



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline ou spécialité : *Gènes, Cellules & Développement*

Présentée et soutenue par *Nicolas MATTIUZZO*
Le *23 octobre 2009*

Titre : *ETUDE DU TRANSCRIPTOME KERATINOCYTAIRE
AU COURS DU PROGRAMME DE DIFFERENCIATION DE L'EPIDERME*

JURY

Frédéric CAUX (Rapporteur)
Judith FISCHER (Rapporteur)
Jérôme LAMARTINE (Rapporteur)
Nathalie JONCA (Examineur)
Guy SERRE (Examineur)
Rémy POUPOT (Président)

Ecole doctorale : *Biologie-Santé-Biotechnologie*
Unité de recherche : *Unité Différenciation Epidermique et Autoimmunité Rhumatoïde
(UMR5165 CNRS-Université de Toulouse)*
Directeur(s) de Thèse : *Marina GUERRIN WEBER*
Rapporteurs : *Frédéric CAUX, Judith FISCHER, Jérôme LAMARTINE*

REMERCIEMENTS

Je tiens à exprimer toute ma gratitude à Guy Serre pour m'avoir accueilli au sein de son laboratoire, où j'ai pu trouver des conditions de travail exceptionnelles tout au long de ces années.

Un grand merci à Judith Fischer, dirigeant l'équipe Maladies Dermatologiques au Centre National de Génotypage (Evry), à Frédéric Caux, Professeur de dermatologie à l'Hôpital Avicenne (Bobigny), et à Jérôme Lamartine, Professeur de génétique humaine à l'Université Claude Bernard (Lyon), qui ont tous les trois accepté d'évaluer mon travail de thèse en tant que rapporteurs et fait le déplacement jusqu'à Toulouse. Merci également à Rémy Poupot, Professeur de biochimie à l'Université Paul Sabatier, d'avoir accepté de présider le jury de cette thèse.

MERCI (mais même en majuscules, le mot ne suffit probablement pas) à Marina Weber, pour tout... Encadrer un pseudo-« génie des alpages » n'a pas du être facile tous les jours, merci de ne t'être pas (trop) vengée dans les Pyrénées espagnoles, même si j'estime que nous sommes quittes vu certains des endroits où tu m'as fait passer ! Plus sérieusement, bravo pour ton encadrement qui a été impeccable quelles que soient les circonstances. En tout cas, c'est quand tu veux pour une petite rando dans les Appalaches ou un méchoui sur les rives du Potomac !

J'adresse un énorme merci à l'ensemble du laboratoire, pas seulement pour le soutien professionnel ou personnel dont j'ai pu bénéficier, mais aussi pour l'ambiance unique, les nombreux sujets de débats autour des grandes tablées de la cafétéria... et les 5-6 kilos que j'ai pris pendant ces années (la gourmandise est contagieuse, et j'ai été bien entouré !)

Bon, pour Eve, je n'ai même pas vraiment besoin de le faire, tu sais toute l'admiration et l'amitié que j'ai pour toi, même si une certaine question d'optique et de Pyrénées reste encore source de discorde... (mais j'aurai le dernier mot !). Merci pour ton amitié ô combien précieuse (au moins autant que le vin de ton mari !) Et je maintiens... le labo sans toi c'est plus tout à fait pareil, mais bon, je m'en fiche, j'ai récupéré ton bureau.

Merci également à Emilie, la seule personne à être parvenue à me faire venir à Ikea un samedi à huit heures du matin, et à avoir vraiment adoré mes sablés à la lavande... Lequel des deux est le plus grand exploit ? (Au passage, pardon à tous pour les nombreux plats étonnants/étranges/suspects/immangeables que j'ai porté à intervalles – trop – réguliers). Merci pour ton entrain, et navré de t'avoir remerciée en tentant de te faire passer pour une personne violente et sanguinaire (à la rigueur, mon temps sur Toulouse étant Comté, j'ai quelques raisons de te craindre désormais). (Profite, Cyril, c'est mon dernier *private* calembour !)

Merci aussi à Mitou, ta bonne humeur en toutes circonstances et tes doigts de fée en toutes expériences en ont impressionnées plus d'un, et je ne fais pas exception... Comme pour Eve, le labo sans toi, ça fait un peu bizarre... Merci vraiment pour tout.

Hors équipe, une pensée spéciale à Mireille et Cyril, pour les trop rares conversations non-professionnelles que nous avons pu avoir, qu'il s'agisse de littérature étrang(èr)e, de philosophie ou... de calembours.

Evidemment je remercie toute ma famille pour m'avoir soutenu. Merci à Lucie, qui est même parvenue un temps à se souvenir de mon sujet... Merci surtout à ma mère, pour pas mal de choses, comme m'avoir supporté (dans tous les sens du terme) pendant toutes ces années, ou avoir trouvé le moyen de satisfaire ma curiosité quand j'étais gamin. Merci à tous

d'être venus le jour J. Je dédie cette thèse à mon grand-père, je suis sûr qu'il aurait adoré être là.

Pour leur soutien et leur amitié qui remontent à bien avant la thèse pour certains, merci à Mag, Ludo, Denis, Thierry, Laeti, Karène, Élé, Rami, Momo, Miriam et tous les autres... Marie-Hélène, Juju et Mathias, vous êtes « hors catégories », votre présence à été mon oxygène, alors « merci » ne suffirait pas, tant pis. Et puis y'a déjà trop de « merci » de toute façon.

