Methods xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Methods

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

A guide for building biological pathways along with two case studies: hair and breast development

Daniel Trindade [a,1], Lissur A. Orsine [a,1], Adriano Barbosa-Silva [b], Elisa R. Donnard [c], J. Miguel Ortega [a,⇑]

[a] Depto. de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-010, Brazil
[b] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette 4362, Luxembourg
[c] Centro de Oncologia Molecular, Instituto de Ensino e Pesquisa, Hospital Sírio-Libanês, São Paulo, SP 01308-060, Brazil

article info

Article history:
Received 5 May 2014
Received in revised form 26 August 2014
Accepted 3 October 2014
Available online xxxx

Keywords:
Pathway
PubMed
PESCADOR
Hair development
Breast development

abstract

Genomic information is being underlined in the format of biological pathways. Building these biological pathways is an ongoing demand and benefits from methods for extracting information from biomedical literature with the aid of text-mining tools. Here we hopefully guide you in the attempt of building a customized pathway or chart representation of a system. Our manual is based on a group of software designed to look at biointeractions in a set of abstracts retrieved from PubMed. However, they aim to support the work of someone with biological background, who does not need to be an expert on the subject and will play the role of manual curator while designing the representation of the system, the pathway. We therefore illustrate with two challenging case studies: hair and breast development. They were chosen for focusing on recent acquisitions of human evolution. We produced sub-pathways for each study, representing different phases of development. Differently from most charts present in current databases, we present detailed descriptions, which will additionally guide PESCADOR users along the process. The implementation as a web interface makes PESCADOR a unique tool for guiding the user along the biointeractions, which will constitute a novel pathway.

1. Introduction

Most of the genomic information studied to date is underlined in the format of biological pathways [1]. In its essence, a typical biological pathway to be promptly remembered in biochemistry is the energy metabolism glycolysis pathway. A personal communication of Minoru Kanehisa, founder of KEGG [2] (www.genome.jp/kegg), in a pizza restaurant in São Paulo, revealed the starting motivation for this database: to represent Escherichia coli glycolysis and other metabolic pathways from other microbes that were being sequenced, in the year of 1995 [3]. As most readers will be aware, the KEGG database kept on creating pathway-like charts and today, an inspection of these provides great help with the interpretation of some phenomena of interest to the user. For instance, people interested in comparative genomics can, even before they start sequencing their favorite genome, begin inspecting, let's say, DNA repair mechanisms (KEGG pathway maps: map03410, map03420 and map03430) which will be either present or absent in a species related to the organism of choice. Other people perhaps interested in comparing gene expression in two situations, can determine which pathways are enriched or moderated after the treatment. Actually, the current interpretation for KEGG and other pathway databases is the representation of systems, therefore contributing for the study of systems biology.

What are the non-usual pathways already depicted in pathway databases? Reactome (named after a collective for reactions, like genome, proteome, etc.) is a database of pathways [4] authored by expert biologists, following their interests and expertise, standardized and integrated with other databases in collaboration with the Reactome editorial staff. Their homepage states: "The rationale behind Reactome is to convey the rich information in the visual representations of biological pathways familiar from textbooks and articles in a detailed, computationally accessible format". In Reactome one can find charts such as: (i) Circadian Clock, comprising 99 proteins; (ii) Reproduction–Fertilization, with 28 proteins; (iii) under the disease category, HIV infection, subdivided in the

Abbreviations: PMID, PubMed ID; LCA, last common ancestor; HF, hair follicle; MX, matrix; MD, medulla; ORS, outer root sheath; IRS, inner root sheath; DP, dermal papilla; CX, cortex; CL, cuticle; EMT, epithelial to mesenchymal transition.

⇑ Corresponding author.
E-mail addresses: daniel-trindade@c-bio.grad.ufmg.br (D. Trindade), lissurorsine@gmail.com (L.A. Orsine), adriano.barbosa@uni.lu (A. Barbosa-Silva), edonnard@mochsl.org.br (E.R. Donnard), miguel@icb.ufmg.br (J.M. Ortega).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ymeth.2014.10.006
1046-2023/ó 2014 Elsevier Inc. All rights reserved.

HIV life cycle, with 340 proteins, and host interactions of HIV factors, with 170 proteins. Therefore, the trend that started out with E. coli glycolysis back in 1995 is moving further.

A third resource deserving of mention is WikiPathways [5], established to facilitate the contribution and maintenance of pathway information by the biology community. In there one can find charts on human processes including spinal cord injury, codeine and morphine metabolism, and cardiac progenitor differentiation, with respectively 118, 9 and 53 gene products.

Thus, building biological pathways is an ongoing demand and might largely benefit from a method for extracting information from biomedical literature with the aid of text-mining tools [6]. For the purpose of designing biological pathways, several computational representation models have been developed: (i) markup languages such as KGML (www.kegg.jp/kegg/xml/) used in KEEG database; (ii) universal models such as Biological Expression Language (BEL, www.openbel.org); (iii) or even completely dedicated and inter-operational languages as is the case of Systems Biology Markup Language (SBML, www.sbml.org). The biological information represented using these models might be loaded to downstream applications for visualization, edition or expansion. For that purpose, software such as Cytoscape [7], CellDesigner [8] or PathVisio [9] are broadly used by the pathways design community.

Here we hopefully guide you in the attempt of building a customized pathway or chart representation of a system. Our manual is based on a group of software designed to look at biointeractions in a set of abstracts extracted from PubMed. However, they aim to support the work of someone with good biological background, who will play the role of manual curator while designing the representation of the system. The interesting characteristic of this approach is that the curator does not need to be an expert on the subject, but otherwise interested in depicting the phenomenon for further studies. We therefore illustrate with two cases. They were chosen to focus on recent acquisitions of human evolution. If one has information on the distribution of orthologues of the listed genes, it is possible to depict the acquisition of genes along evolution that compose the system or process of interest [10,11].

This manual will guide you along with the necessary care to address finding a useful PubMed query, inspecting a comprehensive and not so large set of abstracts and at last, mining gene names associated by biointeraction terms. Further than obtaining an interaction network, this manual aims to teach how to attain a chart for the pathway of desire. Additionally, we will comment on some approaches that might lead to the definition of the evolutionary history of such a pathway.

The approach suggested here has been undertaken in previous work. Our first case study was instigated by a desire to build a KEGG-like pathway for the preimplantation embryo development [10]. We could summarize what happens in the embryo Inner Cell Mass (ICM), where the phosphorylation of Yap transcription coactivator releases it from the association with Tead4, thus avoiding the induction of the Cdx2 kinase. This same kinase is expressed in the trophectoderm, and represses Nanog and Oct4, therefore releasing the action of genes implicated in the differentiation of this follicle. Remarkably, Nanog is more ancient, with homologues shared with all Coelomata organisms including Drosophila melanogaster, while Oct4 humans share just with fish and other Euteleostomi organisms, so it is a recent acquisition of this system [10]. Another previous case was included in the publication of PESCADOR software [12], used in this tutorial. In it, an existing KEGG pathway, Colorectal Cancer, was roughly doubled with additional genes, therefore suggesting that users interested on already published pathways might place some effort on supplementing it.

Here we give all tips that we use in our procedure and present two cases representing challenging subjects, phenomena restricted to recent epochs of human evolution, which are not covered by charts in the available literature.

2. Methods

The initial step for text-mining pathway construction is selecting a set of 3–10 relevant references on the subject of interest. These references should contain highly relevant gene regulation descriptions and will serve as the basis for the subsequent query. For the Hair Development pathway, the initial list consisted of 10 references (PMIDs: 12533516, 11751679, 14962104, 20944652, 15630473, 22506014, 16481354, 15245425, 23602386, 10767081) and for the breast development pathway we selected seven references (PMIDs: 15351091, 17877612, 16807800, 16861925, 18947364, 15351091 and 15886886). Considering that the curator may not be an expert on the subject, this step might be repeated as he/she faces seminal abstracts along the project, including them and repeating the procedure. The best approach for a non-expert in selecting these reference articles is searching for them in the citations of the major reviews in the field. The major reviews will most likely encompass the main interactions established by different research groups, which will form the basis for the abstract ranking step (see below).

The next step is extracting the relevant literature from the PubMed database (www.ncbi.nlm.nih.gov/pubmed). The query should be carefully selected and exclude possibly misleading results that will hinder the text mining approach. The queries used for the two pathways described here were "(hair AND follicle) NOT (ear OR auditory)" and "(breast AND development) NOT cancer". Another suggestion is to start with only few "seed" articles and employing NCBI PubMed's existing "Related Articles" search functionality. The list of PubMed IDs (PMIDs) resulting from the PubMed search should be retrieved as a text file. We obtained 6452 and over 40000 PMIDs for our two queries, respectively.

An efficient and appropriated (i) choice of relevant references and (ii) PubMed query as well is expected to produce many abstracts with marked genes in the final step, while using PESCADOR (below). If this does not occur, the user can adjust both procedures.

The results from the literature search can be overwhelming and difficult to prioritize. The next step makes use of the Medline Ranker software (cbdm.mdc-berlin.de/tools/medlineranker/) to classify the full set of references based on their relevant information content. At this point the initial list of 3–10 references is used as input for the training set. The background or test set is the full PMID list retrieved from your PubMed query. Medline Ranker was used to classify the results from both queries and 1000 references with higher scores (p-value < 0.01) were selected as input to the second software PESCADOR (cbdm.mdc-berlin.de/□pescador/). PESCADOR adds web interface and several features on top of LAITOR software [13] designed to depict biointeractions between terms (genes).

The engine underneath PESCADOR, LAITOR, was designed to be used as command line software, and it depicts co-occurrence analysis of terms (either genes or concepts) with a dictionary of biointeractions (such as induces, represses, regulates, etc.). Co-occurrences are classified into four types: (i) type 1, when the structure is term 1 – biointeraction – term 2, in a single sentence; (ii) type 2, when both terms are in the same sentence but the biointeraction is not between them; (iii) type 3, when both terms are in the same sentence, but no biointeraction is found in the internal dictionary of biointeractions and (iv) type 4, when two terms are found in the abstract, in different sentences.

The PESCADOR online platform allows the user to input not only the list of PMIDs but also customized biological concepts that

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

facilitate the curation process. Therefore, not only biointeractions connecting two genes or proteins are highlighted, but adding "abscisic acid" as a concept, for instance, will lead to identifying co-occurrences such as "Exogenous abscisic acid (ABA) induced the alcohol dehydrogenase gene (Adh) in Arabidopsis roots". Moreover, by adding "trophectoderm" as a concept, one can retrieve co-occurrences such as "GATA3 is selectively expressed in the trophectoderm". Since a concept is treated as gene or protein, one can add as concept a gene name that is absent in official databases.

PESCADOR results can be visualized in different manners, as a starting point, the highlighted abstract list provides an easier view and allows for faster validation of the biointeractions. Conveniently, abstracts comprising terms (genes, proteins or concepts) appear highlighted in green, facilitating the navigation along informative abstracts. The validation is the stage that will require biological knowledge, and all interactions must be carefully verified in the text. In the real world, users pencil a draft sub-pathway while conducting the analysis. PESCADOR eases the navigation along the borders of the sub-pathway, since one can quickly jump to its genes page. It is recommended to start with type 1 biointeractions, then moving to other types. In our experience, much relevant information is retrieved with abstracts containing type 4, where two genes are just in the same abstract but in different sentences.

For the two pathways presented here, an initial step of retrieving interactions from the PESCADOR result is illustrated in Fig. 1 and Supplementary Fig. 1. In the case of the mammary gland development pathway (Fig. 1), the first abstract [14] reveals a type I interaction that indicates the activation of STAT5 by MYC. The abstract also indicates the downregulation of CAV1 by MYC (although the gene name in this form was not recognized by the NLProt dictionary) and these first interactions are selected for drawing the pathway. With the second abstract shown [15] another small hub of interactions is aggregated to the growing pathway, with USF2 leading to an increase in the EIF4E and EIF4G proteins, as well as inducing higher levels of oxytocin. The same logic is followed as the reader analyzes more abstracts in the PESCADOR tool and validates their interaction. The start of construction for the hair development pathway is shown in Supplementary Fig. 1.

After selecting enough information for an initial sub-pathway comprised of several small modules, the interactions can be expanded by parsing through the PESCADOR results with the "Terms" page, which provides a list of all gene symbols found in the abstracts and links them to the biointeractions. The results after validation can be downloaded as an excel file, altered if necessary to include relevant information and used as input for the pathway design software PathVisio (www.pathvisio.org). The input files used for both pathways generated as case studies and presented here can be accessed in the Appendices Table A (for hair development) and Table B (for breast development). While designing the pathway in PathVisio or any presentation software, often some areas of the growing pathway turn out to be overcrowded, so here is a hint. The same tables created to register biointeractions can be used as input to Cytoscape (www.cytoscape.org), which can facilitate the separation of groups of genes involved in the same function and the organization of the pathway landscape. However, we have found PathVisio more useful for producing the final pathway.

PESCADOR matches the terms against a dictionary of genes present in the species which was declared at the setup, to avoid presenting in the gene page annoying synonyms that stand for completely different genes from other organisms, however, this might reduce the marking for species with little information in the Gene database (www.ncbi.nlm.nih.gov/gene). To correct this, one can press the link "highlight all NLPROT terms" to mark all possible gene names.

Building of a pathway will require manual curation along the process. PESCADOR provides a link to UniProt accession for the terms labeled by the NLProt algorithm [16] and filtered against Gene database [17] with the list of Gene entries of the chosen organism, avoiding the known occurrence of synonyms in different species. However, while manually checking UniProt, the curator should pay attention to the existence of subunits (alpha, beta, etc.) and save the accession, which refers to a complete sequence and to a SwissProt revised version, if present.

The co-occurrence to be reported in the pathway will also suffer interference of manual curation. It is possible that the bio-interaction is either inferred upon reading or maybe there is a more standardized way of reporting it. Therefore, together with the pathway, we recommend building a table containing the two UniProt accessions, the edited biointeraction (using standardized words that will easily refer to the connections in the pathway), and the PubMed ID of the most informative abstract for this interaction. This table can be provided as Supplementary material for readers to inspect the work. Thus, precision is improved by curation and can be promptly inspected.

As exemplified with the case studies, it might be convenient to divide the analysis into separate developmental phases, or stages of a response. If necessary, distinct PESCADOR processes may be triggered. Case studies were prepared, each one by an undergrad student on biological sciences, over two months of work, partial time. The number of distinct genes found, respectively for hair and breast development, was 122 and 85. In order to show the product in a didactic way, Section 1 presents the description of the pathways, which is recommended to accompany the chart.

Thus, the implementation as a web interface makes PESCADOR a unique tool for guiding the user along the biointeractions which will constitute a pathway, in some sense, acting as an agent which will interface with human curators to convert text-mined information into a product in the format of a pathway, instead of a network.

The elaborated pathway cannot be deposited in a pathway database, such as KEGG Pathway or, Reactome, but it can be sent to WikiPathways [5]. The pathways built for the case studies in this manuscript were deposited in WikiPathways. The first step for this deposit is to create a login, from which the user can opt to draw the pathway using the online editor or send a pathway previously elaborated in PathVisio. It is also possible to export a pathway directly from PathVisio to WikiPathways through a specific plugin. The graphical representation can be accompanied by a description and bibliographical references, as well as added information on the genes and gene products involved. Each pathway receives in the end an identifier allowing their visualization, editing or download by any WikiPathways user. For the induction, organogenesis and cytodifferentiation stages of hair follicle development the identifiers are, respectively, WP2804, WP2839 and WP2840. The mammary gland development pathways can be accessed through the identifiers WP2813, for the Embryonic Development stage, WP2814 for Puberty, WP2817 for pregnancy and lactation and WP2815 for involution.

