Published in final edited form as:
Mamm Genome. 2014 April ; 25(0): 120–128. doi:10.1007/s00335-013-9494-7.

Ion Torrent sequencing for conducting genome wide scans for mutation mapping analysis

Rama Rao Damerla, **Bishwanath Chatterjee**, **You Li**, **Richard J. B. Francis**, **Sarosh N. Fatakia**, and **Cecilia W. Lo***
Department of Developmental Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15201

Abstract

Mutation mapping in mice can be readily accomplished by genome wide segregation analysis of polymorphic DNA markers. In this study, we showed the efficacy of Ion Torrent next generation sequencing for conducting genome wide scans to map and identify a mutation causing congenital heart disease in a mouse mutant, *Bishu*, recovered from a mouse mutagenesis screen. The *Bishu* mutant line generated in a C57BL/6J (B6) background was intercrossed with another inbred strain, C57BL/10J (B10), and the resulting B6/B10 hybrid offspring were intercrossed to generate mutants used for the mapping analysis. For each mutant sample, a panel of 123 B6/B10 polymorphic SNPs distributed throughout the mouse genome was PCR amplified, bar coded, and then pooled to generate a single library used for Ion Torrent sequencing. Sequencing carried out using the 314 chip yielded >600,000 usable reads. These were aligned and mapped using a custom bioinformatics pipeline. Each SNP was sequenced to a depth >500×, allowing accurate automated calling of the B6/B10 genotypes. This analysis mapped the mutation in *Bishu* to an interval on the proximal region of mouse chromosome 4. This was confirmed by parallel capillary sequencing of the 123 polymorphic SNPs. Further analysis of genes in the map interval identified a splicing mutation in *Dnaic1*$^{c.204+1G>A}$, an intermediate chain dynein, as the disease causing mutation in *Bishu*. Overall, our experience shows Ion Torrent amplicon sequencing is high throughput and cost effective for conducting genome-wide mapping analysis and is easily scalable for other high volume genotyping analyses.

INTRODUCTION

Next generation sequencing has greatly accelerated experimental investigations into the genetic etiology of human diseases. The availability of high throughput sequencing data now makes “systems genetics” a possibility with the use of large-scale forward genetic screens for novel gene discovery (Caruana et al. 2013; Hill et al. 2013). Forward genetic screens in mice and other model organisms with chemical mutagenesis using ethylnitrosourea (ENU) has yielded new insights into the genetic basis for a wide spectrum of biological and disease processes (Arnold et al. 2011; Guenet 2004), including genes that can cause congenital heart disease (CHD) (Shen et al. 2005; Yu et al. 2004; Zhang et al. 2009).

*Corresponding Author: Department of Developmental Biology, Rangos Research Center Rm 8120, 530 45 St., Pittsburgh, PA 15201 Ph: 412-692-9901; cel36@pitt.edu.

Competing Interests
The authors declare that they have no competing interests.

Availability of Supporting Data:
The raw Ion Torrent PGM sequencing data sets supporting the results of this article are available for download from the following link: fhttp://apps.devbio.pitt.edu/Genome/RawData_PGM.zip

While mutagenesis screens have been highly successful in recovering mutants with a variety of defect phenotypes for disease modeling, mutation recovery remains challenging. While this process is becoming more straightforward with easier access to whole exome sequencing analysis (Bull et al. 2013), mutation mapping is often still an important part of the mutation recovery effort to clearly identify the mutation(s) responsible for the given phenotype. This usually entails crossing the mutation generated in one inbred strain into a different inbred strain, and conducting a full genome scan to examine segregation of the disease phenotype with a panel of DNA markers polymorphic between the two strains (Bode et al. 1988; Xia et al. 2010; Zhang et al. 2009). However, a common problem with such mapping intercrosses is the common observation that the mutant phenotype can be modified or even lost due to genetic modifier effects (Nadeau 2001) To avoid the confounding effects arising from genetic modifiers, intercrosses can be carried out with more closely related mouse strains, such as intercrosses with the closely related C57B10 mouse strain to map mutations generated in the C57BL6 background.

Mapping ENU induced mutations in mice can be carried out using a panel of 50–150 polymorphic DNA markers comprising a combination of microsatellite DNA repeats and single nucleotide polymorphic (SNPs) markers distributed through out the genome. A panel of B6/B10 polymorphic markers have been previously described for mapping ENU induced mutations in the B6 background (Xia et al. 2010). While an Illumina high density mouse SNP genotyping array containing approximately 1400 markers and the mega-MUGA array with 77,000 markers are commercially available, these SNP arrays are not suitable for mapping polymorphisms between the closely related B6 and B10 mouse strains. Therefore, in this study, we investigated the feasibility and cost effectiveness of conducting a full genome scan using bulk amplicon sequencing with the Ion Torrent next generation sequencing system for mapping mutations in B6/B10 mice.

To test the efficacy of Ion Torrent sequencing for mutation mapping, we focused our analysis on *Bishu*, a newly recovered mutant mouse model exhibiting complex congenital heart disease (CHD) associated with heterotaxy obtained from an ongoing mouse ENU mutagenesis screen. We conducted a full genome scan with 123 SNPs using amplicon sequencing with the Ion Torrent and developed a custom bioinformatics pipeline for automating the analysis of the resulting next generation sequence data. Through this analysis, we mapped the mutation in *Bishu* to mouse chromosome 4 and identified the mutation as a splicing defect mutation in *Dnaic1*

RESULTS

Bishu, a mutant recovered from a mouse ENU mutagenesis screen, conducted in the NHLBI Bench to Bassinet Cardiovascular Development Consortium, exhibited congenital heart disease (CHD) associated with laterality defects (Figure 1). *Bishu* mutants can have complete reversal of visceral organ situs known as situs inversus totalis (Figure 1B), or they can exhibit heterotaxy where the left-right positioning of organs may be randomized (Figure 1C). A spectrum of complex CHD is usually observed in conjunction with heterotaxy, such as double outlet right ventricle (Figure 1D, E) with multiple ventricular septal defects (Figure 1F, G) and hypoplastic transverse arch (Figure 1H). As *Bishu* was recovered from a C57BL/6J inbred strain background, *Bishu* mice were intercrossed with another inbred strain, C57BL/10J (B10), to generate B6/B10 hybrid offspring for the mapping analysis. These B6/B10 hybrid animals were intercrossed to generate homozygous mutants that were used for conducting a full genome scan. The mutation map interval was identified by tracking segregation of homozygous B6 markers with the *Bishu* disease phenotype.

