A workflow to increase verification rate of chromosomal structural rearrangements using high-throughput next-generation sequencing

Kelly Quek[1], Katia Nones[1], Ann-Marie Patch[1], J. Lynn Fink[1], Felicity Newell[1], Nicole Cloonan[1], David Miller[1], Muhammad Z. H. Fadlullah[1], Karin Kassahn[1], Angelika N. Christ [1], Timothy J. C. Bruxner[1], Suzanne Manning[1], Ivon Harliwong[1], Senel Idrisoglu[1], Craig Nourse[1], Ehsan Nourbakhsh[1], Shivangi Wani[1], Anita Steptoe[1], Matthew Anderson[1], Oliver Holmes[1], Conrad Leonard[1], Darrin Taylor[1], Scott Wood[1], Qinying Xu[1], Australian Pancreatic Cancer Genome Initiative[2], Peter Wilson[1], Andrew V. Biankin[3,4,5,6], John V. Pearson[1], Nic Waddell[1], and Sean M. Grimmond[1,6]

[1]Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, QLD, Australia, [2]Australian Pancreatic Cancer Genome Initiative (for full list of contributors, see www.pancreaticcancer.net.au/apgi/collaborators), [3]The Kinghorn Cancer Centre, Cancer Research Program, Garvan Institute of Medical Research, Sydney, NSW, Australia, [4]Department of Surgery, Bankstown Hospital, Sydney, NSW, Australia, [5]South Western Sydney Clinical School, Faculty of Medicine, University of NSW, Liverpool, NSW, Australia, and [6]Wolfson Wohl Cancer Research Centre, Institute for Cancer Sciences, University of Glasgow, Glasgow, Scotland, United Kingdom

BioTechniques 57:31-38 (July 2014) doi 10.2144/000114189
Keywords: next-generation sequencing; cancer; chromosome breakpoints; structural variation; verification; high-throughput

Supplementary material for this article is available at www.BioTechniques.com/article/114189.

Somatic rearrangements, which are commonly found in human cancer genomes, contribute to the progression and maintenance of cancers. Conventionally, the verification of somatic rearrangements comprises many manual steps and Sanger sequencing. This is labor intensive when verifying a large number of rearrangements in a large cohort. To increase the verification throughput, we devised a high-throughput workflow that utilizes benchtop next-generation sequencing and in-house bioinformatics tools to link the laboratory processes. In the proposed workflow, primers are automatically designed. PCR and an optional gel electrophoresis step to confirm the somatic nature of the rearrangements are performed. PCR products of somatic events are pooled for Ion Torrent PGM and/or Illumina MiSeq sequencing, the resulting sequence reads are assembled into consensus contigs by a consensus assembler, and an automated BLAT is used to resolve the breakpoints to base level. We compared sequences and breakpoints of verified somatic rearrangements between the conventional and high-throughput workflow. The results showed that next-generation sequencing methods are comparable to conventional Sanger sequencing. The identified breakpoints obtained from next-generation sequencing methods were highly accurate and reproducible. Furthermore, the proposed workflow allows hundreds of events to be processed in a shorter time frame compared with the conventional workflow.

Cancer is the result of the accumulation of genetic damage in key genes and pathways, which ultimately leads to uncontrolled growth of mutated cells (1,2). This damage ranges from small point mutations to large chromosome structural rearrangements. Structural rearrangements include deletions, insertions, tandem duplications, inversions, and translocations. Many cancer genomes carry tens to hundreds of structural rearrangements that may

METHOD SUMMARY

To increase the efficiency of verification of chromosomal structural rearrangements, we present a high-throughput workflow utilizing an amplicon pooling strategy combined with benchtop sequencing and standard bioinformatics techniques that facilitates the examination of more than 300 breakpoints, resulting in the identification of breakpoints in more than 80% of events at single base resolution.

have functional consequences such as disruption of tumor suppressor genes, activation of oncogenes, gene copy number alteration, and/or formation of new fusion genes (3–6).

Traditionally, structural rearrangements have been investigated by cytogenetic methods or array-based approaches. However, these techniques are low resolution, low-throughput, and do not fully capture all types of rearrangements. With the advent of next-generation sequencing, structural rearrangements can be identified to base level resolution using whole genome paired sequencing methods. Sequencing data are mapped to a reference genome, and potential structural rearrangements are identified using approaches such as discordant read pairs (7), split read (8), or soft clipping (9). Two large international consortia, the International Cancer Genomic Consortium (ICGC) and The Cancer Genome Atlas (TCGA), have been established to interrogate different cancer types. Both generate high quality comprehensive catalogs of genomic abnormalities (somatic mutations, abnormal expression of genes, and epigenetic modifications) using next-generation sequencing to better understand the molecular pathophysiology of cancers (10,11). To achieve high quality data, it is essential for genomic abnormalities in the tumors to be curated and verified for the catalog.