After the pathway is completed, it is possible to reveal its evolution [10]. Different sources of orthologous groups are available, such as Kegg orthology (www.genome.jp/kegg/ko.html), or programs can be used to build one [18]. Then, by analyzing the taxonomic distribution with NCBI Taxonomy Common Tree (www.ncbi.nlm.nih.gov/taxonomy) or with scripts that deal with NCBI Taxonomy data, retrieved from the Taxonomy FTP, the last common ancestor (LCA) can be determined, as shown in Fig. 4 of Section 1. Briefly, starting with gene symbol marked in PESCADOR one can retrieve the protein ID of a complete amino acid sequence, retrieve the cluster from available databases or create it, determine the Taxonomic IDs of the clustered proteins and with them, find the node of Taxonomy Tree which is the father of the encountered

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

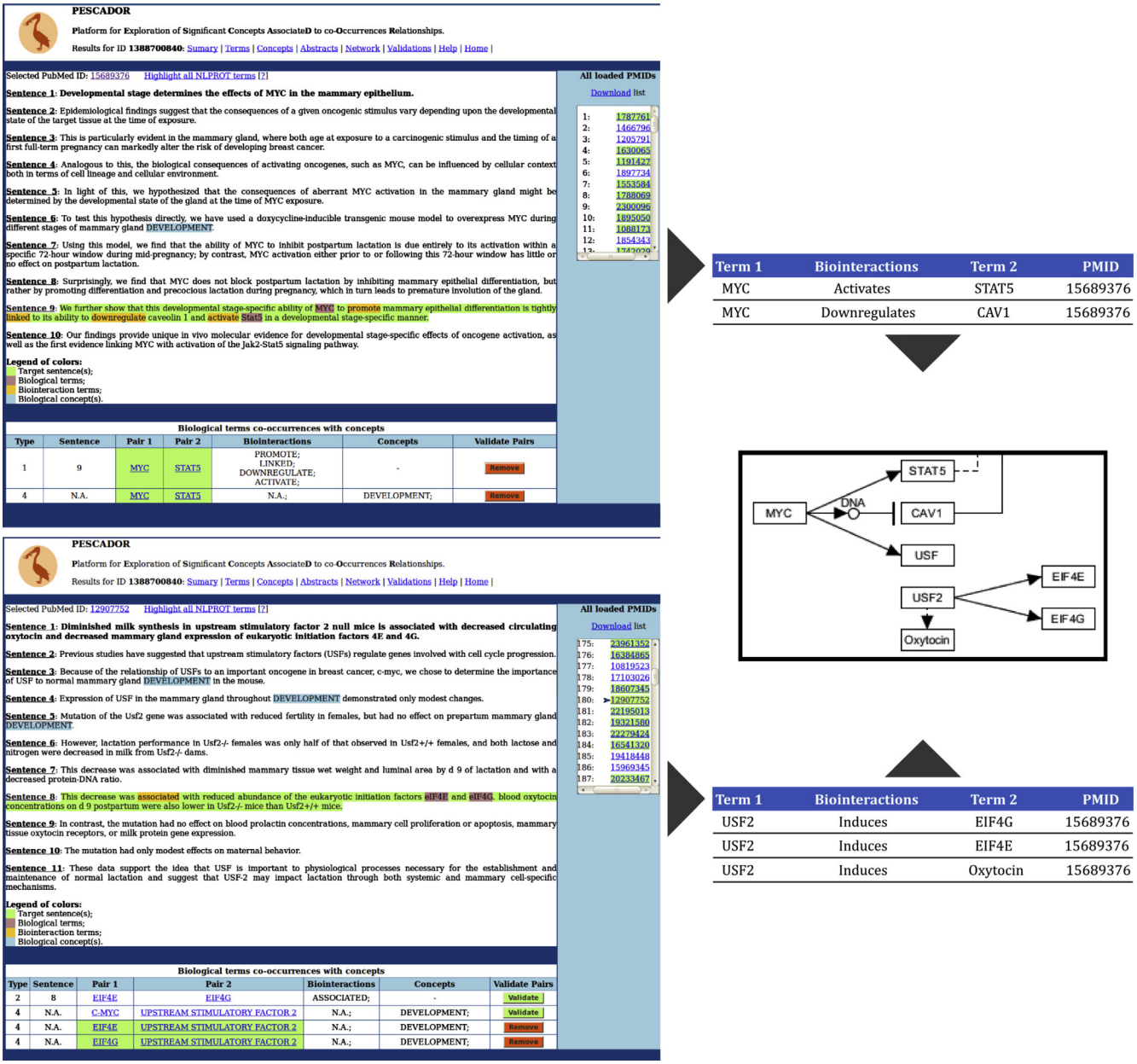

Fig. 1. Starting a pathway design. Two abstracts selected in the PESCADOR interface (left side) show the highlighted interactions found by the text-mining software. These interactions are validated and annotated in text tables, which can then be used as input for the PathVisio and other pathway building software. The result from the two abstracts selected is the small node of interactions shown in the middle right.

species. It might be a family, an order, or a so called "no rank" node of Taxonomy Tree (e.g. Amniota and Eutheria, both from the human lineage).

3. Results

Hair follicle and mammary gland development are processes dependent on epithelial–mesenchymal transitions orchestrated by many signaling pathways [19]. Here we describe the complex regulatory pathways assembled from a text mining approach portraying the development of each of these key mammal-specific structures.

3.1. Hair follicle development pathway

The hair follicle (HF) is the biological unit responsible for producing a single hair shaft. The follicles are arranged with concentric epithelial progenitor layers surrounding the dermal core, which is the dermal papilla (DP) [20]. Classically, the development of the follicle itself can be divided into three stages: induction, organogenesis and cytodifferentiation [21].

In order to try recognizing distinguishable patterns of regulatory interactions, a pathway was created for each of these stages, using the PathVisio software (Fig. 2A–C, respectively). Moreover, the first two pathway representations are divided into two cellular compartments, in an attempt to symbolize more precisely the morphogenesis, during which the dermis and epidermis maintain an intense interplay of regulatory signals [20].

However, the subsequent stage, cytodifferentiation, markedly increases the number of cell layers and regulatory interactions, thereby complicating our purpose of illustrating this process. So for this stage, representing cell compartments would be impractical. Thus, we decided to represent cytodifferentiation as a generalized and schematic phenomenon, in which the main characteristics or events of the process are emphasized.

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

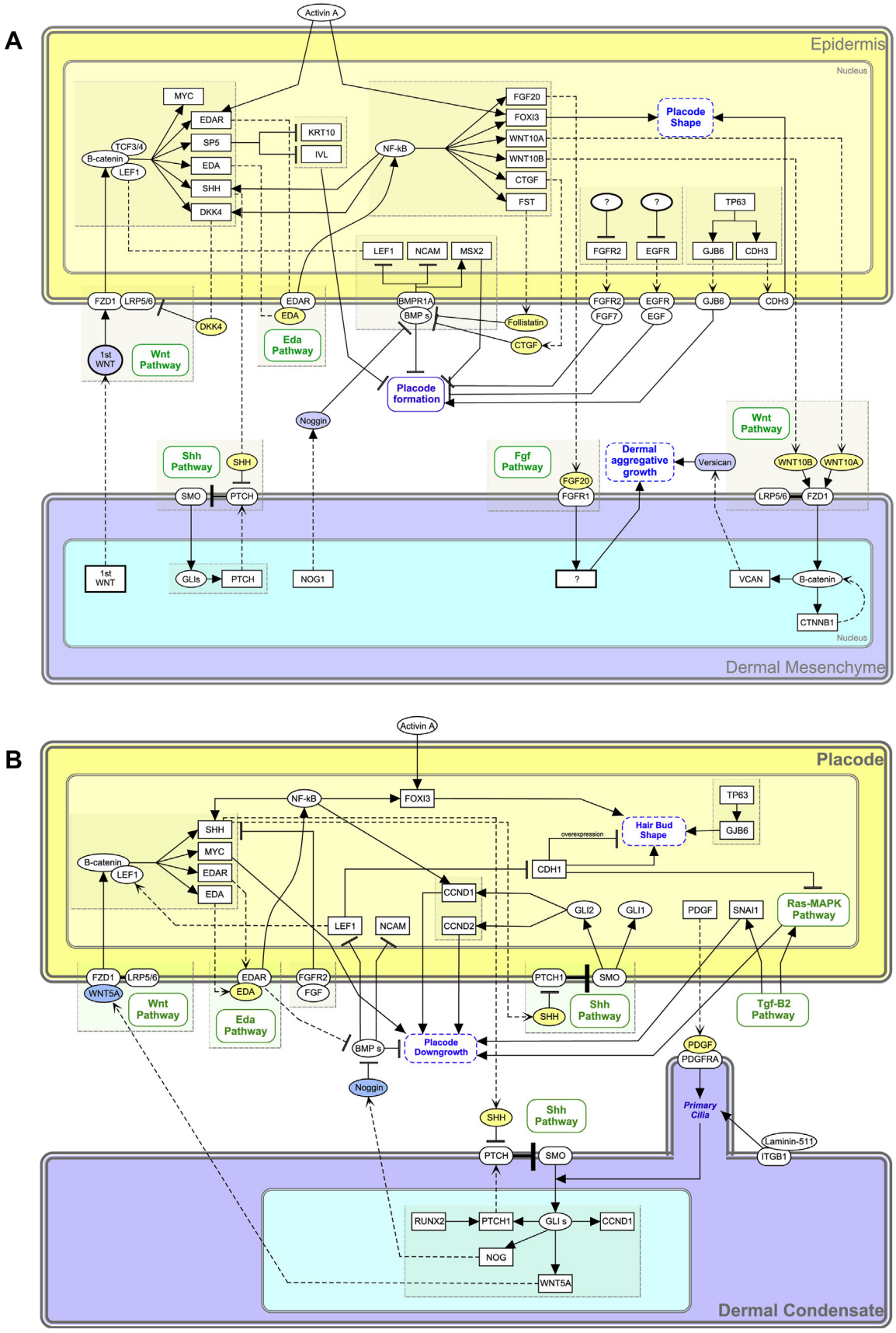

Fig. 2. Pathways for hair follicle development. A pathway was created for each of these stages: (A) induction; (B) organogenesis and (C) cytodifferentiation. Epidermis and dermal mesenchyme and their products are represented respectively in yellow and blue in (A) and (B). For cytodifferentiation (C), the main characteristics or events of the process are emphasized. See text for detailed explanation.

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

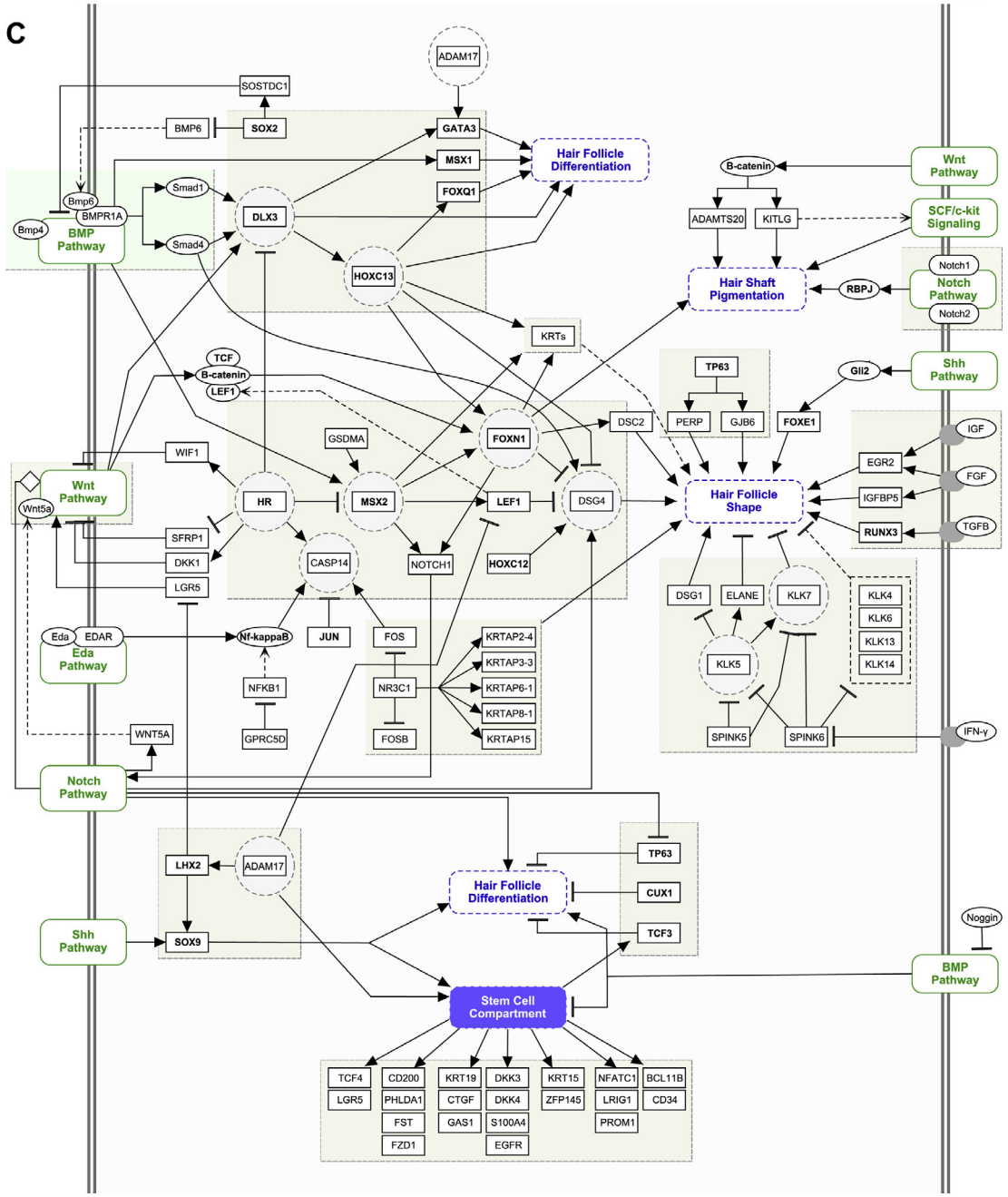

Fig. 2 (continued)

Finally, observing all three figures, it is possible to note that different nodes are put into clusters (represented as boxes), so the more closely related genes or signaling pathway elements are kept together and didactically individualized. Besides this, particularly on Fig. 2A, the most participative found genes are marked with a dashed line circle, while gene symbols of main transcription factors appear in bold.

3.1.1. Induction

3.1.1.1. Mesenchymal Wnt as the first signal. In hair follicle (HF) development, the induction stage marks the beginning of the crosstalk between epithelium and mesenchyme (represented at Fig. 2A as yellow and blue cell compartments, respectively). The first compartment will result in the small focal epithelial thickening, named placode, which, later, will give origin to most of the layers of the HF. Then, the second cell compartment results in the dermal condensate, which thereafter becomes the dermal papilla [21–23]. The first signal for induction is given by mesenchymal cells expressing WNT (Fig. 2A middle left) early in development [22,24–26]. Although it is not known what specific Wnt is responsible for this first signal, some authors consider Wnt10b as a reasonable candidate [27]. Anyhow, activation of Wnt signaling pathway blocks b-catenin cytoplasmic degradation fate, releasing this transcription factor to get in the nucleus and starts expression of placode formation genes.

During embryogenesis, b-catenin in the HF epithelium, together with two transcription factors, Lef1 and Tcf3-4 [26], is capable of activating the expression of several crucial genes for HF induction and morphogenesis. Between them are MYC, SP5, DKK4, SHH, EDA and EDAR [28–33].

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

In mice HF, MYC expression is predominantly expressed in the epithelium and is related to the differentiation and proliferation of keratinocytes and stem cells migration during HF morphogenesis [34]. SP5, in its turn, is a direct target of b-catenin capable of inhibiting expression of KRT10 and IVL. Both products of these genes are associated with epidermal differentiation fate and, as a consequence, they may be suppressed to promote placode formation [29].

Dkk molecules are widely recognized for its inhibitory effects against Lrp5 and 6, which are crucial co-receptors in the activation of Wnt pathway [35]. Strong evidences have shown that DKK4 expression is a Wnt target gene in epidermis, an evidence that Dkk4 acts as Wnt pathway negative feedback loop, fine-tuning this signaling effects on placode shape [30].

3.1.1.2. Shh pathway. SHH expression starts in the induction stage and is also upregulated by Wnt pathway in HF epithelium [36]. After secretion, Shh signal activates Shh signaling pathway in the mesenchyme compartment activating transcription factors Gli, which induces expression of Shh receptor Ptch, increasing cell sensibility to Shh [31,37].

3.1.1.3. Eda pathway. Until this point in HF development, Wnt and Shh pathways are strongly operating between epithelium and mesenchymal compartments. Following MYC activation, epithelial cells proliferate and thus the placode begins to get its form. Co-transfection studies have shown that the Wnt pathway is able to increase EDA expression, whereas Eda pathway does not affect Wnt signal levels [32]. As Fig. 2A shows in the Epidermal compartment, Eda receptor's gene EDAR is a direct Wnt target, and Activin A acts as another upregulator [33,38]. Furthermore, the Eda–Edar interaction activates the transcription factor NF-kappaB, which upregulates a considerable number of genes, including SHH, DKK4, WNT genes as WNT10A and WNT10B, FOXI3, FGF20, CTGF and FST [33,36,38].