Ion Torrent Amplicon Sequencing

Seven homozygous B6/B10 *Bishu* mutant animals with the CHD/laterality defect phenotypes were used for the genome scan mapping analysis. A panel of 123 B6/B10 polymorphic DNA markers previously shown to be effective for mapping mutations were used for this analysis (Xia et al. 2010). Primers were designed to generate short DNA fragments (150 to 300 bp) encompassing each SNP (Supplementary Table S1) to allow the use of the Ion Torrent 200 bp sequencing chemistry. For each mutant, sequence fragments spanning each of the 123 markers were PCR amplified and then bar coded with a unique sequence tag. In parallel, we also tested the efficacy of multiplex PCR in which a single amplification reaction was performed for all 123 DNA markers and the resulting product was similarly bar coded.

The bar coded PCR products encompassing all 123 markers from all 7 mutant animals were pooled and a single library was constructed and processed for sequencing using a single 314 chip on the Ion Torrent Personal Genome Machine (PGM) (Figure 2). The total sequence data output was aligned to the reference genome created by concatenating the sequences of all the amplicons. Overall, we obtained 615,712 reads providing an average ~500× coverage (Table 1). To assess the quality of the sequencing run, we examined the AQ20 score, which corresponds to the longest length at which the error rate is 1% or less. The mean AQ20 was 153 with 99% of the library covered at this length. We also examined the “perfect” length score corresponding to the longest length at which there were no mismatches. We observed 99% of the reads had mean “perfect” length score of 110 bp (Torrent Suite 2.2 User Documentation). Together these assessments show the sequencing data generated are of high quality.

Ion Torrent Amplicon Sequencing Analysis

To analyze the large volume of Ion Torrent sequencing reads and extract the genotype calls for each of the 123 SNPs, a custom bioinformatics pipeline was developed comprising a a PERL script that uses mpileup command from SAM tools (samtools.sourceforge.net/mpileup.shtml) to tabulate the genotype calls (see Supplementary Tables S2 to S9). This automation streamlined the determination of genotype calls for all 123 SNPs (Table 2). Unambiguous genotype calls were obtained for all 123 SNPs that were PCR amplified individually (columns 3 – 9 in Table 2). In the multiplex PCR sample, genotype calls were successfully generated for 85% of SNPs. These exhibited sequencing coverage that was comparable to those obtained with individual PCR amplifications (Column 9 in Table 2). To validate the results obtained by Ion Torrent amplicon sequencing, we carried out a parallel genotyping analysis of 6 mutants with a subset of 62 SNPs using Sanger capillary sequencing analysis. The genotype calls obtained by Sanger sequencing (Supplementary Table S10) were concordant with those identified by the PGM sequencing analysis, showing the efficacy of SNP genotyping by Ion Torrent amplicon sequencing.

Genome Scan Mapping and Sequencing Analysis Identifies the *Bishu* Mutation

The genome scan analysis mapped the *Bishu* mutation to one genomic interval - a 17 MB region situated between SNP rs13477622 (Chr 4: 28249560) and rs49519173 (Chr 4: 45462131; mm9) in the proximal end of chromosome 4 (black box in Table 2). This region was observed to be consistently B6 homozygous in all 7 mutants. This map interval was confirmed with further analysis of the linkage data using recombinant interval haplotype analysis (Neuhaus and Beier 1998). Examination of this chromosome 4 region revealed a gene known to cause CHD and laterality defects, *Dnaic1* (Guichard et al. 2001). *Dnaic1* is an axonemal dynein required for motile cilia function, and is a gene that is also known to be associated with primary ciliary dyskinesia (PCD), a sinopulmonary disease arising from mucus clearance defects due to motile cilia defects in the airway. Similar to *Bishu* mutants,

PCD patients also can exhibit situs inversus totalis or heterotaxy, a reflection of the dual requirement for motile cilia both in left-right patterning and airway clearance.

Sequencing of cDNA obtained from a *Bishu* mutant embryo revealed transcripts with exon 5 deleted (Figure 3). Sequencing of genomic DNA from *Bishu* revealed a G to A substitution flanking the splice donor site of exon 4 (*Dnaic1*$^{c.204+1G>A}$), which would account for the observed exon 5 skipping and would predict a reading frame shift resulting in protein truncation after a 15 amino acid insertion beyond residue 76 of the Dnaci1 protein (p*76Argext*15, Figure 3). Further genotyping analysis confirmed all *Bishu* mutants are homozygous for this *Dnaic1* mutation, validating this as the disease causing mutation. Consistent with *Dnaic1* as the gene harboring the disease causing mutation, we observed *Bishu* mutants have motile cilia defects. Thus instead of the normal rapid synchronous ciliary beat driving fluid flow across the tracheal respiratory epithelia and in brain ependymal tissue, *Bishu* mutants exhibited immotile/slow/dyskinetic cilia with little or no net fluid flow (see Supplemental Movie). Consistent with the compromised ciliary motion in the ependyma, *Bishu* mutants surviving postnatally usually die from hydrocephalus by 5–10 days after birth. Furthermore, analysis by electron microscopy confirmed *Bishu* mutant airway cilia are missing the outer dynein arms (Figure 1 I, J), a cilia ultrastructural defect associated with *Dnaic1* mutations in mice and in PCD patients (Guichard et al. 2001; Ostrowski et al. 2010; Pennarun et al. 1999).