With respect to structural rearrangements, the verification workflow in previous studies utilizes manual primer design of each rearrangement, PCR, gel electrophoresis, and Sanger sequencing (12–14). This approach is low-throughput and requires extensive labor input, especially when verifying a large number of rearrangements; consequently it can take weeks to complete the entire verification process.

To increase verification throughput, we have devised a high-throughput workflow that utilizes benchtop next-generation sequencing (e.g., Ion Torrent PGM or Illumina MiSeq) and bioinformatics tools for primer design, de novo assembly of sequencing data, and BLAT to identify the DNA sequence of breakpoints for hundreds of structural rearrangements. As proof of principle, we employed this high-throughput workflow to verify structural variants in a highly complex cancer genome. We assessed the impact of the verification rate when primers were placed at a greater distance from the breakpoints and also examined the performance of the two benchtop next-generation sequencing approaches to identify breakpoints and compare with those obtained by Sanger sequencing. We conclude that the high-throughput verification workflow incorporating next-generation sequencing methods is comparable to the conventional methods employing Sanger sequencing and can complete the verification workflow in less than half the time.

Materials and methods

Samples, library preparation, and sequencing

A primary pancreatic ductal adenocarcinoma and matching normal tissue were obtained from the Australian Pancreatic Cancer Genome Initiative. DNA was extracted using the Allprep DNA/RNA Mini Kit method (Qiagen, Victoria, Australia). A long mate-pair library for each sample was generated according to the Mate-Paired Library Preparation 5500 Series SOLiD Systems kit protocol (Life Technologies, Foster City, CA) (http://tools.lifetechnologies.com/content/sfs/manuals/cms_093442.pdf). Briefly, 5 μg of genomic DNA was sheared into ~2 kb fragments (Covaris S220 System; Life Technologies) and circularized with linker sequences, followed by digestion of the circularized DNA to generate a template used in emulsion PCR. The template was flanked with adaptor sequences, coupled to beads, and clonally amplified before immobilizing it to a solid surface for a 50 bp sequencing run using the SOLiD v4 (Life Technologies).

Analysis of potential somatic rearrangements

Sequence data were mapped to a reference genome based on the Genome Reference Consortium (www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/) GRCh37 assembly using the Bioscope v1.2.1 software suite (Applied Biosystems, Foster City, CA). The average physical coverage of the analyzed samples was 268× (tumor) and 211× (normal).

