

RESEARCH HIGHLIGHT

Improved coverage and accuracy with strand-conserving sequence enrichment

Toumy Guettouche and Stephan Zuchner*

See related Research: <http://genomemedicine.com/content/5/5/50>

Abstract

Targeted next-generation sequencing is becoming a common tool in the molecular diagnostic laboratory. However, currently available methods to enrich for regions of interest in the DNA sequence suffer from drawbacks such as high cost, complex protocols, lack of clinical-level accuracy and uneven target coverage. A target-enrichment approach using complementary long padlock probes described in a recent article significantly improves on previous methods in most of these areas.

From whole-genome sequencing to target capture

In the almost 13 years since the first whole human genome was sequenced and published [1,2], tremendous advances in technology have enabled the sequencing of human genomes for a fraction of the cost and time. However, although the cost of sequencing has dropped considerably, large-scale whole-genome sequencing remains challenging, particularly in the clinical arena. This is due to the still significant cost of sequencing an entire human genome, and the challenges of analyzing enormous amounts of data with tools that are not standardized to a level acceptable for routine diagnostic use. Consequently, targeted sequencing approaches may be more suitable for clinically actionable genes.

Cheap and high-quality targeted sequencing is key for a number of clinical research applications, including large-scale variant screening in disease genes or as follow-up for genetic markers identified as significant in genome-wide association studies. Various methods have been developed to enable whole-exome sequencing and targeted-region sequencing. Early on, solid-state capture arrays were used, but these were expensive and had

relatively complex protocols [3]. In-solution capture and PCR-based enrichment methods have reduced the cost and complexity of protocols considerably [4]. These improvements led to a wider adoption of next-generation sequencing and, in the past 12 months particularly, an increase in the use of targeted resequencing as a diagnostic tool [5].

Nevertheless, current methods are far from perfect. For example, PCR-based methods require highly multiplexed oligonucleotide pairs targeted to heterogeneous sequences with a range of melting temperatures and CG content to generate hundreds or thousands of amplicons in a single tube. This leads to differences in amplicon presentation and uneven sequence coverage. Hybridization-based methods exhibit significantly more off-target capture than other enrichment methods, do not capture repetitive sequences, and poorly cover GC- and AT-rich regions. Methods employing 'capture by circularization' (Figure 1), such as connector inversion probes (CIPs), also have problems. These methods use single-stranded DNA molecules with gene-specific targeting regions at the 5' and 3' ends that are complementary to the targeted genomic DNA [6]. After hybridization of the targeting ends of the CIP to the genomic DNA, a single-stranded DNA circle is formed and closed by gap filling and ligation. The single-stranded DNA circle is then linearized by restriction digest, and the target region is enriched by PCR and finally sequenced. CIPs require a large backbone for the probes to capture targets efficiently, which makes them expensive and difficult to manufacture [7].

The size of a target region is limited to a few megabases, which restricts the number of genes/exons that can be included in a clinical sequencing panel. In addition, all current capture methods use only one strand of genomic DNA, missing out on an additional level of possible accuracy.

Overcoming current limitations in target enrichment

By contrast with standard capture methods, the complementary long padlock probe (cLPP) approach, as

*Correspondence: szuchner@med.miami.edu

Dr John T MacDonald Department of Human Genetics and Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, FL 33136, USA

