Gene 512 (2013) 16–22

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Large-scale profiling and identification of potential regulatory mechanisms for allelic gene expression in colorectal cancer cells

Robin Dong-Woo Lee [a,b], Min-Young Song [a], Jong-Keuk Lee [a,*]

[a] Asan Institute for Life Sciences, University of Ulsan College of Medicine, Seoul, Korea
[b] Department of Genetics, Cell Biology, and Development, University of Minnesota, MN, USA

ARTICLE INFO

Article history:
Accepted 2 October 2012
Available online 10 October 2012

Keywords:
Single nucleotide polymorphism (SNP)
Alelic gene expression
Monoallelic expression
Allelic imbalance
Allele substitution
Colorectal cancer cells

ABSTRACT

Allelic variation in gene expression is common in humans and this variation is associated with phenotypic variation. In this study, we employed high-density single nucleotide polymorphism (SNP) chips containing 13,900 exonic SNPs to identify genes with allelic gene expression in cells from colorectal cancer cell lines. We found 2 monoallelically expressed genes (*ERAP2* and *MYLK4*), 32 genes with an allelic imbalance in their expression, and 13 genes showing allele substitution by RNA editing. Among a total of 34 allelically expressed genes in colorectal cancer cells, 15 genes (44.1%) were associated with *cis*-acting eQTL, indicating that large portions of allelically expressed genes are regulated by *cis*-acting mechanisms of gene expression. In addition, potential regulatory variants present in the proximal promoter regions of genes showing either monoallelic expression or allelic imbalance were not tightly linked with coding SNPs, which were detected with allelic gene expression. These results suggest that multiple rare variants could be involved in the *cis*-acting regulatory mechanism of allelic gene expression. In the comparison with allelic gene expression data from Centre d'Etude du Polymorphisme Humain (CEPH) family B cells, 12 genes showed B-cell specific allelic imbalance and 1 noncoding SNP showed colorectal cancer cell-specific allelic imbalance. In addition, different patterns of allele substitution were observed between B cells and colorectal cancer cells. Overall, our study not only indicates that allelic gene expression is common in colorectal cancer cells, but our study also provides a better understanding of allele-specific gene expression in colorectal cancer cells.

1. Introduction

Allelic variation in gene expression is common in humans (Lo et al., 2003) and constitutes different phenotypes and complex human traits (Frazer et al., 2009). Allelic gene expression in human cells containing a diploid genome can be categorized into four major patterns: equal gene expression of both alleles, monoallelic gene expression, allelic imbalance in gene expression showing preferentially higher expression of one allele, and allele substitution by RNA editing. Monoallelic expression is well-exemplified by mammalian X-chromosome inactivation, genomic imprinting and allelic exclusion (Yang and Kuroda, 2007). Monoallelic gene expression is widespread in human autosomes (Gimelbrant et al., 2007) and can significantly affect cell fate and physiology (Tarutani and Takayama, 2011). Allelic imbalance is described as one allele having higher expression than the alternative allele, which is very common and frequently observed in many cell types including cancer (Chen et al., 2008; Försti et al., 2001). Allele substitution by RNA editing has also been frequently detected in cells. This phenomenon is likely due to A-to-I RNA editing, where the adenosine base is converted to inosine, catalyzed by adenosine deaminases (Wulff and Nishikura, 2010). Recently, RNA editing has received much attention because the amount of functional RNA molecules can be controlled by this mechanism (Li et al., 2011; Wulff and Nishikura, 2010).

Two different regulatory mechanisms of gene expression, either *cis*- or *trans*-regulations, are responsible for the expression of functional genes in different tissues (Bayele et al., 2006; Mahr et al., 2006). In particular, genome-wide association studies based on single nucleotide polymorphism (SNP) markers showed that large proportions of protein-coding genes are regulated by *cis*-acting mechanisms (Emilsson et al., 2008; Göring et al., 2007). Thus, identifying genetic variants in the nearby region of the allelically expressed gene provides insight into allele-specific expression mechanisms.

To date, many studies of allelic gene expression, including our previous study (Song et al., 2012), have employed lymphoblastoid cell lines (Cheung et al., 2003; Gimelbrant et al., 2007; Pant et al., 2006). However, not many allelic gene expression studies have used other

Abbreviations: ADAR, adenosine deaminase, RNA-specific; AI, allelic imbalance; CEPH, Centre d'Etude du Polymorphisme Humain; eQTL, expression quantitative trait loci; MAE, monoallelic expression; SNP, single nucleotide polymorphism.

* Corresponding author at: Asan Institute for Life Sciences, University of Ulsan College of Medicine, 388-1 Pungnap-2-Dong, Songpa-Gu, Seoul 138-736, Korea. Tel.: +82 2 3010 4142; fax: +82 2 488 3312.
E-mail address: cookie_jklee@hotmail.com (J.-K. Lee).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.10.001

cell types. Thus, in this study, we employed the high-density SNP chip to discover and characterize the potential mechanisms of allele-specific gene expressions in colorectal cancer cells.

2. Materials and methods

2.1. Colorectal cancer cell lines and cell culture

A total of 11 colorectal cancer cell lines that originated from Korean cancer patients (Ku and Park, 2005) were obtained from the Korean Cell Line Bank (http://cellbank.snu.ac.kr). Additionally, genomic DNA from 90 healthy cohort samples was used as the control. Colorectal cancer cell lines were maintained in RPMI-1640 medium with 2 mM L-glutamine and 25 mM HEPES, supplemented with 10% fetal bovine serum and penicillin (1000 U/ml) with streptomycin (100 μg/ml).