Similarly to Wnt, the Eda/NF-kappaB pathway upregulates SHH and DKK4 [39–41]. Eda also activates expression of Wnt genes such as WNT10A and WNT10B in the epithelial cells, which are indirectly and directly stimulated by NF-kappaB, respectively. Those two Wnt proteins activate the Wnt pathway in the mesenchymal cells, where released b-catenin activates expression of other Wnts and further upregulates its own expression, maintaining the tissue crosstalk and enhancing the signaling pathway performance [33,42]. It is also known that mesenchymal b-catenin activity is important for epithelial b-catenin maintenance [36].

3.1.1.4. Placode and dermal condensate formation. Once the initial crosstalk is established and the placode has started forming, the tissues require instructions that guide cell proliferation. Once again the Eda/NF-kappaB pathway has an important role. NF-kappaB is responsible for upregulating FOXI3, a forkhead box family member expressed in teeth and HF during embryogenesis [43]. It has roles in skin appendages shape formation and possibly in placode shape patterning. Furthermore, Activin A is also capable of upregulating Foxi3 in the epidermis [43].

Other genes also related to placode formation include gap junction protein genes CDH3 and GJB6, both induced by transcriptional activator TP63 [44].

After its establishment, the placode emits epidermal signals to the underneath mesenchymal cells, in a way to start their aggregation [22]. The Eda pathway induces FGF20 expression in the epithelium that when secreted binds to its mesenchymal receptor Fgfr1, promoting the dermal condensate aggregative growth [45]. Moreover, this aggregation is also sustained by Versican (VCAN) expression mediated by mesenchymal Wnt pathway [46].

3.1.1.5. Inhibitory signals. At induction, as well as at the two later HF developmental stages, there are countless other signals mediating the existing interface between epithelial and mesenchymal cells. Strong induction suppressor signals are present, such as Bmp, Fgf7 and Egf [47,48]. Molecules such as Noggin, Follistatin and CTGF act as BMP signal suppressors, preventing its inhibitory effects against placode formation [39,47,49]. Crucial placodal genes downregulated by BMP are LEF1 and NCAM, but BMP is also capable of inducing early upregulation of differentiation fate transcription factor Msx2 [47,49,50]. Noggin however, is expressed by mesenchymal cells and acts as a BMP repressor by directly binding the molecules in the extracellular space [49]. Follistatin and CTGF (expressed in epithelial cells) are controlled by Eda/NF-kappaB activity and also repress BMP signals [39,47,51].

Two other signals Fgf7 and Egf direct differentiation toward the epidermal fate, instead of inducing HF placode formation. These signals persist during induction stage, and their receptors (FGFR1 and EGFR, respectively) are downregulated in the placode to allow the correct HF development [48].

3.1.2. Organogenesis

3.1.2.1. Placode downgrowth. After hair follicle placode and dermal condensate are successfully formed, the organogenesis stage proceeds. The placode receives a "second signal" originated from the dermis, which determines the maintenance of some signaling pathways seen on induction. Therefore, the placodal growth is preserved, but this time advancing toward the deep dermis [25]. So, the hair peg is gradually formed by the proliferating cells, which, later on in this process, end up enveloping the dermal papilla [22,52].

As shown inside the blue compartment on Fig. 2B, Shh pathway is the upstream inducer of Wnt5a, which promotes placodal Wnt pathway activation and, thus, the upregulation of the proliferative growth gene MYC [27,28,34].

Several genes have their expression maintained in the organogenesis stage, this can be noticed by comparing Fig. 2A and B. Similar to MYC, other genes also found at induction stage are kept upregulated by the placodal Wnt pathway: SHH, EDA and EDAR [32,33,36]. Although activation by NF-kappaB is maintained, SHH is repressed by FGF signaling, mediated by placodal FGFR2 (Fig. 2B, center) [40,53]. Dermal Shh signaling is influenced by PTCH1, and PTCH1 upregulation is achieved by GLIs as described previously, together with Runx2 [31,54].

BMP signaling still has inhibitory effects for the HF formation during organogenesis, and placode downgrowth is deprived by BMP [55]. Noggin directly suppresses BMP, and NOG gene activation is induced by dermal GLIs [49]. Eda/Edar signaling present in placodes is also responsible for indirectly suppressing BMP [39].

The expression of epidermal genes CCND1 and CCND2 and the Tgf-b2 pathway is specific to the organogenesis stage. CCND1 is induced by NF-kappaB [56], and both CCND1 and CCND2 are induced by Gli2 activation, downstream of Shh signaling [57,58]. In conjunction with cyclins, the Tgf-b2 signaling pathway is necessary for induction of SNAI1 and Ras–MAPK pathway. Both of which have major roles in events that guide hair bud formation within the developing skin [59].

Lef1 transcriptional complexes in the placode were identified as possible downregulators of epidermal CDH1 [55]. The reason for this rests in the observation that high levels of E-cadherin are known to block epidermal invagination and follicle production as well as downregulation of CDH1 resulting in Ras–MAPK activation [59].

The gap junction protein GJB6 is highly expressed in hair follicles [60] and induced by TP63 [60]. Mutations associated with GJB6 lead to hypotrichosis [60].

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

3.1.2.2. Primary cilia and the dermal papilla formation. Primary cilia are sensory organelles that display specific receptors and ion channels, which transmit signals from the extracellular environment via the cilium to the cell, and control tissue homeostasis and function. Additionally, primary cilium plays a critical role in dermal papilla formation and maintenance, mainly by stimulating the Shh pathway via PDGFRa [61,62]. PDGFRa ligand PDGF-A is produced in the epidermis and has a role in stimulating the proliferation of dermal cells that may contribute to the formation of DP [63].

Laminin-511 mutants lack the key BMP inhibitor Noggin and the absence of laminin-511 receptor ITGB1 can disrupt DP primary cilia formation impairing hair development [62]. These functions are a critical to direct ciliary function and maintenance of the dermal papilla.

3.1.3. Cytodifferentiation

The cytodifferentiation starts after HF has achieved its maximum size and when the hair bulb finally surrounds the dermal papilla. From this moment on, the different layers seen in developed hair follicles begin to differentiate. Dermal papilla (DP) is believed to control this process, sending crucial signals to the overlaying matrix (MX), a lineage of compromised germinative cells of the hair bulb [20–22].

In response to DP's signal, MX cells proliferate and daughter cells are pushed up along the HF longitudinal axis toward skin surface. During migration, these cells gradually differentiate, giving rise to the hair shaft and its channel, the inner root sheath. The hair shaft is in the center and consists of a medulla (MD), cortex (CX) and a cuticle (CL) layer. The inner root sheath (IRS) surrounds the hair shaft and consists of cuticle and Henley and Huxley layers [20,21]. Fully outside and surrounding these layers, is the outer root sheath (ORS), which is a continuation of the basal layer of the epidermis [64].

3.1.3.1. Main regulators of hair follicle differentiation. The regulatory cytodifferentiation pathway construction (Fig. 2C) allowed us to find out four distinguishable clusters in which internal elements seem to be more closely related than with outsiders. The first two, are directly related to the differentiation fate. One is located on the middle top of the figure, having DLX3 and HOXC13 as the most participative genes on the analysis. The second cluster is a small one located on the middle bottom of our pathway, being composed by three genes (TP63, CUX1 and TCF3).

Transcription factor Dlx3 has a central role as a crucial transcriptional regulator of hair morphogenesis, differentiation and regeneration [65]. It is expressed in MX cells and activated by the BMP pathway. Evidence of this consists of mutations in Smad1/Smad4-binding sites downregulating DLX3 [66]. Moreover, the same study elucidated that Dlx3 is positioned upstream other transcriptional factors such as GATA3 and HOXC13 products.

Gata3 is crucially involved in epidermis and HF differentiation, integrating diverse signaling networks to regulate the balance between HF and epidermal cell fates [67]. However, this transcription factor is downregulated in the absence of ADAM17, suggesting a critical relation between them [68]. Concerning HOXC13, FOXQ1 is a downstream target and both control medulla differentiation through a common regulatory pathway [69]. In addition, direct involvement of HOXC13 in the control of hair keratin gene expression is seen during differentiation [70].

Another member of the first cluster is Sox2, which acts as a key regulator of hair growth, controlling progenitor migration by fine-tuning BMP-mediated mesenchymal–epithelial crosstalk. Transcriptional profiling of Sox2 null DPs reveals increased Bmp6 and decreased BMP inhibitor Sostdc1, a direct Sox2 transcriptional target [71].

BMPR1A encodes a BMP receptor in the surface epithelium and regulates terminal differentiation in HF. Mutations associated with this receptor markedly downregulate expression of the known transcriptional regulator of follicular differentiation Msx1 [72].

On the second pathway cluster, Notch1 and Tp63 maintain a cross-regulation in keratinocyte differentiation process. TP63 expression is suppressed by Notch1 activation in both mouse and human keratinocytes. In its turn, Tp63 functions as a selective modulator of Notch1-dependent transcription and function [73].

CUX1, also known as Cutl1, is a transcriptional repressor in diverse processes such as terminal HF differentiation [74]. Finally, Tcf3 controls multipotent stem cells to generate the keratinocytes of the epidermis, sebaceous gland and hair follicles. It is also shown that Tcf3 can act independently of its b-catenin interacting domain to promote features of the follicle outer root sheath (ORS) and multipotent stem cells (bulge), the compartments that naturally express Tcf3 [75].

3.1.3.2. Hair shape main actors. Our pathway (Fig. 2C) shows that a lot of different genes act by controlling or participating in the hair follicle shape formation. This time, the related genes are, apparently, more diffused, except for the two major gene clusters located in the middle of the figure (circulated HR, CASP14, MSX2, FOXN1, and DSG4) and the other on its right side (circulated KLK5 and KLK7).

The cluster in the middle contains the gene that encodes Hairless (HR), which is a transcriptional cofactor that plays important roles in HF morphogenesis and is capable of upregulating CASP14. Casp14 is known to play a role in apoptosis, and in addition, it is directly associated with epidermal cell differentiation [76,77]. Casp14 is also upregulated by NF-kappaB and FOS, although downregulated by JUN [78]. However, both upregulators are suppressed by GPRC5D and NR3C1, respectively [79,80]. Similarly, NR3C1 is capable of blocking expression of FOSB too, as well as activating the transcription of five different types of genes encoding keratin-associated proteins (KRTAP2-4, KRTAP3-3, KRTAP6-1, KRTAP8-1 and KRTAP15) [80]. These KRTAPs play an important role on keratin-bundle assembly in the hair cortex and provide important physical properties to the hair shaft [81].

Going back to Hr, our pathway also shows that its expression is associated with the upregulation of Wnt signaling inhibitors such as Wif1 and Dkk1. The only exception is Sfrp1 that also suppresses Wnt, but is downregulated by Hr [76,82]. Since the hair follicle downgrowth has to be stopped at the end of organogenesis, Wnt signaling, as a central stimulator of cell proliferation, has to be indeed stopped.

Another gene downregulated by Hr is MSX2 and its related target genes [83]. Although Hr counteracts against it, Msx2 is stimulated by GDSMA [84]. Among the genes that are upregulated by Msx2 are NOTCH1, FOXN1, LEF1 and keratin genes (KRTs) [83,85]. Notch1 is capable of modulating Wnt signaling, what is important due the fact that the establishment and stability of cell fates during development depend on the integration of multiple signals, which ultimately modulate specific patterns of gene expression [86]. Wnt5a, a specific dermal papilla signature gene is found to be under direct Notch control [87].

Lef1 associates with b-catenin, and participates in the expression controls mediated by Wnt signaling. It is known that LEF1 is upregulated in lack of ADAM17, an important gene that has indirect effects also against another Wnt stimulator, Lgr5. Actually, LGR5 inhibition occurs through Lhx2, which in turn is upregulated by Adam17 stimulation [68,88,89].

Functional rescue experiments establish Wnt5a as an essential downstream mediator of Notch-CSL signaling, resulting in the sustained expression of FoxN1 [87]. Inactivation of FOXN1 in mice causes defective assembly of the hair medulla, what reveals its

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

critical role in proper hair morphogenesis. In addition, Dsc2 is a downstream effector of Foxn1 activity and its expression is restricted to the hair medulla [90]. Desmocollins (DSCs) interact extracellularly to link neighboring epithelial cells.

Desmoglein 4 mutations are known to cause hypotrichosis in humans. DSG4 protein is expressed in the more highly differentiated layers of the epidermis. It is an important adhesion molecule maintaining the integrity of the hair shaft [91,92]. Multiple transcriptional factors seem to aim for Dsg4 production. DSG4 expression decrease is provided by upstream counteractions controlled by three transcriptional factors already mentioned: Lef1, Foxn1 and Hoxc13 [93]. Still, three other regulatory elements seem to act as stimulants: Hoxc12, Bmp and Notch signals [92,93].

Concerning the hair shape, important related signaling pathways are FGF, IGF, TGF-b and SHH, whose effectors control shape formation. They are shown in the middle right of the pathway. The first two pathways activate EGR2 expression, though FGF also activates IGFBP5 [94]. TGF-b and SHH stimulate, respectively, Runx3 and Foxe1 [95–97].

Other elements that closely participate in hair shape formation are the Desmoglein 1 (Dsg1), Elane and Kallikreins (encoded by genes such as KLK4, KLK5, KLK6, KLK13 and KLK14) [98–102]. KLK5 upregulates Elane and KLK7, and also inhibits DSG1 [101]. Both KLK5 and KLK7 are downregulated by SPINK5 and SPINK6, which is particularly inhibited by INF-gamma [101,102].

Finally, the same TP63, which participates in HF differentiation as an inhibitor, acts this time as an activator of PERP and GJB6 [60,103]. Gjb6 has crucial roles in morphogenesis and development of skin and its appendages. Immunostaining in normal human skin sections demonstrated predominant expression of GJB6 in hair follicles [60]. Perp in turn is a critical component of the desmosome, a cell–cell adhesion complex whose constituents are frequently mutated in human diseases affecting the skin and hair [103].

3.1.3.3. Some controllers of hair shaft pigmentation. Hair color depends on pigment-producing melanocytes that remain stationary at the base of matrix (MX) and transfer melanin-containing organelles to adjacent keratinocytes that are pushed upward as they proliferate. During development, however, the nonpigmentary components of the hair follicle are built first, and melanocyte movement is a secondary and stepwise process that proceeds from the dermis to the epidermis and then into developing follicles [104].

Concerning this process, SCF/c-kit signaling and FoxN1 play an important role. Stem cell factor (SCF) and its receptor c-kit are important for melanocyte survival during development and mutations in these genes result in unpigmented hairs [105]. Foxn1, in turn, already seen as a regulator of HF keratinocyte differentiation (Section 3.1.3.1) is also fundamental in signaling specific transfer of pigment from melanocytes to keratinocytes of the hair cortex [104,106].

Pigmentation regulators Adamts20 and Kit ligand (KITLG) are believed to act as mediators of the effects of b-catenin signaling on pigmentation [29]. This finding provides evidence for a new paradigm in which, in addition to promoting hair follicle placode and hair shaft fate, b-catenin signaling actively directs pigmentation.

Additionally, Notch signaling has a role in the regeneration of the melanocyte population during hair follicle cycles. The deletion of NOTCH1 and NOTCH2 or RBPJ in the melanocyte lineage results in the elimination of melanoblasts and melanocyte stem cells [107].

3.1.3.4. Stem cells compartment expressions. Hair follicles are self-renewing structures that cycle and reconstitute themselves throughout life because they contain keratinocyte stem cells. During the hair follicle development, a reservoir of those cells is created. It resides in a region of the outer root sheath surrounding the hair shaft and is termed bulge [108]. During the hair cycle, the dermal papilla and the regressed hair follicle are located adjacently to the hair bulge and signal to the stem cells that they should enter a differentiation program and form a new hair germ.

Here we list some genes that are overexpressed in these stem cells: CD200, PHLDA1, FST and FZD1 [108]; NFATC1, PROM1 (also known as CD133), CD34, BCL11B and LRIG1 [109]; and KRT15 and ZFP145 [110]. In addition, targeted transcriptional profiling indicates that KRT19, DKK3, DKK4, TCF3, S100A4, GAS1, EGFR and CTGF are also preferentially expressed by human bulge cells, compared to differentiated HF keratinocytes [111]. In the same way, so are SOX9, TCF4 and LGR5 [88]. Sox9 is upregulated by the Shh pathway and Lhx2 [88].

3.2. Mammary gland development pathway

The breast is composed by a connective tissue matrix, in which the glandular tissue is inserted. It is composed by many independent glands, each one constituted by ducts and lobules. These independent glands form a network, which opens to the outside in the nipple for milk secretion.