Et parce qu'avec une vie sociale plutôt famélique il n'y a pas que des personnes qui m'ont tenu compagnie, merci (pèle-mèle) à : ceux qui tombent en dehors, aux habitants d'Hyrule, aux seize colosses, aux ombres d'Ico, à la Princesse qui n'est jamais dans le château et au pinceau des Dieux ; aux groupes islandais mondialement connus, tels que Sigur Ros ou Múm ; à JJG ; aux pandas roux (réels ou virtuels), au nombre 404, au BSoD et au RRoD ; à la poussière qui rentre dans mon appareil photo à chaque changement d'objectif ; aux Pyrénées que l'on peut apercevoir (ou pas) depuis Toulouse et au seul vrai pic du Midi, celui d'Ossau évidemment... (oui, je sais, quelques uns attendaient encore plus de références de *geek*... je fais ce que je peux $\neg \neg$)

Enfin, merci à ma thèse, pour avoir répondu à la lancinante question « Peut-on faire un thèse et avoir une vraie vie sociale ? »... La réponse est quelque part au dessus on dirait.

LISTE DES ABREVIATIONS

aa : acides aminés
ADN : Acide DésoxyriboNucléique
ADNc : ADN complémentaire
ARN : Acide RiboNucléique
ARNm : ARN messenger
ATP : Adenosine TriPhosphate
gb/kb/pb : gigabases/kilobases/paires de bases
CGH : Comparative Genomic Hybridization
ChIP : Chromatin ImmunoPrecipitation
CNV : Copy Number Variation
Cre-Lox : système recombinase Cre - sites *LoxP*
Ct : Cycle threshold
dNTP : désoxyribonucléotides triphosphates
EDC : Epidermal Differentiation Complex
EGF : Epidermal Growth Factor
EPU : Epidermal Proliferative Unit
EST : Expressed Sequence Tag
FACS : Fluorescent-Activated Cell Sorting
FDR : False Discovery Rate
FT : Facteur de transcription
GFP : Green Fluorescent Protein
GO : Gene Ontology
(HN)SCC : (Head and Neck) Squamous Cell Carcinoma
kDa : kiloDalton
LOH : Loss Of Heterozygosity
LOWESS : LOcally WEighted Scatterplot Smoothing
LRC : Label-Retaining Cells
MAPK : Mitogen-Activated Protein Kinase
MIAME : Minimal Information About Microarray Experiment
ORESTE : Open Reading Frame EST
PCR : Polymerase Chain Reaction
PKC : Protein Kinase C
qRT-PCR : quantitative Real-Time PCR
RT : Reverse Transcription
SAGE : Serial Analysis of Gene Expression
SNP : Single Nucleotide Polymorphism
TA : Transit Amplifying cells
FT : Facteur de Transcription
TFBS : Transcription Factor Binding Site
TNF : Tumor Necrosis Factor
vs : versus

RÉSUMÉ

Au sein de l'épiderme, la fonction de barrière est assurée par la partie la plus superficielle, la couche cornée, constituée de cornéocytes, des cellules mortes très cohésives entourées d'une matrice lipidique. Un processus complexe de mort cellulaire programmée, la cornification, aboutit à la transformation des dernières cellules vivantes, les kératinocytes granuleux, en cornéocytes dépourvus d'activité transcriptionnelle et traductionnelle. Les kératinocytes granuleux constituent donc à la fois le stade ultime de la différenciation épidermique, mais aussi celui où sont produits les différents acteurs de la cornification et de la desquamation, donc de la fonction barrière.

Très peu d'études moléculaires systématiques ont été consacrées à l'épiderme humain, et seulement quatre facteurs de transcription spécifiques requis pour la fonction barrière ont été identifiés. Le travail présenté ici combine plusieurs techniques à grande échelle afin d'identifier de nouveaux gènes impliqués dans les étapes tardives de ce programme. Un procédé de purification de cellules à partir d'épiderme humain normal a été développé afin d'obtenir des fractions enrichies en kératinocytes de différents stades de différenciation. Ce procédé, couplé à la technique des ORESTEs, a permis dans un premier temps de produire 22 000 séquences et d'identifier 3387 gènes exprimés dans la couche granuleuse, parmi lesquels de nouveaux marqueurs de la différenciation impliqués dans diverses fonctions (protéine structurale, métabolisme et transport des lipides, protéase et inhibiteurs potentiellement impliqués dans la desquamation, etc). Les profils d'expression génique des kératinocytes granuleux ont également été comparés avec ceux des kératinocytes basaux au moyen de puces à ADN pangénomiques, mettant en évidence 200 candidats. Afin d'identifier des facteurs de transcription supplémentaires, une recherche bioinformatique de sites de fixation conservés évolutivement a été réalisée sur les promoteurs de 52 marqueurs de la différenciation.

Les résultats issus de ces trois techniques ont été combinés et analysés afin de réduire à 300 gènes le nombre de candidats à valider. Pour 155 d'entre eux, leur profil d'expression au sein de l'épiderme a pu être analysé par PCR quantitative, permettant d'identifier 49 nouveaux marqueurs de la différenciation. De plus, sur 94 gènes quantifiés codant pour des facteurs de transcription, 37 apparaissent positivement ou négativement régulés, suggérant un rôle dans le contrôle des étapes tardives du programme de différenciation.

Ces résultats ouvrent de nombreuses voies et constituent donc une avancée importante pour la compréhension des mécanismes complexes mis en jeu dans ce réseau de régulation génique, tant dans le cadre physiologique que pour certaines pathologies caractérisées par une altération de la différenciation, comme dans le cas du psoriasis ou des ichthyoses.

ABSTRACT

The barrier function of the epidermis is provided by its outermost part, the cornified layer, formed by corneocytes, strongly interconnected dead cells surrounded by a lipid-rich extracellular environment. A complex programmed cell death, named cornification, turns the last living cells, granular keratinocytes, into dead corneocytes devoid of transcriptional and translational activity. Since all the constituents of corneocytes as well as the enzymes regulating their release leading to the desquamation are produced by granular keratinocytes, they do not only represent the last step of epidermal differentiation, but also its climax.

Very few large-scale studies have been performed on human epidermis, and until now only four transcription factors required for the barrier function were described. The present work combines several large-scale techniques in order to identify new genes, especially transcription factor (TF)-coding genes, potentially involved in the last step of epidermal differentiation.