2.2. Genomic DNA and RNA extraction, and cDNA synthesis

Genomic DNA was extracted from colorectal cancer cell lines using the QIAamp DNA Mini kit (Qiagen, Germany). Total RNA was also isolated from colorectal cancer cells using TRIZOL reagent (Invitrogen, Carlsbad, CA) according to the protocols provided by the manufacturer. cDNA was synthesized from 10 μg of total RNA using the SuperScript Preamplification System for First Strand cDNA Synthesis kit (Invitrogen). cDNA was precipitated with 96% ethanol with 20 μg glycogen and 12 μmol sodium acetate to use as a template for SNP chip genotyping.

2.3. SNP genotyping

For SNP genotyping, 200 ng of genomic DNA or cDNA derived from 10 μg of total RNA was used. SNP genotyping was performed using the Human NS-12 BeadChip (Illumina, San Diego, CA) according to the manufacturer's instructions. The Human NS-12 BeadChip contains more than 11,000 SNPs located across exons and untranslated mRNA regions of 6310 genes and 2000 SNPs in introns and flanking regions of genes (Evans et al., 2008). Genotyping clustering and calling were performed using BeadStudio software (Illumina).

2.4. Detection of allelic gene expression

Using the genotype clustering and calling, we looked for allelically expressed genes. Allelic gene expression was determined by calculating the ratio of fluorescence signal between two alleles (allele1/allele2) in cDNA and genomic DNA for heterozygous SNPs, as described in our previous study (Song et al., 2012). The signal ratio from cDNA was divided by the corresponding ratio in genomic DNA to obtain a measure for allelic gene expression. Specifically, SNPs where the fluorescence signal ratio of two alleles in cDNA was ≤ 0.1 or ≥ 0.9 were defined as monoallelic expression. Allelic imbalance was considered when the fluorescence signal ratio of two alleles for heterozygous SNPs in cDNA was ≤ 0.3 or ≥ 0.7, while the signal ratio of two alleles in genomic DNA was approximately equal (usually 0.4–0.6). SNPs where the heterozygote and/or homozygote genotypes in genomic DNA samples were altered into new homozygote genotypes in cDNA were defined as allele substitution.

2.5. Analysis and interpretation of allelic gene expression

Additionally, genotype clusters were inspected visually and allelically expressed genes were subject to eQTL database search (http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi). eQTL was used to identify genomic loci that regulate gene expression. SNPs with the highest p value from the eQTL database were selected as candidates for further analyses. Allele substitutions were identified when the allele of the cDNA was different from the allele of the gDNA from the same cell line sample. In order to predict the potential role of the allele substitution in nonsynonymous SNPs, we performed *in silico* prediction of functionality of the genetic variants using PolyPhen program (http://genetics.bwh.harvard.edu/pph/). The DARNED database (http://darned.ucc.ie), a database for RNA editing in humans, was used to determine the novelty of the allele substitutions discovered. The pairwise correlation coefficient (r^2) between coding SNPs and regulatory SNPs was determined by using HapAnalyzer program (Jung et al., 2004). Additionally, to examine the difference of allelic gene expression between B cells and colorectal cancer cells, we compared colorectal cancer cell gene expression with our previous allelic gene expression data in B cells (Song et al., 2012). In particular, SNP data showing allelic gene expression, either monoallelic, imbalanced or allele substitution, in one of cell types (B cells & colon cells) were collected and compared to determine if the SNP site showed any difference in allelic gene expression. The genes expressing in both cell types (B cells and colon cells) were used in comparison. Genes with no cDNA expression, poor clustering, or non-heterozygote samples were excluded from allelic gene expression analysis in B cells and colon cancer cells.

2.6. Promoter sequencing of allelically expressed genes

To identify *cis*-acting regulatory SNPs, we sequenced the promoter region of 34 allelically expressed genes of colorectal cancer cells using capillary sequencing. For each allelically expressed gene, we initially sequenced 2 heterozygote samples showing allelic gene expression in the NS-12 genotyping chip data. Once we found variants in the initial samples, all samples were used to determine the genotype of the genetic variants. Approximately more than 800 base pairs (mean ± s.d., 864.03 ± 70.4 bp) of the upstream promoter regions of the corresponding gene were screened by sequencing. PCR products were sequenced with an ABI Prism 3730 sequencer (Applied Biosystems, Foster City, CA) and the Polyphred Program (http://www.droog.gs.washington.edu/PolyPhred.html) was used to detect genetic variation. Genetic variation was identified by searching heterozygote variants as well as comparing sample sequences to the reference sequence.

3. Results

3.1. Profiling of allelic gene expression in colorectal cancer cells

The genotype call rates from the NS-12 Genotyping BeadChip for genomic DNA and cDNA were 98.42 ± 1.18% (mean ± s.d.) and 52.86 ± 4.11% (mean ± s.d.), respectively (Fig. 1). Genotype plots of the NS-12 genotyping BeadChip showed three distinct patterns of allelic gene

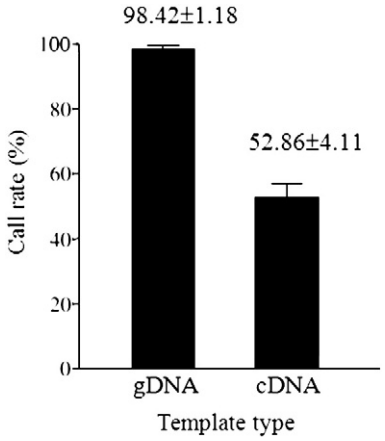

Fig. 1. Genotype call rates of the Human NS-12 Genotyping BeadChip in colorectal cancer cell lines ($n = 11$). gDNA and cDNA were prepared from colorectal cancer cell lines as described in Materials and methods.