Minireview

SAGE profiling of the forelimb and hindlimb

Malcolm Logan

Address: Division of Developmental Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.
E-mail: mlogan@nimr.mrc.ac.uk

Published: 27 February 2002

Genome ***Biology*** 2002, **3(3)**:reviews1007.1–1007.3

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/3/reviews/1007

Abstract

A recent study has used serial analysis of gene expression to compare mouse forelimb and hindlimb gene-expression profiles. The method successfully identified known regulators of limb identity and has generated a candidate set of differentially expressed genes that may regulate limb identity.

Patterning of tissues during embryonic development requires the coordination of cell-fate specification with rates of cell proliferation and apoptosis. The developing vertebrate limb is a favored model system for studying the fundamental processes involved in normal embryo development. The developing limb bud consists of multiple cell types that ultimately form the intricate structure of interconnected bones, tendons, muscles and nerves that make up the mature limb. From initial embryological observations and manipulations and more recent molecular techniques, we are beginning to understand the genetic control of limb development.

The vast majority of genes known to play a role in limb development are expressed in apparently identical patterns in both the forelimb and hindlimb. In the forelimb, these genes coordinate the patterning of characteristic structures such as fingers, whereas in the hindlimb the same genes lead to the formation of toes. It is only relatively recently that candidate genes have been identified that control this differential response. The limb expression pattern of *Tbx5*, which encodes a T-box transcription factor, is restricted to the developing forelimb, whereas the closely related T-box gene *Tbx4* and the gene encoding the homeodomain transcription factor Pitx1 are restricted to the developing hindlimb [1-4]. Functional studies have determined that *Tbx5*, *Tbx4* and *Pitx1* regulate limb-type identity [5,6].

Innis and colleagues [7] have recently reported the use of serial analysis of gene expression (SAGE) to generate a comprehensive gene-expression profile of mouse forelimbs and hindlimbs and to allow identification of genetic differences between them. SAGE involves the production of short sequence tags of around 10 base-pairs (bp) from cDNAs derived from the tissue of interest. Each tag contains sufficient information to identify it as representing a particular transcript. Many transcript tags are concatenated and cloned to generate a SAGE library. Library clones are then sequenced to reveal the identity of multiple tags simultaneously [8]. The presence and abundance of tags representing a particular gene in a SAGE library provides both a qualitative and a quantitative readout of gene expression in the original source. Moreover, the comparison of tags from one library to another can, in the most comprehensive examples, provide accurate gene-expression profiles and enable the genetic regulators of cellular difference to be determined.

The forelimb and hindlimb share many similar features. For example, the hand and foot, wrist and ankle and forearm and shank can be considered homologous structures in the forelimb and hindlimb, respectively. As might be expected from a comparison of homologous structures, and in agreement with the vast majority of functional and expression analyses completed to date, the comparison of forelimb and hindlimb SAGE libraries [7] indicated that most genes were similarly expressed in the two developing limb types. Over 90% of the SAGE tags displayed a two-fold or smaller difference in representation, and only 0.2% of tags were differentially represented with statistical significance. All of these 0.2% of tags represented either *Pitx1* or *Tbx4* and were restricted to the hindlimb library, consistent with previously published gene-expression data. The majority of differentially expressed SAGE tags were present at very low frequency and therefore

could not be analyzed accurately by statistical methods or by fold difference analysis; *Tbx5* and four *Hox* genes previously demonstrated to be differentially expressed in forelimbs and hindlimbs were not found to be differentially expressed with statistical significance. When the SAGE tags corresponding to these particular genes were identified, however, they were found only in the SAGE library that would be expected from previous expression analyses.

Low tag frequency in the libraries makes conclusive identification of forelimb-specific or hindlimb-specific genes difficult. A list of 251 tags were found that were expressed with the same fold difference as *Tbx5* [7]; the genes corresponding to these tags may have a similar differential gene expression pattern in the limb or they may represent genes expressed in both forelimbs and hindlimbs. The authors tackled the problem of low tag frequency by carrying out a 'virtual' subtraction approach to identify the subset of genes expressed exclusively in the limb libraries. The tags from six previously published and publicly available non-limb SAGE libraries were pooled and compared with the forelimb and hindlimb SAGE libraries. This method enabled the authors to reduce significantly the list of candidate genes that might be differentially expressed.

After virtual subtraction [7], a list of candidate genes remained that should, in theory, contain limb-specific genes; it did indeed retain some genes known to be differentially expressed in the forelimb or hindlimb. The original list of 251 transcription factors and homologous ESTs was reduced to 43 potentially limb-specific genes following virtual subtraction. This significant reduction in the number of candidate genes increases the feasibility of subsequent expression analysis and functional studies. But this list contains some errors, and these have most probably occurred because of the lack of depth in sequencing of SAGE tags (each tag is represented only a few times) and the low number of available SAGE libraries from other tissues. For example, the transcription factors *Mox2*, *Six2*, and *Dlx3* are in the list of limb-specific genes of Innis and colleagues [7], but all three are known to be expressed in regions outside the developing limb buds. Such errors could be corrected by the creation of a variety of larger and more completely sequenced SAGE libraries from various non-limb tissues. The authors point out, however, that given that SAGE sequencing efforts are demanding, resources could be used more effectively in verifying the SAGE-generated candidate list by other methods.

