design (Fig. 4). Nevertheless, including non-coding exons in the design might improve the yield of NGS as variants affecting splicing of non-coding exons have been reported to be disease-causing [27].

Many congenital disorders do not have unique and exclusive features, and the presentations may be non-specific. Even for syndromic disorders, there are overlapping features, and the phenotypic features in some patients may be atypical, making it challenging for the clinical geneticists to come to a diagnosis based on clinical history and examination. All the 15 patients in this study have constitutional disorders and suspicion of chromosomal disorders, but CMA did not find any pathogenic copy number abnormality. With this targeted panel,

**Table 4** Pathogenic variants identified and the respective patients’ associated clinical features

| Patient | Gender | Age | Gene       | Nucleotide change | Amino acid change | Clinical features                                      |
|---------|--------|-----|------------|-------------------|-------------------|-------------------------------------------------------|
| 1       | M      | 1d  | CHD7       | NM_017780.3:c.7891C > T | p.R2631X         | Hypoplastic left heart, choanal atresia, oesophageal atresia |
| 2       | F      | 1y4m| CHD7       | NM_017780.3:c.601C > T | p.Q201X          | PDA, aortic stenosis, coloboma, hypotonia              |
| 3       | F      | 3y9m| MECP2      | NM_004992.3:c.763C > T | p.R255X          | Developmental delay, hypotonia, neurodevelopmental regression, epilepsy |
| 4       | F      | 2w  | SHH        | NM_000193.3:c.413C > A | p.S138Y          | Alobar HPE, PDA, hypertelorism, single nostril, choanal atresia, overlapping fingers |
| 5       | M      | 5y11m| TCF4       | NM_001083962.1:c.1793G > A | p.R580Q         | GDD, microcephaly, epicranial folds, hypertelorism, drooling, no speech |
| 6       | F      | 5y8m| TSC2       | NM_000548.3:c.3364delC | p.R1121Vfs*69    | Bilateral large renal cysts, ballotable left kidney, cardiac rhabdomyoma, iris pigmentation & hamartomas, epilepsy |

\textit{GDD} global developmental delay, \textit{HPE} holoprosencephaly, \textit{PDA} patent ductus arterio

*Age at enrollment (d = day, y = year, m = month)
we were able to reach a molecular diagnosis for six patients after reviewing the results with their primary physicians (Table 4). Pathogenic CHD7 variants were detected in two patients with clinical features consistent with CHARGE syndrome. Both CHD7 variants identified (p.R2613X and p.Q201X) have been previously reported in other CHARGE patients [28]. A pathogenic p.R255X MECP2 variant was detected in a patient with clinical features of Rett syndrome. This variant has also been reported previously [29]. The patients with the truncating TSC2 variant and the missense SHH variant also showed clinical features consistent with the respective causative genes. These two variants are novel and the missense variant is predicted to be pathogenic according to both SIFT and Polyphen. Similarly, the clinical features of the patient with the TCF4 variant are found to be consistent with Pitt-Hopkins syndrome upon retrospective review of the patient’s progressive features by the attending physician. This p.R580Q TCF4 variant has been reported as pathogenic in patients with Pitt-Hopkins syndrome [30].

The identification of a patient’s causative mutation has the translational benefit of providing the parents with an answer for their child’s condition. In addition, it provides a guide to the attending clinician on the management and prognosis of the patient. A molecular diagnosis would also facilitate access to clinical trials and programs for special needs children. The use of appropriate gene panels obviates the need for subjective clinical decision on which gene(s) to test in each patient, and may lead to a standard testing workflow for each group of disorders. Generally for those whose diagnosis can be narrowed down to a few suspected genetic syndromes, targeted gene panels would be superior to exome sequencing which has more limitations in the diagnostic setting due to coverage deficiencies in some genes and longer turnaround time. Higher-average read depth could be attained at a lower cost, making it superior to exome sequencing in terms of cost, sensitivity, and expected diagnostic yield [31, 32].

Conclusions

The Haloplex ICCG panel had good coverage except for ten of the target genes. Consideration would have to be made for the low coverage for some regions in several genes which might have to be supplemented by Sanger sequencing. However, comparing the cost, ease of analysis, and shorter turnaround time, it is a good alternative to exome sequencing for patients whose features are suggestive of a genetic etiology involving one of the genes in the panel.

Competing interests

All authors declare that they have no competing interests.

Authors’ contributions

ECT designed the study and obtained the funding. EL and SPL carried out the sequencing experiments. EL, MB, and SJ performed the analysis and interpretation of the sequencing data. MB, AL, EST, and IN carried out the selection of patients, clinical assessment, and phenotype correlation. EL and ECT prepared the manuscript. All authors read and approved the manuscript.

Acknowledgements

This work is supported by NMRC/CG/006/2013 and NMRC/PPG/KKH12010-Theme3 from the National Medical Research Council, Ministry of Health, Republic of Singapore. We thank K Chang and CH Kuick for the use of the Ion Torrent PGM.

Author details

1 KK Research Centre, KK Women’s and Children’s Hospital, 100 Bukit Timah Road, Singapore 229899, Singapore. 2 Genetics Service, Department of Paediatrics, KK Women’s and Children’s Hospital, Singapore 229899, Singapore. 3 Paediatrics Academic Clinical Programme, SingHealth Duke-NUS Graduate Medical School, Singapore 169857, Singapore.