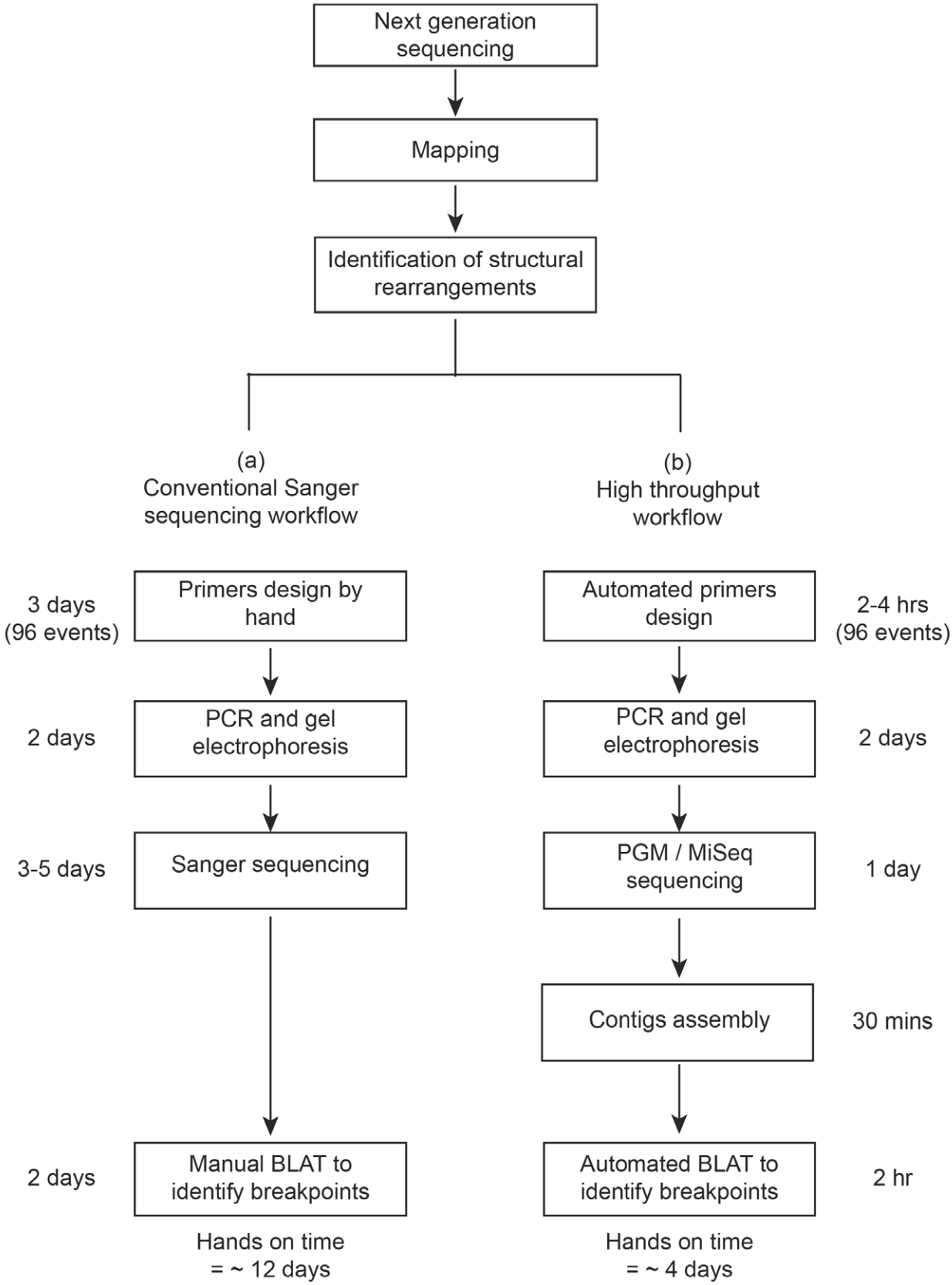

Figure 1. Conventional and high-throughput workflows for the verification of somatic rearrangements and identification of breakpoints. (a) Conventional workflow for verifying somatic rearrangements and their breakpoints using Sanger sequencing. (b) High-throughput verification workflow using either PGM or MiSeq sequencing together with in-house bioinformatics tools to accelerate verification rate. The time to perform each step to verify true somatic rearrangements (assuming 96 events are tested) and determine the sequence of breakpoints is shown. Note that this workflow is applicable for long mate-pair and paired-end whole genome sequencing.

The base coverage was 32× (tumor) and 26× (normal). After mapping, the insert size range for each sequence run was determined to be 630–2780 bp and 670–2900 bp in the tumor and normal tissue, respectively. Discordant read pairs were clustered to identify potential somatic rearrangements using an in-house tool, qSV (Patch et al., manuscript in preparation). A potential rearrangement is supported by at least 10 read pairs. Rearrangements were classified as intra-chromosomal (events within a single chromosome) or inter-chromosomal (events between two chromosome, such as translocations). In cases, where intra-chromosomal rearrangements could be associated with copy number change, together with pair orientation information, this allowed a more specific categorization of the events into deletions, tandem duplications, and inversions. The latter events were not associated with a copy number change and thus were broadly grouped as intra-chromosomal rearrangements. Potential somatic rearrangements were verified using PCR and sequencing (such as benchtop next-generation sequencing and Sanger sequencing).

Automated primer design

The primers were designed using the qAmplicon tool (http://sourceforge.net/p/adamajava/wiki/qAmplicon/) (Lynn et al., manuscript in preparation), which can automatically design primers for each event type (deletions, tandem duplications, intra-chromosomal rearrangements, inversions, and translocations). To do this, qAmplicon extracts a user-defined number of bases with a sequence that spans the predicted breakpoint, it then employs Primer 3 (15) to suggest primer pairs that surround the breakpoint. qAmplicon uses BLAST to match the primers against the reference genome and selects the best primer pairs according to a defined set of criteria, including unique alignment to the reference genome, inability of the primers to form dimers, and an acceptable predefined melting temperature. The input files for qAmplicon include a reference genome and a list of two genomic ranges for each structural rearrangement. Primer pairs for Sanger sequencing were designed to generate a maximum amplicon size of either 1 kb (short amplicon category) or 3 kb (long amplicon category) using benchtop next-generation sequencing.

PCR verification

Each candidate somatic rearrangement was verified by PCR amplification of tumor and matching normal tissue DNA using a 25 µL reaction and was set up in 96 well plates using a robot (Bravo; Agilent Technologies, Victoria, Australia). PCR reactions were performed using the following parameters: 94°C × 2 min, (94°C × 30 s, 60°C × 30 s, 68°C × 1 min) for 40 cycles, 68°C × 15 min. PCR products were visualized by gel electrophoresis (Supplementary Figure S1). A single clear PCR band specific to the tumor at the expected size range was classified as true somatic; PCR bands amplified in both tumor and normal tissue DNA were classified as germline; no band or multiple bands were classified as negative (Supplementary Figure S2).

Sanger sequencing and identification of breakpoints

PCR products of somatic events were purified using AMPure XP (Agencourt, New South Wales, Australia) using a bead:DNA volume ratio of 1.8:1. The products were then quantified using Qubit Flurometer (Invitrogen, Victoria, Australia). Each amplicon was individually prepared for sequencing using forward and reverse primers to increase the chance of identifying the breakpoints. Sequences of verified somatic rearrangements were aligned to the reference genome using BLAT (27) to resolve the breakpoint sequences to base pair level.