Efficiency of Amplicon Sequencing

To assess the efficacy of Ion Torrent amplicon sequencing for mapping analysis, we considered the cost of conducting genome scan on different sample sizes on the Ion Torrent (Table 3). Based on the Ion Torrent PGM run report and coverage statistics, we expect 100× coverage can be achieved with the multiplexing of 40 samples in a single run using the Ion 314 chip (Table 3). With the Ion 316 chip, 128 samples can be simultaneously sequenced using the available bar codes (Table 3), and many more can be accommodated as additional bar codes become available. With amplicon sequencing, there is also savings in personnel time with the automation of genotype calling, made possible with the bioinformatics pipeline we have developed for analysis of the Ion Torrent PGM short sequence reads. Thus PGM amplicon sequencing is easily scalable to maximize the efficiency and cost effectiveness of high volume custom genotyping and genome wide mapping analyses.

DISCUSSION

We show the efficacy of next generation amplicon sequencing with the Ion Torrent PGM for genome scan mapping. Using this approach, we mapped an ENU induced mutation causing CHD associated with laterality defects in the *Bishu* mutant mouse line to a 15 Mb interval on mouse chromosome 4. This made it possible to identify the disease causing mutation as a splicing defect mutation in *Dnaic1*. This would be predicted to yield a loss of function mutation given the splicing defect is predicted to generate a frame shift after amino acid 76, and consistent with this, *Bishu* mutants exhibit phenotypes similar to that found in the *Dnaic1* knockout mice (Francis et al. 2012). Using this same amplicon sequencing method, we have performed genome scan analysis to map the mutations in two other mutant lines with congenital heart defects. This allowed the mapping of the mutation in each line to a 35 MB region on mouse chromosome 15, confirming the general utility of this strategy for genome wide mapping analysis (unpublished observations; Damerla et al. 2013).

While our genome scan was conducted using 123 polymorphic DNA markers, the number of markers can be increased or decreased to accommodate the mapping resolution desired. The bioinformatics pipeline we developed for analysis of the next generation sequencing data allowed for rapid automated genotype calling that streamlined the mapping analysis. It

should be noted while there are commercial mouse SNP genotyping arrays, these are only available for more commonly used inbred mouse strains, none for the closely related B6/B10 inbred mice used in our studies. This points to the utility of amplicon sequencing for conducting custom genotyping analysis. Although our study was focused on mapping mutation in the mouse genome, this same approach can be applied for genome scanning and genotyping analysis in any organism, including human clinical studies.

The emergence of bench top sequencing machines such as the PGM for next generation sequencing (Chan et al. 2012) has made it possible to scale sequencing projects to accommodate higher sample throughput while decreasing cost. Our studies showed the efficacy and scalability of amplicon sequencing for conducting genome scans and genotyping analysis. We showed the cost for amplicon sequencing can be significantly reduced by multiplexing the PCR amplifications and also using sample bar coding to reduce the time and cost for library construction. At present, up to 128 unique bar codes are available, and when combined with the use of the higher capacity 316 chip, this allows significant scale up of the sequencing run on the PGM. The addition of more bar codes is expected in the near future, which will further facilitate scale up of amplicon sequencing.

Previous reports have shown amplicon resequencing can be used to identify novel mutations with targeted resequencing of selected genes (Daum et al. 2012; Otto et al. 2011). These studies also reported significant savings in costs and time in utilizing next-generation sequencing platforms over Sanger sequencing for targeted resequencing analysis. One previous concern in using next generation sequencing for resequencing analysis is its lower sequencing accuracy. This concern is abated with the much higher sequencing depth afforded by the current next generation sequencing platforms. Using data generated in this study, we found sequencing depth with 100× coverage is sufficient to provide 100% accuracy in genotyping calls.

Our findings show amplicon sequencing using the PGM is cost effective for high throughput genome scan mapping analysis and is customizable and easily adapted for small or large-scale studies. The flexibility of amplicon sequencing will make it possible to undertake custom genotyping analysis, whether in studies involving model organisms or clinical studies involving unique patient population. Overall, this approach has general applicability for a wide variety of large-scale genotyping analyses and can be employed clinically, such as in HLA genotyping (Wang et al. 2012), pharmacogenomics studies, or other custom genotyping analysis such as those required for the clinical practice of personalized medicine.

METHODS

Institutional Approval for Animal Studies

All mouse experiments were carried out using protocols approved by the Institutional Animal Care and Use Committee of the University of Pittsburgh.

DNA Samples and PCR amplification

The *Bishu* mutant analyzed in this study was recovered from a mouse ENU mutagenesis screen. The mutagenesis and breeding of mice were carried out as previously described (Yu et al. 2004). Skin tissue was collected from 7 *Bishu* mutants and genomic DNA was extracted for the mapping analysis using 123 SNPs. Each SNP was amplified in 50-μl PCR reaction containing 10× Amplitaq Gold buffer, MgCl2, 0.15 mM dNTP mix, 0.1 μM each of forward and reverse primer, 1 unit Amplitaq Gold polymerase and 50 ng genomic DNA. PCR amplification was performed at DNA denaturation at 95°C for 5 min, 40 cycles of 95°C for 30 s, 55°C for 30, 72°C for 1 minute, and finally 5 min at 72°C. Multiplex PCR

involved amplification of all the SNPs simultaneously in a single PCR reaction with all the primers mixed into the same PCR reaction mixture. All PCRs were performed in either DNA Engine Tetrad® 2 (Bio-Rad) or Mastercycler® nexus (Eppendorf) thermal cyclers.

