Next-generation sequencing of multiple individuals per barcoded library by deconvolution of sequenced amplicons using endonuclease fragment analysis

Jeppe D. Andersen, Vania Pereira, Carlotta Pietroni, Martin Mikkelsen, Peter Johansen, Claus Børsting, and Niels Morling
Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

BioTechniques 57:91-94 (August 2014) doi 10.2144/000114200

Keywords: next-generation sequencing; restriction endonucleases; DNA library preparation; amplicons; melanocortin 1 receptor (MC1R)

Supplementary material for this article is available at www.BioTechniques.com/article/114200.

The simultaneous sequencing of samples from multiple individuals increases the efficiency of next-generation sequencing (NGS) while also reducing costs. Here we describe a novel and simple approach for sequencing DNA from multiple individuals per barcode. Our strategy relies on the endonuclease digestion of PCR amplicons prior to library preparation, creating a specific fragment pattern for each individual that can be resolved after sequencing. By using both barcodes and restriction fragment patterns, we demonstrate the ability to sequence the human melanocortin 1 receptor (*MC1R*) genes from 72 individuals using only 24 barcoded libraries.

Current next-generation sequencing (NGS) platforms generate thousands of millions of reads per investigation. Often, the number of sequences obtained exceeds the required sequencing depth for a given sample. This is especially true if small genomes or small target regions are sequenced. Despite considerable effort, the sequencing cost per NGS run remains relatively high. Therefore, simultaneous sequencing of many samples is often preferred.

Multiplexing with short, identifying sequences (barcodes) is a widely used strategy in which different barcodes can be used to identify different samples. Fragments of DNA are first ligated to adapters with barcodes, followed by emulsion or bridge PCR depending on the platform used. Sequencing is carried out using the pooled barcoded libraries. Barcoding allows for the post-sequencing sorting of data and the assignment of sequencing reads to specific samples. However, this also means the maximum

number of samples per sequencing run is, therefore, determined by the number of barcodes available. Yet another multiplexing strategy involves the sequencing of dissimilar DNA targets that can be easily distinguished using bioinformatics tools. In this way, multiple samples with the same barcodes can be sequenced, and the results can still be assigned to the correct sample.

Here we present a novel next-generation sequencing strategy using endonuclease digestion of PCR products prior to library preparation that allows post-sequencing sorting and assignment of sequences to the correct individual through a combination of (i) barcodes and (ii) restriction fragment patterns (Figure 1).

To demonstrate the effectiveness of our approach, we sequenced the human melanocortin 1 receptor (*MC1R*) gene from 72 individuals using only 24 barcoded libraries. *MC1R* is a pigmentation gene of 954 bp consisting of only a single exon (chr16: 89,985,667–89,986,620). It is a major determinant of the human red hair color phenotype (red hair and pale skin color) and is also associated with the risk of malignant melanoma (1). A region encompassing both the *MC1R* exon and the gene promoter was amplified in a single PCR, resulting in an amplicon of 1963 bp (Supplementary Material). Each amplicon sample was digested with one of three mixtures of endonuclease enzymes (mixes A, B, and C) after PCR (Table 1). Prior to library preparation, the three samples digested with each different endonuclease mix were pooled in equal amounts. The endonucleases in each mix were selected using NEBcutter (2), with the following criteria: (i) the endonuclease combination should cut the amplicon into fragments with different lengths such that each fragment can be separated from the fragments generated by the other endonuclease mixtures; and (ii) the fragment lengths should be shorter than the read lengths of the sequencing equipment used. By taking alignment position into consideration, it is possible to separate sequences with the same length. For

METHOD SUMMARY

Here we present a novel strategy to further enhance barcoded library output using endonuclease digestion of the PCR products prior to next-generation sequencing library construction. The approach enables post-sequencing sorting and assignment of sequences to the correct individual through a combination of both (i) barcodes and (ii) restriction fragment digest patterns.

Figure 1: Flowchart of method allowing multiplexing of samples per barcoded library. After PCR, the samples were treated with either endonuclease mix A, B, or C. Samples were pooled such that each pool contained a sample treated with endonuclease mix A, B, and C. The pooled samples were prepared for sequencing using barcoded libraries. Post-sequencing, barcode-separated pairs of FASTQ files were aligned against the human melanocortin 1 receptor gene (*MC1R*) reference sequence. Each BAM file was divided into three BAM files based on the lengths and alignment positions of the expected fragments.

example, endonuclease mixes A and C both produce a fragment of length 384 bp, but because the alignment positions are different for each fragment, these can be resolved. Libraries were prepared for sequencing with the TruSeq DNA Sample Preparation kit (Illumina, San Diego, CA) and paired-end sequenced (2 x 250 cycles) on the Illumina MiSeq platform

using the Illumina MiSeq Reagent Nano Kit v2 (500 cycles). Barcode-separated pairs of FASTQ files were obtained from the Illumina MiSeq Reporter. Illumina adaptors were trimmed using Flexbar (3). The FASTQ files were aligned to the *MC1R* reference sequence assembly Feb.2009 GRCh37/hg19 (UCSC Genome Browser; <http://genome.ucsc.edu>) with

the Burrows-Wheeler Aligner (BWA) algorithm (4) to generate BAM files. In total, 1,529,670 sequences passed the Illumina software quality filter. Of these, 25,790 sequences (1.7%) were discarded due to filtering criteria in adaptor removal and alignment to the reference sequence. Using a custom designed Python script (Supplementary Material), each BAM file

horizon
precision genome editing

GENASSIST™ A range of gene-editing kits and reagents to enable easier, robust implementation of CRISPR and rAAV gene-editing experiments.

