Available online at www.sciencedirect.com
ScienceDirect
Analytical Biochemistry 356 (2006) 117–124
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Identification of coding single nucleotide polymorphisms and mutations by combination of genome tiling arrays and enrichment/depletion of mismatch cDNAs

Meng-Min Liu [a,*], Sherman M. Weissman [a], Ling Tang [b]

[a] *Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA*
[b] *Department of Animal Sciences, Auburn University, Auburn, AL 36849, USA*

Received 15 March 2006
Available online 5 June 2006

Abstract

Genome tiling array technology combined with a method for both enrichment and depletion of mismatch-containing cDNA fragments offers a useful approach for detecting coding single nucleotide polymorphisms (cSNPs) and mutations in pooled cDNA samples. Enriched mismatch and perfect match cDNA samples from human primary melanoma cells and normal melanocytes were obtained by selection using mismatch repair thymine DNA glycosylase-bound beads. These cDNA samples were then labeled and hybridized to Encyclopedia of DNA Elements genome tiling arrays. The results revealed that the hybridization intensity values of potential cDNA variation regions of the enriched mismatch samples increased, whereas the hybridization intensity values of corresponding regions of the enriched perfect match samples decreased. Six potential mutations were confirmed by polymerase chain reaction product sequencing, including two novel heterozygous mutations in melanoma cells. We suggest that this strategy should increase the efficiency of both cSNP and mutation detection throughout the entire human genome and decrease the cost and complexity of genomewide analysis of cDNA variations.

Keywords: Mutation/single nucleotide polymorphisms; Genome tiling arrays; Encyclopedia of DNA elements; Melanoma; Thymine DNA glycosylase

With completion of sequencing the human genome, an important challenge in the postgenomic era is to efficiently use genetic information to associate individual DNA variations with medically important parameters such as disease susceptibility, individual response to drugs, and prognostic variables. Single nucleotide polymorphisms (SNPs),[1] the most common DNA variants [1–3], are of particular value in whole-genome association studies for identifying the genes involved in complex trait variation [4,5]. They occur, on average, once every 300 bp with a minor allele frequency greater than 1% [6–8]. It is estimated that 1.42 million SNPs are distributed throughout the human genome and about 60,000 of these SNPs fall within exons [9]. DNA variations in coding regions, together with changes in regulatory regions, are believed to have the highest impact on phenotype. Thus, particular attention is directed toward the identification of coding SNPs (cSNPs) [10].

A number of methods have been developed and widely applied for the purpose of identifying DNA variations of a known gene or candidate mutation site [11]. Two of the main approaches for identifying unknown sequence variations from complex genomes are hybridization-based and enzyme-based procedures [12]. The hybridization-based approach utilizes DNA microarray technology which provide a simple, accurate, and high-throughput typing

* Corresponding author. Fax: +1 203 785 6350.
E-mail address: meng-min.liu@yale.edu (M.-M. Liu).

[1] *Abbreviations used:* SNPs, single nucleotide polymorphisms; cSNPs, coding SNPs; hTDG, human thymine DNA glycosylase; Mig.MthI or mTDG, Methanobacterium *thermoautotropicum* DNA mismatch *N*-glycosylase; ENCODE, Encyclopedia of DNA Elements; IGB, Integrated Genome Browser; DTT, dithiothreitol; BSA, bovine serum albumin; Mes, 4-morpholineethanesulfonic acid.

0003-2697/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2006.05.013

method due to its solid surface hybridization reaction [13,14]. This approach was successful in identifying a large fraction of, but not all polymorphisms in the regions examined [15–17]. The enzyme-based approaches involve preparing restriction digests of the test DNA(s), followed by denaturing and annealing the resulting DNA fragments. Enzymatic treatments are then used to distinguish perfectly matched fragments from those containing internal mismatches derived from sequence variations in the input DNA pools. Genomic mismatch scanning, which enriches for identical-by-descent regions between individuals using mismatch repair proteins, showed some promise for gene mapping when combined with DNA microarray technology [18–21]. However, there are limited data with regard to its use for mammalian DNA analysis.

Recently, whole-genome tiling arrays have emerged as a preferred platform for genomic analysis beyond simple gene expression profiling. These arrays offer an opportunity to comprehensively investigate novel functional elements of any species using an unbiased approach. Potential uses for this new technology include empirical annotation of the transcriptome, identification of transcription factor binding sites, analysis of alternative splicing, characterization of the methylome (the methylation state of the genome), polymorphism discovery and genotyping, comparative genome hybridization, and genome resequencing [22–25].

