ARCHIVES OF ORAL BIOLOGY 58 (2013) 42–49

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: http://www.elsevier.com/locate/aob

Leader genes in osteogenesis: a theoretical study

Bruno Orlando[a,c,*], Luca Giacomelli[b], Massimiliano Ricci[a], Antonio Barone[c], Ugo Covani[c]

[a] Laboratories of Biophysics and Nanobiotechnology, Department of Medical Science, University of Genova, Italy
[b] Istituto Stomatologico Tirreno, Lido di Camaiore, Italy
[c] Department of Surgery, University of Pisa, Italy

ARTICLE INFO

Article history:
Accepted 22 July 2012

Keywords:
Osteogenesis
Genes
Bone formation
Bioinformatics

ABSTRACT

Little is still known about the molecular mechanisms involved in the process of osteogenesis. In this paper, the leader genes approach, a new bioinformatics method which has already been experimentally validated, is adopted in order to identify the genes involved in human osteogenesis.

Interactions among genes are then calculated and genes are ranked according to their relative importance in this process. In total, 167 genes were identified as being involved in osteogenesis. Genes were divided into 4 groups, according to their main function in the osteogenic processes: skeletal development; cell adhesion and proliferation; ossification; and calcium ion binding. Seven genes were consistently identified as leader genes (i.e. the genes with the greatest importance in osteogenesis), while 14 were found to have slightly less importance (class B genes). It was interesting to notice that the larger part of leader and class B genes belonged to the cell adhesion and proliferation or to the ossification subgroups. This finding suggested that these two particular sub-processes could play a more important role in osteogenesis. Moreover, among the 7 leader genes, it is interesting to notice that RUNX2, BMP2, SPARC, PTH play a direct role in bone formation, while the 3 other leader genes (VEGF, IL6, FGF2) seem to be more connected with an angiogenetic process. Twenty-nine genes have no known interactions (orphan genes).

From these results, it may be possible to plan an ad hoc experimentation, for instance by microarray analyses, focused on leader, class B and orphan genes, with the aim to shed new light on the molecular mechanisms underlying osteogenesis.

1. Introduction

Skeletal biology is characterized by two main features. The first feature is that, unlike most organs, such as the liver, heart, and brain, the skeleton is not confined to one structure or one location in the body but is composed of more than 200 elements spread throughout the body; moreover, every one of these elements has a unique shape. The second characteristic of the skeleton is that it is an organ made up of two distinct tissues in dynamic equilibrium, and of three specific cell types, chondrocytes in cartilage, osteoblasts and osteoclasts in bone.

In most skeletal elements, bone formation occurs through endochondral ossification, a process in which cells of the mesenchymal condensations differentiate into chondrocytes that form a cartilage template for future bone. Subsequently, this template is replaced following vascular invasion by bone cells. In a minority of skeletal elements, such as part of the clavicle and part of the skull, cells of the mesenchymal condensation differentiate directly into bone-forming osteoblasts, a process called intramembranous ossification.[1]

---

[*] Corresponding author at: Via Aurelia 335, Lido di Camaiore (LU), Italy. Tel.: +39 0584 6059889.
E-mail address: b_orlando@virgilio.it (B. Orlando).

0003–9969/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.archoralbio.2012.07.010

Furthermore, different cell types play a different role in skeletal biology. Chondrocytes are required for longitudinal growth and synthesize the cartilaginous scaffolds on which osteoblasts can deposit the bone matrix in the endochondral ossification model. In few other bones, mesenchymal cells differentiate into osteoblasts without any cartilaginous intermediary step (intramembranous ossification). The most important function of osteoblasts is to form mineralized bones, in the process called osteogenesis. This is a highly controlled developmental process in which numerous extrinsic factors, including hormones and growth factors, activate osteoblast-specific signalling proteins and transcription factors (TFs) required for osteoblast differentiation. The function of the osteoclasts is to reabsorb the mineralized matrix, allowing this rigid tissue to remodel.[2–5]