There are four morphological types of mammary gland: immature, mature inactive, mature active and in process of involution. This definition is according to the stage of development. It is possible to distinguish four phases in the mammary gland development, each one having specific pathways associated. The phases are embryonic development, puberty, pregnancy and lactation, and involution.

3.2.1. Embryonic development

The mammary gland has its embryonic origin in the ectoderm. In human, there is a formation of two ectoderm lines, called mammary lines, from the anterior to posterior limb in the ventral surface [112]. These mammary lines give rise to two placodes, located in the thoracic region [113]. The ectoderm then penetrates the underlying mesoderm to form the first mammary ductal system [112]. The immature mammary gland, or the primary mammary gland, persists from birth to puberty.

Two processes that should be highlighted at this stage are epithelial to mesenchymal transition (EMT) and placode formation.

3.2.1.1. Epithelial-to-mesenchymal transition (EMT). The epithelial to mesenchymal transition (EMT) in the embryonic development occurs during gastrulation and neurulation. Zinc finger E-box binding homeobox 2 (ZEB2) is an EMT regulator (Fig. 3A). It represses the expression of CCND1, SFRP1, MIR200A, MIR200B, MIR429, TERT, CDH1, CLDN4 and ALPL, and also upregulates the expression of mesenchymal markers [114]. Both cyclin D1 (CCND1) and secreted frizzled-related protein 1 (SFRP1) are related to cell proliferation, and SFRP1 is a modulator of the WNT signaling pathway. The microRNA genes MIR200A, MIR200B and MIR429 are repressed as well, and unable to regulate their target genes expression. TERT encodes the telomerase reverse transcriptase, responsible for elongation of telomere ends. Cadherin-1 (CDH1) and claudin 4 (CLDN4) are involved in cell adhesion, and alkaline phosphatase, tissue-nonspecific isozyme (ALPL) may play a role in skeletal mineralization. Many genes modulate ZEB2. Hypoxia signals, TGFB1, TNF, IL1 and AKT1 upregulate ZEB2, and Hedgehog signals upregulate ZEB2 via TGFB1 [114]. Transforming growth factor beta-1 (TGFB1), tumor necrosis factor (TNF) and interleukin-1 (IL1) are cytokines and AKT1 is a kinase that plays a role in many processes like proliferation, cell survival and growth.

3.2.1.2. Placode formation. Conversely, in embryonic development, neuregulin 3 (NRG3) appears to influence cell fate. NRG3 binds to

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

its receptor, ERBB2. This association results in MYC induction which downregulates ITGA6 (integrin alpha-6) and ITGB1 (integrin beta-1). The result is a change in cell adhesion and proliferation and consequent exit from the stem cell compartment. NRG3 is also a signal for placode formation along the mammary line [115]. Invagination of cells within placode into the mesenchyme forms the mammary primordium.

3.2.2. Puberty

In this stage of development, the mammary gland network grows and becomes branched, giving arise to the secondary mammary gland. This is followed by development of the surrounding stroma. The result is the mature inactive mammary gland. Some genes involved in this phase are AP1, ESR1, NRIP1 and estrogen signaling pathway genes (Fig. 3B).

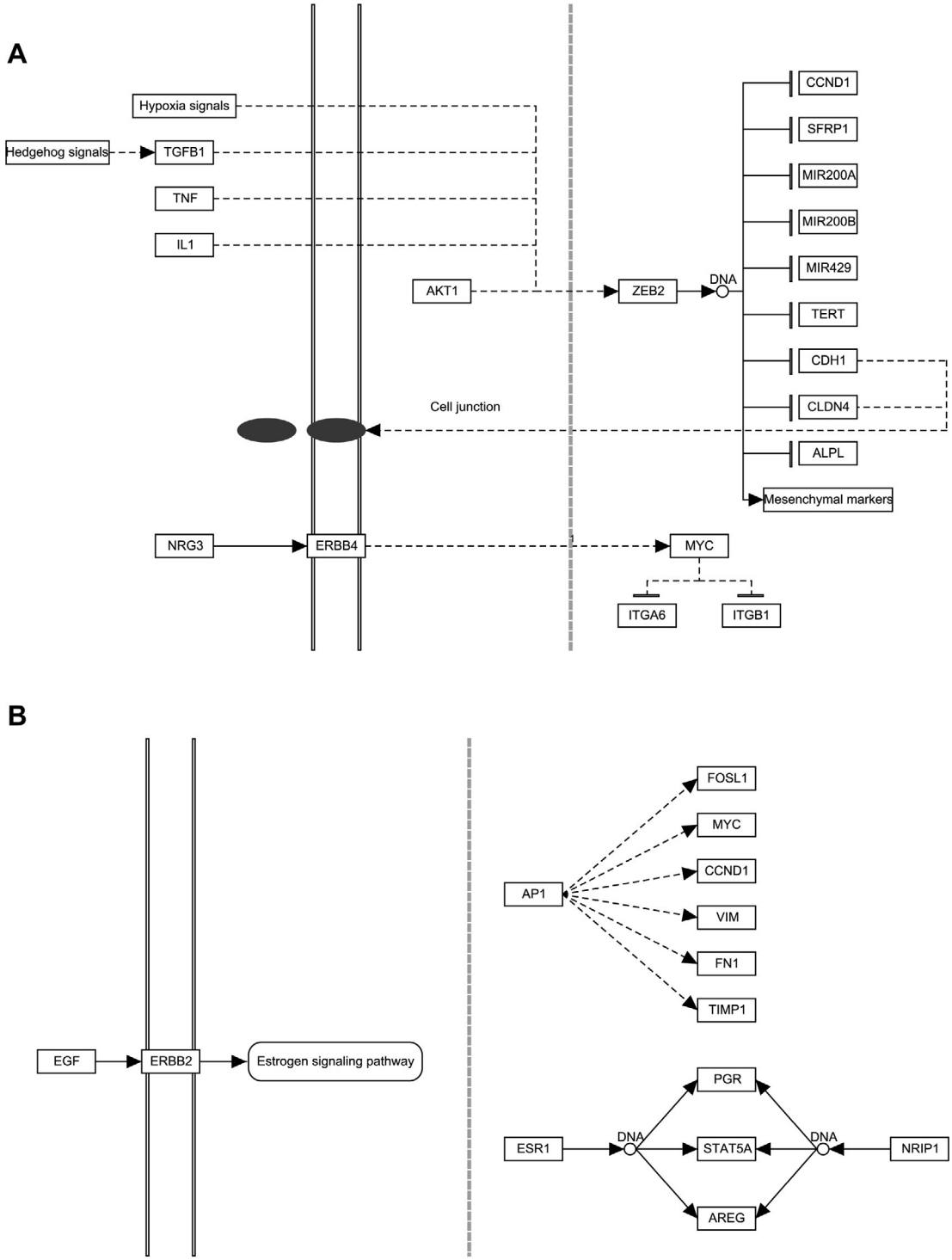

Fig. 3. Pathways for mammary gland differentiation. A pathway was created for each of these stages: (A) embryonic development; (B) puberty; (C) pregnancy and lactation and (D) involution. See text for detailed explanation.

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

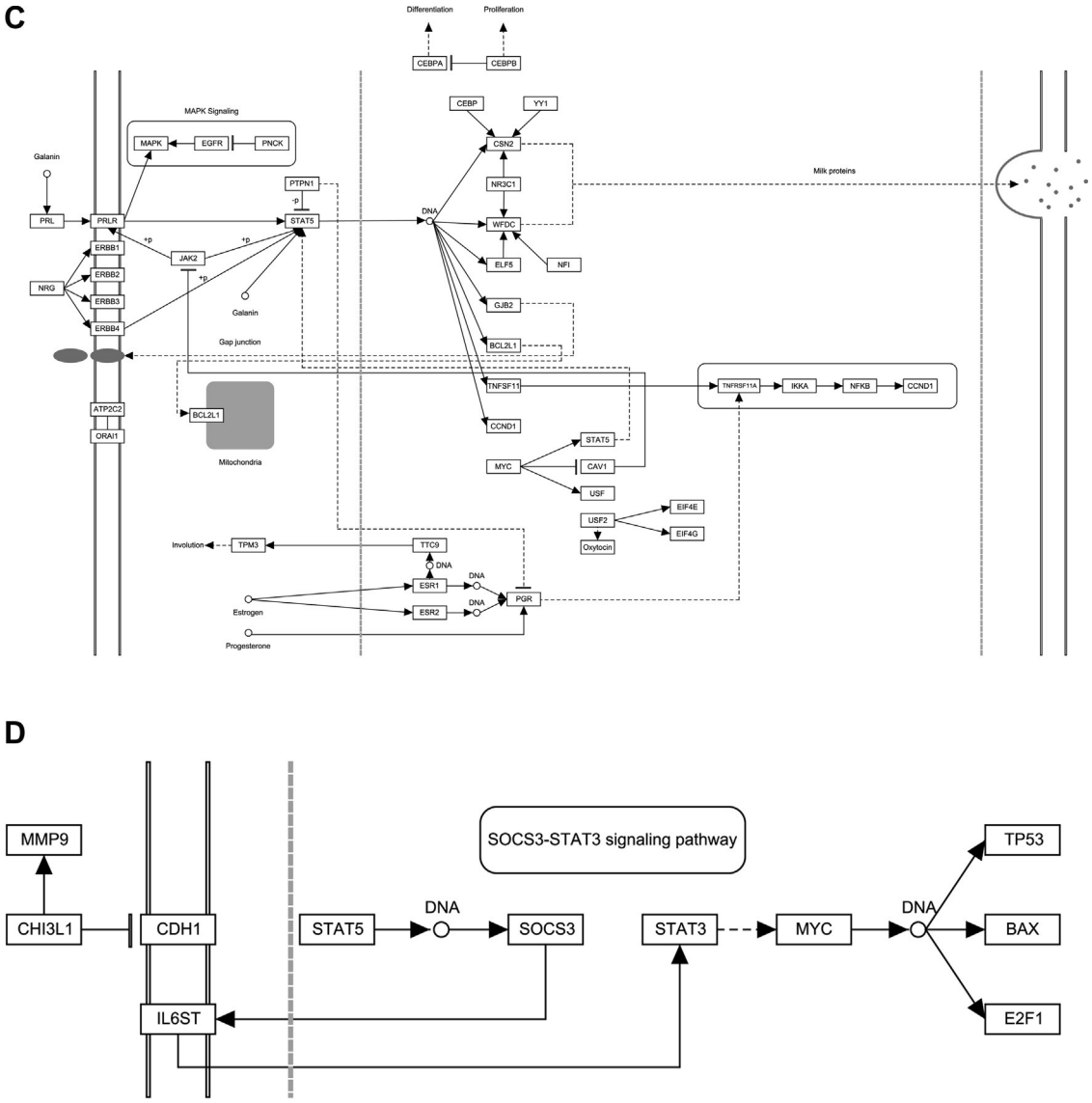

Fig. 3 (continued)

AP1 is an important regulator of mammary epithelial cell proliferation along postnatal development, mainly in puberty. FOSL1, MYC, CCND1, VIM, FN1 and TIMP1 are examples of AP1-dependent genes [116]. FOSL1 encodes a protein of the FOS family. Proteins of FOS family dimerize with proteins of the JUN family to form AP1 complex. MYC and CCND1 are involved in cell proliferation. VIM encodes vimentin, a component of intermediate filaments. FN1 encodes fibronectin, a protein related to adhesion and migration. TIMP1 encodes a protease inhibitor.

ESR1 (estrogen receptor) and the nuclear receptor interacting protein 1 (NRIP1) are co-regulators of progesterone receptor (PGR), STAT5A and amphiregulin (AREG), which control proliferation and differentiation during mammary gland development. In puberty, NRIP1 expression is necessary in the epithelial and stromal compartments for ductal elongation [117].

One pathway used by estrogen to execute its growth effect in the mammary gland is through epidermal growth factor (EGF) and EGF receptors (ERBB). In this case, ERBB2 in special may activate or inhibit the growth according to the development stage. In puberty, ERBB2 impairs the mammary epithelial cell proliferation [118].

3.2.3. Pregnancy and lactation

The mature active mammary gland appears at this stage. Its main characteristic is the presence of alveoli, the structure responsible for milk production. There are four inputs for signaling the breast development during pregnancy and lactation (Fig. 3C): through prolactin receptor (PRLR), epidermal growth factor receptors (ERBB1, ERBB2, ERBB3 and ERBB4), estrogen receptors (ESR1 and ESR2) and progesterone receptor (PGR).

3.2.3.1. Via PRLR and JAK2/STAT5 pathway. In this first case, prolactin (PRL) binds to its receptor, leading to its phosphorylation by Janus kinase 2 (JAK2) and subsequent coupling of the signal transducer and activator of transcription 5 (STAT5) [113]. JAK2 then phosphorylates STAT5, inducing its migration to the nucleus. STAT5 targets are related to proliferation (TNFSF11, CCND1), differentiation (CSN2, WFDC, ELF5, and GJB2) and survival (BCL2L1) [113]. CSN2 and WFDC encode for milk proteins beta-casein and WFDC protein, respectively (WFDC is also known as whey acidic protein). CSN2 promoter has binding sites for CEBP and YY1; WFDC promoter has sites for ELF5 and NFI, and both have binding sites for the glucocorticoid receptor, encoded by NR3C1 [113,119]. CEBP

product isoforms alpha and beta are both related to the balance between proliferation and differentiation of epithelial cells [119]. GJB2 expresses the gap junction beta-2 protein, the major element of this junction type. Moreover, CCND1 and TNFSF11 signalize for proliferation. STAT5 induces proliferation directly by CCND1, and indirectly through TNFSF11; TNFSF11 (also known as RANK-L) binds to its receptor (TNFRSF11A, also known as RANK); activated by TNFRSF11A, IKKA relieves inhibition over NFKB, resulting in its activation; NFKB then also induces transcription of CCND1 [113]. Lastly, Bcl-2-like protein 1 (BCL2L1) is related to survival, and as an inhibitor of cell death. PRLR can also propagate the signal through the MAPK pathway [113].

3.2.3.2. Via ERBB and STAT5. Another route for signaling is through epidermal growth factor receptors. In this case, neuregulin (NRG) binds to one of its receptors (ERBB1, ERBB2, ERBB3 or ERBB4) and the signal is transmitted by ERBB4 to STAT5, from where it follows as described above [113]. ERBB4 substitutes JAK2 by phosphorylating STAT5.

3.2.3.3. Regulation of JAK2/STAT5 pathway. MYC induces STAT5 directly and indirectly through CAV1. CAV1 negatively regulates JAK2. By negatively regulating CAV1, MYC positively regulates JAK2 and, therefore, STAT5 [14,113]. MYC is also associated with a group of transcription factors implicated in cell cycle control: UFS. UFS2, a member of this group, influences the expression of translation initiation factors EIF4E and EIF4G, as well as the concentration of oxytocin in the blood [15].

Another STAT5 modulator is galanin. Galanin is a hormone, which stimulates PRL and therefore STAT5. Furthermore, galanin also stimulates STAT5 directly [113].

Conversely, PTPN1 seems to negatively regulate STAT5 [120]. Being a phosphatase, PTPN1's action might be impairing STAT5 migration to the nucleus. PTPN1 also appears to negatively regulate PGR [120].

3.2.3.4. Via ESR1, ESR2 and PGR. Estrogen binds to its receptors ESR1 and ESR2. Both receptors stimulate the progesterone receptor, which is associated with the TNFSF11 receptor (TNFSF11A) and to its downstream pathway. TNFSF11 pathway is related to induction of cellular proliferation, as previously described [113].

Estrogen receptor 1 (ESR1) also affects adhesion through induction of TTC9, which interacts with TPM3 that expresses a protein associated with actin filaments, playing a role in involution [121].

3.2.3.5. PNCK and differentiation along pregnancy. The calcium/calmodulin-dependent protein kinase type 1B, expressed by PNCK, appears to negatively regulate EGFR during pregnancy. Similarly the MAPK signaling pathway also seems to be negatively regulated in response to the effect over EGFR. Consequently proliferation decreases and terminal differentiation of epithelial cells is triggered [122].

3.2.3.6. Transport of Ca^{2+} and differentiation along lactation. The ATPase 2C2 encoded by ATP2C2 is co-expressed with the component of Calcium influx channel encoded by ORAI1. Together, they regulate Ca^{2+} uptake, influencing differentiation and supporting the large calcium transport requirements for milk secretion along lactation [123].