Library Preparation and Ion Torrent Amplicon Sequencing

The libraries were generated using the Ion Plus Fragment Library Kit (Cat. no. 4471252, Life Technologies). Amplicons generated by PCRs from each mutant were mixed in equal volumes. A total volume of 500 μl of mixed amplicons was used for a single purification reaction using Agencourt® AMPure® SP Reagent (Beckman Coulter) at 1:1.8 ratio of DNA to beads in a 2 ml eppendorf tube. Each tube was placed on a magnetic rack for 2 minutes followed by 2 washes with 70% ethanol. After ethanol was removed and the tube air dried, the beads were resuspended in 50 μl low TE buffer. 1 μl of the pooled amplicons were analyzed on an Agilent® Bioanalyzer® using the Agilent® High Sensitivity DNA Kit, and Bioanalyzer® software to determine the concentration of the amplicon pools. 100 ng of pooled amplicons from each sample were end-repaired using 20 μl End repair buffer and 1 μl end repair enzyme supplied in Ion Plus Fragment Library Kit and incubated at room temperature for 20 minutes. Agencourt® AMPure® SP Reagent was used to purify the end-repaired pooled amplicons as described above.

Ligation of Adapters and Barcodes

Barcodes 1 through 8 from Ion Xpress 1–16 barcoding kit were used for each mutant respectively and combined with reagents supplied in the Ion Plus Fragment Library Kit were incubated in a thermal cycler at 25°C for 15 minutes followed by 72°C for 5 minutes. These barcoded pooled amplicons were further purified using the Agencourt® AMPure® SP Reagent.

Library Quantification & Template Preparation

The library was quantified by qPCR using the Ion Library Quantitaion Kit (Cat. No. 4468802) to determine a suitable template dilution factor so as to be clonally amplified in ion sphere particles,. Template dilution factors for each of the barcoded libraries were established according to the protocol in Ion Library Quantitaion Kit (Cat. No. 4468802). Briefly, 1:200 and 1:2000 dilutions of each barcoded library were analyzed by qPCR by validating them on a standard curve generated by a pre-quantified standard *E. coli* library. qPCR reactions were carried out in 7900 HT System (Applied Biosystems). Template dilution factors were calculated and the barcoded libraries were combined in equimolar ratios for template preparation. We used the automated procedure for template preparation using the Ion One Touch System and the Ion OneTouch 200 template kit using the manufacturer's instructions.

Ion Torrent PGM Sequencing and DNA Sequence Data Analysis

Templates prepared from pooling all the barcoded libraries were sequenced on a 314 chip using the Ion Torrent PGM and the Ion PGM™ Sequencing Kit according to the manufacturer's instructions. Ion Torrent reads were aligned to amplicon reference sequences using CLCBio Genomic Workbench software. Genotype calling for all the marker positions were automated using samtools mpileup (http://samtools.sourceforge.net) together with custom scripts. Positions with insufficient coverage (less than 20×) are denoted as "genotype unknown".

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH Grants U01-HL098180 and P30-HL101322. We thank members of the Lo laboratory, Ashok Srinivasan, Mark Kimak and Yang Li for technical support. We thank Dr. Bruce Beutler and Yu Xia for helpful discussions and providing information on the panel of B6/B10 SNP markers ahead of publication.

References

Arnold CN, Xia Y, Lin P, Ross C, Schwander M, Smart NG, Muller U, Beutler B. Rapid identification of a disease allele in mouse through whole genome sequencing and bulk segregation analysis. Genetics. 2011; 187:633–641. [PubMed: 21196518]

Bode VC, McDonald JD, Guenet JL, Simon D. hph-1: a mouse mutant with hereditary hyperphenylalaninemia induced by ethylnitrosourea mutagenesis. Genetics. 1988; 118:299–305. [PubMed: 3360305]

Bull KR, Rimmer AJ, Siggs OM, Miosge LA, Roots CM, Enders A, Bertram EM, Crockford TL, Whittle B, Potter PK, Simon MM, Mallon AM, Brown SD, Beutler B, Goodnow CC, Lunter G, Cornall RJ. Unlocking the bottleneck in forward genetics using whole-genome sequencing and identity by descent to isolate causative mutations. PLoS Genet. 2013; 9:e1003219. [PubMed: 23382690]

Caruana G, Farlie PG, Hart AH, Bagheri-Fam S, Wallace MJ, Dobbie MS, Gordon CT, Miller KA, Whittle B, Abud HE, Arkell RM, Cole TJ, Harley VR, Smyth IM, Bertram JF. Genome-wide ENU mutagenesis in combination with high density SNP analysis and exome sequencing provides rapid identification of novel mouse models of developmental disease. PLoS One. 2013; 8:e55429. [PubMed: 23469164]

Chan M, Ji SM, Yeo ZX, Gan L, Yap E, Yap YS, Ng R, Tan PH, Ho GH, Ang P, Lee AS. Development of a Next-Generation Sequencing Method for BRCA Mutation Screening: A Comparison between a High-Throughput and a Benchtop Platform. J Mol Diagn. 2012; 14:602–612. [PubMed: 22921312]

Damerla, R.; Cui, C.; Gabriel, G.; Liu, X.; Gibbs, B.; Francis, R.; LI, Y.; Chatterjee, B.; Michaud, J.; Pazour, G.; Lo, C. Mutation in the mouse homolog of C5ORF42 disrupts ciliogenesis and causes cerebellar defects and other Joubert Syndrome phenotypes associated with the disruption of Shh signaling. 63rd Annual meeting of the American Society of Human Genetics; October 23rd, 2013; Boston, USA. 2013. Abstract#3162W

Daum LT, Rodriguez JD, Worthy SA, Ismail NA, Omar SV, Dreyer AW, Fourie PB, Hoosen AA, Chambers JP, Fischer GW. Next-generation ion torrent sequencing of drug resistance mutations in Mycobacterium tuberculosis strains. J Clin Microbiol. 2012; 50:3831–3837. [PubMed: 22972833]