An approach for separating oligonucleotide mismatch (Mm) duplexes from a pool of oligonucleotide perfect match (Pm) duplexes using both immobilized human thymine DNA glycosylase (hTDG) and Methanobacterium *thermoautotropicum* DNA mismatch *N*-glycosylase (Mig.MthI or mTDG) beads was reported [26]. This approach employs DNA glycosylases (hTDG and mTDG) that specifically recognize nucleotide mismatches and generate abasic sites. The immobilized bead-bound hTDG/mTDG can bind reversibly to these abasic sites and thus can be used for affinity purification of mismatch-containing DNA fragments through the procedure of binding, washing, and elution [26]. Here, we present a useful approach to localize cSNPs and mutations in a large fraction of human genes using a combination of genome tiling arrays and a method for both enrichment and depletion of mismatch-containing cDNA fragments by TDG selection. This approach enables the identification of cSNPs and somatic mutations in human primary melanoma cells.

Materials and methods

Preparation and enrichment of mismatch and perfect match oligonucleotide duplexes/complexes

The following gel-purified oligonucleotides were used for preparation of mismatch and perfect match oligonucleotide duplexes and complexes. Oligodeoxynucleotide L1, 5′*-cgcggatccgcgaggaatcggtctacgtgcgtgaatccacaccg**cg**agctattctctcgatgtgtgcaggtcgacg-3′; oligodeoxynucleotide L2, 5′-gacttacgtcgacctgcacacatcgagagaatagcttg**tg**gtgtggattcacgcacgtagaccgattcctcgcggatccgcg-3′ (5′-phosphorate); oligodeoxynucleotide L3, 5′- taagtcatggctcagtggacaagagcacccgactgcttccgaagggagctcaaatcccagcaaccaggcggatccggc-3′ (5′-phosphorate); oligodeoxynucleotide L4, 5′- gccggatccgcctggttgctgggatttgagctcccttcggaagagcagtcgggtgctcttgtccactgagccat-3′; oligodeoxynucleotide L5, 5′*- cgcggatccgcgaggaatcggtctacgtgcgtgaatcctctcgatgtgtgcaggtcgacg-3′ (* = [γ-P^{32}]ATP); oligodeoxynucleotide L6, 5′- gacttacgtcgacctgcaccacgagaggatcacgcacgtagaccgattcctcgcggatccgcg-3′ (5′-phosphorate). L1 and L5 were 5′-end-labeled with approximately 10 pmol of [γ-P^{32}]ATP using T4 polynucleotide kinase and purified using microchromatography columns containing Sephadex G-50. L1 and L2 were annealed to form 80-bp mismatch duplexes (L1L2), which contain an internal CG/GT pairing (the boldfaced letters indicate the mispaired nucleotides). L3 and L4 were annealed to form 74-bp perfect match duplexes (L3L4). L5 and L6 were annealed to form 66-bp perfect match duplexes (L5L6). Annealing reactions were performed in 3× EEN buffer (30 mM *N*-(2-hydroxyethyl) piperazine-*N*′-3-propanesulfonic acid, pH 8.25 (Sigma), 3 mM EDTA, pH 8.0, 1 M NaCl) at 68 °C for 20 h. The L1L2 duplex was then ligated with the L3L4 duplex to form a 154-bp mismatch complex (L1L2L3L4). The L3L4 duplex was ligated with the L5L6 duplex to form a 140-bp perfect match complex (L3L4L5L6) by T4 DNA ligase.

The previously published protocol for enrichment of oligonucleotide duplexes containing a mismatch from a pool of perfect match oligonucleotide duplexes using fused glutathione-*S*-transferase–TDG immobilized to glutathione Sepharose 4B beads [26] was modified as follows. Briefly, after preincubating 20 μg of *Cot*I DNA with 100 μl solution containing bead-bound hTDG and mTDG in binding buffer (50 mM Tris–HCl (pH 8.2), 50 mM KCl, 5 mM EDTA, 0.2 mM Zn_2SO4, 1 mM DTT, 0.25 mg/ml BSA) at room temperature for 4 h, 154-bp mismatch and 140-bp perfect match oligonucleotide complexes were added into the mixture and incubated for 8 h on a rotating platform (45 rpm rotation). Next, the bead-bound material was subjected to three cycles of washing using 1 ml washing buffer (50 mM Tris–HCl (pH 8.2), 50 mM KCl, 5 mM EDTA, 0.2 mM Zn_2SO4, 1 mM DTT, 0.25 mg/ml BSA, 50–400 mM LiCl), while gradually increasing the LiCl concentrations from 50 to 400 mM. Each wash cycle included 3 h of incubation at room temperature on a rotating platform followed by collection of the supernatants after centrifugation at 1500 rpm for 5 min. The bead-bound material was then eluted using 1 ml elution buffer (50 mM Tris–HCl (pH 7.5), 50 mM KCl, 0.2 mM Zn_2SO4, 1 mM DTT, 0.25 mg/ml BSA, 30–120 mM $MgCl_2$), while gradually increasing the $MgCl_2$ concentrations from 30 to 120 mM at room temperature for 3 h on a rotating platform. The eluted supernatants contained the enriched mismatch oligonucleotide fraction, while the washing supernatants contained the enriched perfect match oligonucleotide fraction due to depletion of mismatch oligonucleotide complexes which remained bound to the beads. Each fraction was phenol/chloroform-extracted, ethanol-precipitated, and viewed on a 6% sequencing gel.