Based on a purification process using iterative trypsinizations of chilled epidermis fragments, we recovered cell fractions enriched in keratinocytes from various differentiation states. This process, used together with the ORESTEs technique, allowed the production of more than 22000 sequences corresponding to 3387 genes expressed in the granular layer, including new differentiation markers involved in various functions (structural protein, lipid metabolism and transport, protease and inhibitors, etc). Gene expression profile of granular keratinocytes was also compared with basal keratinocytes' one using pangenomic microarrays, highlighting 200 candidates. To identify some additional TF-coding genes, a study of conserved TF-binding site was performed by bioinformatics on the promoters of 52 differentiation markers. The results produced by these three techniques were combined and analyzed to select 300 candidates to be validated by real-time PCR. Among them, 155 could be reliably quantified, including 49 new differentiation-associated genes. Moreover, 37 of 94 TF-coding genes appeared to be up- or downregulated in granular keratinocytes, suggesting a function in the regulation of the differentiation.

These advances made in the field of normal keratinocyte differentiation should improve the knowledge of the complex regulatory networks driving the formation of the epidermal barrier, and may also contribute to the deciphering of complex diseases such as atopic dermatitis, psoriasis and ichthyoses, which are characterized by deregulated differentiation programs.

Table des matières

I. INTRODUCTION	13
1. ETUDE DU TRANSCRIPTOME ET DE LA REGULATION GENIQUE	15
1.1. ETUDES STATIQUES DE TRANSCRIPTOMES	15
1.1.1. BANQUES D'ESTS	15
1.1.2. TECHNIQUE DES ORESTES	17
1.2. ETUDES COMPARATIVES DE TRANSCRIPTOMES	19
1.2.1. PUCES A ADN	19
1.2.1.1. Origines de la technique et principe	19
1.2.1.2. Types de support	20
1.2.1.3. Types de conception des puces d'expression génique.....	22
1.2.1.4. Applications	24
1.2.1.5. Normalisation et tests statistiques.....	27
1.2.2. PCR QUANTITATIVE EN TEMPS REEL	36
1.2.2.1. Origine et principe.....	36
1.2.2.2. Normalisation	37
1.2.2.3. Applications	38
1.2.3. SAGE & SEQUENÇAGE NOUVELLE GENERATION	40
1.3. BIOINFORMATIQUE.....	45
1.3.1. BANQUES DE DONNEES	45
1.3.2. OUTILS D'ANNOTATION	46
1.3.3. IDENTIFICATION IN SILICO DE SEQUENCES REGULATRICES	48
1.3.3.1. Recherche de sites de fixation pour facteurs de transcription.....	49
1.3.3.2. Recherche de régions conservées.....	50
1.3.3.3. Outils d'analyse des régions régulatrices.....	51

2. LA DIFFERENCIATION EPIDERMIQUE.....54

2.1.	PRESENTATION GENERALE	54
2.1.1.	LA PEAU.....	54
2.1.2.	L'HYPODERME ET LE DERME	55
2.1.3.	LA JONCTION DERMO-EPIDERMIQUE	55
2.1.4.	L'EPIDERME	56
2.1.5.	LES ANNEXES EPIDERMiques	56
2.2.	LA DIFFERENCIATION KERATINOCYTAIRE.....	58
2.2.1.	ASPECTS MORPHOLOGIQUES	58
2.2.1.1.	La couche basale (stratum basale).....	59
2.2.1.2.	La couche épineuse (stratum spinosum).....	59
2.2.1.3.	La couche granuleuse (stratum granulosum).....	60
2.2.1.4.	La couche cornée (stratum corneum).....	60
2.2.2.	ASPECTS FONCTIONNELS	62
2.2.2.1.	Homéostasie de l'épiderme	62
2.2.2.1.1.	<i>Couche basale et auto-renouvellement</i>	<i>62</i>
2.2.2.1.2.	<i>Couche cornée et desquamation</i>	<i>67</i>
2.2.2.2.	Fonction barrière de l'épiderme	72
2.2.2.2.1.	<i>Imperméabilité</i>	<i>73</i>
2.2.2.2.1.1.	Lipides intercornéocytaires.....	73
2.2.2.2.1.2.	Jonctions serrées	77
2.2.2.2.2.	<i>Résistance mécanique</i>	<i>81</i>
2.2.2.2.2.1.	Matrice fibreuse.....	81
2.2.2.2.2.2.	Enveloppe cornée	84
2.2.2.2.2.3.	Cornéodesmosomes.....	92
2.2.3.	REGULATION DES ETAPES TARDIVES DE LA DIFFERENCIATION .	95
2.2.3.1.	Voies de signalisation	95
2.2.3.1.1.	<i>Voie calcique</i>	<i>95</i>
2.2.3.1.2.	<i>Voie de la vitamine D3</i>	<i>99</i>
2.2.3.1.3.	<i>Voie de l'acide arachidonique</i>	<i>101</i>
2.2.3.2.	Facteurs de transcription	102
2.2.3.2.1.	<i>Facteurs ubiquistes.....</i>	<i>102</i>
2.2.3.2.2.	<i>Facteurs d'expression tissulaire restreinte</i>	<i>105</i>
2.2.3.2.2.1.	DLX3.....	105

2.2.3.2.2.2.	KLF4	106
2.2.3.2.2.3.	GRHL3.....	109
2.2.3.2.2.4.	GATA3.....	112
2.2.3.2.2.5.	PRDM1 (BLIMP-1)	113

2.3. PATHOLOGIES ASSOCIEES AU PROGRAMME DE DIFFERENCIATION EPIDERMIQUE 116

2.3.1.	LES GENODERMATOSES.....	116
2.3.2.	LE PSORIASIS	119

3. ETUDES A GRANDE ECHELLE DU TRANSCRIPTOME EPIDERMIQUE 123

3.1. ETUDES DU TRANSCRIPTOME LORS DE PATHOLOGIES EPIDERMQUES123

3.1.1.	TRANSCRIPTOME ET PSORIASIS.....	123
3.1.2.	TRANSCRIPTOME ET INFLAMMATION	124
3.1.3.	TRANSCRIPTOME ET CARCINOGENESE	125