Next-generation sequencing and identification of breakpoints

Somatic PCR products ranging up to 3 kb were pooled in equal volumes and purified using AMPure XP bead clean-up at a bead:DNA volume ratio of 1.8:1. Pools of PCR products were quantified using a Qubit Flurometer (Invitrogen). Pooled amplicons were sequenced using the PGM (Life Technologies) and MiSeq (Illumina, San Diego, CA) platforms.

For PGM sequencing, the pool of PCR products was sheared using Ion Shear Plus as per the manufacturer's instructions, then analyzed on an Agilent Bioanalyzer DNA High-Sensitivity LabChip to verify size and purity. Ion libraries were prepared using 50–100 ng of sheared material with the Ion Xpress Plus

Table 1. Summary of designed primers and verification rates for a pancreatic cancer genome using qAmplicon and PCR analysis.

SV	# of potential events	# of long* primers	# of short* primers	# of overlapped primers	# of verified events—long primers (verification rate)	# of verified events—short primers (verification rate)	# of verified events—overall (verification rate)
Deletion	41	41	38	38	17 (39.0%)	19 (50.0%)	19 (46.3%)
Tandem dup	5	5	5	5	1 (20.0%)	1 (20.0%)	1 (20.0%)
Intra-chr	243	226	220	205	147 (65.0%)	166 (75.5%)	183 (75.3%)
Inversion	23	22	21	20	9 (40.9%)	10 (47.6%)	11 (47.8%)
Translocation	46	46	43	43	33 (71.7%)	30 (69.8%)	34 (73.9%)
Total (average rate)	358	340 (95.0%)	327 (91.3%)	311 (86.9%)	207 (60.6%)	226 (69.1%)	248 (69.3%)

* Long primers are suitable for next-generation sequencing and allow flexibility in PCR product size; short primers are suitable for Sanger sequencing restricted to a PCR product size < 1kb.

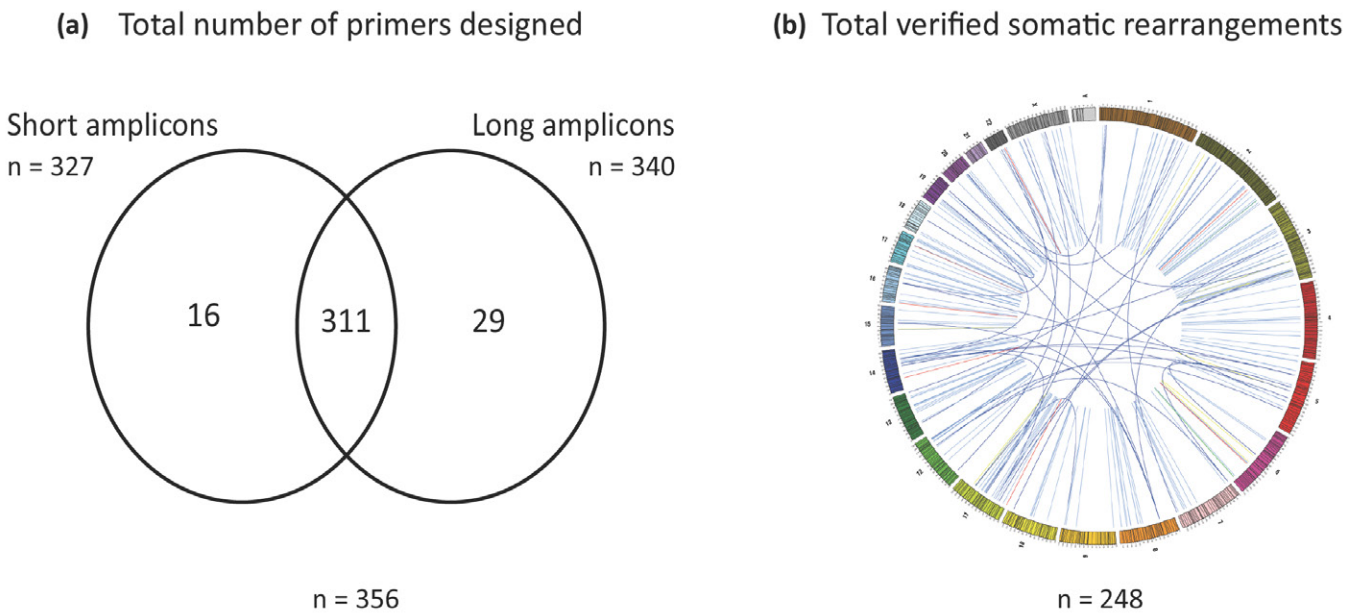

Figure 2. Verification of somatic structural rearrangements from a highly rearranged cancer genome. A total of 356 candidate events had primers designed. (a) A Venn diagram shows the number of primers designed for Sanger sequencing and benchtop next-generation sequencing. (b) A circos image displaying the 248 somatic rearrangements that were verified by Sanger sequencing and/or next-generation sequencing methods. The chromosomes are displayed in the outer ring and the structural rearrangements are shown by the inner connecting lines: green = deletion, red = tandem duplication, orange = inversion, light blue = intra-chromosomal rearrangement, dark blue = translocation.

