Published in final edited form as:

Anal Biochem. 2015 February 1; 470: 48–51. doi:10.1016/j.ab.2014.10.010.

Hi-Plex targeted sequencing is effective using DNA derived from archival dried blood spots

T Nguyen-Dumont¹, M Mahmoodi¹, F Hammet¹, T Tran¹, H Tsimiklis¹, Kathleen Cuningham Foundation Consortium for research into Familial Breast Cancer², GG Giles³, JL Hopper⁴, Australian Breast Cancer Family Registry⁴, MC Southey¹, and DJ Park¹

¹Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia

²The Peter MacCallum Cancer Centre, St. Andrew's Place, East Melbourne, Victoria 3002, Australia

³Cancer Epidemiology Centre, The Cancer Council Victoria, 615 St. Kilda Road, Melbourne, Victoria 3004, Australia

⁴Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, 207 Bouverie Street, Parkville, Victoria 3010, Australia

Abstract

Many genetic epidemiology resources have collected dried blood spots (predominantly as Guthrie Cards) as an economical and efficient means of archiving sources of DNA, conferring great value to genetic screening methods that are compatible with this medium. We applied Hi-Plex to screen the breast cancer predisposition gene *PALB2* in 93 Guthrie Card-derived DNA specimens previously characterised for *PALB2* genetic variants via DNA derived from lymphoblastoid cell lines, whole blood and buffy coat. 92 of the 93 archival Guthrie Card-derived DNAs (99%) were processed successfully and sequenced using approximately half of a MiSeq run. From these 92, all 59 known variants were detected and no false-positive variant calls were yielded. 98.13% of amplicons (5417/5520) were represented within 15-fold of the median coverage (2786 reads) and 99.98% of amplicons (5519/5520) were represented at a depth of 10 read-pairs or greater. With Hi-Plex, we show for the first time that a high-plex amplicon based MPS system can be applied effectively to DNA prepared from dried blood spot archival specimens and, as such, dramatically increase the scopes of both method and resource.

© 2014 Elsevier Inc. All rights reserved.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TN-D, TT, FH, MM and HT carried out the Hi-Plex experiments; TN-D carried out the bioinformatics analyses; kConFab, the ABCFR, GGG and JLH contributed research resources; MCS participated in the study design and coordination; DJP conceived the study and participated in its design and coordination. All authors helped to draft the manuscript. All authors read and approved the final manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Hi-Plex; massively parallel sequencing; targeted sequencing; dried blood spot; Guthrie Card; archival DNA

Introduction

Dried blood spots, or Guthrie Cards specimens, provide a long-term, cost-effective and convenient alternative to freezing blood [1]. In many developed countries, they are obtained routinely from newborns to screen for metabolic disorders [2]. Dried blood spots have been collected by large epidemiological studies such as the Breast Cancer Family Registry [3] and the Melbourne Collaborative Cohort Study [4] as well as numerous others [5]. The ability to use dried blood-derived DNA in the context of such large study resources would allow researchers to make a considerable contribution to our understanding of the genetics of human diseases. However, there is evidence for DNA fragmentation over time with storage of dried blood spots [6], and this can influence the efficiency of downstream applications. Dried blood spots have been reported as a source of DNA suitable for downstream SNP genotyping via prior whole-genome amplification [7, 8] and, more recently, without the requirement for pre-amplification [9]. Archived neonatal dried blood spot samples have also been used, following whole-genome amplification, for accurate whole-genome and exome-targeted massively parallel sequencing [10]. A ‘low amplicon-plexity’-based approach has been published to show that massively parallel sequencing (MPS) of dried blood spot specimens can offer a novel approach to HIV drug-resistance surveillance [11]. To our knowledge, no prior study has been published that demonstrates or validates the accuracy of ‘high amplicon-plexity’ targeted enrichment applied to dried blood spot-derived DNA for genetic screening via MPS.

