Gene 576 (2016) 312–318

Contents lists available at ScienceDirect

Gene

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

Research paper

Identification of differentially methylated regions in new genes associated with knee osteoarthritis

Carolina A. Bonin [a], Eric A. Lewallen [a], Saurabh Baheti [b], Elizabeth W. Bradley [a], Michael J. Stuart [a], Daniel J. Berry [a], Andre J. van Wijnen [a,c], Jennifer J. Westendorf [a,c,*]

[a] Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN, United States
[b] Department of Biomedical Statistics and Informatics, Rochester, MN, United States
[c] Department of Biochemistry & Molecular Biology, Mayo Clinic, Rochester, MN, United States

ARTICLE INFO

Article history:
Received 22 September 2015
Accepted 15 October 2015
Available online 17 October 2015

Keywords:
Articular cartilage
DNA methylation
Differentially methylated regions
FOXP4
SHROOM1

ABSTRACT

Epigenetic changes in articular chondrocytes are associated with osteoarthritis (OA) disease progression. Numerous studies have identified differentially methylated cytosines in OA tissues; however, the consequences of altered CpG methylation at single nucleotides on gene expression and phenotypes are difficult to predict. With the objective of detecting novel genes relevant to OA, we conducted a genome-wide assessment of differentially methylated sites (DMSs) and differentially methylated regions (DMRs). DNA was extracted from visually damaged and normal appearing, non-damaged human knee articular cartilage from the same joint and then subjected to reduced representation bisulfite sequencing. DMRs were identified using a genome-wide systematic bioinformatics approach. A sliding-window of 500 bp was used for screening the genome for regions with clusters of DMSs. Gene expression levels were assessed and cell culture demethylation experiments were performed to further examine top candidate genes associated with damaged articular cartilage. More than 1000 DMRs were detected in damaged osteoarthritic cartilage. Nineteen of these contained five or more DMSs and were located in gene promoters or first introns and exons. Gene expression assessment revealed that hypermethylated DMRs in damaged samples were more consistently associated with gene repression than hypomethylated DMRs were with gene activation. Accordingly, a demethylation agent induced expression of most hypermethylated genes in chondrocytes. Our study revealed the utility of a systematic DMR search as an alternative to focusing on single nucleotide data. In particular, this approach uncovered promising candidates for functional studies such as the hypermethylated protein-coding genes *FOXP4* and *SHROOM1*, which appear to be linked to OA pathology in humans and warrant further investigation.

1. Introduction

Osteoarthritis (OA) is a progressive multi-factorial disease that commonly affects joints of the hand, hip and knee. Nearly 10% of men and 20% of women >60 years of age worldwide present OA symptoms (Woolf and Pfleger, 2003) translating into a tremendous global socio-economic burden. In the United States, the average total direct cost associated with OA from 2008 to 2011 totalled U$339.7 billion, a number that will increase as the population ages (Yelin et al., 2014).

It is established that mechanical loading of joints directly affects chondrocyte metabolism, and acute injury or persistently altered mechanical stimuli can lead to OA (Goldring and Marcu, 2009; Goldring and Goldring, 2010). Early stage disease is characterized by increased bone remodeling and invasion of deeper cartilage layers by blood vessels, allowing contact with chondrolytic enzymes. This is often followed by secondary inflammation and disruption of bone homeostasis with a shift toward tissue sclerosis and thickening of the calcified cartilage layer (Burr and Gallant, 2012; Goldring and Goldring, 2007). Microscopically, an initial repair response causes chondrocytes to proliferate and increase cartilage matrix production, while at a later stage they become hypertrophic and increase synthesis of catabolic cytokines and matrix degrading proteases. This cascade of events that degrades collagens and proteoglycans within the cartilage extracellular matrix (ECM) in OA has been well documented (Goldring, 2000). Although phenotypic changes are well characterized, the identification of genes and molecular pathways involved are still underway.