Our knowledge of genetic bases regulating the differentiation of osteoclasts and chondrocytes is particularly advanced. In contrast with these two lineages, until recently, little was known about osteoblast lineage. Therefore, considering the high degree of complexity in the various processes, a key question in skeletal biology is to identify the genes and genetic pathways that regulate osteogenesis and skeletal patterning, i.e. the definition of the shape and the location of a particular skeletal element. It is only with the characterization of the function of one transcription function, *Cbfa1* (core binding factor alpha 1) also called *Runx2* (runt-related gene 2), that our knowledge has made significant progress. However, *Runx2* is not sufficient for osteogenesis to take place, since osteoblast differentiation and skeletal development cannot be induced by *Runx2* alone.[6] This relative paucity of transcription factors known to affect osteoblast differentiation, contrasts with the role of these factors in other cell differentiation processes. Therefore, it is likely that other transcription factors affecting osteoblasts differentiation and proliferation, positively or negatively, remain to be identified.

In recent years, the leader gene approach was proposed, and it has already given very promising results when applied to human T lymphocytes cell cycle.[7] This bioinformatics ab initio method is based on the study of gene networks, as derived from sophisticated web-available software, such as STRING (Search Tool for the Retrieval of Interacting Genes, Heidelberg, Germany)[8] and the subsequent consideration that a very low number of genes are actually playing the most important role in a particular cellular process. In particular, the leader gene approach allows the identification, through advanced data-mining techniques, of every gene involved in a given process. The genes are then ranked according to their number of interactions with the other identified genes, in order to recognize those genes with the highest number of interactions, termed ''leader genes''. Moreover, it is possible to draw and analyse maps of interactions among genes and therefore, a detailed qualitative (interaction maps) and quantitative (ranking) analysis of every gene involved in a given process.

For instance, the application of this algorithm to human T lymphocytes cell cycle was successful in identifying only 6 genes, among the several hundred involved, with a significantly higher number of interactions if compared to the others. Due to their very high number of interactions, they were supposed to play a central role in the regulation of the cell cycle in human T lymphocytes. It was also very interesting to notice how different leader genes are actually involved in different key points of the cell cycle. An experimental analysis,[9,10] which analysed human T lymphocytes cell cycle through time by means of microarray technology and of a new, simple technology to acquire microarray images, the DNA-SER[11] fully confirmed the bioinformatics predictions.

The leader gene approach, validated by experimental analysis on a model system, can suggest a more rational approach to experimental techniques and methods, as DNA microarray. This experimental methodology, which allows the study of a whole genome with only a few experiments, is largely used in mass-scale molecular genomics, but results are often altered by experimental complication, due to the very large amount of genes displayed on the array. Targeted arrays, dedicated to a particular cellular process, are commercially available, but the number of genes displayed on these arrays is still quite high. The application of bioinformatics studies and the identification of leader genes can predict the most important genes in a particular cellular process. In this way, it becomes possible to design smaller microarrays, which display only the most interesting genes for a specific cellular process and thus are much easier to interpret. These arrays are supposed to display a very low number of genes (from 20 to 200), therefore reducing experimental noise and complications and targeting analysis only of those genes which are theoretically predicted to be the most important ones.

Moreover, through leader genes identification and the subsequent analysis of interaction maps, it becomes also possible to identify some potentially promising pharmaceutical targets, by going deep into biochemical processes at a molecular level. In addition, the basic concept of ranking genes according to their number of interactions in a given cellular process can also lead to the identification of genes with no known interactions (''orphan genes''), which should be studied with a targeted experimental analysis in order to investigate their biological function more deeply.

In this study, the leader gene approach was applied to the osteogenic process, in order to identify the most important genes involved in this process and to discuss their role in osteogenesis.

2. Materials and methods

The leader gene approach was described in detail elsewhere.[7] In brief, this method can be summarized in three steps:

2.1. *Identification of gene set*

Several search strategies were implemented and iteratively repeated until the newly identified genes coincided with the previous search. In particular, the research was performed in several databases, such as PubMed, GeneBank, Geneatlas,[12] Genecards[13] and commercial microarray websites, such as www.superarray.com.

The search was performed in the above-mentioned databases by operating Boolean logic, with proper operators (*AND*, *OR*) and on pertinent keywords, such as ''osteogenesis'', ''ossification'', ''skeletal development'', ''gene'', in all sequences and possible permutations. In addition, in order