We previously developed and reported Hi-Plex, a streamlined and cost-effective highly multiplexed PCR approach for MPS library preparation. In addition to superior cost-effectiveness and accuracy [12] and Hsu *et al.* (submitted manuscript), Hi-Plex confers mechanistic advantages over alternative amplicon-based targeted enrichment systems for application to fragmented DNA since it can define a small and uniform size of amplicons [13, 14] and Nguyen-Dumont *et al.* (submitted manuscript). In this study, we assess the performance of Hi-Plex applied to archival dried blood spot-derived DNA.

Materials and Methods

DNA samples

Our sample set comprised 93 Guthrie Card-derived DNAs from women affected by breast cancer that had been screened previously for mutations in the coding and flanking intronic regions of *PALB2* via Hi-Plex and Sanger sequencing and/or high resolution melting curve analysis of lymphoblastoid cell line, whole blood or buffy coat-derived DNA [15–17]. All participants provided written informed consent for participation in the study. This study was approved by The University of Melbourne Human Research Ethics Committee.

Guthrie Card samples were provided by the Australian Breast Cancer Family Registry [3] (ABCFR, 89 specimens, including one duplicated sample) and the Kathleen Cuninghame Foundation Consortium for research into Familial Breast cancer (kConFab, Melbourne, Australia, four specimens). The samples were archived between six and 21 years prior to this study (mean: 12 years, median: 10 years, standard deviation: 4 years). DNA extractions from 2 mm diameter circular punches were performed using the QIAamp® 96 DNA blood kit 4 (Qiagen, Hilden, Germany) according to the manufacturer's instructions, including a proteinase K incubation step. Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies) was used for quantification.

Mutation Screening using Hi-Plex

This Hi-Plex assay was designed to target the *PALB2* and *XRCC2* genes. However, genotyping aspects of this study focus on *PALB2* only, as we did not have a similar test set with genotyping data for *XRCC2*. Sixty primer pairs targeting the protein coding and some flanking intronic and untranslated regions of *PALB2* and *XRCC2*, dual-indexing hybrid adapter primer sets and 'Bridge' primers are described in [13], [15] and Nguyen-Dumont *et al.* (submitted manuscript), respectively and listed in Supplementary Table 1. Supplementary Figure 1 schematically indicates how gene-specific primers target *PALB2*. All oligonucleotides used in this study were manufactured to standard desalting grade by Integrated DNA Technologies (Coralville, IA, USA). 94 individual PCR reactions (93 specimens and one no-template control) were conducted in wells of a skirted PCR plate in a final volume of 25µl with 1xPhusion® HF PCR buffer (ThermoScientific, Waltham, MA, USA), 1 unit of Phusion Hot Start II High-Fidelity DNA Polymerase (ThermoScientific), 400µM dNTPs (Bioline, London, UK), approximately 1 µM gene-specific primer pool aggregate (individual gene-specific primer concentrations vary and are described in [14] – those deviating from 4nM final reaction concentration are listed in Supplementary Table 2), 1µM 'Bridge' primers F8-bridge and R5-bridge (Nguyen-Dumont *et al.*, submitted), 2.5mM MgCl₂ (ThermoScientific) and 25 ng input genomic DNA. The following steps were then applied: 98°C for 1 min, 20 cycles of [98°C for 30 sec, 50°C for 1 min, 55°C for 1 min, 60°C for 1 min, 65°C for 1 min, 70°C for 1 min], addition of 1 µM dual-indexed hybrid adapter primers, then a further four cycles of [98°C for 30 sec, 68°C for 1 min, 70°C for 1 min], followed by incubation at 68°C for 20 min. Pooled library size-selection, quantification and sequencing were performed as detailed in [15], except that only ~53% of the sequencing run was dedicated to this experiment. Briefly, equal volumes of Hi-Plex products from each specimen were pooled and 40µl of the pooled library was resolved on a single wide lane (~2 cm) by 2% (w/v) agarose TBE gel electrophoresis using HR-agarose (Life Technologies, Carlsbad, CA, USA). The ~275bp library was excised from the gel and purified using the QIAEXII gel extraction kit (Qiagen, Dusseldorf, Germany). The library was sequenced on a MiSeq instrument using the MiSeq Reagent kit v2 300 cycles (Illumina). Prior to performing the sequencing run, 3.4µl of 100µM sequencing primers TSIT_Read1, TSIT_Read2 or TSIT_i7_read (Supplementary Table 1) were added respectively to the read1, read2 and i7 primer reservoirs in the MiSeq reagent cartridge. Mapping to hg19 was performed using bowtie-2-2.1.0 [18] with default parameters except for –trim5 20 and –trim3 20. ROVER variant caller [12] was applied using a variant proportion threshold of 0.15 and minimum required variant depth of two read-pairs.