3.2.4. Involution

After pregnancy and lactation, the breast undergoes a tissue remodeling process, known as involution. This process is characterized by the alveoli degeneration through apoptosis. Some genes implicated in this stage of development are CHI3L1, SOCS3 and STAT3 (Fig. 3D).

In the presence of lactogenic hormones, such as prolactin, hydrocortisone and insulin, the chitinase-3-like protein 1 encoded by CHI3L1 mediates the inhibition of mammary epithelial cell differentiation and polarization. Moreover, CHI3L1 suppress the expression of CDH1 and increases the expression of MMP9 (matrix metalloproteinase-9, a remarkable protein related to extracellular matrix degradation). This is an important process for involution [124].

Another significant pathway to involution is the SOCS3/STAT3 signaling pathway. STAT5 induces the suppressor of cytokine signaling 3 (SOCS3), which binds to and activates its receptor, the interleukin-6 receptor (IL6ST). Activation of IL6ST induces the activation of STAT3 [113]. SOCS3/STAT3 signaling pathway is associated with MYC and, thereafter, with MYC target genes such as TP53, BAX and E2F1. Changes in the SOCS3/STAT3 signaling pathway affect MYC. As SOCS3 has an anti-apoptotic function and MYC an apoptotic function, SOCS3 is considered an important regulator of involution [125].

3.3. Deciphering a pathway origin

Although it is not the aim of this manual to guide users on accessing or creating a cluster of homologues for the gene of interest, this is the secret underneath the cracking of a pathway origin along evolution. Whenever a researcher understands the scenario of regulation of a phenomenon such as hair and breast development, exemplified here, it is possible to verify that proteins building the pathway in a complex organism, are actually playing similar roles, although not equal, in less complex organisms. Furthermore, it becomes clear that some key partners sometimes have originated in very distinct epochs. We have exemplified this with Nanog and Oct-4, partners in keeping the totipotency of cells in the embryonic Inner Cell Mass [10], which have originated, in the human lineage, together with, respectively, the rising Coelomata and Theria organisms, remarkably separate evolutionary periods.

How can one reveal the origin of a gene and, furthermore, the pathway evolution? We will initially exemplify with the gene Hr, one of the hair shape main actors (Section 3.1.3.2). Using the protein of interest as Seed, the first step is to obtain a cluster of orthologues for it (Fig. 4). With those in hand, one can easily retrieve the Taxonomy IDs (man is 9609, mouse is 10,090 and so on). A simple way of discovering the origin is to send all TaxIDs to a tool in NCBI named Taxonomy Common Tree, and the output will present the clade Theria as the Lowest Common Ancestor (LCA) for the Hr gene. A complete tutorial on looking for orthologues escapes the scope of this manual, however we present an example that gives a perspective for understanding the evolution of a system, constructed with text-mining, using a subset of the hair development cytodifferentiation pathway as a case study (Fig. 5).

Similarly to what was explained for Hr gene, we searched for orthologues of the human genes of interest. We used a combination of searches with software from our group [11] and KEGG orthology clusters enriched with UniProt clusters under a procedure developed by us [126], however any other approach as finding the organisms that comprise a pFam family [127], or looking at Panther families [128], OrthoMCL-DB [129], KEGG orthology [2], etc. will provide the putative taxonomic distribution of the gene. To determine the clade in the human lineage when the gene is first seen, we use a program that walks along the NCBI taxonomy tree and determines the node that first concentrates the leaves containing the orthologues, but this can be done in Taxonomy Common Tree as shown above. The results are shown in Fig. 5.

Each gene presents a numeric notation that represents the node, counted from root, in the human lineage where it is first seen. Therefore the number corresponds to a Lowest Common Ancestor level in the human lineage. Circles with number were

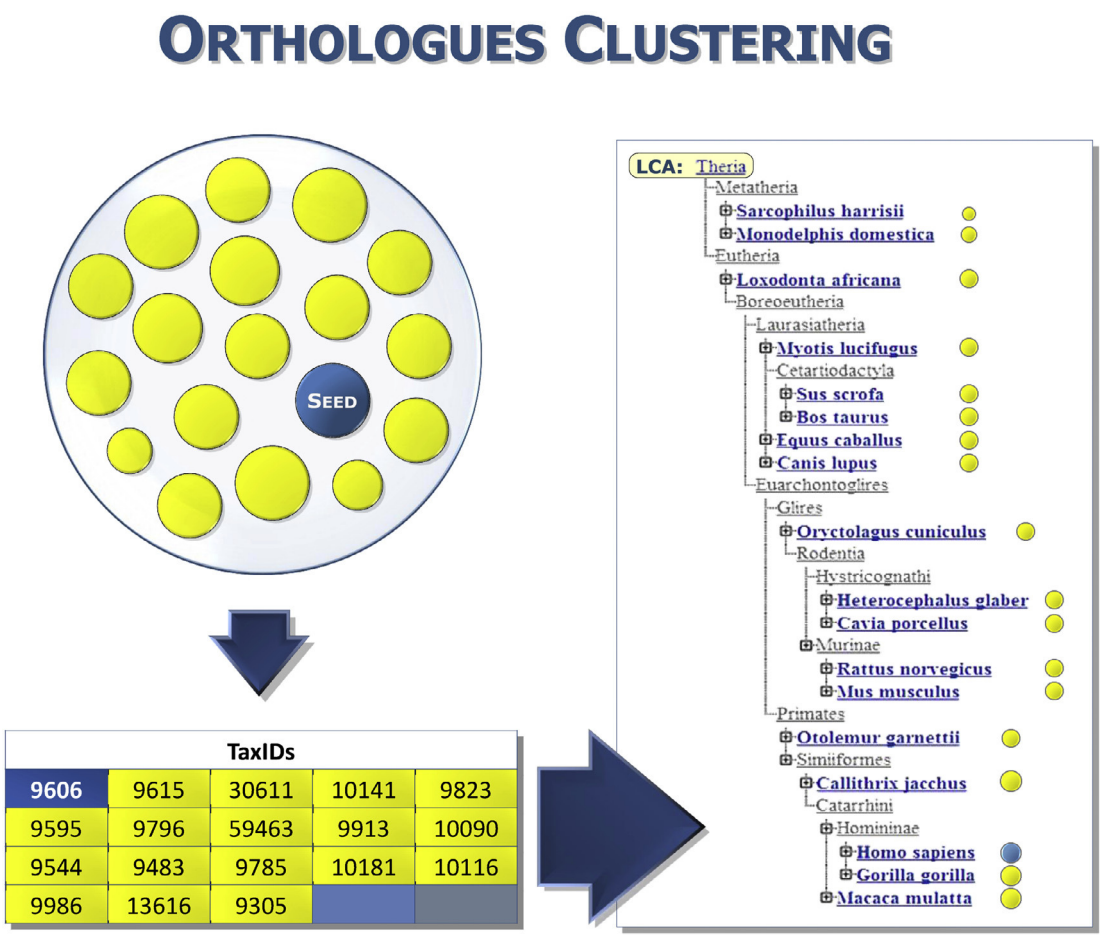

Fig. 4. Discovering the last common ancestor (LCA). The UniProt ID obtained for the gene symbol mined from literature is used to search for orthologues with proper software, and the taxonomic IDs (TaxIDs) of the organisms that bear them are collected (lower left panel). TaxIDs are used as input to the NCBI tool Taxonomy Common Tree. The resulting tree is obtained, in which the LCA is determined as the lowest taxon achieved (Theria).

colored with different colors according to periods of human evolution. From the clade Cellular Organisms (a gene is mapped in this ancestrality level if it is shared for example with bacteria) up to Gnathostomata, circles are not colored. In this example we found genes originated in the epoch of Metazoa (5), Eumetazoa (3), Bilateria (1), Deuterostomia (1), Chordata (2), and Gnathostomata (1). In the Euteleostomi period, many genes have been originated in human (not shown), so we depicted those separately, coloring the circles in yellow. In this example five genes appeared. Another set of genes were first seen in Tetrapoda (2) and Aminiota (1) and those are colored in green. For example, human and chicken lineages separate at the Amniota clade. Completing the zoomed pathway, some genes appeared in the epoch of Mammalia (2), Theria (6) and Eutheria (4) (circles labeled in blue). A complete analysis of all genes in two case studies will be reported elsewhere, but this zoom exemplifies how the present pathway has originated.

A brief analysis of Fig. 5 reveals that before the appearance of Euteleostomi organisms (zebrafish as a reference), a considerable portion of this machinery already existed, like the WNT and HOTCH signaling pathways. WIF1, SFRP1 and DKK1 already could suppress the WNT pathway, impairing the association of b-catenin with TCF in the nucleus. As described before, this suppression is required for the cytodifferentiation to proceed. Conversely, it is plausible that b-catenin and TCF could activate expression of FOXN1. EDA was already present, although its specific receptor EDAR appears later, as well as the NF-kappaB subunits RelA and RelB. Three transcription factors were already present, HOXC12, FOXN1 and MSX2; although the first did not have a role in this machinery, the other two could be activating NOTCH1, as well as participating in the synthesis of keratins (KRTs). FOXN1 could already be involved in the transference of melanin between melanocytes and keratinocytes.

In Euteleostomi (Fig. 5, yellow circles), the EDA pathway might have started becoming functional, after the origin of EDAR and RelA, which could be acting with a remote ancestor of RelB or alone. Moreover, WNT signaling would become more complex with the appearing of LEF1, leading to a more efficient activation of FOXN1 by b-catenin. IFN-c is already present, however without a role in the present pathway.

After this and up to the origin of Mammalia (Fig. 5, green circles), three more acquisitions are accomplished: RelB, GSDMA and ELANE, RelB can then work with its counterpart RelA as the NF-kappaB complex. GSDMA is required for placode differentiation into the hair shaft, regulates MSX2 and its absence is known to cause hairless phenotype [84], so this important dependence might have originated in Amniota organisms. Moreover, ELANE expression is associated with abnormalities in hair development [130], and might be impairing its origin in this period.

Furthermore, in the period that encompasses the origin of Mammalia organisms and subsequent evolution, a set of relevant genes appears and the complete regulation is accomplished (Fig. 5, blue circles). A series of kallikrein-related peptidases (KLKs) are originated. KLKs are related to skin desquamation and, when in high levels, cause hair abnormalities in hair shape [130]. However,

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

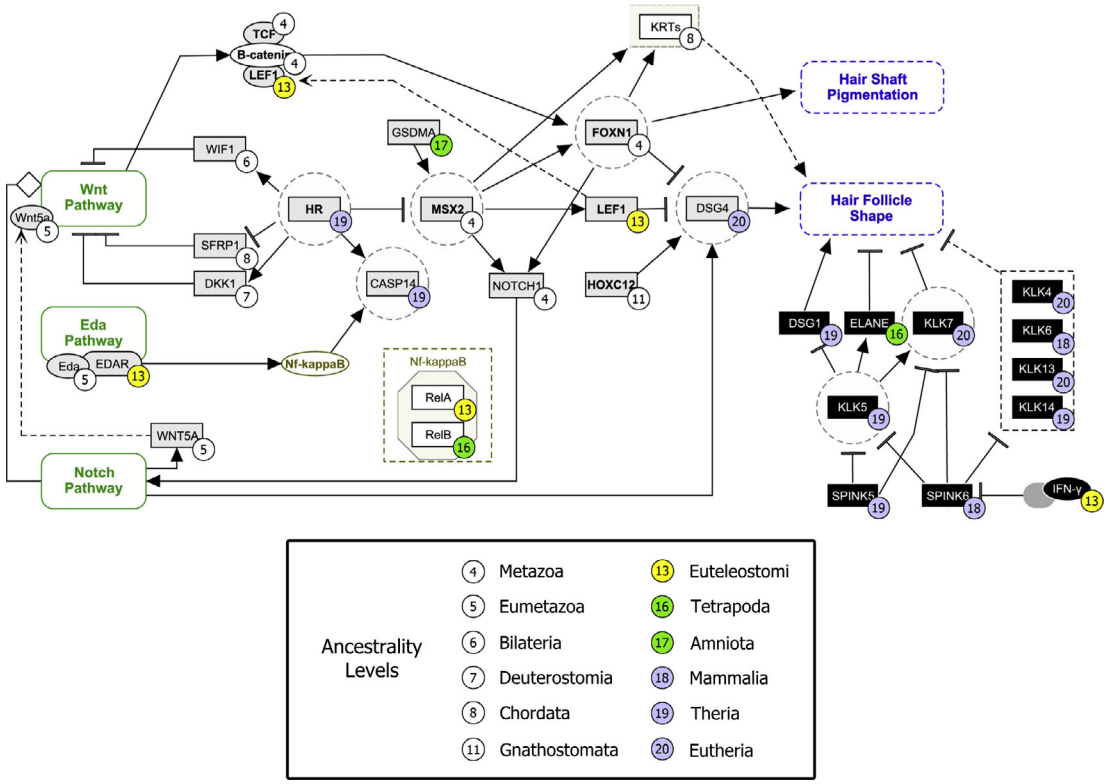

Fig. 5. Deciphering a pathway origin. A zoomed section of Fig. 2C was used to illustrate the evolution of the cytodifferentiation stage of hair follicle differentiation. For each gene, the orthologues as in [11] and the LCA determined as exemplified in Fig. 4. Numbers and colors in circles represent the clades of human lineage respectively shown in the legend where they have putatively originated.

in this epoch the specific inhibitors of KLKs also arise, known as serine protease inhibitor of Kazal-type (Spink), which are able to control KLK activity and therefore, provide a micro-habitat favorable to hair development [102]. Thus, with the presence of Spinks, production of ELANE and KLK7 would be impaired due to inhibition of KLK5, as shown in Fig. 5. This regulation might also moderate the suppression of DSG1, a desmoglein required for normal cellular adhesion between hair follicle cells [100]. The sub-regulatory network could have originated with Spink6 already present, controlling KLK6 (since they are present in Mammalia) and suffering modulation according to IFN-c levels, already present since Euteleostomi epoch.

Remarkably, in organisms that give birth to live young without using a shelled egg, the most important elements of cytodifferentiation process arise: HR and CASP14 in Theria, and DSG4 in Eutheria. As shown in Fig. 5, HR networks with several genes, suggesting a central importance for hair development. HR controls the inhibitors of WNT pathway and promotes activation of CASP14. CASP14 is present in hair follicles and is also stimulated by the EDA pathway, playing a role in the terminal differentiation of keratinocytes [131]. Furthermore, HR regulates the suitable expression of MSX2 and its target genes, implicated in cellular differentiation [83]. DSG4 is another desmoglein whose function is implicated with desmosome integrity in the hair shaft [132]. DSG4 is stimulated by the NOTCH pathway and by HOXC12, more ancient components, while it is inhibited by FOXN1 and LEF1, originated in Euteleostomi, therefore exerting a role in the fine-tuning between proliferation and differentiation in the hair follicle [93].

Thus, adding information of ancestrality to a pathway, one is allowed to depict the scenario of its evolution up to the present. Hair development relies on more ancient counterparts, and the key components for the machinery function become apparent with this simple analysis.

4. Discussion

PESCADOR works as a web-based tool for users to study biointeractions text-mined from a specific PubMed query. Although several online databases describe protein–protein interactions, such as iHop [133] or STRING [134], it is hard to associate the interaction with a specific biological event of interest [135]. This is accomplished with PESCADOR, which allows users to add customized biological concepts as search terms.

The success relies partially on a successful query as reviewed by Jensen and colleagues [136]. For instance, while searching for abstracts on breast development, it was difficult to exclude papers on breast cancer. Hair development was also a challenge. Often the PubMed query is easy to formulate, such as “embryo preimplantation development” [10]. In cases where the query is not straightforward, it is important to have a superficial knowledge about the phenomenon and start with a couple of more specific queries, such as (i) “(hair[Title/Abstract] AND follicle[Title/Abstract]) NOT (ear[Title/Abstract] OR auditory[Title/Abstract])”; (ii) “hair AND cortex AND matrix AND (inner OR IRS)”; or (iii) “(hair AND follicle) AND (organogenesis OR morphogenesis OR prenatal OR neonatal OR initiation OR primordia OR onset OR induction)”. These were used for deeply querying hair development.

Subsequently, the success also relies on selecting a few articles that describe the phenomenon in a particular way, mentioning genes, regulation of the phenomenon or its particularities. These abstracts are represented as vector of terms and used to classify the thousands of PubMed abstracts returned with the query mentioned above [137]. The software used for this is MedlineRanker [138]. It is recommended to work with the top 500–1000 abstracts which pass the threshold of 1e-2. Whenever the PESCADOR process is ready, it is possible to inspect the number of abstracts with at least one gene-pair extracted, since they appear highlighted in

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

Table A
Hair follicle development.