Francis RJ, Christopher A, Devine WA, Ostrowski L, Lo C. Congenital heart disease and the specification of left-right asymmetry. Am J Physiol Heart Circ Physiol. 2012; 302:H2102–2111. [PubMed: 22408017]

Guenet JL. Chemical mutagenesis of the mouse genome: an overview. Genetica. 2004; 122:9–24. [PubMed: 15619957]

Guichard C, Harricane MC, Lafitte JJ, Godard P, Zaegel M, Tack V, Lalau G, Bouvagnet P. Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome). Am J Hum Genet. 2001; 68:1030–1035. [PubMed: 11231901]

Hill JT, Demarest BL, Bisgrove BW, Gorsi B, Su YC, Yost HJ. MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. Genome Res. 2013; 23:687–697. [PubMed: 23299975]

Nadeau JH. Modifier genes in mice and humans. Nat Rev Genet. 2001; 2:165–174. [PubMed: 11256068]

Neuhaus IM, Beier DR. Efficient localization of mutations by interval haplotype analysis. Mamm Genome. 1998; 9:150–154. [PubMed: 9457677]

Ostrowski LE, Yin W, Rogers TD, Busalacchi KB, Chua M, O'Neal WK, Grubb BR. Conditional deletion of dnaic1 in a murine model of primary ciliary dyskinesia causes chronic rhinosinusitis. Am J Respir Cell Mol Biol. 2010; 43:55–63. [PubMed: 19675306]

Otto EA, Ramaswami G, Janssen S, Chaki M, Allen SJ, Zhou W, Airik R, Hurd TW, Ghosh AK, Wolf MT, Hoppe B, Neuhaus TJ, Bockenhauer D, Milford DV, Soliman NA, Antignac C, Saunier S,

Johnson CA, Hildebrandt F. Mutation analysis of 18 nephronophthisis associated ciliopathy disease genes using a DNA pooling and next generation sequencing strategy. J Med Genet. 2011; 48:105–116. [PubMed: 21068128]

Pennarun G, Escudier E, Chapelin C, Bridoux AM, Cacheux V, Roger G, Clement A, Goossens M, Amselem S, Duriez B. Loss-of-function mutations in a human gene related to Chlamydomonas reinhardtii dynein IC78 result in primary ciliary dyskinesia. Am J Hum Genet. 1999; 65:1508–1519. [PubMed: 10577904]

Shen Y, Leatherbury L, Rosenthal J, Yu Q, Pappas MA, Wessels A, Lucas J, Siegfried B, Chatterjee B, Svenson K, Lo CW. Cardiovascular phenotyping of fetal mice by noninvasive high-frequency ultrasound facilitates recovery of ENU-induced mutations causing congenital cardiac and extracardiac defects. Physiol Genomics. 2005; 24:23–36. [PubMed: 16174781]

Wang C, Krishnakumar S, Wilhelmy J, Babrzadeh F, Stepanyan L, Su LF, Levinson D, Fernandez-Vina MA, Davis RW, Davis MM, Mindrinos M. High-throughput, high-fidelity HLA genotyping with deep sequencing. Proc Natl Acad Sci U S A. 2012; 109:8676–8681. [PubMed: 22589303]

Xia Y, Won S, Du X, Lin P, Ross C, La Vine D, Wiltshire S, Leiva G, Vidal SM, Whittle B, Goodnow CC, Koziol J, Moresco EM, Beutler B. Bulk segregation mapping of mutations in closely related strains of mice. Genetics. 2010; 186:1139–1146. [PubMed: 20923982]

Yu Q, Shen Y, Chatterjee B, Siegfried BH, Leatherbury L, Rosenthal J, Lucas JF, Wessels A, Spurney CF, Wu YJ, Kirby ML, Svenson K, Lo CW. ENU induced mutations causing congenital cardiovascular anomalies. Development. 2004; 131:6211–6223. [PubMed: 15548583]

Zhang Z, Alpert D, Francis R, Chatterjee B, Yu Q, Tansey T, Sabol SL, Cui C, Bai Y, Koriabine M, Yoshinaga Y, Cheng JF, Chen F, Martin J, Schackwitz W, Gunn TM, Kramer KL, De Jong PJ, Pennacchio LA, Lo CW. Massively parallel sequencing identifies the gene Megf8 with ENU-induced mutation causing heterotaxy. Proc Natl Acad Sci U S A. 2009; 106:3219–3224. [PubMed: 19218456]