Results and Discussion

Of the 93 specimens, 92 (99%) were sequenced successfully. The remaining specimen conferred low polymerase fidelity during amplification and a very low yield, indicative of PCR inhibition. The failed specimen was unexceptional in that it had been archived for the median time prior to this study. It is likely that this specimen was affected by unusual handling at some point during collection, storage or DNA extraction, although the timing of collection, transport and processing to produce the Guthrie Card specimen was routine.

For the sequenced 92 specimens, using ~53% of the MiSeq run, the median coverage depth for all specimens and all amplicons was 2786 reads. Figure 1 illustrates a high degree of coverage uniformity; shown are the median coverage and median absolute deviation from the median for all specimens for each of the 60 amplicons. The median coverage for the lowest represented amplicon (419 reads) was 6.7-fold lower than the overall median (2786 reads), while the median for the highest represented amplicon (11759 reads) was 4.3-fold higher than overall. As another way of representing the data, 98.13% (5417/5520) of amplicons were represented within 15-fold of the overall median (2786 reads). 99.98% (5519/5520) of amplicons were covered at a depth of 10 read-pairs or greater. All 59 known genetic variants were detected by the ROVER variant calling software (Table 1). Furthermore, no false-positive calls were made, resulting in 100% sensitivity and 100% specificity. The uniformity of amplicon coverage across the targeted regions was high and consistent with the profile observed previously following application of Hi-Plex to matched specimens derived from other blood-based sources of DNA, including freshly cultured lymphoblastoid cell lines. This high performance is probably at least partly a consequence of Hi-Plex enabling the use of relatively small and highly uniform amplicon lengths. Other amplicon-based target enrichment systems are more constrained in this regard and would be predicted to struggle for coverage uniformity with increasing DNA fragmentation. As such, it is expected that Hi-Plex will offer performance advantages in other contexts in which DNA integrity is compromised, e.g. DNA-derived from formalin-fixed, paraffin-embedded tumour specimens or ancient DNA specimens. The accuracy and stringent artefact filtering afforded by Hi-Plex are expected to confer advantages in applications of genetic variant detection in sub-populations, such as identifying emerging drug resistance in heterogeneous tumours, for example.

If we use the MiSeq performance metrics for the two genes targeted in this study and assume a target mean coverage depth of 200 reads per specimen amplicon and factor in the lower cost per base of HiSeq2500 sequencing compared with MiSeq sequencing, we can realistically project that for large-scale screening, the cost per specimen would currently be ~65 Australian cents or ~36 British pence per specimen.

The ability to apply Hi-Plex in the context of dried blood spot material opens a wide variety of possibilities for genetic epidemiology and diagnostic applications.

Conclusions

With Hi-Plex, we show for the first time that highly multiplex amplicon-based target enrichment for MPS can produce robust and highly accurate sequence screening in the context of archival dried blood spot-derived DNA. This empowers genetic epidemiologists and diagnosticians with the ability to use this very important bioresource for a broad range of applications to address many research questions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