Despite the difficulty in identifying true limb-specific genes by this method, the identification of genes required to specify forelimb versus hindlimb development should not necessarily be restricted to limb-specific genes. In fact, *Tbx5*, *Tbx4*, and *Pitx1* are all expressed in non-limb tissues, where they also play important roles: for example, *Tbx5* is required for normal heart development [9]. Therefore, some genes that potentially regulate forelimb versus hindlimb development may have been lost during the virtual subtraction step because these genes are also expressed in other tissues in addition to the limbs. *Tbx5* would have been eliminated from the final list of genes, for example, if a SAGE library made from heart tissue had been used for the virtual subtraction. Virtual subtraction could thus be a valuable tool for reducing the numbers in SAGE-generated candidate gene lists, although it needs to be applied judiciously.

The analysis by Innis and colleagues [7] is the first attempt to compare directly the expression profiles of forelimbs and hindlimbs. The approach was successful in identifying factors that had previously been identified as limb-type-specific and has generated a list of candidate differentially expressed genes worthy of further investigation using other methods. Importantly, the SAGE data generated in this study will become more informative as the number and content of mouse EST databases expand, as the mouse genome sequence reaches completion, and as the number of SAGE libraries increases.

This effort has generated SAGE libraries of developing mouse limbs from embryonic day 11.5, providing a transcriptional 'snapshot' of forelimb and hindlimb development. To better understand the distinct transcriptional programs controlling forelimb and hindlimb development, it will be necessary to produce a developmental time course of SAGE libraries, starting at limb-bud initiation stages and continuing through limb formation.

This type of SAGE analysis has been applied to the study of other developmental processes and disease states, most recently with impressive effect to the study of retina development and retinal disease genes [10]. Strikingly, of the 264 genes identifed as specific to or highly enriched for rod photoreceptor cells, 86 had human orthologs that mapped to 37 different genetic loci associated with retinal disease. Some of the human orthologs of the differentially expressed genes identified in the study by Innis and colleagues [7] may similarly map to disease loci associated with limb abnormalities. Identifying the genetic mutations responsible for human congenital limb defects will undoubtedly be assisted by current, and continued, gene-expression analyses in the limb.

Although we know from this and other earlier studies [1-7] that some genes are differentially expressed between the forelimb and hindlimb, the number of genes expressed exclusively in either the forelimb or hindlimb may not be large. A relatively unexplored area is the potential for differential post-transcriptional control in achieving the subtle differences between the limbs. There is evidence for conserved translational regulation of the patterning transcription factor HoxC6, for example: although transcripts are found in both forelimbs and hindlimbs, the protein product is expressed exclusively in the forelimb (or fin) of mouse, chick, *Xenopus* and zebrafish [11-13]. The characteristic

morphologies of the forelimb and hindlimb are unlikely to be generated solely by differential gene expression, but also by differences in expression levels or subtle differences in the timing of gene expression. Whereas differential screening approaches such as library subtraction or differential display would fail to identify genes with subtle differences in expression profiles, SAGE analysis should allow detection of both qualitative and quantitative differences in gene expression. As the results of whole-genome sequencing continue to flood into databases the need for good annotation of this raw data becomes all the more pressing. The challenge is shifting away from cataloging genes and towards understanding when and where all these genes are expressed, as a first step toward understanding their function in normal development and physiology.

References

1. Gibson-Brown JJ, Agulnik SI, Silver LM, Papaioannou VE: **Involvement of T-box genes Tbx2-Tbx5 in vertebrate limb specification and development.** *Development* 1998, **125:**2499-2509.
2. Isaac A, Rodriguez-Esteban C, Ryan A, Altabef M, Tsukui T, Patel K, Tickle C, Izpisua-Belmonte JC: **Tbx genes and limb identity in chick embryo development.** *Development* 1998, **125:**1867-1875.
3. Logan M, Simon HG, Tabin C: **Differential regulation of T-box and homeobox transcription factors suggests roles in controlling chick limb-type identity.** *Development* 1998, **125:**2825-2835.
4. Gibson-Brown JJ, Agulnik SI, Chapman DL, Alexiou M, Garvey N, Silver LM, Papaioannou VE: **Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity.** *Mech Dev* 1996, **56:**93-101.
5. Rodriguez-Esteban C, Tsukui T, Yonei S, Magallon J, Tamura K, Izpisua Belmonte JC: **The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity.** *Nature* 1999, **398:**814-818.
6. Logan M, Tabin CJ: **Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity.** *Science* 1999, **283:**1736-1739.
7. Margulies EH, Kardia SL, Innis JW: **A comparative molecular analysis of developing mouse forelimbs and hindlimbs using serial analysis of gene expression (SAGE).** *Genome Res* 2001, **11:**1686-1698.
8. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270:**484-487.
9. Bruneau BG, Nemer G, Schmitt JP, Charron F, Robitaille L, Caron S, Conner DA, Gessler M, Nemer M, Seidman CE, Seidman JG: **A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease.** *Cell* 2001, **106:**709-721.
10. Blackshaw, S, Fraioli RE, Furukawa T, Cepko CL: **Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes.** *Cell* 2001, **107:**579-589.
11. Molven A, Wright CV, Bremiller R, De Robertis EM, Kimmel CB: **Expression of a homeobox gene product in normal and mutant zebrafish embryos: evolution of the tetrapod body plan.** *Development* 1990, **109:**279-288.
12. Nelson CE, Morgan BA, Burke AC, Laufer E, DiMambro E, Murtaugh LC, Gonzales E, Tessarollo L, Parada LF, Tabin C: **Analysis of Hox gene expression in the chick limb bud.** *Development* 1996, **122:**1449-1466.
13. Oliver G, Wright CV, Hardwicke J, De Robertis EM: **A gradient of homeodomain protein in developing forelimbs of *Xenopus* and mouse embryos.** *Cell* 1988, **55:**1017-1024.