3.2. ETUDES DU TRANSCRIPTOME DANS UN CONTEXTE PHYSIOLOGIQUE 128

3.2.1.	TRANSCRIPTOME ET CICATRISATION	128
3.2.2.	ETUDES DE LA DIFFERENCIATION EPIDERMIQUE IN VITRO	130
3.2.3.	ETUDES DE LA DIFFERENCIATION EPIDERMIQUE IN VIVO	133

II. RESULTATS EXPERIMENTAUX..... 139

Large-scale identification of human genes implicated in epidermal barrier function 141

A large-scale, multi-technique approach identifies 49 new players of human epidermal differentiation 167

III. DISCUSSION	203
Technique des ORESTEs.....	205
Puces à ADN.....	208
PCR quantitative	215
Nouveaux acteurs de la différenciation	218
Facteurs de transcription	228
IV. ANNEXES	233
Annexe 1 : Stratification épidermique	235
Annexe 2 : le séquençage nouvelle génération.	235
Annexe 2 : le séquençage nouvelle génération.	237
Annexe 3 : Vers l'intégration des données d'expression : de la connaissance à la modélisation	245
Annexe4 : Rôle de la cornéodesmosine in vivo : modèles murin	246
V. BIBLIOGRAPHIE	261

I. INTRODUCTION

1. Etude du transcriptome et de la régulation génique

1.1. Etudes statiques de transcriptomes

1.1.1. *Banques d'ESTs*

La production d'EST est la technique d'analyse systématique du transcriptome la plus ancienne (Adams et al. 1991; Boguski et al. 1994). Les ARNm sont convertis en ADNc, clonés dans un vecteur puis séquencés de façon systématique à une de leurs extrémités. Chaque lecture de séquenceur, généralement de quelques centaines de nucléotides, est nommée EST pour Expressed Sequence Tag. Selon que le clonage a été fait de façon orientée ou non, et que le séquençage a été fait en utilisant des amorces situées sur le vecteur, les EST correspondent à la partie 5' ou 3' de l'ARNm. Les EST 3' correspondent habituellement à la région 3' non traduite de l'ARNm. Cette partie du transcrit, qui fait en moyenne 750 nucléotides de long chez l'homme, est en général la moins conservée au sein des familles multigéniques, et permet ainsi une identification univoque du gène correspondant. Les EST 5' couvrent plus fréquemment la région codante car la région 5' non traduite est courte (en moyenne 300 pb). De plus, il arrive que les ADNc synthétisés avec une amorce oligo(dT) soient incomplets à leur extrémité 5' en raison de la faible efficacité de la reverse transcriptase et de la présence de structures secondaires au sein de l'ARNm. Les EST 5' peuvent ainsi se révéler utiles pour la mise en évidence d'épissages alternatifs ou pour l'alignement de séquences codantes conservées entre espèces différentes.

La production d'EST permet l'identification des séquences codantes à l'échelle d'un génome et ainsi la découverte de nouveaux gènes. Elle est aujourd'hui plus particulièrement utilisée pour la constitution de banques spécifiques (tissus, stade embryonnaire...). La principale limite de cette technique réside dans son manque de sensibilité pour la détection des ARNm de faible abondance, qui constituent 85 à 95% des transcrits (Bishop et al. 1974). Afin d'éviter le séquençage répétitif des transcrits les plus fréquents, plusieurs méthodes de normalisation ont été mises au point. C'est le cas par exemple des banques soustraites, ou de la technique des ORESTES décrite ci-après.

Au printemps 2009, la division dbEST de GenBank contient plus de 60 millions de séquences d'EST dont plus de 8 millions sont d'origine humaine, espèce la plus largement représentée dans cette banque de données

(http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). Viennent ensuite la souris, avec environ 5 millions de séquences, puis le riz (2 millions), le cochon, *Arabidopsis thaliana*, le bœuf et le poisson-zèbre avec 1,5 millions de séquences environ chacun. De nombreuses autres espèces d'intérêt scientifique, agronomique ou médical y sont aussi diversement représentées. Pour une grande partie de ces espèces, la base de données UniGene du NCBI rassemble les EST chevauchants ou correspondant à un même ARNm dans un « cluster », c'est-à-dire un groupe de séquences correspondant à un gène unique, et indique l'origine tissulaire de chaque EST lorsqu'elle est connue.

Tous les EST annotés correspondent à autant de données d'expression : origine cellulaire ou tissulaire de l'ARNm, stade de différenciation... Plusieurs initiatives de développement de programmes d'analyse virtuelle du transcriptome à partir de ces données ont vu le jour. C'est par exemple le cas de ExQuest (Expressional Quantification of EST) qui permet d'identifier pour un gène donné le nombre d'EST provenant de chaque tissu ou organe et propose une normalisation de

EST Profile

Hs.556031 - CDSN: Corneodesmosin

pituitary gland	0	0 / 16736
placenta	197	56 / 284142
prostate	0	0 / 190888
salivary gland	0	0 / 20291
skin	255	54 / 211668
spleen	0	0 / 54070
stomach	0	0 / 97199
testis	0	0 / 331421
thymus	12	1 / 81245

Type d'échantillon

Transcrit par million (TPM)

Intensité relative (d'après TPM)

EST du gène / EST de l'échantillon

Figure 1. Extrait du profil des EST du gène *CDSN* dans la banque UniGene.

ces profils par rapport au nombre total de séquences (Brown et al. 2004). Il est également possible de visualiser pour un tissu donné la répartition des EST sur une région chromosomique, correspondant à un locus d'intérêt par exemple, sous la forme d'un diagramme. La fonction EST Profile (figure 1) de la base de données UniGene permet elle aussi de visualiser la provenance des séquences d'un gène selon le tissu, l'état, le stade de développement...