TN-D is a Susan G. Komen for the Cure Postdoctoral Fellow. MCS is a National Health and Medical Research Council (NHMRC) Senior Research Fellow. The Australian Breast Cancer Family Registry (ABCFR; 1992–1995) was supported by the Australian NHMRC, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia). We wish to thank Margaret McCredie for a key role in the establishment and leadership of the ABCFR in Sydney, Australia, and the families who donated their time, information and biospecimens. This work was supported by grant UM1 CA164920 from the National Cancer Institute (NCI). The content of this manuscript does not necessarily reflect the views or policies of the NCI or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products or organizations imply endorsement by the US Government or the BCFR. We wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (funded 2001–2009 by NHMRC and currently by the National Breast Cancer Foundation (NBCF) and Cancer Australia #628333) for their contributions to this resource, and the many families who contribute to kConFab. kConFab is supported by grants from the NBCF, the NHMRC, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia and the Cancer Foundation of Western Australia.

This work was supported by the Australian NHMRC (APP1025879 and APP1029974), the National Institute of Health (USA) (RO1CA155767) and by a Victorian Life Sciences Computation Initiative (VLSCI) grant (#VR0182) on its Peak Computing Facility, an initiative of the Victorian Government.

References

1. Mei JV, Alexander JR, Adam BW, Hannon WH. Use of filter paper for the collection and analysis of human whole blood specimens. *The Journal of nutrition*. 2001; 131:1631S–1636S. [PubMed: 11340130]
2. Audrezet MP, Costes B, Ghanem N, Fanen P, Verlingue C, Morin JF, Mercier B, Goossens M, Ferec C. Screening for cystic fibrosis in dried blood spots of newborns. *Molecular and cellular probes*. 1993; 7:497–502. [PubMed: 8145780]
3. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, Ziogas A, Andrulis IL, Anton-Culver H, Boyd N, Buys SS, Daly MB, O'Malley FP, Santella RM, Southey MC, Venne VL, Venter DJ, West DW, Whittemore AS, Seminara DR. Breast Cancer Family, . The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast cancer research : BCR*. 2004; 6:R375–389. [PubMed: 15217505]
4. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC scientific publications*. 2002; 156:69–70. [PubMed: 12484128]
5. McEwen JE, Reilly PR. Stored Guthrie cards as DNA “banks”. *Am J Hum Genet*. 1994; 55:196–200. [PubMed: 8023846]
6. Makowski GS, Davis EL, Hopfer SM. The effect of storage on Guthrie cards: implications for deoxyribonucleic acid amplification. *Annals of clinical and laboratory science*. 1996; 26:458–469. [PubMed: 8879364]