Fragment Library kit and the AB Library Builder System as per the manufacturer's instructions. Emulsion PCR, emulsion breaking, and enrichment were performed using the Ion OneTouch System to obtain Ion Sphere Particles (ISPs). Enriched ISPs were loaded on an Ion 318 chip on the PGM machine for 200 bp single-read sequencing.

For MiSeq sequencing, the pool of PCR products (1 ng of DNA) was prepared using the Nextera XT kit. The DNA was simultaneously fragmented and tagged with adaptors using a tagmentation enzymatic reaction, followed by PCR amplification. Sequencing was performed on the MiSeq system with paired-end 2 × 150 bp reads.

After sequencing, the short reads were assembled to create consensus contigs using an in-house de novo assembly tool (qNovo4.pl; Supplementary File 1), which uses an overlap consensus method. However, any other assembly tool, such as EBARDenovo (16), Oases (17), or Trinity (18), could be used in this workflow. The first step of the qNovo4 assembly trimmed the reads based on a user-defined Phred score and ambiguous base (N) content. The reads were then filtered by nucleotide complexity using user-defined Shannon's entropy for mono-, di-, and tri-nucleotide content. Reads that did not pass the filters or reads that were shorter than the user-defined minimum after trimming were not used in the assembly. New contigs generated by sequence artifacts were avoided by first collapsing all reads that were within Hamming distance based on the user-defined mismatch rate. The construction of each contig began with a randomly selected read. All reads with a minimum overlap length and less than the maximum Hamming distance were aligned to construct a consensus sequence. The number of reads (based on user defined threshold) used to construct the consensus sequence justified the contig extension. Extension occurs from both the 5′ and 3′ ends of the contig. The built contig was used as a seed for another round of extension, until there were no reads justifying an extension. Reads used to generate the contig were removed from the memory and were not available to form other contigs. qNovo4 reiterated these steps for every read in the memory. The assembled contigs generated in this workflow used the following parameters: minimum length of the tags = 30 bp, mismatch rate = 3, and length of seed = 20 bp. Resultant contigs were automatically aligned to the human reference genome (GRCh37/hg19) using a local installation BLAT. Aligned results were searched for matching blocks with the correct orientation in the predicted genomic region. The identification of breakpoints was based on the highest quality BLAT match that was selected to maximize the length of the contig and the score of matched bases.

Results and discussion

The aim of this study was to develop a workflow of standard techniques to increase the throughput process for verifying structural rearrangement events and identifying breakpoint locations (Figure 1). To evaluate the workflow, we sequenced DNA from a tumor and matched normal tissue, identifying a total of 358 potential somatic rearrangements in this cancer genome using a threshold of at least 10 read pairs supporting an event. Of these, 41 (11.5%) were deletions, 5 (1.4%) tandem duplications, 243 (67.9%) intra-chromosomal rearrangements, 23 (6.4%) inversions, and 46 (12.9%) translocations (Table 1).