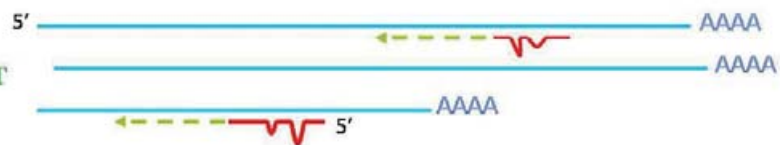
1.1.2. Technique des ORESTEs

La production d'ORESTEs (Open Reading Frame EST) repose sur la construction de minibanques d'EST par RT-PCR à faible stringence à l'aide d'amorces choisies de façon arbitraire (figure 2). Il peut s'agir de n'importe quelle amorce d'une vingtaine de nucléotides, pourvu que sa séquence n'ait d'homologie parfaite avec aucun des transcrits potentiels du tissu étudié : séquence choisie au hasard, séquence génomique non codante (intergénique, intronique ou brin non transcrit), ou séquence correspondant à une espèce différente. Initialement dénommés RAP-EST (RNA arbitrarily primed-EST) (Dias Neto et al. 1997), ces ORESTEs correspondent à des séquences distribuées au hasard sur les ARNm et donc susceptibles de correspondre aux régions codantes, d'où leur nom.

Transcription inverse :

Faible stringence (37°C)

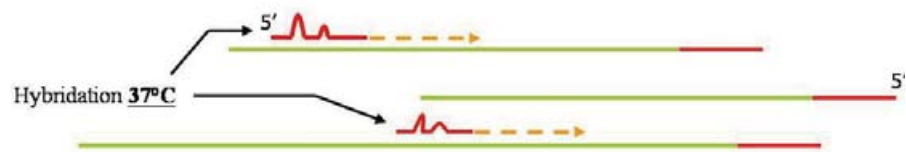
Amorce arbitraire + ARN poly(A) + RT



PCR :

1er cycle à faible stringence

Amorce + ADNc + Taq



39 cycles classiques

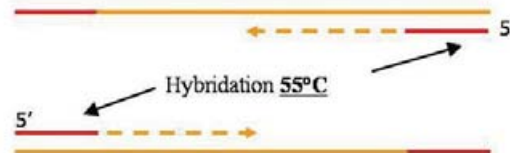


Figure 2. Principe de la technique des ORESTEs.

Après extraction des ARNm de l'échantillon, une transcription inverse est réalisée à une température de 37°C, permettant l'hybridation imparfaite de l'amorce choisie à une sous-population des ces ARNm. La première étape de la PCR qui suit, se fera de même à faible stringence, à 37°C (figure 1). Cette RT-PCR constitue le principe de la normalisation conceptuelle de la méthode : l'hybridation de l'amorce arbitraire ne dépend pas de l'abondance relative des différents ARNm, comme c'est le cas avec des amorces oligo(dT), et favorise ainsi l'amplification des transcrits rares. Les produits d'amplification sont ensuite déposés sur gel et les bandes visibles, correspondant à l'amplification préférentielle d'une molécule donnée, sont éliminées,

constituant une étape supplémentaire de normalisation. Avec l'utilisation d'amorces différentes, la construction de nombreuses minibanques d'une centaine de clones rend possible l'établissement d'un « catalogue » des gènes exprimés et l'identification de gènes spécifiques du type cellulaire considéré.

Le premier avantage de la technique des ORESTEs est son principe de normalisation : l'amorce s'hybridera lors de la première étape de transcription inverse à faible stringence (diminution de la température de la réaction à 37°C) avec la même probabilité sur un ARNm abondant ou sur un ARNm rare. L'amplification d'un transcrit donné dépendra donc principalement de sa séquence et non de son abondance. La production de minibanques d'EST avec des amorces différentes ainsi que la purification sur gel *a posteriori* ajoutent encore à cette normalisation conceptuelle. De fait, une normalisation du nombre de séquences par rapport à l'abondance relative des messagers a été démontrée (Dias Neto et al. 2000). Cette technique permet en outre la production de séquences distribuées statistiquement selon une courbe de Gauss centrée au milieu de la séquence de l'ARNm, donc correspondant en majorité à sa partie codante. Elle est ainsi complémentaire des techniques classiques de séquençage en 5' et en 3', pour l'obtention de segments contigus (contigs) d'EST, et pour la mise en évidence d'épissages alternatifs affectant le cadre de lecture et par conséquent la protéine produite (Sakabe et al. 2003).

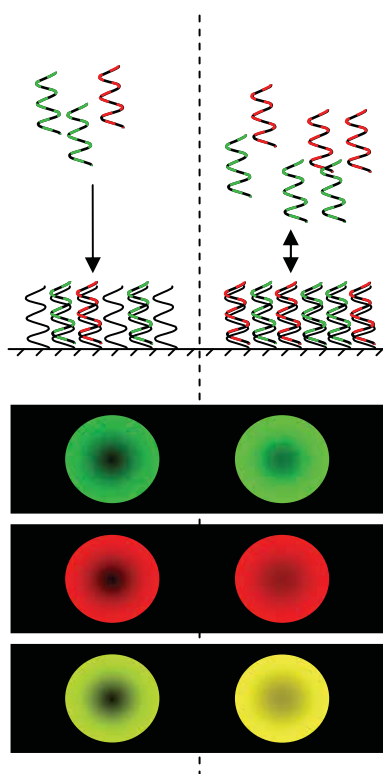
L'amplification des séquences se faisant au hasard parmi toutes les molécules d'ARN disponibles, une attention toute particulière doit être portée à la purification des ARNm poly-adenylés qui ne constituent en fait que 1 à 5% des ARN cellulaires. Une étude par la technique des ORESTEs du transcriptome d'ARNm poly-adenylés purifiés à partir de prélèvements de tumeurs mammaires humaines a montré que parmi 10 000 séquences, environ 10% correspondent toutefois à des contaminations ribosomiques, mitochondriales et bactériennes (Dias Neto et al. 2000).