Term 1	Biointeractions	Term 2	PMID
"Eda/Edar pathway"	Indirectly inhibits	BMPs	17164417
"Notch/RBPJ pathway"	Related to	"HF pigmentation"	18283646
"Tgf-B2 pathway"	Induction of	"Ras–MAPK pathway"	15630473
"Tgf-B2 pathway"	Induction of	Snail	15630473
?	Inhibition of	Egfr	19474150
?	Inhibition of	Fgfr2	19474150
"Primary cilia"	Stimulation of	"Shh pathway"	18676816
"Ras–MAPK pathway"	Induction of	"Placode downgrowth"	15630473
"SC compartment"	Upregulation of	BCL11B	22383956
"SC compartment"	Upregulation of	CD200	16395407
"SC compartment"	Upregulation of	CD34	22383956
"SC compartment"	Upregulation of	CTGF	20050020
"SC compartment"	Upregulation of	DKK3	20050020
"SC compartment"	Upregulation of	DKK4	20050020
"SC compartment"	Upregulation of	EGFR	20050020
"SC compartment"	Upregulation of	FST	16395407
"SC compartment"	Upregulation of	GAS1	20050020
"SC compartment"	Upregulation of	KRT15	20050020
"SC compartment"	Upregulation of	KRT15	22383956
"SC compartment"	Upregulation of	KRT19	20050020
"SC compartment"	Upregulation of	LGR5	22028024
"SC compartment"	Upregulation of	LRIG1	22383956
"SC compartment"	Upregulation of	NFATC1	21892602
"SC compartment"	Upregulation of	PHLDA1	16395407
"SC compartment"	Upregulation of	PROM1 (CD133)	22383956
"SC compartment"	Upregulation of	S100A4	20050020
"SC compartment"	Upregulation of	Sox9	21892602
"SC compartment"	Upregulation of	TCF3	20050020
"SC compartment"	Upregulation of	TCF4	22028024
"SC compartment"	Upregulation of	Zfp145	21892602
"SCF/c-kit signaling"	Induction of	"HF pigmentation"	11259383
"Wnt pathway"	Upregulation of	Dlx3	18684741
Activin A	Actv. Expression of	Edar	11973284
Activin A	Actv. Expression of	Foxi3	23441037
Activin A	Upregulation of	Foxi3	23441037
ADAM17	Regulation of	"SC compartment"	22696231
ADAM17	Upregulation of	Gata3	22696231
ADAM17	Downregulation of	Lef1	22696231
ADAM17	Upregulation of	Lhx2	22696231
ADAM17	Upregulation of	Sox9	22696231
ADAMTS20	Related to	"HF pigmentation"	18480165
B-catenin	Actv. Expression of	CTNNB1	24451143
B-catenin	Upregulation of	Edar	24451143/19619491
B-catenin	Binds	Lef1	12646922
B-catenin	Actv. Expression of	MYC	22971575
B-catenin	Actv. Expression of	Shh	12917489
B-catenin	Actv. Expression of	Sp5	18480165
B-catenin	Upregulation of	Edar	24451143/19619491
B-catenin	Actv. Expression of	MYC	22971575
B-catenin	Actv. Expression of	Shh	12917489
Bmp	Inhibition of	"Placode downgrowth"	12646922
Bmp	Downregulation of	Lef1	12646922
Bmp	Downregulation of	Ncam	10559902
Bmp4	Upregulation of	Gata3	14610062
BMPR1A	Actv. Expression of	MSX2	15102710
BMPs	Inhibition of	"Placode formation"	16769906
BMPs	Inhibition of	Lef1	10559902
BMPs	Upregulation of	Msx2	17301089
BMPs	Inhibition of	NCAM	10559902
BMPs	Inhibition of	Lef1	10559902
BMPs	Inhibition of	NCAM	10559902
CDH1	Induction of	"Hair Bud Shape"	12646922
CDH1	Downregulation of	"Ras–MAPK pathway"	15630473
CDH1 (overexp.)	Inhibition of	"Hair Bud Shape"	12646922
CDH3	Induction of	"Placode shape"	18199584
CTGF	Inhibition of	BMPs	17164417
Cux1 (Cutl1)	Inhibition of	"HF differentiation"	11544187
Cyclin D1 (CCND1)	Induction of	"Placode downgrowth"	16481354
Cyclin D2 (CCND2)	Induction of	"Placode downgrowth"	12533516
Dkk4	Inhibition of	"Wnt Pathway"	18508042
Dkk4	Inhibition of	LRP5/6	16672341
Dlx3	Induction of	"HF differentiation"	18684741
Dlx3	Regulation	Gata3	18684741
Dlx3	Regulation	Hoxc13	18684741

(continued on next page)

Table A (continued)

Term 1	Biointeractions	Term 2	PMID
DSC2	Related to	"HF shape"	15739220
DSG1	Related to	"HF shape"	15606502/15482466
Dsg4	Related to	"HF shape"	18677469
Eda	Induction of	CTGF	17164417
Eda	Induction of	Fgf20	23431057
Eda	Upregulation of	Foxi3	23441037
Eda	Upregulation of	Foxi3	23441037
Eda	Actv. Expression of	Shh	21926481
Eda1	Actv. Expression of	Dkk4	18508042
Eda1	Binds	Edar	11973284
Eda1	Actv. Expression of	Shh	21926481
Egf	Inhibition of	"Placode formation"	19474150
Egfr	Inhibition of	"Placode formation"	19474150
EGFR (Krox20)	Actv. Expression of	Igfbp5	16875803
EGR2	Inhibition of	"HF shape"	16875803
ELANE (Ela2)	Inhibition of	"HF shape"	23344365
FGF	Inhibition of	Shh	23123965
FGF	Modulates	Igfbp5	16875803
Fgf20	Induction of	"Dermal aggregative growth"	23431057
Fgf20	Binds	FGFR1	24451143
Fgf7	Inhibition of	"Placode formation"	19474150
Fgfr2	Inhibition of	"Placode formation"	19474150
Follistatin	Inhibition of	Bmps	17164417
Fos	Actv. Expression of	Casp14	18424262
FOXE1	Related to	"HF shape"	15367491
Foxi3	Induction of	"Placode Shape"	23441037
Foxi3	Induction of	"Hair Bud Shape"	23441037
Foxn1	Induction of	"Hair pigmentation"	17803901
Foxn1	Induction of	"HF differentiation"	15739220
Foxn1	Inhibition of	Dsg4	19683850
Foxn1	Actv. Expression of	KRTs	19103190
Foxn1	Actv. Expression of	Notch1	19103190
Foxq1	Induction of	"HF differentiation"	11309849
Gata3	Induction of	"HF differentiation"	17151017
GJB6	Induction of	"Hair Bud Shape"	23219093
GJB6	Related to	"HF shape"	23219093
Gli2	Actv. Expression of	Cyclin D1 (CCND1)	12533516
Gli2	Actv. Expression of	Cyclin D2 (CCND2)	12533516
Gli2	Actv. Expression of	Foxe1	15367491
GLIs	Actv. Expression of	PTCH1	12917489
GPRC5D	Downregulation of	Nf-kappaB	15746257
Hoxc12	Actv. Expression of	Dsg4	19683850
Hoxc13	Induction of	"HF differentiation"	9420327
Hoxc13	Inhibition of	Dsg4	19683850
Hoxc13	Actv. Expression of	Foxn1	21191399
Hoxc13	Actv. Expression of	Foxq1	16835220
HR	Upregulation of	Casp14	21083932
HR	Upregulation of	Dkk1	24447645
HR	Downregulation of	Dlx3	22442153
HR	Downregulation of	Msx2	23702391
HR	Downregulation of	SFRP1	21083932
HR	Upregulation of	Wif1	21083932
IGFBP5	Inhibition of	"HF shape"	16024235
Jun	Inhibition of	Casp14	18424262/19747408
KITLG	Related to	"HF pigmentation"	18480165
Klk5	Actv. Expression of	ELANE (ELA2)	23344365
Klk5	Actv. Expression of	KLK7	23344365
KLK7	Inhibition of	"HF shape"	23344365
Laminin-511 (laminin-10)	Binds	ITGB1	18676816
Laminin-511/ITGB1	Induction of	"Primary Cilia"	18676816
Lef1	Downregulation of	CDH1	12646922
Lef1	Inhibition of	Dsg4	19683850
Lhx2	Actv. Expression of	Sox9	22028024
Msx2	Inhibition of	"Placode formation"	17301089
Msx2	Actv. Expression of	Foxn1	23702391
Msx2	Actv. Expression of	KRTs	19103190
Msx2	Actv. Expression of	Lef1	23702391
Msx2	Actv. Expression of	Notch1	19103190
MYC	Induction of	"Placode downgrowth"	21621827
Nf-kappaB	Actv. Expression of	Dkk4	19619491
Nf-kappaB	Actv. Expression of	FST	17164417
Nf-kappaB	Upregulation of	Wnt10a	19619491
Nf-kappaB	Actv. Expression of	Wnt10b	19619491
NF-kappaB	Upregulation of	Shh	16481354
Nf-kappaB	Actv. Expression of	Casp14	18424262

Table A (continued)

Term 1	Biointeractions	Term 2	PMID
Noggin	Inhibition of	BMP4	10559902
Noggin	Inhibition of	BMPs	10559902
Notch	Modulates	"Wnt Pathway"	15772135
Notch	Inhibition of	TP63	16618808
Notch	Upregulation of	Wnt5a	20634318
NR3C1 (GR)	Downregulation of	FOS	17935973
NR3C1 (GR)	Downregulation of	FOSB	17935973
NR3C1 (GR)	Upregulation of	KRT73	17935973
NR3C1 (GR)	Upregulation of	KRTAP15	17935973
NR3C1 (GR)	Upregulation of	KRTAP2-4	17935973
NR3C1 (GR)	Upregulation of	KRTAP3-3	17935973
NR3C1 (GR)	Upregulation of	KRTAP6-1	17935973
NR3C1 (GR)	Upregulation of	KRTAP8-1	17935973
PDGF	Binds	PDGFRA	10331973
PDGFRA	Induction of	"Primary cilia"	10331973
PERP	Related to	"HF shape"	15970683
Runx2	Actv. Expression of	PTCH1	18262513
Runx2	Upregulation of	Shh	1391898414
RUNX3	Related to	"HF shape"	15937937
Shh	Actv. Expression of	Ptch	12917489/9882493
Shh	Induction of	"Placode downgrowth"	9768360
Shh	Upregulation of	Cyclin D1 (CCND1)	14623238
Shh	Activates	GLI1	9882493
Shh	Actv. Expression of	Noggin (NOG)	9882493
Shh	Upregulation of	PTCH1	9882493
Shh	Actv. Expression of	WNT5A	11520664
Shh	Actv. Expression of	Sox9	16085486
Smad1	Actv. Expression of	Dlx3	11788714
Smad4	Actv. Expression of	Dlx3	11788714
Smad4	Upregulation of	Dsg4	16533311
Snail	Regulation of	"Placode downgrowth"	15630473
Sox2	Downregulation of	Bmp6	23153495
Sox2	Upregulation of	Sostdc1	23153495
Sox9	Induction of	"HF differentiation"	16085486
Sox9	Induction of	"SC compartment"	16085486
Sp5	Inhibition of	"Epidermal fate"	18480165
Sp5	Inhibition of	Involucrin (IVL)	18480165
Sp5	Inhibition of	Krt10	18480165
Spink5	Inhibition of	ELANE (ELA2)	23344365
Spink5	Actv. Expression of	KLK5	23344365
Spink5	Inhibition of	KLK7	23344365
SPINK6	Inhibition of	KLK12	24352040
SPINK6	Inhibition of	KLK13	24352040
SPINK6	Inhibition of	KLK14	24352040
SPINK6	Inhibition of	KLK2	24352040
SPINK6	Inhibition of	KLK4	24352040
SPINK6	Inhibition of	KLK5	24352040
SPINK6	Inhibition of	KLK6	24352040
SPINK6	Inhibition of	KLK7	24352040
Tcf3	Inhibition of	"HF differentiation"	11445543
TGFB2	Actv. Expression of	RUNX3	15156178
TNFA/IFNgamma	Downregulation of	SPINK6	24352040
TP63	Actv. Expression of	CDH3	18199584
TP63	Actv. Expression of	GJB6	23219093
TP63	Upregulation of	GJB6	23219093
TP63	Inhibition of	"HF differentiation"	16618808
TP63	Actv. Expression of	Gjb6	23219093
TP63	Activates	PERP	15970683
Versican	Induction of	"Dermal aggregative growth"	23099107
Wnt5a	Activates	"Placode downgrowth"	23099107
Wnt5a	Upregulation of	Foxn1	20634318
WNTs	Stabilizes	B-catenin	24451143
WNTs	Actv. Expression of	Dkk4	17397822
WNTs	Actv. Expression of	Eda	12039047
WNTs	Actv. Expression of	Versican	23099107
WNTs	Actv. expression of	Eda	12039047

green, therefore sensing the amount of information that might be mined from this particular PESCADOR run. In cases where the phenomenon might comprise several aspects, it is always possible to add up the result of several jobs.

One important point is that the user might have a background in Biological Sciences, which may not be so uncommon. More importantly, it is not required to be an expert. Actually the procedure described here drives the user on the knowledge of the phenomenon and its details. PESCADOR does this by guiding the curator from one gene to the other. Usually after depicting a core regulatory interaction, it is possible to promptly check ramifications upstream and downstream, consequently clarifying for the

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006

Table B
Mammary gland development.

Term 1	Biointeractions	Term 2	PMID
AKT1	Upregulates	ZEB2	19424592
AP1	Regulates	TIMP1	16678816
AP1	Regulates	VIM	16678816
AP1	Regulates	FOSL1	16678816
AP1	Regulates	FN1	16678816
AP1	Regulates	MYC	16678816
AP1	Regulates	CCND1	16678816
CAV1	Sequesters	JAK2	16231422
CEBP	Binds	CSN2	9513715
CEBPA	Suppress	CEBPB	9513715
CHI3L1	Increases	MMP9	21991364
CHI3L1	Suppress	CDH1	21991364
EGF	Binds	ERBB2	11146549
EGFR	Activates	MAPK	18562482
ELF5	Binds	WFDC	16231422
ERBB2	Associated	Estrogen signaling pathway	11146549
ERBB4	Stimulates	STAT5	16231422
ESR1	Regulates	AREG	23404106
ESR1	Regulates	PGR	23404106
ESR1	Regulates	STAT5A	23404106
ESR1	Stimulates	PGR	16231422
ESR1	Induces	TTC9	22917536
ESR2	Stimulates	PGR	16231422
Estrogen	Binds	ESR1	16231422
Estrogen	Binds	ESR2	16231422
Galanin	Regulates	PRL	16231422
Galanin	Activates	STAT5	16231422
Hedgehog signals	Upregulates	TGFB1	19424592
Hypoxia signals	Upregulates	ZEB2	19424592
IKKA	Activates	NFKB	16231422
IL1	Upregulates	ZEB2	19424592
IL6ST	Activates	STAT3	16231422
JAK2	Phosphorylates	PRLR	16231422
MYC	Downregulates	ITGA6	17880691
MYC	Downregulates	ITGB1	17880691
MYC	Activates	STAT5	15689376
MYC	Downregulates	CAV1	15689376
MYC	Associated	USF	12907752
MYC	Binds	TP53	17377501
MYC	Binds	BAX	17377501
MYC	Binds	E2F1	17377501
NFI	Binds	WFDC	9513715
NFKB	Binds	CCND1	16231422
NR3C1	Binds	CSN2	9513715
NR3C1	Binds	WFDC	9513715
NRG	Binds	ERBB1	16231422
NRG	Binds	ERBB2	16231422
NRG	Binds	ERBB3	16231422
NRG	Binds	ERBB4	16231422
Nrg3	Binds	ERBB4	17880691
Nrg3	Induces	MYC	17880691
NRIP1	Co-regulates	AREG	23404106
NRIP1	Co-regulates	PGR	23404106
NRIP1	Co-regulates	STAT5A	23404106
ORAI1	Co-expressed	ATP2C2	23840669
PGR	Induces	TNFRSF11A	16231422
PNCK	Downregulates	EGFR	18562482
PRL	Binds	PRLR	16231422
PRLR	Activates	STAT5	16231422
PRLR	Activates	MAPK	16231422
Progesterone	Binds	PGR	16231422
PTPN1	Inhibits	STAT5	23154416
PTPN1	Inhibits	PGR	23154416
SOCS3	Binds	IL6ST	16231422
STAT3	Associated	MYC	17377501
STAT5	Binds	CSN2	16231422
STAT5	Binds	WFDC	16231422
STAT5	Binds	ELF5	16231422
STAT5	Binds	GJB2	16231422
STAT5	Binds	TNFSF11	16231422
STAT5	Binds	CCND1	16231422
STAT5	Controls	BCL2L1	16231422
STAT5	Controls	SOCS3	16231422
TGFB1	Upregulates	ZEB2	19424592
TNF	Upregulates	ZEB2	19424592
TNFRSF11A	Activates	IKKA	16231422
TNFSF11	Binds	TNFRSF11A	16231422
TTC9	Interacts	TPM3	22917536
USF2	Induces	EIF4E	12907752
USF2	Induces	EIF4G	12907752
USF2	Induces	Oxytocin	12907752
YY1	Binds	CSN2	9513715
ZEB2	Represses	CDH1	19424592
ZEB2	Represses	CLDN4	19424592
ZEB2	Represses	CCND1	19424592
ZEB2	Represses	TERT	19424592
ZEB2	Represses	SFRP1	19424592
ZEB2	Represses	ALPL	19424592
ZEB2	Represses	MIR200A	19424592
ZEB2	Represses	MIR200B	19424592
ZEB2	Represses	MIR429	19424592
ZEB2	Upregulates	Mesenchymal markers	19424592

curator the details of the studied phenomenon and expanding the pathway.

