Molecular Genetics and Metabolism 100 (2010) 184–192

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme

Next generation sequencing in research and diagnostics of ocular birth defects

Gordana Raca [a,b,*], Craig Jackson [a], Berta Warman [d], Tom Bair [c], Lisa A. Schimmenti [d]

[a] UW Cytogenetic Services, Wisconsin State Laboratory of Hygiene, 465 Henry Mall, Madison WI 53706, USA
[b] Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, 13000 University Avenue, Madison WI 53706, USA
[c] The University of Iowa DNA Facility, John W. Eckstein Medical Research Building, 431 Newton Rd, Iowa City, IA, USA
[d] Department of Pediatrics, Division of Genetics and Metabolism, University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 18 January 2010
Received in revised form 8 March 2010
Accepted 8 March 2010
Available online 15 March 2010

Keywords:
Next generation sequencing
Sequence capture
GS FLX
Anophthalmia
Microphthalmia
Coloboma

ABSTRACT

Sequence capture enrichment (SCE) strategies and massively parallel next generation sequencing (NGS) are expected to increase the rate of gene discovery for genetically heterogeneous hereditary diseases, but at present, there are very few examples of successful application of these technologic advances in translational research and clinical testing. Our study assessed whether array based target enrichment followed by re-sequencing on the Roche Genome Sequencer FLX (GS FLX) system could be used for novel mutation identification in more than 1000 exons representing 100 candidate genes for ocular birth defects, and as a control, whether these methods could detect two known mutations in the *PAX2* gene. We assayed two samples with heterozygous sequence changes in *PAX2* that were previously identified by conventional Sanger sequencing. These changes were a c.527G > C (S176T) substitution and a single basepair deletion c.77delG. The nucleotide substitution c.527G > C was easily identified by NGS. A deletion of one base in a long polyG stretch (c.77delG) was not registered initially by the GS Reference Mapper, but was detected in repeated analysis using two different software packages. Different approaches were evaluated for distinguishing false positives (sequencing errors) and benign polymorphisms from potentially pathogenic sequence changes that require further follow-up. Although improvements will be necessary in accuracy, speed, ease of data analysis and cost, our study confirms that NGS can be used in research and diagnostic settings to screen for mutations in hundreds of loci in genetically heterogeneous human diseases.

Introduction

The goal of our study was to evaluate the capacity of array based sequence capture target enrichment (SCE) and massively parallel, next generation sequencing (NGS) to successfully identify mutations in candidate genes for the developmental ocular birth defects anophthalmia, microphthalmia and coloboma.

The advent of NGS technologies is expected to transform the practice of medical genetics [1–3]. With the high throughput and decreased sequencing costs achieved by NGS, it is no longer impossible to sequence hundreds or even thousands of exons and other genomic sequences in an individual with a suspected genetic disease. It is predicted that in the near future NGS might replace array based techniques and Sanger sequencing in their current clinical applications for the detection of mutations [1,3]. Additionally, NGS provides entirely new research and diagnostic capabilities, including whole genome screening for novel mutations and sequencing biological specimens for the genomic signature of novel infectious agents [4,5].

NGS could be particularly advantageous in research and testing for genetically heterogeneous hereditary conditions. Common disorders evaluated by clinical geneticists are caused by heterogeneous Mendelian loci and lend themselves to enrichment strategies followed by NGS. Examples include intellectual disability [6], deafness [7], familial cardiomyopathy [8] and retinitis pigmentosa [9]. In these conditions there are often very subtle phenotypic differences between affected patients to guide molecular diagnostics by indicating which gene is likely to be mutated in a particular individual. Current diagnostic evaluation proceeds by sequencing a series of genes, individually or in small sets, based on the relative frequency of the mutations and the sensitivity of available assays. If there is no predominant mutation(s) causing the disease, the pathogenic change often remains unknown even after very extensive and expensive molecular testing. With enrichment strategies followed by NGS, sequencing of all genes implicated in a particular genetic disorder could be performed simultaneously, efficiently and at low cost.

While clearly superior to traditional Sanger sequencing, NGS has had little impact on clinical testing to date. There are very few examples of successful application of NGS in translational research and diagnostics. Clinical testing using NGS is currently offered for Hypertrophic Cardiomyopathy (HCM), Dilated

* Corresponding author at: UW Cytogenetics Service, Wisconsin State Laboratory of Hygiene, 465 Henry Mall, Madison, WI 53706, USA. Fax: +1 608 265 7818.
E-mail address: racago@mail.slh.wisc.edu (G. Raca).

1096-7192/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymgme.2010.03.004

Cardiomyopathy (DCM) and Long QT syndrome (http://www.genedx.com/). NGS has also been explored as a method to perform rapid human leukocyte antigen (HLA) typing for high resolution allele identification [10,11], and to develop assays for Neurofibromatosis Type 1 [12], autosomal recessive ataxia [13] and mitochondrial disorders caused by mutations in the mitochondrial genome and 362 nuclear genes controlling mitochondrial function [14]. Ng et al. applied targeted sequencing of all coding regions (''exome") to show the presence of causative mutations in four unrelated individuals with a rare dominantly inherited disorder, Freeman–Sheldon syndrome (FSS) [15] and to discover the gene for a rare recessive disorder of previously unknown cause, Miller disease [16]. Exciting applications have also been described in cancer research, where NGS has been applied in discovering new candidate genes for acute myeloid leukemia [17,18], glioblastoma multiforme [19] and other malignancies [20].