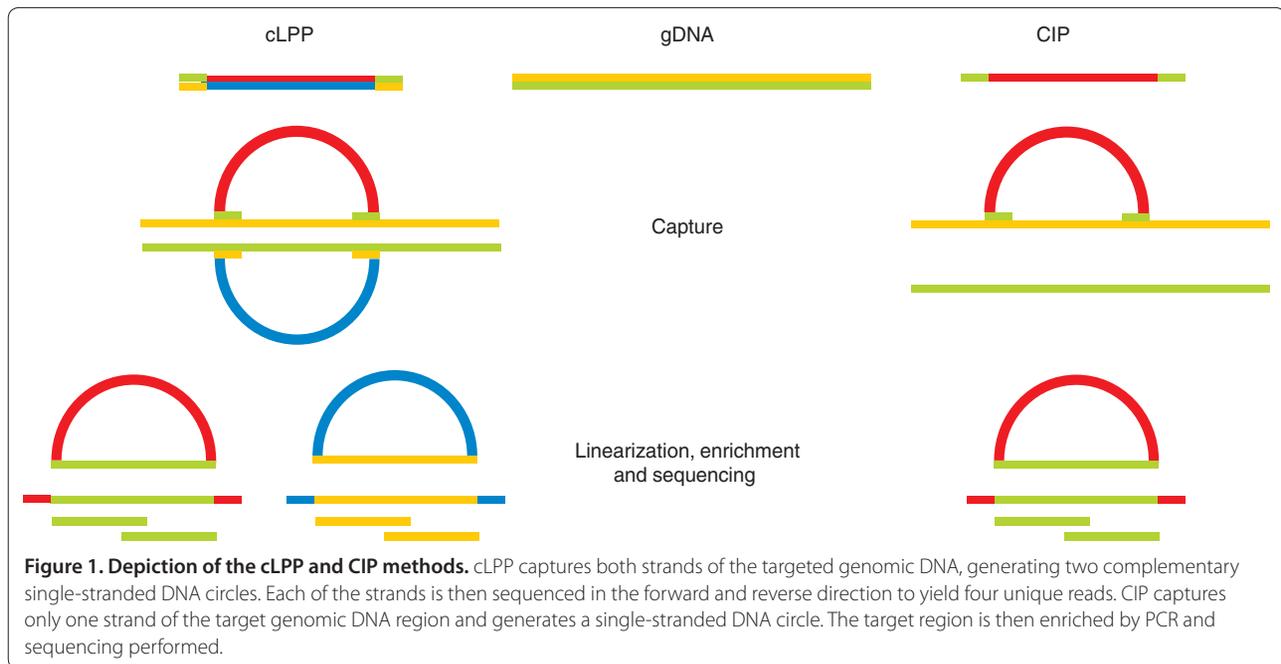


Figure 1. Depiction of the cLPP and CIP methods. cLPP captures both strands of the targeted genomic DNA, generating two complementary single-stranded DNA circles. Each of the strands is then sequenced in the forward and reverse direction to yield four unique reads. CIP captures only one strand of the target genomic DNA region and generates a single-stranded DNA circle. The target region is then enriched by PCR and sequencing performed.

presented by Shen *et al.* in a recent article [8], captures both strands of the target region, effectively doubling the target sequence information compared with other capture methods. This is achieved by generating double-stranded CIPs that are incubated at high temperatures to create single DNA strands, and then hybridized to the sense and antisense strands of genomic DNA, effectively forming two complementary single-stranded DNA circles. In addition, cLPP enables the sequencing of both strands in both the forward and reverse direction (Shen *et al.* call this reciprocal paired-end sequencing), resulting in a total of four unique sequence reads per template. This redundancy reduces uneven coverage due to differences in the amplification efficiencies of the target regions, and increases coverage and accuracy. This should lead to increased confidence in variant calls in the downstream bioinformatics analysis, and might allow for a reduced average depth of sequence coverage resulting in less sequencing per sample - thus lowering cost. Shen *et al.* also demonstrate that copy number variation (CNV) detection can be improved with this enrichment method owing to its significantly better discrimination of high- and low-covered targets.

An additional interesting potential application for cLPP is the targeted resequencing of problematic DNA samples derived from formalin-fixed paraffin-embedded (FFPE) tissues. DNA extracted from FFPE samples frequently contains lesions such as abasic sites that lead to a significant increase in sequencing errors when using traditional single-strand sequence capture methods [9]. Owing to the ability of cLPP to capture both strands, it

could become a compelling option for targeted resequencing of these sample types. Although cLPP appears to be better suited than traditional CIPs for clinical use, both methods require a large sample size to be economical because of the initial cost of assay development. Furthermore, to our knowledge, reagents based on cLPP are not yet commercially available, which poses a challenge to its widespread adoption.

Conclusion

cLPP is an innovative new approach for high-throughput target enrichment for next-generation sequencing. It improves on a number of shortcomings of current targeted sequencing methods such as accuracy, CNV detection and cost. Most compelling is its ability to preserve strand information and separately sequence sense and antisense strands. Beyond the resulting improvement of variant detection fidelity, other applications that rely on double-strand targeting could benefit. Such applications include problematic DNA samples, where redundancy is important to retrieve as much information as possible because of damage to a single DNA strand.

List of abbreviations

CIPs, connector inversion probes; cLPP, complementary long padlock probes; CNV, copy number variation; FFPE, formalin-fixed paraffin-embedded.

Competing interests

The authors declare that they have no competing interests.

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **6822**:860-921.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, *et al.*: **The sequence of the human genome.** *Science* 2001, **550**: 1304-1351.
3. Rowida A, van der Heijden J, Ariyurek Y, Lai Y, Bakker E, van Galen M, Breuning MH, den Dunnen J: **Experiences with array-based sequence capture; toward clinical applications.** *Eur J Hum Genet* 2011, **19**:50-55.
4. Hedges DJ, Guettouche T, Yang S, Bademci G, Diaz A, Andersen A, Hulme WF, Linker S, Mehta A, Edwards YJ, Beecham GW, Martin ER, Pericak-Vance MA, Zuchner S, Vance JM, Gilbert JR: **Comparison of three targeted enrichment strategies on the SOLiD sequencing platform.** *PLoS One* 2011, **6**:e18595.
5. Boyd SD: **Diagnostic applications of high-throughput DNA sequencing.** *Annu Rev Pathol* 2013, **8**:381-410.
6. Akhras MS, Unemo M, Thiyagarajan S, Nyrén P, Davis RW, Fire AZ, Pourmand N: **Connector inversion probe technology: a powerful one-primer multiplex DNA amplification system for numerous scientific applications.** *PLoS One* 2007, **2**:e195.
7. Krishnakumar S, Zheng J, Wilhelmy J, Faham M, Mindrinos M, Davis R: **A comprehensive assay for targeted multiplex amplification of human DNA sequences.** *Proc Natl Acad Sci U S A* 2008, **27**:9296-9301.
8. Shen P, Wang W, Chi A-K, Fan Y, Davis RW, Scharfe C: **Multiplex target capture with double-stranded DNA probes.** *Genome Med* 2013, **5**:50.
9. Kerick M, Isau M, Timmermann B, Sülthmann H, Herwig R, Krobitch S, Schaefer G, Verdorfer I, Bartsch G, Klocker H, Lehrach H, Schweiger MR: **Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity.** *BMC Med Genomics* 2011, **4**:68.

doi:10.1186/gm450

Cite this article as: Guettouche T, Zuchner S: Improved coverage and accuracy with strand-conserving sequence enrichment. *Genome Medicine* 2013, **5**:46.