1.2. Etudes comparatives de transcriptomes

1.2.1. Puces à ADN

1.2.1.1. Origines de la technique et principe

Les premières techniques de quantification de l'expression génique dérivent de celles développées pour l'ADN, telles que le Southern blot (Southern 1974), basées sur l'hybridation moléculaire entre des cibles d'intérêts et des sondes complémentaires marquées (radioactives ou fluorescentes). Ainsi, dans les années 70 fut développée la technique du buvardage de Northern, ou Northern blot (Alwine et al. 1977), consistant à séparer des ARN totaux par électrophorèse, à les transférer sur une membrane de nitrocellulose ou un filtre de nylon, et à employer des sondes (ARN ou ADN) radiomarquées pour quantifier les transcrits d'intérêt. Les techniques du RNA dot blot (Kafatos et al. 1979) ou slot blot sont des variantes simplifiées du Northern blot, où les ARN sont déposés directement sur les membranes sans électrophorèse préalable, ce qui les rend plus adaptées à des comparaisons simultanées entre de multiples conditions, mais ne permet pas d'identifier des différences de longueur entre les messagers produits par ces cellules. Ces méthodes comportent peu d'étapes susceptibles d'introduire des biais, mais les risques d'hybridations croisées, non détectables en absence de migration, peuvent conduire à une détection non spécifique.



L'idée de concevoir des puces à ADN découle de la volonté de développer un outil permettant de quantifier de façon fiable le niveau d'expression d'un grand nombre de gènes simultanément (parallélisme à l'origine de l'emploi du terme « array »). Toutes les puces à ADN reposent sur le même mécanisme (figure 3), celui de fixer sur un support un grand nombre de sondes (sous la forme de dépôts distincts),

Figure 3. Principe de l'hybridation de cibles marquées avec des sondes immobilisées.

Dans le cas de deux conditions marquées par deux fluorochromes distincts, la co-hybridation est généralement non compétitive (gauche), seuls les gènes très fortement exprimés saturant alors le support (droite). Dans les deux cas, le rapport des fluorescences est proportionnel au ratio d'expression.

complémentaires des acides nucléiques que l'on souhaite analyser (généralement des ADNc ou de l'ADN génomique, plus rarement des ARN). En plus d'une simplification et transposition à grande échelle du principe du Northern blot, cette technique s'en distingue donc par le fait que les sondes sont fixées et non marquées, alors que les cibles sont marquées et en solution.

Une autre particularité de la technique des puces à ADN est le fait que les résultats bruts ne sont pas directement analysables et nécessitent une étape cruciale de normalisation. Plusieurs techniques de normalisation existent, et le choix dépend à la fois du type de puces employé et de la nature des données. Le but de la normalisation est de s'affranchir au maximum des biais inhérents à la technique des puces, afin de rendre les données comparables entre elles. Malgré tout, les données issues de puces sont dites « bruitées », et si la normalisation permet de réduire le bruit « systématique » (par exemple une différence de concentration entre deux populations d'ADNc différentiellement marquées et co-hybridées), elle ne permet pas de prendre en compte la variabilité dans les mesures pour un gène ou une sonde donnée. Seules la répétition de l'expérience (à partir d'échantillons biologiques totalement indépendants) et une analyse statistique permettent d'évaluer la solidité des résultats. Le nombre de répétitions nécessaires pour avoir des mesures « fiables » dépend essentiellement de la finesse d'analyse souhaitée (ex : seuil de ratio pour déclarer un gène comme surexprimé fixé à 1,5 ou 2) et la variabilité entre chaque mesure (variabilité biologique + expérimentale).

Les différences entre puces à ADN résident principalement dans leur type de support et leur mode d'emploi (hybridation simple ou co-hybridation de deux échantillons), mais aussi dans leur champ d'application et donc leur « design » (choix des sondes présentes sur le support).

1.2.1.2. Types de support

L'origine des premières puces à ADN est imprécise : réalisées sur membrane de nylon, les « macroarrays » constituaient simplement un changement d'échelle dans la technique du dot blot inverse. La robotisation de l'étape de dépôt permit d'augmenter la densité et donc le nombre de sondes présentes sur le support, passant de 96 à plusieurs centaines. Dans un premier temps, les sondes employées avec ce type de support étaient des ADN complémentaires clonés dans des vecteurs, puis l'emploi de produits de PCR a permis une plus grande plasticité dans la conception

« à façon » des puces à ADN. Toutefois, les supports en nylon présentant une très forte autofluorescence, les acides nucléiques à analyser doivent être marqués radioactivement, ce qui représente une contrainte pratique évidente, mais aussi technique. En effet, l'autoradiographie n'autorise pas une résolution aussi importante – donc une densité de dépôts – qu'un scanner en fluorescence ; de plus la fluorescence permet l'hybridation simultanée de deux populations d'acides nucléiques marquées avec des fluorochromes différents (hybridation comparative), tandis que les macroarrays imposent d'hybrider un seul échantillon par puce (hybridation simple), ce qui impose des contraintes supplémentaires pour la normalisation et la comparaison des résultats de puces (cf. paragraphe 1.2.1.5). Cette technologie a été rapidement supplantée par les microarrays en raison des contraintes moindres et de leurs performances sans cesse améliorées.

La nécessité de changer de support pour atteindre une plus grande échelle d'analyse a conduit à l'émergence de deux technologies radicalement différentes dans le milieu des années 90 : les microarrays fabriqués par dépôt de sondes sur lame de verre, et les microarrays réalisés par synthèse *in situ* (figure 4).