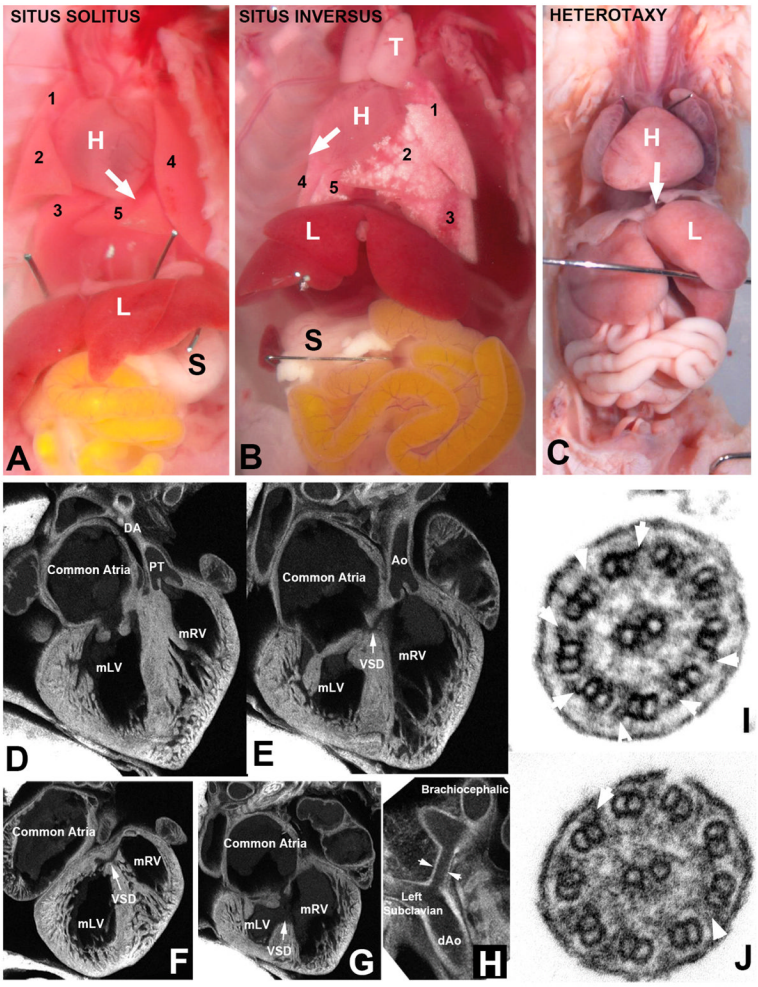

Figure 1. Congenital heart disease and left-right patterning defects in *Bishu* mutants

(A–C). *Bishu* mutants can exhibit abnormal visceral organ situs such as situs inversus (B) in which all organs show mirror symmetry or heterotaxy (C) with disocardant organ situs. Normal situs solitus animal (A) show levocardia with heart apex pointing to the left (denoted by white arrow), normal pattern of three lung lobes on the right and two on the left, and with stomach positioned on the left. In contrast, *Bishu* mutant with situs inversus show dextrocardia (white arrow pointing to the right) with reversed lung lobation and stomach positioned on the right (B). In panel C is *Bishu* mutant with heterotaxy with the heart showing a mid-line placement or mesocardia (white arrow pointing down the midline). This mutant had stomach on the left (not visible in this image), and left pulmonary isomerism with bilateral single lung lobes.

(D–H). *Bishu* mutant showing complex congenital heart disease involving malalignment of the great arteries. This mutant was stillborn and exhibited dextrocardia such that the morphological right ventricle (mRV) was positioned on the left. The atrial septum failed to form, giving a common atrium (F, G), and both the pulmonary outflow (PT) and the aorta were connected to the mRV, indicating a double outlet right ventricle. This was accompanied by multiple ventricular septal defects (VSD) and also narrowing of the transverse arch (denoted by arrowheads in H).

(I, J). Cilia in tracheal epithelia of *Bishu* mutant showed missing out dynein arm defect. Cilia of the mouse tracheal airway epithelia are shown in cross section view. In control

animal (I), the outer microtubule doublets showed abundant outer dynein arms (white arrowheads), but in the *Bishu* mutant cilia, very few outer dynein arms are observed (J).

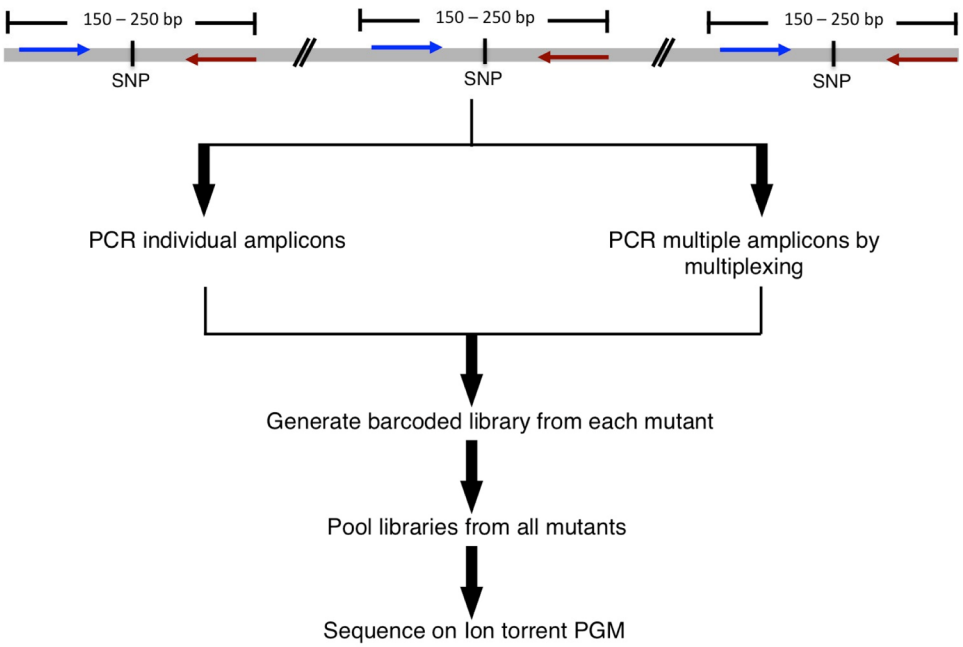

Figure 2. Ion Torrent amplicon sequencing workflow

A schematic diagram of the steps involved in amplicon sequencing for genome scan analysis are shown. 150 to 200 bp around each SNP are amplified either individually or in a single multiplex PCR. Barcoded libraries generated from each mutant are then pooled to make a single library that is then amplified for sequencing on the Ion Torrent PGM.