Producing a pathway of a simple biological event or a complex one, such as those presented here, allows for an understanding of the contribution of individual genes to the system. Charts are often seen in seminal papers, or in reviews, but the tendency of accessing them in databases such as KEGG and Reactome is remarkable. Often, only partial pathways are found in independent publications.

The absence of certain pathways in these databases inspired us to develop software which could resolve that. Although some implementations try to computationally automate the networking of genes related to a given phenomenon, our experience shows that interactions of type 4, as qualified by PESCADOR (those where two genes or biological terms are found in different sentences in the abstract), contribute with much important information. Thus, manual curation that does not disregard type 4 interactions, yields denser networking and speeds up the procedure. PESCADOR supports jumping from one gene to another very quickly, in such a way that the curator reveals the biological aspects of the phenomenon while learning about the biointeractions implicated in it. PESCADOR therefore relies on curation to greatly expand the usage of the underlying software LAITOR [13].

Curators might be concerned about how good, both in terms of precision and completeness, the produced pathway is. One feasible way of looking for feedback is depositing the pathway in the WikiPathways database [5]. Experts in the field might spot them and either edit it or send feedback. However, often the end user is in the same curator's research group or institution. Building a pathway is a useful approach for those engaged in differential gene expression approaches, since sometimes the subject of investigation does not have a pathway in any database. And as more differentially expressed genes appear in the pathway, most likely the pathway will be useful for this immediate goal. Furthermore, what might make a pathway useful for general researchers is the production of an accompanying detailed text as exemplified here and in previous work [10]. Authors encourage users of this manual to also generate similar descriptions, thus inspection of its level of details, strategy for covering all genes and biointeractions, etc., might help on guiding the creation of novel pathways.

Although existing information about orthologues are retrievable from databases such as KEGG Orthology, within KEGG [139], or Panther [128] and this can be used to feed NCBI Taxonomy Common Tree, clustering of orthologues with software is also possible. OrthoMCL [129], Inparanoid [140] require comparisons of complete genomes. Our group contributed with software that clusters

cognate proteins among distinct proteomes, which are linked to a seed sequence [18]. Combined with KEGG orthology (KO) and other additions, we searched for enzymes implicated in the biosynthesis of amino acids [11]. Thus, with some extra bioinformatics analysis it is feasible to understand the evolution of the pathway, as we did for the preimplantation embryo development [10] and exemplified here for part of the hair development pathway that presented several recent genes. The case studies shown here will be further explored in respect to the origin of their participating genes, since they involve pathways or systems which are acquired relatively recently in human evolution.

5. Conclusions

Text mining was applied to two challenging issues, hair and breast development. A work in partial time conducted by non-experts for two months resulted in comprehensive pathways that covered the distinct stages of development. The methodological guide provided and exemplified with the two case studies presented, aim to support future initiatives. The final product, pathways along with descriptions, may assist biological studies such as comparative transcriptome analysis.

Appendix A

(See Tables A and B)

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2014.10.006.

References

[1] M.D. Stobbe, G.A. Jansen, P.D. Moerland, A.H.C. van Kampen, Brief. Bioinform. (2012), http://dx.doi.org/10.1093/bib/bbs060.
[2] M. Kanehisa, S. Goto, Nucleic Acids Res. 28 (2000) 27–30. http://dx.doi.org/gkd027.
[3] M. Kanehisa, Trends Genet. 13 (1997) 375–376.
[4] D. Croft, A.F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, et al., Nucleic Acids Res. 42 (2014) D472–D477, http://dx.doi.org/10.1093/nar/gkt1102.
[5] T. Kelder, M.P. van Iersel, K. Hanspers, M. Kutmon, B.R. Conklin, C.T. Evelo, et al., Nucleic Acids Res. 40 (2012) D1301–D1307, http://dx.doi.org/10.1093/nar/gkr1074.
[6] R. Hoffmann, M. Krallinger, E. Andres, J. Tamames, C. Blaschke, A. Valencia, Sci. STKE 2005 (2005) pe21, http://dx.doi.org/10.1126/stke.2832005pe21.
[7] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al., Genome Res. 13 (2003) 2498–2504, http://dx.doi.org/10.1101/gr.1239303.
[8] A. Funahashi, M. Morohashi, H. Kitano, N. Tanimura, BIOSILICO 1 (2003) 159–162, http://dx.doi.org/10.1016/S1478-5382(03)02370-9.
[9] M.P. van Iersel, T. Kelder, A.R. Pico, K. Hanspers, S. Coort, B.R. Conklin, et al., BMC Bioinform. 9 (2008) 399, http://dx.doi.org/10.1186/1471-2105-9-399.
[10] E. Donnard, A. Barbosa-Silva, R.L.M. Guedes, G.R. Fernandes, H. Velloso, M.J. Kohn, et al., BMC Genomics 12 (2011) S3, http://dx.doi.org/10.1186/1471-2164-12-S4-S3.
[11] R. Guedes, F. Prosdocimi, G. Fernandes, L. Moura, H. Ribeiro, J. Ortega, BMC Genomics 12 (2011) S2, http://dx.doi.org/10.1186/1471-2164-12-S4-S2.
[12] A. Barbosa-Silva, J.-F.F. Fontaine, E.R. Donnard, F. Stussi, J.M. Ortega, M.A. Andrade-Navarro, BMC Bioinform. 12 (2011) 435 (http://dx.doi.org/1471-2105-12-435, pii: 10.1186/1471-2105-12-435).
[13] A. Barbosa-Silva, T.G. Soldatos, I.L. Magalhaes, G.A. Pavlopoulos, J.F. Fontaine, M.A. Andrade-Navarro, et al., BMC Bioinform. 11 (2010) 70 (http://dx.doi.org/1471-2105-11-70, pii: 10.1186/1471-2105-11-70).
[14] C.M. Blakely, L. Sintasath, C.M. D'Cruz, K.T. Hahn, K.D. Dugan, G.K. Belka, et al., Development 132 (2005) 1147–1160, http://dx.doi.org/10.1242/dev.01655.
[15] D.L. Hadsell, S. Bonnette, J. George, D. Torres, Y. Klimentidis, Y. Klementidis, et al., Mol. Endocrinol. 17 (2003) 2251–2267, http://dx.doi.org/10.1210/me.2002-0031.
[16] S. Mika, B. Rost, Nucleic Acids Res. 32 (2004) W634–W637, http://dx.doi.org/10.1093/nar/gkh427.
[17] D. Maglott, J. Ostell, K.D. Pruitt, T. Tatusova, Nucleic Acids Res. 39 (2011), http://dx.doi.org/10.1093/nar/gkq1237.
[18] A. Barbosa-Silva, V.P. Satagopam, R. Schneider, J.M. Ortega, BMC Bioinform. 9 (2008) 141. http://dx.doi.org/1471-2105-9-141, (pii: 10.1186/1471-2105-9-141).
[19] M.L. Mikkola, Semin. Cell Dev. Biol. 18 (2007) 225–236, http://dx.doi.org/10.1016/j.semcdb.2007.01.007.
[20] R. Sennett, M. Rendl, Semin. Cell Dev. Biol. 23 (2012) 917–927, http://dx.doi.org/10.1016/j.semcdb.2012.08.011.
[21] M.R. Schneider, R. Schmidt-Ullrich, R. Paus, Curr. Biol. 19 (2009) R132–R142, http://dx.doi.org/10.1016/j.cub.2008.12.005.
[22] X. Wang, E.E. Tredget, Y. Wu, Stem Cells Dev. 21 (2012) 7–18, http://dx.doi.org/10.1089/scd.2011.0230.
[23] E. Fuchs, Nature 445 (2007) 834–842, http://dx.doi.org/10.1038/nature05659.
[24] T. Andl, S.T. Reddy, T. Gaddapara, S.E. Millar, Dev. Cell 2 (2002) 643–653.
[25] S.E. Millar, J. Invest. Dermatol. 118 (2002) 216–225, http://dx.doi.org/10.1046/j.0022-202x.2001.01670.x.
[26] R. DasGupta, E. Fuchs, Development 126 (1999) 4557–4568.
[27] S. Reddy, T. Andl, A. Bagasra, M.M. Lu, D.J. Epstein, E.E. Morrisey, et al., Mech. Dev. 107 (2001) 69–82.
[28] R.A. Ridgway, B. Serrels, S. Mason, A. Kinnaird, M. Muir, H. Patel, et al., Carcinogenesis 33 (2012) 2369–2376, http://dx.doi.org/10.1093/carcin/bgs284.
[29] Y. Zhang, T. Andl, S.H. Yang, M. Teta, F. Liu, J.T. Seykora, et al., Development 135 (2008) 2161–2172, http://dx.doi.org/10.1242/dev.017459.
[30] H. Bazzi, K.A. Fantauzzo, G.D. Richardson, C.A.B. Jahoda, A.M. Christiano, Dev. Biol. 305 (2007) 498–507, http://dx.doi.org/10.1016/j.ydbio.2007.02.035.
[31] C. Niemann, A.B. Unden, S. Lyle, C.C. Zouboulis, R. Toftgård, F.M. Watt, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 11873–11880, http://dx.doi.org/10.1073/pnas.1834202100.
[32] M.C. Durmowicz, C.Y. Cui, D. Schlessinger, Gene 285 (2002) 203–211.
[33] Y. Zhang, P. Tomann, T. Andl, N.M. Gallant, J. Huelsken, B. Jerchow, et al., Dev. Cell 17 (2009) 49–61, http://dx.doi.org/10.1016/j.devcel.2009.05.011.
[34] N. Wang, T. Yang, J. Li, M. Lei, J. Shi, W. Qiu, et al., Acta Histochem. 114 (2012) 199–206, http://dx.doi.org/10.1016/j.acthis.2011.04.009.
[35] M. Mukhopadhyay, M. Gorivodsky, S. Shtrom, A. Grinberg, C. Niehrs, M.I. Morasso, et al., Development 133 (2006) 2149–2154, http://dx.doi.org/10.1242/dev.02381.
[36] P. Rishikaysh, K. Dev, D. Diaz, W.M.S. Qureshi, S. Filip, J. Mokry, Int. J. Mol. Sci. 15 (2013) 1647–1670, http://dx.doi.org/10.3390/ijms15011647.
[37] C. Chiang, R.Z. Swan, M. Grachtchouk, M. Bolinger, Y. Litingtung, E.K. Robertson, et al., Dev. Biol. 205 (1999) 1–9, http://dx.doi.org/10.1006/dbio.1998.9103.
[38] J. Laurikkala, J. Pispa, H.-S. Jung, P. Nieminen, M. Mikkola, X. Wang, et al., Development 129 (2002) 2541–2553.
[39] M. Pummila, I. Fliniaux, R. Jaatinen, M.J. James, J. Laurikkala, P. Schneider, et al., Development 134 (2007) 117–125, http://dx.doi.org/10.1242/dev.02708.
[40] C.-Y. Cui, M. Kunisada, V. Childress, M. Michel, D. Schlessinger, Cell Cycle 10 (2011) 3379–3386, http://dx.doi.org/10.4161/cc.10.19.17669.
[41] I. Fliniaux, M.L. Mikkola, S. Lefebvre, I. Thesleff, Dev. Biol. 320 (2008) 60–71, http://dx.doi.org/10.1016/j.ydbio.2008.04.023.
[42] S.-Y. Tsai, R. Sennett, A. Rezza, C. Clavel, L. Grisanti, R. Zemla, et al., Dev. Biol. 385 (2013) 179–188, http://dx.doi.org/10.1016/j.ydbio.2013.11.023.
[43] V. Shirokova, M. Jussila, M.K. Hytönen, N. Perälä, C. Drögemüller, T. Leeb, et al., Dev. Dyn. 242 (2013) 593–603, http://dx.doi.org/10.1002/dvdy.23952.
[44] Y. Shimomura, M. Wajid, L. Shapiro, A.M. Christiano, Development 135 (2008) 743–753, http://dx.doi.org/10.1242/dev.006718.
[45] S.-H. Huh, K. Närhi, P.H. Lindfors, O. Häärä, L. Yang, D.M. Ornitz, et al., Genes Dev. 27 (2013) 450–458, http://dx.doi.org/10.1101/gad.198945.112.
[46] Y. Yang, Y. Li, Y. Wang, J. Wu, G. Yang, T. Yang, et al., J. Dermatol. Sci. 68 (2012) 157–163, http://dx.doi.org/10.1016/j.jdermsci.2012.09.011.
[47] C. Mou, B. Jackson, P. Schneider, P.A. Overbeek, D.J. Headon, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 9075–9080, http://dx.doi.org/10.1073/pnas.0600825103.
[48] G.D. Richardson, H. Bazzi, K.A. Fantauzzo, J.M. Waters, H. Crawford, P. Hynd, et al., Development 136 (2009) 2153–2164, http://dx.doi.org/10.1242/dev.031427.
[49] V.A. Botchkarev, Nat. Cell Biol. 1 (1999) 158–164, http://dx.doi.org/10.1038/11078.
[50] J.R. Hens, P. Dann, J.-P. Zhang, S. Harris, G.W. Robinson, J. Wysolmerski, Development 134 (2007) 1221–1230, http://dx.doi.org/10.1242/dev.000182.
[51] J.G. Abreu, N.I. Ketpura, B. Reversade, E.M. De Robertis, Nat. Cell Biol. 4 (2002) 599–604, http://dx.doi.org/10.1038/ncb826.
[52] M.H. Hardy, Trends Genet. 8 (1992) 55–61.
[53] A. Mukhopadhyay, S.R. Krishnaswami, C. Cowing-Zitron, N.-J. Hung, H. Reilly-Rhoten, J. Burns, et al., Dev. Biol. 373 (2013) 373–382, http://dx.doi.org/10.1016/j.ydbio.2012.10.024.
[54] D.J. Glotzer, E. Zelzer, B.R. Olsen, Dev. Biol. 315 (2008) 459–473, http://dx.doi.org/10.1016/j.ydbio.2008.01.005.
[55] C. Jamora, R. DasGupta, P. Kocieniewski, E. Fuchs, Nature 422 (2003) 317–322, http://dx.doi.org/10.1038/nature01458.
[56] R. Schmidt-Ullrich, D.J. Tobin, D. Lenhard, P. Schneider, R. Paus, C. Scheidereit, Development 133 (2006) 1045–1057, http://dx.doi.org/10.1242/dev.02278.
[57] P. Mill, R. Mo, H. Fu, M. Grachtchouk, P.C.W. Kim, A.A. Dlugosz, et al., Genes Dev. 17 (2003) 282–294, http://dx.doi.org/10.1101/gad.1038103.
[58] K. Michno, K. Boras-Granic, P. Mill, C.C. Hui, P.A. Hamel, Dev. Biol. 264 (2003) 153–165.
[59] C. Jamora, P. Lee, P. Kocieniewski, M. Azhar, R. Hosokawa, Y. Chai, et al., PLoS Biol. 3 (2005) e11, http://dx.doi.org/10.1371/journal.pbio.0030011.