Received: 8 September 2015 Accepted: 22 November 2015
Published online: 14 December 2015

References

1. Dolk H, Loane M, Garne E. The prevalence of congenital anomalies in Europe. Adv Exp Med Biol. 2010;686:634–64.
2. Birth defects [http://www.cdc.gov/ncbddd/birthdefects/index.html]. Accessed 4 Aug 2015.
3. Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, et al. Trends in the prevalence of developmental disabilities in US children, 1997–2008. Pediatrics. 2011;127:1034–42.
4. Miller DT, Adam MP, Aradhya S, Baraekger LG, Brothman AR, Carpenter 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;86:740–64.
5. Bannik M, Wsienowicka-Kowalnik B, Nowakowska B, Smyk M, Kedzior M, Sobieka K, et al. The usefulness of array comparative genomic hybridization in clinical diagnostics of intellectual disability in children. Dev Period Med. 2014;18:307–17.
6. Battaglia A, Doccini V, Bernardini L, Novelli A, Loddo S, Capalbo A, et al. Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features. Eur J Paediatr Neurol. 2013;17:589–95.
7. Chong WW, Lo F, Lam ST, Wang CC, Luk HM, Leung TY, et al. Performance of chromosomal microarray for patients with intellectual disabilities/developmental delay, autism, and multiple congenital anomalies in a Chinese cohort. Mol Cytogenet. 2014;7:34.
8. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15:565–74.
9. ACMG Board of Directors. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. Genet Med. 2015;17:688–9.
10. Botkin JR, Belmont JW, Berg JS, Berkman BE, Bombard Y, Holm IA, et al. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. Am J Hum Genet. 2015;97:6–21.
11. Haloplex disease research panels [http://www.chem.agilent.com/Library/ datasheets/Public/HaloplexDiseaseResearchPanels5991-2526Ena.pdf]. Accessed 4 Aug 2015.
12. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med. 2011;13:777–84.
13. Moreno-De-Luca D, Sanders SJ, Willsey AJ, Mulle JG, Lowe JK, Geschwind DH, et al. Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. Mol Psychiatry. 2013;18:1090–5.
14. Yourshaw M, Taylor SP, Rao AR, Martin MG, Nelson SF. Rich annotation of DNA sequencing variants by leveraging the Ensembl Variant Effect Predictor with plugins. Brief Bioinform. 2014;16:235–64.
15. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15:733–47.

16. Yohe S, Hauge A, Bunjer K, Kemmer T, Bower M, Schomaker M, et al. Clinical validation of targeted next-generation sequencing for inherited disorders. Arch Pathol Lab Med. 2015;139:204–10.

17. Bragg LM, Stone G, Butler MK, Hugenholz P, Tyson GW. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLoS Comput Biol. 2013;9:e1003031.

18. Coonrod EM, Durtschi JD, VanSant WC, Voelkerding KV, Kumanovics A. Next-generation sequencing of custom amplicons to improve coverage of HaloPlex multigene panels. Biotechniques. 2014;57:204–7.

19. Bragg LM, Stone G, Butler MK, Hugenholz P, Tyson GW. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLoS Comput Biol. 2013;9:e1003031.

20. Brett M, McPherson J, Zang ZJ, Lai A, Tan ES, Ng I, et al. Massively parallel sequencing of patients with intellectual disability, congenital anomalies and/or autism spectrum disorders with a targeted gene panel. PLoS One. 2014;9:e93409.

21. Nijman IJ, van Montfrans JM, Hoogstraat M, Boes ML, van de Corput L, Renner ED, et al. Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies. J Allergy Clin Immunol. 2014;133:1279–38.

22. Kammermeier J, Drury S, James CT, Dzubak R, Ocola L, Elawad M, et al. Targeted gene panel sequencing in children with very early onset inflammatory bowel disease—evaluation and prospective analysis. J Med Genet. 2014;51:748–55.

23. Human gene mutation database [https://portal.biobase-international.com/hgmd/pro/gene.php?gene=CFC1]. Accessed 2 Sep 2015.

24. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27:182–9.

25. Hu H, Wrogemann K, Kalscheuer V, Tzschach A, Richard H, Haas SA, et al. Mutation screening in 86 known X-linked mental retardation genes by droplet-based multiplex PCR and massive parallel sequencing. Hugo J. 2009;3:9–10.

26. Porreca GJ, Zhang K, Li JH, Xie B, Austin D, Vassallo SL, et al. Multiplex amplification of large sets of human exons. Nat Methods. 2007;4:931–6.

27. Eisenberger T, Neuhaus C, Khan AO, Decker C, Preising MN, Friedburg C, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. PLoS One. 2013;8:e78496.

28. Bartels CF, Scacheri PC, White L, Scacheri PC, Bale S. Mutations in the CHD7 gene: the experience of a commercial laboratory. Genet Test Mol Biomarkers. 2010;14:881–91.

29. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999;23:185–8.

30. Arniel J, Rio M, de Pontual L, Redon R, Malan V, Boddard N, et al. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet. 2007;80:988–93.

31. Ankala A, da Silva C, Gualandi F, Ferlini A, Bean LJ, Collins C, et al. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. Ann Neurol. 2015;77:206–14.

32. Consugar MB, Navarro-Gomez D, Place EM, Bujakowska KM, Sousa ME, Fonseca-kelly JD, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. Genet Med. 2015;17:253–61.