Abbreviations: 5-AzaC, 5-azacytidine; DMRs, differentially methylated regions; DMS, differentiated methylated sites; DMSs, differentially methylated CpG sites; ECM, extracellular matrix; GO, gene ontology; MMP, matrix metalloproteinases; OA, osteoarthritis; qRT-PCR, quantified using semi-quantitative real-time reverse transcriptase PCR; RRBS, representation bisulfite sequencing technique; SAAP-RRBS, streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing; TSS, transcription start sites.

* Corresponding author at: Mayo Clinic, 200 First Street SW, Rochester, MN 55905, United States.
E-mail address: westendorf.jennifer@mayo.edu (J.J. Westendorf).

http://dx.doi.org/10.1016/j.gene.2015.10.037
0378-1119/

Epigenetic regulation of gene expression in chondrocytes is a key mechanism behind the phenotypic changes observed in OA (Goldring and Marcu, 2012). In particular, DNA methylation surveys, which measure presence of a methyl group on cytosine residues (5meC) at CpG di-nucleotides, have been a focus of OA studies because they can associate epigenetic states of genes and pathways with disease progression. CpG methylation is typically a repressive mark for gene expression, especially when sites are located in the gene promoter region, but can be increased in actively transcribed genes if located within intragenic regions (Laird, 2010). Thus, the significance of CpG methylation in regulating gene expression is highly variable when observed in the context of CpG locus density (Ziller et al., 2013). An extensive study of more than 42 whole-genome methylation datasets and 30 human cell and tissue types revealed that only ~20% of CpGs are dynamically regulated. Most of these differentially methylated regions in the genome span 1 kb and are often distal to gene transcription start sites (TSS) (Ziller et al., 2013).

Array-based surveys have revealed significant differences in DNA methylation between OA and non-OA bone and cartilage in the human hip and knee at hundreds to thousands of CpG sites throughout the genome (Delgado-Calle et al., 2013; Fernandez-Tajes et al., 2014; Rushton et al., 2014; Jeffries et al., 2014; Moazedi-Fuerst et al., 2014; den Hollander et al., 2015). In some of these studies, differences among sample groups were most pronounced in genes involved in inflammation and immune-responses (Fernandez-Tajes et al., 2014; Jeffries et al., 2014), while others reported an over-representation of genes involved in developmental pathways (Moazedi-Fuerst et al., 2014; den Hollander et al., 2015). These studies corroborated previous findings for some identifiable genes associated with the onset and progression of OA. Among them are matrix metalloproteinases (*MMP-3*, -9, -13, -14) and aggrecanases (e.g., a disintegrin and metalloproteinase with thrombospondin motifs; *ADAMTS-4*), whose expression are both directly (Roach et al., 2005) and indirectly (e.g., via leptin expression) (Iliopoulos et al., 2007) associated with the demethylation of CpG sites in promoter regions. Demethylation has also been reported for genes related to the inflammatory process of OA, such as *ILB1B* (Hashimoto et al., 2009, 2013) and *NF-κB* (de Andres et al., 2013). Other epigenetic changes have been detected in *GDF5* (Reynard et al., 2014) and *SOX9* (Kim et al., 2013).

The inherent complexity in interpreting CpG methylation often produces weak correlations between CpG methylation and gene expression data (Delgado-Calle et al., 2013). Therefore, in addition to focusing on individual CpG sites, the genome can be scanned for clusters of differentially methylated CpG sites spanning a short region (den Hollander et al., 2015). Indeed, differentially methylated regions (DMRs) have been invaluable for tissue characterization (Song et al., 2005), with broad applications in developmental (Song et al., 2009) and aging studies (Bell et al., 2012). Importantly, DMRs have been used as markers for cancer diagnosis and progression (Ushijima, 2005). Yet, the majority of genome-wide OA methylation studies have focused on single-nucleotide CpG sites, rather than their clusters.