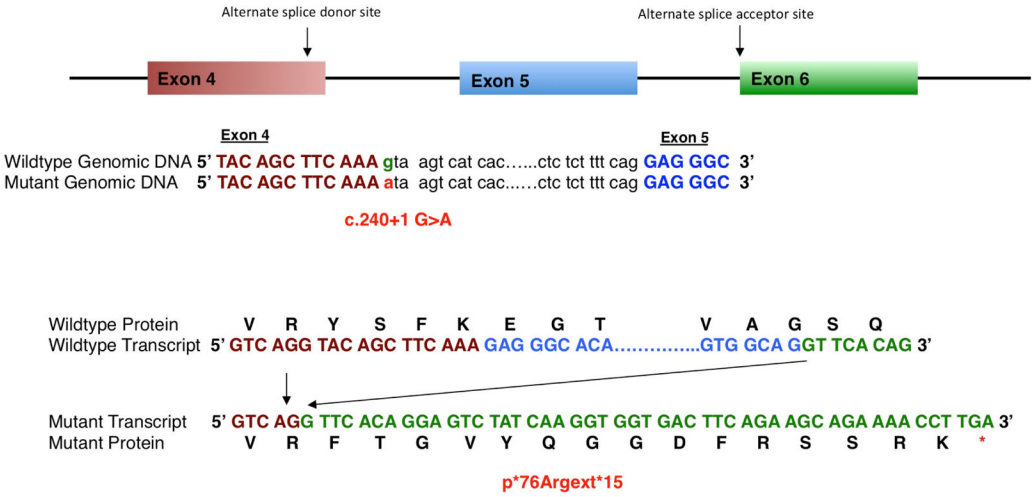

Figure 3. *Dnaic1* splicing defect mutation identified in *Bishu* mutant

Schematic of the mouse *Dnaic1* gene is depicted with a point mutation in exon 4 positioned at the 3′ splice junction, resulting in the use of an alternative splice donor and acceptor site in exons 4 and 6, respectively. The mutation c.240+1 G>A in the genomic region is highlighted in red and the wild type is highlighted in green. Also shown are the wild type and mutant transcripts, and the mutant polypeptide generated from the mutant transcripts without exon 5 (blue).

Table 1

Sequencing Metrics

	Mutants							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**7-Multiplex**
Barcode ID	1	2	3	4	5	6	7	8
Total reference length	25040	25040	25040	25040	25040	25040	25040	25040
Total read count	76452	65899	68278	64002	112660	81941	80033	42459
Mean read length	184.3	185.08	183.21	182.56	183.02	182.31	183.72	176.38
Maximum coverage	2390	2126	1872	1439	2709	1954	2020	1927
Average coverage	556.92x	481.68x	494.24x	461.88x	814.51x	590.46x	581x	296x

Table 2

Genome Scan Results from Ion Torrent Amplicon Sequencing[¶]

				Mutants							
Ch	**Position**	**SNP**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**7-M**
1	4905893	C	A	6	6	10	6	H	6	6	6
1	43063842	T	A	6	6	10	H	6	6	H	H
1	61228463	T	C	6	6	10	H	6	6	H	H
1	75483331	G	A	H	6	10	H	6	6	H	H
1	95571814	T	C	H	H	H	H	H	6	H	H
1	118565405	C	A	H	H	6	6	H	6	H	H
1	131470919	C	T	10	H	6	6	H	6	H	N/A
1	152988092	C	T	10	H	6	6	H	6	H	N/A
1	173342737	T	A	10	H	6	H	H	6	H	H
1	188081175	A	G	10	H	H	H	H	6	H	H
2	3582646	A	G	H	H	H	6	10	10	10	10
2	25034194	C	G	H	H	10	6	10	10	H	H
2	50942492	C	T	H	H	10	6	10	10	H	H
2	67362739	G	A	H	H	10	6	10	10	H	H
2	84313007	A	G	H	10	H	H	10	10	H	N/A
2	103735349	C	A	H	10	H	H	10	10	H	H
2	146096051	C	T	H	10	H	H	10	10	H	N/A
2	164025160	A	G	H	10	H	H	H	10	H	N/A
3	4365615	T	C	H	H	6	H	10	H	H	H
3	25440271	C	G	10	6	6	6	6	6	6	6
3	51646346	T	C	H	6	H	H	10	H	H	H
3	67807084	G	A	H	6	10	H	10	H	H	H
3	84907155	T	C	H	6	10	H	10	10	H	H
3	107273295	A	G	H	6	10	H	10	10	H	H
3	122116882	C	T	H	6	10	H	10	10	H	H
3	146593495	C	T	H	H	H	H	10	H	6	N/A

				Mutants							
Ch	**Position**	**SNP**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**7-M**
4	7149610	C	T	10	H	H	6	H	6	6	N/A
4	28249560	T	C	H	6	6	6	6	6	6	6
4	45462131	T	G	6	6	6	6	6	6	6	6
4	66040938	A	G	6	H	H	H	H	6	6	6
4	87804727	C	A	6	H	H	H	H	6	6	6
4	112653568	G	T	6	H	H	H	H	H	6	6
4	126253529	G	C	6	H	H	H	H	6	6	6
4	142217407	T	C	6	6	6	10	6	6	6	6
5	23345000	A	G	H	6	H	10	H	6	10	10
5	44059873	A	G	H	6	H	10	H	H	10	10
5	67382789	A	G	H	6	H	H	H	H	10	N/A
5	88988451	G	A	H	6	H	H	H	10	10	10
5	109001097	C	A	10	6	H	H	H	10	10	10
5	129799057	T	C	10	H	6	6	10	H	10	10
5	148405352	C	T	10	H	6	6	10	H	H	H
6	4944845	C	A	H	H	H	H	H	6	10	10
6	23322761	T	A	H	H	N/A	H	H	6	10	10
6	42395982	C	T	H	H	H	10	H	H	10	10
6	67742189	C	A	H	H	H	10	H	H	10	10
6	84280591	C	T	H	10	H	10	H	H	10	10
6	104875242	A	C	H	10	10	10	H	H	10	N/A
6	125060262	A	G	6	10	10	10	H	H	10	10
6	147720182	A	G	6	10	H	10	6	10	10	10
7	7014667	T	C	H	H	H	H	6	H	H	H
7	28467081	T	A	H	H	H	10	6	H	6	6
7	38216957	A	C	H	H	10	6	6	H	6	N/A
7	54410823	C	T	H	H	10	H	6	6	6	6
7	71519895	G	T	H	H	10	H	6	6	6	6