7. Hollegaard MV, Grove J, Thorsen P, Norgaard-Pedersen B, Hougaard DM. High-throughput genotyping on archived dried blood spot samples. *Genetic testing and molecular biomarkers*. 2009; 13:173–179. [PubMed: 19371215]
8. Hollegaard MV, Grove J, Grauholm J, Kreiner-Moller E, Bonnelykke K, Norgaard M, Benfield TL, Norgaard-Pedersen B, Mortensen PB, Mors O, Sorensen HT, Harboe ZB, Borglum AD, Demontis D, Orntoft TF, Bisgaard H, Hougaard DM. Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet*. 2011; 12:58. [PubMed: 21726430]
9. St Julien KR, Jelliffe-Pawlowski LL, Shaw GM, Stevenson DK, O'Brodovich HM, Krasnow MA, Stanford BPDSG. High quality genome-wide genotyping from archived dried blood spots without DNA amplification. *PloS one*. 2013; 8:e64710. [PubMed: 23737996]
10. Hollegaard MV, Grauholm J, Nielsen R, Grove J, Mandrup S, Hougaard DM. Archived neonatal dried blood spot samples can be used for accurate whole genome and exome-targeted next-generation sequencing. *Molecular genetics and metabolism*. 2013; 110:65–72. [PubMed: 23830478]
11. Ji H, Li Y, Graham M, Liang BB, Pilon R, Tyson S, Peters G, Tyler S, Merks H, Bertagnolio S, Soto-Ramirez L, Sandstrom P, Brooks J. Next-generation sequencing of dried blood spot specimens: a novel approach to HIV drug-resistance surveillance. *Antiviral therapy*. 2011; 16:871–878. [PubMed: 21900719]
12. Pope BJ, Nguyen-Dumont T, Hammet F, Park DJ. ROVER variant caller: read-pair overlap considerate variant-calling software applied to PCR-based massively parallel sequencing datasets. *Source code for biology and medicine*. 2014; 9:3. [PubMed: 24461215]
13. Nguyen-Dumont T, Pope BJ, Hammet F, Southey MC, Park DJ. A high-plex PCR approach for massively parallel sequencing. *BioTechniques*. 2013; 55:69–74. [PubMed: 23931594]
14. Nguyen-Dumont T, Pope BJ, Hammet F, Mahmoodi M, Tsimiklis H, Southey MC, Park DJ. Cross-platform compatibility of Hi-Plex, a streamlined approach for targeted massively parallel sequencing. *Anal Biochem*. 2013; 442:127–129. [PubMed: 23933242]
15. Nguyen-Dumont T, Teo ZL, Pope BJ, Hammet F, Mahmoodi M, Tsimiklis H, Sabbaghian N, Tischkowitz M, Foulkes WD, Giles GG, Hopper JL, Southey MC, Park DJ. Hi-Plex for high-throughput mutation screening: application to the breast cancer susceptibility gene PALB2. *BMC Med Genomics*. 2013; 6:48. [PubMed: 24206657]
16. Teo ZL, Park DJ, Provenzano E, Chatfield CA, Odefrey FA, Nguyen-Dumont T, ConFab K, Dowty JG, Hopper JL, Winship I, Goldgar DE, Southey MC. Prevalence of PALB2 mutations in Australasian multiple-case breast cancer families. *Breast cancer research : BCR*. 2013; 15:R17. [PubMed: 23448497]
17. Southey MC, Teo ZL, Dowty JG, Odefrey FA, Park DJ, Tischkowitz M, Sabbaghian N, Apicella C, Byrnes GB, Winship I, Baglietto L, Giles GG, Goldgar DE, Foulkes WD, Hopper JLR. kConFab for the Breast Cancer Family, A PALB2 mutation associated with high risk of breast cancer. *Breast cancer research : BCR*. 2010; 12:R109. [PubMed: 21182766]
18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012; 9:357–359. [PubMed: 22388286]

Figure 1. Median coverage (reads) and median absolute deviation from the median for all specimens and for each of the 60 amplicons. The 60 amplicons are plotted in increasing order of the median coverage depth. The solid horizontal line shows the overall median coverage. The dashed horizontal lines indicate values 25-fold higher and 25-fold lower than the overall median coverage, respectively.

Table 1

59 *PALB2* coding region genetic variant occurrences in the 92 sequenced specimens.

Variant type	Nucleotide change	Protein change	rs number	Number of carriers
Nonsense	c.196C>T	p.Gln66*	rs180177083	2 heterozygotes
	c.3113G>A	p.Trp1038*	rs180177132	4 heterozygotes
Frameshift	c.1947_1948insA	p.Glu650fs*13	-	1 heterozygote
	c.2982_2983insT	p.Ala995fs*16	rs180177127	1 heterozygote
Missense	c.1010T>C	p.Leu337Ser	rs45494092	5 heterozygotes
	c.1676A>G	p.Gln559Arg	rs152451	13 heterozygotes
				1 homozygote
	c.2014G>C	p.Glu672Gln	rs45532440	8 heterozygotes
Synonymous				1 homozygote
	c.2590C>T	p.Pro864Ser	rs45568339	1 heterozygote
	c.2993G>A	p.Gly998Glu	rs45551636	8 heterozygotes
	c.1470C>T	p.Pro490Pro	rs45612837	1 heterozygote
	c.1572A>G	p.Ser524Ser	rs45472400	3 heterozygotes
	c.3300T>G	p.Thr1100Thr	rs45516100	8 heterozygotes
				1 homozygote
	c.3495G>A	p.Ser1165Ser	-	1 heterozygote