We describe the first study which showed the feasibility of using genomic enrichment by sequence capture followed by NGS to investigate genetic causes of the ocular birth defects anophthalmia, microphthalmia and coloboma. These eye anomalies are among the most prevalent causes of childhood blindness, affecting annually ~2 per 10,000 newborns worldwide [21]. Although they can be of different origins, the majority are caused by defects in genes which regulate normal eye development [22–25]. There is increased evidence that mutations in large numbers (possibly hundreds) of different genes can cause congenital eye malformations, but no single gene is responsible for a high percentage of cases [23–25]. Anophthalmia, microphthalmia and coloboma therefore represent disorders where simultaneous sequencing of large numbers of candidate genes by NGS is an ideal approach to study genetic causes and demonstrate the feasibility of NGS for clinical diagnostics.

We showed that the combination of array based SCE with resequencing on the GS FLX instrument using Titanium chemistry allows concurrent sequencing of more than 100 candidate genes for anophthalmia, microphthalmia and coloboma. However, improvements will be necessary in several areas including accuracy, speed, ease of data analysis and cost to allow successful diagnostic implementation of NGS for simultaneous mutation testing in hundreds of genes in genetically heterogeneous human diseases.

Materials and methods

We tested whether two known sequence changes in the renal-coloboma syndrome (a.k.a. Papillorenal syndrome, OMIM #120330) associated gene *PAX2*, which were previously characterized by Sanger sequencing, can reliably be detected by NGS. The first variant was a missense change in exon 5, c.527G > C, which resulted in serine to threonine amino-acid change S176T. This base substitution was identified in one of our previous studies in a father of a proband with ocular birth defects, but not in his affected child. Since detailed clinical information for the parent was not available, the change was described as a variant of unknown clinical significance.

Since short stretches of mono-, di-, and trinucleotide repeats represent hotspots for disease causing frameshift mutations in genomic DNA, we wanted to determine if this mutation type is detectable by NGS. Therefore, the second sequence change selected for the study was a deletion of one base in a polyG stretch in exon 2 of the *PAX2* gene (c.77delG), which has previously been described by Schimmenti et al. [26].

We designed a custom 385,000 probe SCE array with more than 100 candidate genes for eye malformations. The list of selected candidate genes is provided in Table 1. The genes were chosen based on reports of mutations identified in patients with coloboma, microphthalmia and anophthalmia [22,23,25]. Additionally, a comprehensive literature search was performed for published mutations associated with ocular phenotypes in animal models [27,28]. Some genes were included based on their role in signaling and developmental pathways which are known as important for eye formation and function [24].

Appropriate SCE probes for the regions of interest were chosen in collaboration with Roche-NimbleGen design team (Roche-NimbleGen, Madison, WI). Only protein coding regions (coding exons) of the 112 candidate genes were targeted on the array. We selected 385,000 long oligonucleotide probes (>60 bp) to tile the exons of the genes of interest with a very high density. All the probes had the uniqueness score of one (defined as having no match in the genome other than itself longer than 38 bp, allowing up to five insertions/deletions/mismatches in that match), to exclude repetitive regions from probe selection and avoid capturing pseudogene sequences [29]. Upon completion of the design and manufacturing of the array, the SCE on samples of genomic DNA from two patients with known mutations in the *PAX2* gene was performed at the NimbleGen service laboratory (Roche-NimbleGen, Madison, WI), following previously described protocols [30–32]. Briefly, ~20 μg of each patient's genomic DNA were randomly fragmented by nebulization to an average size of 500 bp. Linkers were ligated to the DNA fragments to provide a priming site for post-enrichment amplification of the eluted fragment pool. The fragments were denatured and hybridized to the custom SCE array. After a 72-h hybridization, unbound material was removed by stringent washing. The arrays were transferred to the NimbleGen elution system, and the enriched fragment pool was eluted and recovered from the array. The enriched fragments were amplified with 22mer linkers to generate enough DNA template for downstream applications. After amplification, the amount of captured DNA was measured by spectrophotometry and the product was tested for enrichment level by quantitative PCR with four proprietary QC control loci. These QC loci are conserved in both human and mouse genomes and have been empirically determined to accurately predict enrichment with several different array designs.

Sequencing of the two SCE prepared samples was performed on the GS FLX instrument using Titanium chemistry, at the University of Iowa DNA Facility following standard protocols. Two samples, separated by gaskets, were sequenced independently on two regions of the picotiter-plate. Briefly, amplified fragments from SCE were end-repaired and ligated to adapter oligonucleotides. The library was diluted based on the results of a previously performed titration, so that upon denaturation single DNA fragments hybridized to individual beads containing sequences complementary to adapter oligonucleotides. The beads were compartmentalized into water-in-oil microvesicles to allow clonal expansion of separate DNA molecules bound to the beads by emulsion PCR. After amplification, the emulsion was disrupted, and the beads containing clonally amplified template DNA were enriched. The beads were again separated by limiting dilution, deposited into individual picotiter-plate wells, and combined with sequencing enzymes. Iterative pyrosequencing was performed on the picotiter-plate by successive flow addition of the four dNTPs. A nucleotide-incorporation event in a well containing clonally amplified template produced pyrophosphate release and picotiter-plate well-localized luminescence, which recorded by a charge-coupled device (CCD) camera. With the flow of each dNTP reagent, wells were imaged, analyzed for their signal-to-noise ratio, filtered according to quality criteria, and subsequently algorithmically translated into a linear sequence output [33–35].

Data analysis was performed using the Roche proprietary software package for the GS FLX system. Image acquisition, image processing and signal processing were performed during the run. Post run analysis was conducted using the GS Reference Mapper.