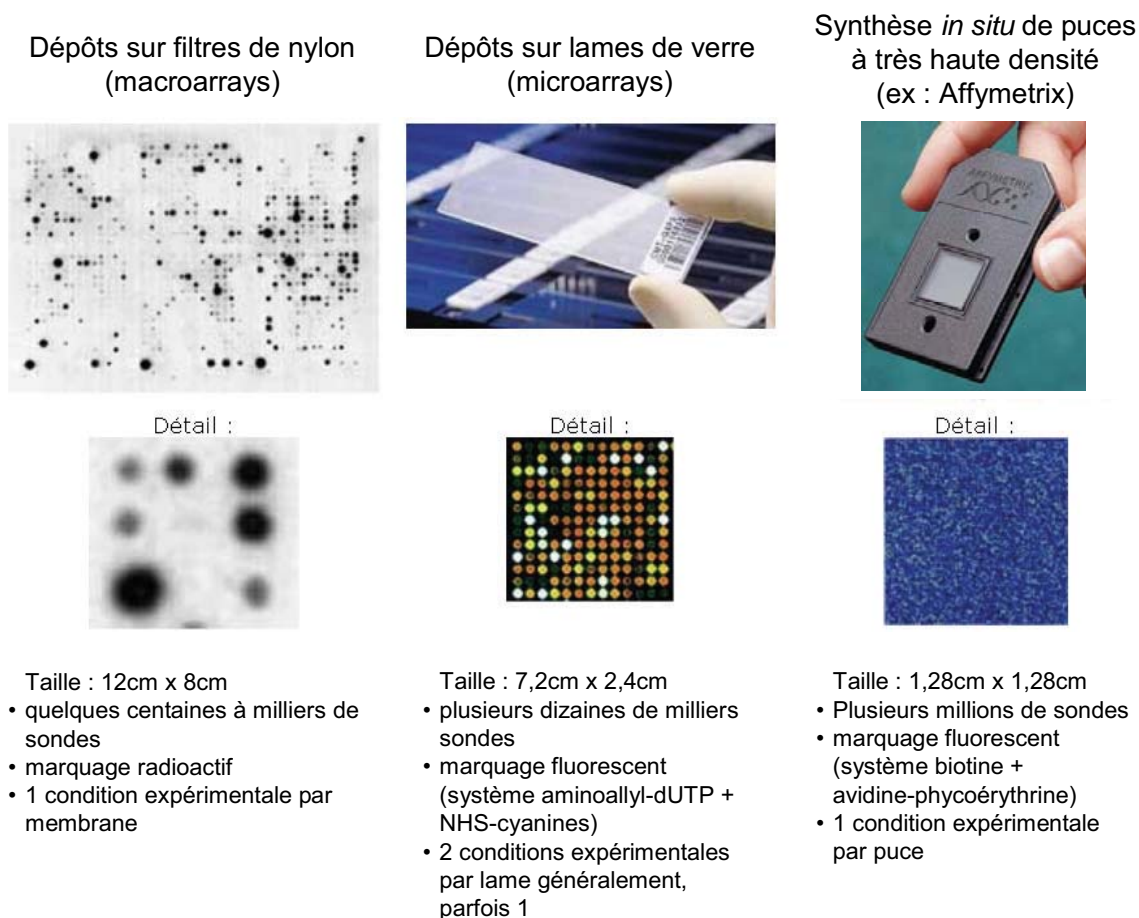


Figure 4. Différentes sortes de puces à ADN.

Les premiers microarrays réalisés sur des lames de verre comportaient 48 sondes différentes contre des gènes exprimés par *Arabidopsis thaliana* et furent employés en cohybridant des ADNc marqués avec deux fluorochromes distincts (Schena et al. 1995). La fonctionnalisation de la surface des lames, étape nécessaire pour la fixation ultérieure des sondes sur le verre, peut se faire par un traitement à la poly-L-lysine comme décrit par Schena et collaborateurs, ou par des dérivés du silane. Les sondes sont ensuite déposées sous la forme de « spots » (généralement ronds) par un robot (le plus souvent au moyen un système de dépôt de type stylo-plume, les sondes en milieu liquide étant déposées par simple capillarité lorsque la pointe touche le support), puis fixées de façon covalente au support.

En parallèle des puces sur lame de verre réalisées au moyen de dépôts, s'est développée une approche alternative, consistant à synthétiser directement sur un support des oligonucléotides de longueur variable. Plusieurs types de synthèse *in situ* ont été développés à ce jour, principalement basées sur des procédés de photolithographie (Fodor et al. 1993; Pease et al. 1994) ou de dépôt du type jet d'encre (Hughes et al. 2001). Ce type de technique permet d'atteindre un parallélisme (une densité de sondes) extrêmement élevé (jusqu'à plusieurs millions par cm²). Au delà des avantages techniques que présente une forte densité de sondes (surface diminuée = hybridation homogène plus facile à obtenir), leur nombre sans cesse croissant autorise de nouvelles approches à l'échelle du transcriptome entier (puces « pangénomiques »), voire du génome entier, notamment grâce aux « tiling arrays » (cf. paragraphe suivant).

1.2.1.3. Types de conception (*design*) des puces d'expression génique

Les puces sur lame de verre, de par leur support et méthode de fabrication, peuvent contenir un nombre limité de sondes. D'une façon comparable aux macroarrays sur filtres de nylon, les premières puces sur lame de verre étaient des puces dites « à façon » ou « dédiées », c'est-à-dire fabriquées spécialement pour la quantification d'un petit nombre (quelques centaines) de gènes spécifiques du tissu ou du processus biologique étudié (apoptose, voie de signalisation, *etc.*). Les avantages de ce type de puces sont multiples : la fabrication de puces à façon est globalement rapide et peu coûteuse, exception faite de la banque de sondes elle-même (fragments PCR ou oligonucléotides synthétisés) ; la fabrication des puces peut

facilement se faire au niveau d'un laboratoire ou d'une plateforme équipés d'un robot de dépôt, et de nombreuses alternatives commerciales sont disponibles ; l'analyse « graphique » des dépôts après hybridation est plus rapide que pour des puces à très haute densité. L'inconvénient majeur concerne l'étape de normalisation qui fait appel à des étalons internes (gènes de référence) ou externes (cf. paragraphe 1.2.1.4).