We were able to design primers to verify 327/358 (91.3%) of the candidate

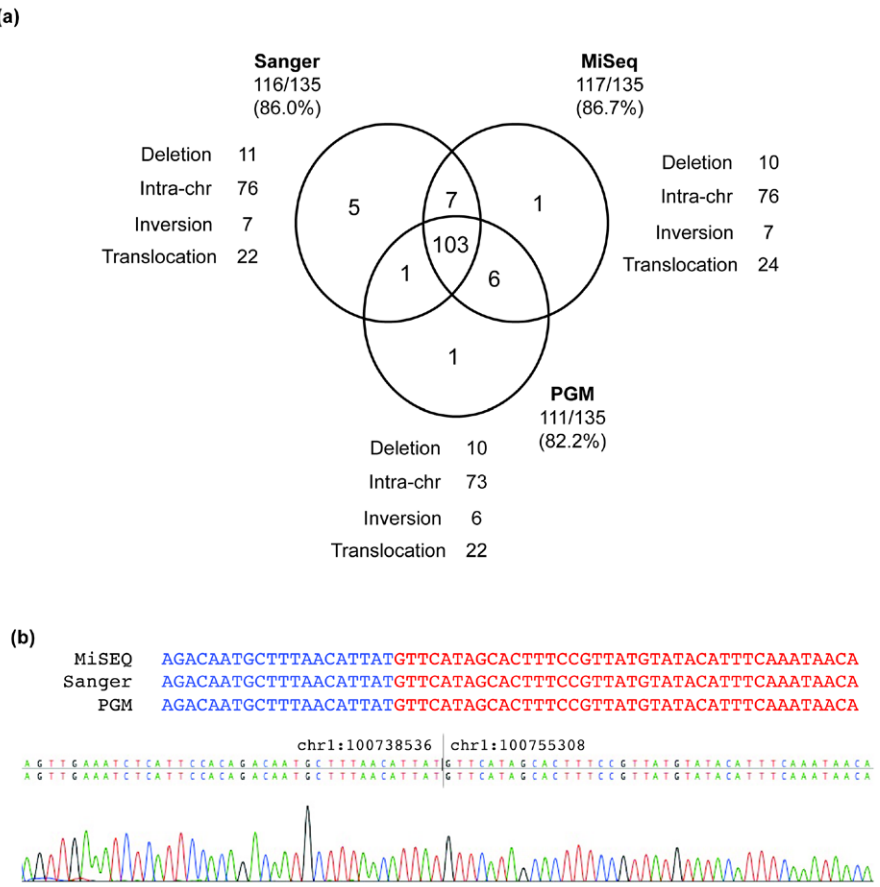

Figure 3. The ability of different sequencing approaches to resolve the sequences of breakpoints. A total of 135 events were tested using Sanger and next-generation sequencing platforms (PGM and MiSeq). In all cases, the short amplicon (~1 kb PCR product size) were used. (a) Number of identified breakpoints across the three sequencing platforms. Venn diagram shows the respective number of identified breakpoints from the individual platform and the overlap breakpoints. (b) Representative breakpoint sequence from an intrachromosomal rearrangement is confirmed by PCR and sequencing showing 100% identity when aligning sequences from the three platforms. The identified breakpoint is chr1:100738536-100755308 with a resolution of 16,773 bp.

rearrangement events by Sanger sequencing. However, due to the low complexity of DNA sequences near the breakpoints (19), qAmplicon was not able to generate primer pairs for all predicted events. Similar problems were also encountered when primers pairs were designed manually using Primer 3 without the filtering performed by qAmplicon. In contrast, the advantage of amplicon sequencing with a next-generation sequencer (PGM and MiSeq), is that unlike Sanger sequencing where primer design is limited to an amplicon size of <1 kb (short amplicons), there is no limitation on amplicon size. This is because the amplicons are fragmented, sequenced, and assembled prior to aligning to the reference genome. Therefore, by allowing PCR products of up to 3 kb, an extra set of 29 primers could be designed (classified as long amplicons), allowing more events to be tested (Figure 2a, Table 1). Altogether, qAmplicon generated a total of 356 (99.4%) primers from 358 potential somatic rearrangements (Supplementary Table S1). There were 311/356 designed primers shared between the short amplicon and long amplicon categories (Figure 2a).

The verification rate of the Sanger primers (short amplicons) was assessed by PCR and gel electrophoresis. Of the 327 primers, a total of 226 (69.1%) events were confirmed as somatic, 13 were confirmed as germline (4.0%), and 88 (27.0%) events were negative. However, an extra 22 events were confirmed as somatic using next-generation sequencing primers (long amplicons). Collectively, a total of 248 (69.0%) somatic events were verified, which included 19 deletions, 1 tandem duplication, 183 intra-chromosomal rearrangements, 11 inversions, and 34 translocations (Figure 2b, Table 1). Interestingly, the PCR results revealed that 62 of the events tested with both long and short amplicon primers yielded different results (Supplementary Figure S3). This suggests that primer design is crucial, and a negative result may not necessarily indicate that an event is false; thus, some PCRs may require individual optimization in order to verify an event.

The sequences surrounding the breakpoints of somatic rearrangements can reveal insights into their formation by identifying potential mechanisms of DNA damage repair (20). Therefore, we compared the ability of the conventional Sanger and high-throughput next-generation sequencing workflow to resolve the sequences of breakpoints using a subset of 135 events. Due to the limitations of Sanger sequencing, only the short amplicons were used to generate the amplicons for sequencing with each platform. The assembly generated a total of 5906 contigs from PGM sequencing data (range 101–1289 bp) and 7866 contigs from MiSeq sequencing data (range 101–6507 bp). Of these, a total of 124 (91.9%) breakpoints were identified by at least 1 sequencing method. Sanger sequencing identified 116 (86.0%), MiSeq sequencing identified 117 (86.7%), and PGM sequencing identified 111 (82.2%) (Figure 3a). Overall, 103 (76.3%) breakpoints were identified by all 3 methods.