[60] A. Fujimoto, M. Kurban, M. Nakamura, M. Farooq, H. Fujikawa, A.-G. Kibbi, et al., J. Dermatol. Sci. 69 (2013) 159–166, http://dx.doi.org/10.1016/j.jdermsci.2012.11.005.
[61] L. Schneider, C.A. Clement, S.C. Teilmann, G.J. Pazour, E.K. Hoffmann, P. Satir, et al., Curr. Biol. 15 (2005) 1861–1866, http://dx.doi.org/10.1016/j.cub.2005.09.012.
[62] J. Gao, M.C. DeRouen, C.-H. Chen, M. Nguyen, N.T. Nguyen, H. Ido, et al., Genes Dev. 22 (2008) 2111–2124, http://dx.doi.org/10.1101/gad.1689908.
[63] L. Karlsson, C. Bondjers, C. Betsholtz, Development 126 (1999) 2611–2621.
[64] O. Duverger, M.I. Morasso, Birth Defects Res. C, Embryo Today 87 (2009) 263–272, http://dx.doi.org/10.1002/bdrc.20158.
[65] J. Hwang, T. Mehrani, S.E. Millar, M.I. Morasso, Development 135 (2008) 3149–3159, http://dx.doi.org/10.1242/dev.022202.
[66] G.T. Park, M.I. Morasso, Nucleic Acids Res. 30 (2002) 515–522.
[67] D. Kurek, G.A. Garinis, J.H. van Doorninck, J. van der Wees, F.G. Grosveld, Development 134 (2007) 261–272, http://dx.doi.org/10.1242/dev.02721.
[68] K. Nagao, T. Kobayashi, M. Ohyama, H. Akiyama, K. Horiuchi, M. Amagai, Stem Cells 30 (2012) 1781–1785, http://dx.doi.org/10.1002/stem.1153.
[69] C.S. Potter, R.L. Peterson, J.L. Barth, N.D. Pruett, D.F. Jacobs, M.J. Kern, et al., J. Biol. Chem. 281 (2006) 29245–29255, http://dx.doi.org/10.1074/jbc.M603646200.
[70] L.F. Jave-Suarez, H. Winter, L. Langbein, M.A. Rogers, J. Schweizer, J. Biol. Chem. 277 (2002) 3718–3726, http://dx.doi.org/10.1074/jbc.M101616200.
[71] C. Clavel, L. Grisanti, R. Zemla, A. Rezza, R. Barros, R. Sennett, et al., Dev. Cell 23 (2012) 981–994, http://dx.doi.org/10.1016/j.devcel.2012.10.013.
[72] T. Andl, K. Ahn, A. Kairo, E.Y. Chu, L. Wine-Lee, S.T. Reddy, et al., Development 131 (2004) 2257–2268, http://dx.doi.org/10.1242/dev.01125.
[73] B.-C. Nguyen, K. Lefort, A. Mandinova, D. Antonini, V. Devgan, G. Della, Genes Dev. 20 (2006) 1028–1042, http://dx.doi.org/10.1101/gad.1406006.
[74] T. Ellis, L. Gambardella, M. Horcher, S. Tschanz, J. Capol, P. Bertram, et al., Genes Dev. 15 (2001) 2307–2319, http://dx.doi.org/10.1101/gad.200101.
[75] B.J. Merrill, U. Gat, R. DasGupta, E. Fuchs, Genes Dev. 15 (2001) 1688–1705, http://dx.doi.org/10.1101/gad.891401.
[76] B.-K. Kim, I.-C. Baek, H.-Y. Lee, J.-K. Kim, H.-H. Song, S.K. Yoon, BMC Genomics 11 (2010) 640, http://dx.doi.org/10.1186/1471-2164-11-640.
[77] L. Eckhart, W. Declercq, J. Ban, M. Rendl, B. Lengauer, C. Mayer, et al., J. Invest. Dermatol. 115 (2000) 1148–1151, http://dx.doi.org/10.1046/j.1523-1747.2000.00205.x.
[78] C. Ballaun, S. Karner, P. Mrass, M. Mildner, M. Buchberger, J. Bach, et al., Biochem. Biophys. Res. Commun. 371 (2008) 261–266, http://dx.doi.org/10.1016/j.bbrc.2008.04.050.
[79] J.L. Cascallana, A. Bravo, E. Donet, H. Leis, M.F. Lara, J.M. Paramio, et al., Endocrinology 146 (2005) 2629–2638, http://dx.doi.org/10.1210/en.2004-1246.
[80] E. Donet, P. Bayo, E. Calvo, F. Labrie, P. Pérez, J. Steroid Biochem. Mol. Biol. 108 (2008) 8–16, http://dx.doi.org/10.1016/j.jsbmb.2007.05.033.
[81] S. Fujimoto, T. Takase, N. Kadono, K. Maekubo, Y. Hirai, J. Dermatol. Sci. 74 (2014) 39–47, http://dx.doi.org/10.1016/j.jdermsci.2013.12.006.
[82] B.-K. Kim, H.-Y. Lee, I. Kim, K. Choi, J. Park, S.K. Yoon, J. Dermatol. Sci. 74 (2014) 81–87, http://dx.doi.org/10.1016/j.jdermsci.2013.12.007.
[83] B.-K. Kim, S.K. Yoon, J. Dermatol. Sci. 71 (2013) 203–209, http://dx.doi.org/10.1016/j.jdermsci.2013.04.019.
[84] J. Li, Y. Zhou, T. Yang, N. Wang, X. Lian, L. Yang, Biochem. Biophys. Res. Commun. 403 (2010) 18–23, http://dx.doi.org/10.1016/j.bbrc.2010.10.094.
[85] J. Cai, J. Lee, R. Kopan, L. Ma, Dev. Biol. 326 (2009) 420–430, http://dx.doi.org/10.1016/j.ydbio.2008.11.021.
[86] P. Hayward, K. Brennan, P. Sanders, T. Balayo, R. DasGupta, N. Perrimon, et al., Development 132 (2005) 1819–1830, http://dx.doi.org/10.1242/dev.01724.
[87] B. Hu, K. Lefort, W. Qiu, B.-C. Nguyen, R.D. Rajaram, E. Castillo, et al., Genes Dev. 24 (2010) 1519–1532, http://dx.doi.org/10.1101/gad.1886910.
[88] A.N. Mardaryev, N. Meier, K. Poterlowicz, A.A. Sharov, T.Y. Sharova, M.I. Ahmed, et al., Development 138 (2011) 4843–4852, http://dx.doi.org/10.1242/dev.070284.
[89] B. Kinzel, M. Pikiolek, V. Orsini, J. Sprunger, A. Isken, S. Zietzling, et al., Dev. Biol. 390 (2014) 181–190, http://dx.doi.org/10.1016/j.ydbio.2014.03.009.
[90] S.A. Johns, S. Soullier, P. Rashbass, V.T. Cunliffe, Dev. Dyn. 232 (2005) 1062–1068, http://dx.doi.org/10.1002/dvdy.20278.
[91] M.-C. Zhang, H. Furukawa, K. Tokunaka, K. Saiga, F. Date, Y. Owada, et al., Immunogenetics 60 (2008) 599–607, http://dx.doi.org/10.1007/s00251-008-0320-4.
[92] H. Bazzi, A. Getz, M.G. Mahoney, A. Ishida-Yamamoto, L. Langbein, J.K. Wahl, et al., Differentiation 74 (2006) 129–140, http://dx.doi.org/10.1111/j.1432-0436.2006.00061.x.
[93] H. Bazzi, S. Demehri, C.S. Potter, A.G. Barber, A. Awgulewitsch, R. Kopan, et al., Differentiation 78 (2009) 292–300, http://dx.doi.org/10.1016/j.diff.2009.06.004.
[94] T. Schlake, Mech. Dev. 123 (2006) 641–648, http://dx.doi.org/10.1016/j.mod.2006.06.001.
[95] K. Miyazono, S. Maeda, T. Imamura, Oncogene 23 (2004) 4232–4237, http://dx.doi.org/10.1038/sj.onc.1207131.
[96] E. Raveh, S. Cohen, D. Levanon, Y. Groner, U. Gat, Dev. Dyn. 233 (2005) 1478–1487, http://dx.doi.org/10.1002/dvdy.20453.
[97] A. Brancaccio, A. Minichiello, M. Grachtchouk, D. Antonini, H. Sheng, R. Parlato, et al., Hum. Mol. Genet. 13 (2004) 2595–2606, http://dx.doi.org/10.1093/hmg/ddh292.
[98] J.A. McGrath, V. Wessagowit, Keio J. Med. 54 (2005) 72–79.
[99] D. Brennan, Y. Hu, A. Kljuic, Y. Choi, S. Joubeh, M. Bashkin, et al., Differentiation 72 (2004) 434–449, http://dx.doi.org/10.1111/j.1432-0436.2004.07208009.x.
[100] Y. Hanakawa, H. Li, C. Lin, J.R. Stanley, G. Cotsarelis, J. Invest. Dermatol. 123 (2004) 817–822, http://dx.doi.org/10.1111/j.0022-202X.2004.23479.x.
[101] A. Hovnanian, Cell Tissue Res. 351 (2013) 289–300, http://dx.doi.org/10.1007/s00441-013-1558-1.
[102] J. Fischer, Z. Wu, T. Kantyka, M. Sperrhacke, O. Dimitrieva, Y. Koblyakova, et al., J. Invest. Dermatol. 134 (2014) 1305–1312, http://dx.doi.org/10.1038/jid.2013.502.
[103] R.A. Ihrie, L.D. Attardi, Cell Cycle 4 (2005) 873–876.
[104] G. Barsh, G. Cotsarelis, Cell 130 (2007) 779–781, http://dx.doi.org/10.1016/j.cell.2007.08.032.
[105] N.V. Botchkareva, M. Khlgatian, B.J. Longley, V.A. Botchkarev, B.A. Gilchrest, FASEB J. 15 (2001) 645–658, http://dx.doi.org/10.1096/fj.00-0368com.
[106] L. Weiner, R. Han, B.M. Scicchitano, J. Li, K. Hasegawa, M. Grossi, et al., Cell 130 (2007) 932–942, http://dx.doi.org/10.1016/j.cell.2007.07.024.
[107] K. Schouwey, F. Beermann, Histol. Histopathol. 23 (2008) 609–619.
[108] M. Ohyama, A. Terunuma, C.L. Tock, M.F. Radonovich, C.A. Pise-Masison, S.B. Hopping, et al., J. Clin. Invest. 116 (2006) 249–260, http://dx.doi.org/10.1172/JCI26043.
[109] X. Liang, S. Bhattacharya, G. Bajaj, G. Guha, Z. Wang, H.-S. Jang, et al., PLoS ONE 7 (2012) e29999, http://dx.doi.org/10.1371/journal.pone.0029999.
[110] A. Vollmers, L. Wallace, N. Fullard, T. Höher, M.D. Alexander, J. Reichelt, Stem Cell Rev. 8 (2012) 402–413, http://dx.doi.org/10.1007/s12015-011-9314-y.
[111] L. Rittié, S.W. Stoll, S. Kang, J.J. Voorhees, G.J. Fisher, Aging Cell 8 (2009) 738–751.
[112] M. Smalley, A. Ashworth, Nat. Rev. Cancer 3 (2003) 832–844, http://dx.doi.org/10.1038/nrc1212.
[113] L. Hennighausen, G.W. Robinson, Nat. Rev. Mol. Cell Biol. 6 (2005) 715–725, http://dx.doi.org/10.1038/nrm1714.
[114] M. Katoh, M. Katoh, Int. J. Oncol. 34 (2009) 1737–1742.
[115] H. Panchal, O. Wansbury, S. Parry, A. Ashworth, B. Howard, BMC Dev. Biol. 7 (2007) 105, http://dx.doi.org/10.1186/1471-213X-7-105.
[116] Q. Shen, Y. Zhang, I.P. Uray, J.L. Hill, H.-T. Kim, C. Lu, et al., Dev. Biol. 295 (2006) 589–603, http://dx.doi.org/10.1016/j.ydbio.2006.03.042.
[117] J. Nautiyal, J.H. Steel, M.R. Mane, O. Oduwole, A. Poliandri, X. Alexi, et al., Development 140 (2013) 1079–1089, http://dx.doi.org/10.1242/dev.085720.
[118] S. Mukherjee, S.G. Louie, M. Campbell, L. Esserman, G. Shyamala, Oncogene 19 (2000) 5982–5987, http://dx.doi.org/10.1038/sj.onc.1203964.
[119] J.M. Rosen, C. Zahnow, A. Kazansky, B. Raught, Biochem. Soc. Symp. 63 (1998) 101–113.
[120] E.S. Milani, H. Brinkhaus, R. Dueggeli, I. Klebba, U. Mueller, M. Stadler, et al., Development 140 (2013) 117–125, http://dx.doi.org/10.1242/dev.082941.
[121] S. Shrestha, S. Cao, V.C.L. Lin, Biochem. Biophys. Res. Commun. 426 (2012) 65–70, http://dx.doi.org/10.1016/j.bbrc.2012.08.036.
[122] T.B. Deb, C.M. Coticchia, R. Barndt, H. Zuo, R.B. Dickson, M.D. Johnson, Am. J. Physiol. Cell Physiol. 295 (2008) C365–C377, http://dx.doi.org/10.1152/ajpcell.00449.2007.
[123] B.M. Cross, A. Hack, T.A. Reinhardt, R. Rao, PLoS ONE 8 (2013) e67348, http://dx.doi.org/10.1371/journal.pone.0067348.
[124] S. Scully, W. Yan, B. Bentley, Q.J. Cao, R. Shao, PLoS ONE 6 (2011) e25819, http://dx.doi.org/10.1371/journal.pone.0025819.
[125] K.D. Sutherland, G.J. Lindeman, J.E. Visvader, Cell Cycle 6 (2007) 799–803.
[126] G.R. Fernandes, Genet. Mol. Res. 7 (2008) 910–924 (http://dx.doi.org/ X-Meeting008).
[127] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, et al., Nucleic Acids Res. 42 (2014) D222–D230, http://dx.doi.org/10.1093/nar/gkt1223.
[128] H. Mi, B. Lazareva-Ulitsky, R. Loo, A. Kejariwal, J. Vandergriff, S. Rabkin, et al., Nucleic Acids Res. 33 (2005) D284–D288, http://dx.doi.org/10.1093/nar/gki078.
[129] L. Li, C.J. Stoeckert Jr., D.S. Roos, C.J. Stoeckert, Genome Res. 13 (2003) 2178–2189, http://dx.doi.org/10.1101/gr.1224503 13/9/2178.
[130] L. Furio, S. de Veer, M. Jaillet, A. Briot, A. Robin, C. Deraison, et al., J. Exp. Med. 211 (2014) 499–513, http://dx.doi.org/10.1084/jem.20131797.
[131] I.D. Gkegkes, K. Aroni, G. Agrogiannis, E.S. Patsouris, A.E. Konstantinidou, Arch. Dermatol. Res. 305 (2013) 379–387, http://dx.doi.org/10.1007/s00403-013-1319-8.
[132] M. Farooq, M. Ito, M. Naito, Y. Shimomura, Br. J. Dermatol. 165 (2011) 425–431, http://dx.doi.org/10.1111/j.1365-2133.2011.10373.x.
[133] R. Hoffmann, A. Valencia, Bioinformatics 21 (Suppl. 2) (2005) ii252–ii258, http://dx.doi.org/10.1093/bioinformatics/bti1142.
[134] C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, B. Snel, Nucleic Acids Res. 31 (2003) 258–261.
[135] S. Ananiadou, S. Pyysalo, J. Tsujii, D.B. Kell, Trends Biotechnol. 28 (2010) 381–390, http://dx.doi.org/10.1016/j.tibtech.2010.04.005 (pii: S0167-7799(10)00072-7).
[136] L.J. Jensen, J. Saric, P. Bork, Nat. Rev. Genet. 7 (2006) 119–129, http://dx.doi.org/10.1038/nrg1768.
[137] W.J. Wilbur, Y. Yang, Comput. Biol. Med. 26 (1996) 209–222.
[138] J.F. Fontaine, A. Barbosa-Silva, M. Schaefer, M.R. Huska, E.M. Muro, M.A. Andrade-Navarro, Nucleic Acids Res. 37 (2009) W141–W146 (http://dx.doi.org/gkp353, pii: 10.1093/nar/gkp353).
[139] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, Nucleic Acids Res. 27 (1999) 29–34, http://dx.doi.org/10.1093/nar/28.1.27.
[140] K.P. O'Brien, M. Remm, E.L.L. Sonnhammer, Nucleic Acids Res. 33 (2005) D476–D480, http://dx.doi.org/10.1093/nar/gki107.

Please cite this article in press as: D. Trindade et al., Methods (2014), http://dx.doi.org/10.1016/j.ymeth.2014.10.006