Here, reduced representation bisulfite sequencing technique (RRBS) was used to search for novel differentially methylated CpG sites and DMRs between visually damaged and non-damaged knee cartilage. An advantage of this method over array-based methods is that it extracts methylation information directly from the converted DNA sequence rather than read counts. Direct genome sequencing also provides more complete and unbiased genomic coverage with higher accuracy, even in comparison to the most advanced high-density gene arrays such as the Infinium 450 Beadchip array (Illumina), which detects only 1.5% of CpGs in the human genome (Bock, 2012). The expression of genes associated to the identified DMRs was further evaluated by qPCR and cell culture experiments. The overarching goal of this work was to generate and prioritize hypotheses that can be directly investigated by mechanistic and functional manipulations.

2. Methods

2.1. Sample collection

Articular cartilage samples were collected during total knee arthroplasties prompted by late-stage OA. Paired samples from 10 patients were collected from sections of visually damaged ("OA damaged (D)") and visually intact ("non-damaged (ND)") cartilage on femoral condyles. Five additional patients were sampled for data validation purposes. Tissues were frozen and stored at −80 °C until DNA/RNA isolation. All surgeries were conducted at Mayo Clinic (Rochester, MN) and all tissues were de-identified according to approvals granted by the Mayo Clinic Institutional Review Board.

2.2. DNA isolation and preparation for sequencing

DNA was isolated from mortar-and-pestle-ground frozen specimens with the DNeasy Kit (Qiagen). After isolation, DNA was adenylated and column purified using a 5′ DNA adenylation kit (New England Biolabs). Illumina adapters (Illumina) were ligated to DNA using T4 DNA ligase. Purified, ligated DNA was treated with sodium bisulfite and purified once more using AMpure XP beads (New England Biolabs). The final libraries from RRBS were prepared for sequencing as per the manufacturer's instructions in the Illumina cBot and HiSeq Paired end cluster kit v.3. The samples were placed onto seven lanes of a paired-end flow cell at concentrations of 7–8 pM and the control sample, PhiX, was placed in the eighth lane to allow the sequencer to account for the unbalanced representation of cytosine bases. The flow cell was then loaded into the Illumina cBot for generation of cluster densities. After cluster generation, flow cells were sequenced as 51 × 2 paired end reads using Illumina HiSeq 2000 with TruSeq SBS sequencing kit v.3. Data were collected using HiSeq data collection v.1.5.15.1 software, and the bases were called using Illumina's RTA v.1.13.48.

2.3. Sequence alignment and quality control

The raw sequence FASTQ files obtained from the sequencing instrument were processed using a Streamlined Analysis and Annotation Pipeline for Reduced Representation Bisulfite Sequencing (SAAP-RRBS) (Sun et al., 2012). Initial data processing included the removal of adaptor sequences from sequencing files, which were then mapped to a reference human genome (version Hg19) using the bisulfite sequence mapping program BSMAP (Xi and Li, 2009). All samples had >50% aligned reads and only the status of CpG sites with >5× coverage were considered for downstream analyses. All samples had similar methylation patterns, presenting a bimodal distribution (Supplementary Fig. 1S). In addition, CpG call rates were calculated for all samples to verify that they surpassed an 80% call rate threshold (Supplementary Fig. 2S).

2.4. Differential methylation

CpGs with consistent methylation ratios of either 0 or 1 across all samples were removed from the datasets. Differentially methylated CpG sites (DMSs) between OA and non-OA tissues were detected using paired t statistics (p values adjusted after Bonferroni correction). Both differentially methylated individual CpG sites (DMS) and differentially methylated regions (DMRs) in the genome were identified. To identify DMRs, we applied a sliding-window approach with a window size of 500 bp and a variable slide step (depending on next CpG location) across identified significant DMS ($p \leq 0.05$) between the sample groups. Windows that contained more than two CpG sites and were supported with at least five reads were kept and merged to define a DMR. The p values were averaged across the CpG sites within a region and were used to weight the significance of DMRs. The resulting gene lists associated with DMRs were analyzed using the