To determine the accuracy of the benchtop next-generation sequencing platforms (PGM and MiSeq) to resolve the breakpoints, we compared the breakpoint sequences of 103 events that were identified by both benchtop next-generation sequencers to Sanger sequencing (Figure 3b). The comparison showed that MiSeq sequencing accurately identified 100 of 103 breakpoints (97.1%), while PGM sequencing identified 95 breakpoints (92.2%). The reason the exact locations for some breakpoints were not identified is likely due to the gapped alignment of the contigs with homopolymers (Supplementary Figure S4). In this context, the

assembled contigs obtained by MiSeq sequencing were slightly more accurate than those obtaing by PGM sequencing, possibly due to the paired-end information and fewer sequence errors around homopolymeric sequences. The observed performance of the two benchtop next-generation sequencing platforms is consistent with previous studies (21,22). However, MiSeq and PGM identified eight additional breakpoints that were not identified by Sanger sequencing (Figure 3a). These results showed that, with respect to the number of identified breakpoint sequences, the high-throughput workflow using MiSeq and PGM sequencing methods showed high accuracy and was comparable to conventional Sanger sequencing.

In conclusion, we have described a high-throughput workflow for verifying somatic structural rearrangements in cancer genomes. The advantages of this high-throughput workflow include (*i*) the utilization of benchtop next-generation sequencing (Ion Torrent PGM and Illumina MiSeq) to replace conventional Sanger sequencing, which mitigates the concern of PCR product size, allowing primers to be designed for more events; (*ii*) the integration of bioinformatics tools and next-generation sequencing based methods greatly increases the speed and volume of the verification process; and (*iii*) the accuracy of the next-generation sequencing methods is comparable to that of the conventional Sanger sequencing method. Although the results presented here were based on the detection of structural rearrangements from SOLiD sequencing data, this workflow is also applicable to Illumina HiSeq data as the qSV tool has the ability to identity structural rearrangements in both platforms (Supplementary Table S2).

We expect that this workflow will enable routine verification of somatic rearrangements and breakpoints, enhancing large-scale screening projects. This workflow can be useful in characterizing rearrangement breakpoints that may lead to disrupted repair pathways and could help prioritize treatment options for patients in future personalized medicine applications (23). The rapid verification of somatic events can also allow the identification of tumor-specific breakpoints as potential candidate biomarkers in cancer patients to assess disease progression in real time (24–26).

Author Contributions

All authors contributed to the work: K.Q., K.N., and N.W. wrote the manuscript; K.Q., K.N., and N.W. conceived the experiments; A.M.P., J.L.F., F.N., N.C., and K.K. contributed to the qAmplicon software; K.Q., D.M., M.F., A.C., T.B., S.M., I.H., S.I., C.N., E.N., S.W., and A.S. performed the laboratory experiments and next-generation sequencing; M.A., O.H., C.L., D.T., S.W., Q.X., and J.V.P. performed variant calling and data management of the next-generation sequence data; the APGI and A.V.B. contributed tumour samples; K.N., P.W., N.W., and S.M.G. oversaw the work.

Acknowledgments

We would like to thank D. Gwynne for central coordination at the Queensland Centre for Medical Genomics. We wish to thank all of the APGI members as well as patients who contribute to APGI. This research has been supported by the National Health and Medical Research Council of Australia (NHMRC: 631701, 535903, 427601, APP1047334); Australian Government: Department of Innovation, Industry, Science and Research (DIISR); Australian Cancer Research Foundation (ACRF); Queensland Government (NI-RAP); University of Queensland; Cancer Council NSW: (SRP06-01); Cancer Institute NSW: (10/ECF/2-26; 06/ECF/1-24; 09/CDF/2-40; 07/CDF/1-03; 10/CRF/1-01, 08/RSA/1-15, 07/CDF/1-28, 10/CDF/2-26,10/FRL/2-03, 06/RSA/1-05, 09/RIG/1-02, 10/TPG/1-04, 11/REG/1-10, 11/CDF/3-26); Garvan Institute of Medical Research; Avner Nahmani Pancreatic Cancer Research Foundation; R.T. Hall Trust; Petre Foundation; Philip Hemstritch Foundation; Gastroenterological Society of Australia (GESA); American Association for Cancer Research (AACR) Landon Foundation INNOVATOR Award; Royal Australasian College of Surgeons (RACS); Royal Australasian College of Physicians (RACP); Royal College of Pathologists of Australasia (RCPA). S.G. is a recipient of a NHMRC Principal Research Fellowship.

Competing Interests

The authors declare no completing interests.