				Mutants							
Ch	**Position**	**SNP**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**7-M**
7	104565567	A	G	H	H	10	H	6	6	6	6
7	130492865	A	G	H	6	H	H	6	6	6	N/A
7	141188625	A	G	H	6	H	6	6	6	H	H
8	4235918	C	T	H	10	H	H	H	H	H	H
8	24145986	C	T	6	10	6	H	H	H	6	6
8	48705306	G	A	6	10	6	10	H	6	6	6
8	64124535	G	A	6	10	6	10	H	6	6	6
8	126154896	A	G	6	H	6	10	6	H	6	6
9	8307298	T	C	H	10	H	6	6	H	H	H
9	25697557	A	G	6	10	H	H	6	H	H	N/A
9	44891469	G	A	H	H	H	H	6	H	6	6
9	65069297	A	T	H	6	H	H	6	10	6	6
9	87513872	T	G	H	6	6	10	10	10	6	6
9	106375830	T	C	H	6	H	10	6	H	6	6
9	122496139	A	T	H	H	H	H	6	H	H	N/A
10	25103948	A	T	10	H	H	10	6	10	6	6
10	43717117	C	T	10	H	H	10	6	10	6	6
10	62547650	C	T	10	6	H	10	6	10	H	H
10	83411614	C	T	10	H	H	10	6	10	H	H
10	104474937	A	C	H	H	10	H	6	10	H	H
10	126774961	C	A	H	6	H	6	H	10	H	H
11	5927550	A	G	10	10	H	H	H	H	H	H
11	24352635	T	C	10	10	H	H	H	H	H	N/A
11	46027800	G	T	10	H	H	H	H	H	H	H
11	65237281	A	G	H	H	H	H	H	H	10	10
11	83361835	C	T	H	H	H	H	H	6	10	10
11	105208941	A	C	10	H	H	H	10	10	H	H
12	4035991	C	T	H	H	H	10	H	6	6	6

				Mutants							
Ch	**Position**	**SNP**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**7-M**
12	25626291	A	G	H	H	H	10	6	6	6	6
12	33783681	T	G	H	H	6	H	6	6	6	6
12	55293511	C	G	H	H	6	H	H	6	6	6
12	69022193	T	C	H	H	6	H	H	6	6	10
12	85893620	T	C	H	H	6	H	H	6	6	6
12	106196191	G	C	H	H	6	H	H	6	H	H
13	23458896	T	A	H	H	H	6	H	10	H	N/A
13	44605368	G	A	H	H	H	6	H	10	H	N/A
13	60294898	T	C	H	10	6	6	6	10	H	H
13	83826961	T	G	10	10	6	H	6	10	H	H
13	107699046	G	A	H	H	6	H	H	10	H	H
14	13021464	C	T	6	10	10	6	10	6	H	H
14	30518255	C	T	6	10	10	6	10	6	H	H
14	55673056	T	G	6	10	H	H	10	6	H	H
14	77195309	C	T	6	10	H	H	10	H	H	H
14	116124073	A	G	H	H	H	H	10	H	H	H
15	7117980	A	G	6	H	10	H	6	H	H	H
15	25941148	C	T	6	H	10	H	6	6	6	N/A
15	42051029	A	G	H	10	10	H	6	6	H	H
15	63931032	G	A	H	10	10	H	6	6	H	H
15	83618701	G	A	H	10	H	H	6	H	H	H
15	100309099	G	A	H	10	10	10	H	H	H	H
16	4972820	A	G	H	H	6	10	H	H	H	H
16	29856457	A	G	H	6	6	10	10	H	10	10
16	44261352	A	G	6	6	6	10	10	H	10	10
16	54901204	G	A	H	6	6	10	10	H	10	10
16	87736505	C	T	H	H	6	H	10	H	10	10
17	3690688	G	A	H	H	H	10	H	10	10	N/A

				Mutants							
Ch	**Position**	**SNP**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**7-M**
17	24310623	T	A	H	6	H	10	10	10	10	10
17	47451200	G	A	H	6	6	H	10	10	H	H
17	66263080	G	A	H	6	6	H	10	6	6	6
17	84624403	A	G	H	6	6	H	H	6	6	6
18	15408257	C	T	6	10	6	H	H	H	H	H
18	35366160	A	T	6	10	H	H	H	H	H	H
18	46803584	T	C	6	10	6	H	H	H	H	N/A
18	65425288	C	T	6	10	H	H	H	H	H	H
18	83497271	G	A	6	10	H	10	H	H	6	6
19	46875560	T	G	H	6	6	6	H	6	H	H
X	55120804	A	G	6	10	6	10	6	10	6	6
X	147904667	A	G	6	10	10	10	10	H	6	6
X	158414344	G	T	6	10	10	10	10	H	6	N/A

[¶]6 = homozygous C57BL/6J; 10= homozygous C57BL/10J; H= heterozygous C57BL/6J:C57BL/10J; 7-M= Mutant 7 multiplex.

Table 3

Cost Comparison for Full Genome Scan Linkage Analysis

Ion Torrent Sequencing	8 Samples	8 Samples Multiplex[1]	40 Samples Multiplex[1]	128 Samples Multiplex[2]
PCR	$1,016	$10	$40	$130
Library and Template Prep.	$630	$630	$2,172	$7,898
Ion Chips	$99	$99	$99	$299
Sequencing	$125	$125	$125	$125
Total Cost	$1,870	$864	$2,436	$8,452
Cost Per Sample	$234	$108	$61	$66

[1] Sequencing cost using Ion 314 chip.

[2] Sequencing cost using Ion 316 chip.