- ✓ Custom guide RNA manufacturing **and Validation**
- ✓ Custom **Donor Design** and manufacturing
- ✓ Both **CRISPR and rAAV** technologies for the right solution to your challenge
- ✓ **Adenovirus and rAAV** packaging for hard to transfect cells
- ✓ Consultation with our **Gene-Editing Expert Scientists**
- ✓ Over 550 genetically defined **isogenic Cell Lines** available for further genomic modification
- ✓ **And** an extensive plasmid library!

Visit www.horizondiscovery.com/GENASSIST to learn more about GENASSIST™ and Horizon's many other products and services

GENASSIST™
PRECISION GENOME EDITING TOOLS

FREE CRISPR Reagents
for Knockout Generation

Join the Program Now
and Learn More at:

horizondiscovery.com/GENASSIST

Fighting to reduce small RNA-seq bias

The NEXTflex™ Small RNA Sequencing Kit v2 provides an easy, flexible, cost-effective solution for generating single and multiplexed libraries from small RNA with reduced bias using Illumina sequencing platforms. Bioo Scientific's patent pending approach to overcoming ligation bias in small RNA-seq libraries involves using adapters containing randomized sequences at the ligation site.

INCREASED ACCURACY

Randomized adapters reduce ligation bias

GREATER SEQUENCING DEPTH

AIR™ Ligase increases ligation efficiency

MULTIPLEXING

Up to 48 unique barcodes

FLEXIBLE

Compatible with RIP-Seq and CLIP-Seq

Visit BiooNGS.com to learn more.

THE NGS EXPERTS™

Amplicon-Seq • Rapid DNA-Seq • Pre-Target Capture
Bisulfite-Seq • Methyl-Seq • Rapid RNA-Seq
Directional RNA-Seq • Small RNA-Seq
Rapid Directional RNA-Seq • PCR-Free DNA-Seq
ChIP-Seq • qRNA-Seq • Cell Free DNA-Seq
Multiple Platform Compatibility

Table 1: Restriction endonuclease enzyme mixes and DNA fragment sizes.

	Combination of restriction enzymes	Size of DNA fragments
Endonuclease mix A	<i>RsaI</i> and <i>StyI</i>	468, 384, 325, 399, and 395
Endonuclease mix B	<i>Tsp45I</i> , <i>TfiI</i> , and <i>BamHI</i>	373, 338, 263, 457, 260, and 294
Endonuclease mix C	<i>HpyCH4III</i> , <i>PfiFI</i> , and <i>PfiMI</i>	383, 228, 284, 404, 275, and 384

was divided into three BAM files based on the lengths and alignment positions of the expected restriction fragments.

All 72 samples were successfully sequenced using the 24 barcoded libraries. A total of 150 SNPs were identified, ranging from 0 to 7 SNPs per individual. This strategy decreased the library preparation cost by two-thirds per sample. The median sequence coverage ranged from 2641 to 4555 for samples treated with endonuclease mix A, 1682 to 2518 for samples treated with endonuclease mix B, and 208 to 539 for samples treated with endonuclease mix C. The low coverage of endonuclease mix C treated samples was a result of lower endonuclease activity. Incomplete digestion of the PCR amplicons was observed with mix C, resulting in fragments of longer, unpredicted lengths. Clearly, this demonstrates the importance of selecting endonucleases such that the activity of each enzyme is preserved when used in combination with other enzymes.

Our study identifies a new multiplexing approach for NGS that can significantly increase the number of individuals per barcoded library. We provide a Python script for sorting aligned sequences based on length and alignment position. The crucial step in this approach is identifying combinations of endonucleases that create different fragments from the amplicons. For the *MC1R* amplicon, we found three combinations of endonucleases that gave different fragment patterns, allowing three samples to be sequenced per barcode. However, for other sequencing experiments, it may be possible to multiplex even more individuals. In fact, the maximum number of individuals that can be sequenced per library is limited only by the number of endonuclease mixtures that can give rise to different fragment patterns. This new approach has a wide range of possible applications, including sequencing of human mitochondria DNA that can be PCR amplified in a few amplicons (5).

Author contributions

J.D.A. wrote the manuscript and analyzed the data. J.D.A., C.B., C.P., and V.P. designed and carried out the experiments. C.B., P.J., M.M., and N.M. had the original idea for the study. J.D.A., V.P., C.P., F.K., C.B., P.J., M.M., and N.M. edited the manuscript. C.B. and N.M. supervised the study.

Acknowledgments

Technical assistance was provided by Anja Jørgensen and bioinformatic assistance was provided by Francisc Kantor. A part of this project was supported by Ellen and Aage Andersen's foundation.

Competing interests

The authors declare no competing interests.

References

1. Raimondi, S., F. Sera, S. Gandini, S. Iodice, S. Caini, P. Maisonneuve, and M.C. Fargnoli. 2008. MC1R variants, melanoma and red hair color phenotype: a meta-analysis. *Int. J. Cancer* 122:2753-2760.
2. Vincze, T., J. Posfai, and R.J. Roberts. 2003. NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* 31:3688-3691.
3. Dodt, M., J.T. Roehr, R. Ahmed, and C. Dieterich. 2012. FLEXBAR--Flexible barcode and adapter processing for next-generation sequencing platforms. *Biology (Basel)* 1:895-905.
4. Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
5. Mikkelsen, M., R. Frank-Hansen, A.J. Hansen, and N. Morling. 2014. Massively parallel pyrosequencing of the mitochondrial genome with the 454 methodology in forensic genetics. *Forensic Sci. Int. Genet.* 12C:30-37.

Received 25 June 2014; accepted 14 July 2014.

Address correspondence to Jeppe D. Andersen, Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. E-mail: jeppe.dyrberg.andersen@sund.ku.dk

To purchase reprints of this article, contact: biotechniques@fosterprinting.com