References

1. **Stratton, M.R., P.J. Campbell, and P.A. Futreal.** 2009. The cancer genome. Nature *458*:719-724.
2. **Negrini, S., V.G. Gorgoulis, and T.D. Halazonetis.** 2010. Genomic instability–an evolving hallmark of cancer. Nat. Rev. Mol. Cell Biol. *11*:220-228.
3. **Leary, R.J., J.C. Lin, J. Cummins, S. Boca, L.D. Wood, D.W. Parsons, and S. Jones. T. Sjöblom, et al.** 2008. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci U S A. *105*:16224-16229.
4. **Nowell, P.C. and D.A. Hungerford.** 1960. Minute chromosome in human chronic granulocytic leukemia. Science *142*:1497.
5. **Sadikovic, B., K. Al-Romaih, J.A. Squire, and M. Zielenska.** 2008. Cause and consequences of genetic and epigenetic alterations in human cancer. Curr. Genomics *9*:394-408.
6. **Xing, E.P., G.Y. Yang, L.D. Wang, S.T. Shi, and C.S. Yang.** 1999. Loss of heterozygosity of the Rb gene correlates with pRb protein expression and associates with p53 alteration in human esophageal cancer. Clin Cancer Res. *5*:1231-1240.
7. **Chen, K., J.W. Wallis, M.D. McLellan, D.E. Larson, J.M. Kalicki, C.S. Pohl, S.D. McGrath, M.C. Wendl, et al.** 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods *6*:677-681.
8. 8. **Ye, K., M.H. Schulz, Q. Long, R. Apweiler, and Z. Ning.** 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*:2865-2871.
9. **Wang, J., C.G. Mullighan, J. Easton, S. Roberts, S.L. Heatley, J. Ma, M.C. Rusch, K. Chen, et al.** 2011. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat. Methods *8*:652-654.
10. **The International Cancer Genome Consortium.** 2010. International network of cancer genome projects. Nature *464*:993-998.
11. **Cancer Genome Atlas Research Network.** 2011. Integrated genomic analyses of ovarian carcinoma. Nature *474*:609-615.
12. **Campbell, P.J., P.J. Stephens, E.D. Pleasance, S. O'Meara, H. Li, T. Santarius, L.A. Stebbings, C. Leroy, et al.** 2008. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat. Genet. *40*:722-729.
13. **Hillmer, A.M., F. Yao, K. Inaki, W.H. Lee, P.N. Ariyaratne, A.S.M. Teo, X.Y. Woo, Z. Zhang, et al.** 2011. Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. Genome Res. *21*:665-675.
14. **Stephens, P.J., D.J. McBride, M.-L. Lin, I. Varela, E.D. Pleasance, J.T. Simpson, L.A. Stebbings, C. Leroy, et al.** 2009. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature *462*:1005-1010.
15. **Rozen, S. and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. *132*:365-386.
16. **Chu, H.T., W.W. Hsiao, J.C. Chen, T.J. Yeh, M.H. Tsai, H. Lin, Y.W. Liu, S.A. Lee, et al.** 2013. EBARDenovo: highly accurate de novo assembly of RNA-Seq with efficient chimera-detection. Bioinformatics *29*:1004-1010.
17. **Schulz, M.H., D.R. Zerbino, M. Vingron, and E. Birney.** 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics *28*:1086-1092.
18. **Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, et al.** 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*:644-652.
19. **Wei, Y., M. Sun, G. Nilsson, T. Dwight, Y. Xie, J. Wang, Y. Hou, O. Larsson, et al.** 2003. Characteristic sequence motifs located at the genomic breakpoints of the translocation t(X;18) in synovial sarcomas. Oncogene *22*:2215-2222.
20. **Onishi-Seebacher, M. and J.O. Korbel.** 2011. Challenges in studying genomic structural variant formation mechanisms: the short-read dilemma and beyond. Bioessays. *33*:840-850.
21. **Loman, N.J., R.V. Misra, T.J. Dallman, C. Constantinidou, S.E. Gharbia, J. Wain, and M.J. Pallen.** 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. *30*:434-439.
22. **Quail, M.A., M. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, A. Bertoni, H.P. Swerdlow, and Y. Gu.** 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics *13*:341.
23. **Stephens, P.J., C.D. Greenman, B.Y. Fu, F.T. Yang, G.R. Bignell, L.J. Mudie, E.D. Pleasance, K.W. Lau, et al.** 2011. Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. Cell. *144*:27-40.
24. **Diehl, F., K. Schmidt, M.A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, et al.** 2008. Circulating mutant DNA to assess tumor dynamics. Nat. Med. *14*:985-990.
25. **Leary, R.J., I. Kinde, F. Diehl, K. Schmidt, C. Clouser, C. Duncan, A. Antipova, C. Lee, et al.** 2010. Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing. Sci. Transl. Med. *2*:20ra14.
26. **McBride, D.J., A.K. Orpana, C. Sotiriou, H. Joensuu, P.J. Stephens, L.J. Mudie, E. Hamalainen, L.A. Stebbings, et al.** 2010. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. Genes Chromosomes Cancer *49*:1062-1069.
27. **Kent, W.J.** 2002. BLAT--the BLAST-like alignment tool. Genome Res. *12*:656-664

Received 22 April 2014; accepted 12 June 2014.

Address correspondence to Nic Waddell or Sean M. Grimmond, Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, QLD, Australia. E-mail: n.waddell@imb.uq.edu.au; s.grimmond@imb.uq.edu.au or Sean.Grimmond@glasgow.ac.uk