Bien que les puces dédiées soient encore parfois employées, la multiplication des offres – commerciales ou non – et l'augmentation de la densité des dépôts ont conduit à l'émergence ces dernières années de puces dites pangénomiques, c'est-à-dire contenant au minimum une sonde pour chacun des gènes connus ou supposés (EST, ARNm, prédictions bioinformatiques...) d'une espèce. Toutefois, le terme pangénomique peut sembler abusif : sur la plupart des puces de ce type, le nombre de sonde par gène est généralement limité (1 à 4 sondes maximum). Ainsi, ces puces ne livreront pas d'information sur d'éventuels nouveaux gènes pour lesquels il n'y aurait pas de sonde, ou encore sur la complexité de certaines unités de transcription possédant plusieurs sites d'initiation de la transcription, variants d'épissage, exons terminaux alternatifs, *etc.*

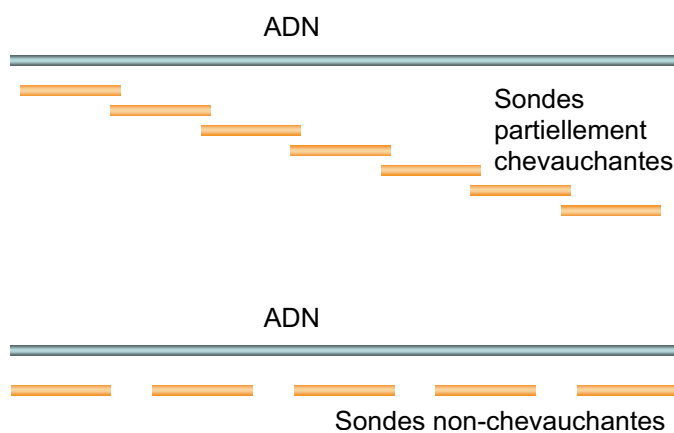


Figure 5. Types de tiling arrays.

Le développement des puces à très haute densité par synthèse *in situ* permet de pallier en partie l'inconvénient des variants, avec jusqu'à plusieurs dizaines de sondes par exon (5,5 millions de sondes au total pour le GeneChip® Exon Array d'Affymetrix), permettant ainsi une « cartographie » très fine des zones transcrites du génome. Pour un nombre limité d'espèces de

mammifères (humain, souris, rat), l'identification de nouvelles régions transcrites est possible en employant des tiling arrays (figure 5) couvrant tout le génome et permettant de s'affranchir de l'annotation du génome (et donc d'identifier de nouveaux gènes, transcrits antisens, etc.). Les tiling arrays proposés par Affymetrix se présentent sous la forme de 14 puces contenant 45 millions de sondes de 25 nucléotides, espacées de 10 paires de bases, soit une résolution de 35pb.

L'inconvénient des tiling arrays est principalement son coût, extrêmement élevé pour des génomes complets de mammifères, mais aussi la sensibilité de la méthode qui conduit à identifier énormément de régions « transcrites » correspondant probablement à du bruit expérimental ou éventuellement du « bruit transcriptionnel » dont la fonction demeure encore énigmatique - pour revue, (Mockler et al. 2005; Yazaki et al. 2007).

1.2.1.4. Applications

Avec l'augmentation de la densité de dépôts, un nombre croissant d'applications a vu le jour. La première est évidemment la mesure de l'expression des gènes, que ce soit l'étude de profils d'expression génique (hybridations simples), ou l'identification de gènes différentiellement exprimés (par co-hybridation sur un même support ou par comparaison de profils). Les premières puces pangénomiques ont constitué un bond technologique mais aussi conceptuel, en permettant de mesurer simultanément le niveau d'expression de l'ensemble des gènes d'un organisme, y compris des gènes non liés *a priori* à la question scientifique posée mais pouvant ouvrir de nouvelles voies de recherche. Les exemples de ce type d'étude sont trop nombreux pour être cités ou même recensés (pour la seule année 2005, les articles citant des données de puces à ADN représentent plus de 5300 publications). Citons tout de même plusieurs exemples, dans l'identification de gènes régulés au cours de certains processus comme la gamétogenèse (Chalmel et al. 2007), ou sous l'effet d'une blessure d'un nerf périphérique, permettant d'identifier des cibles potentielles pour de nouveaux traitements antidouleurs (Xiao et al. 2002). Les puces ont également été employée comme moyen de caractérisation de la fonction des gènes en recherchant une co-expression dans plusieurs tissus (Zhang et al. 2004). Toutefois, un très grand nombre d'études ont été menées dans l'optique d'une comparaison de profils d'expression entre plusieurs types de prélèvement et la recherche de signature, notamment dans certains cancers (cf. paragraphe 1.2.1.5). Ainsi, Ash A. Alizadeh et ses collaborateurs sont parvenu à montrer que chez des patients souffrant de lymphomes B diffus à grandes cellules (DLBCL), maladie dont le pronostic est très variable (40% des patients répondent bien au traitement et ont une survie plus élevée), les tumeurs présentent également une hétérogénéité moléculaire. Les profils d'expression génique différents reflètent notamment une variabilité dans le potentiel prolifératif ou l'état de différenciation, et sont corrélées avec le taux de

survie des patients (Alizadeh et al. 2000). Un travail comparable a été réalisé sur des personnes atteintes de cancers du sein (van 't Veer et al. 2002), et ce travail a abouti en 2006, à partir d'une liste de 70 gènes, à la conception de MammaPrint, le premier microarray spécialement dédié au pronostic qui soit approuvé par la FDA (US Food and Drug Administration) (figure 6). Malgré tout, la fiabilité de certaines études, et la difficulté pour comparer des études similaires mais aux procédures non-standardisées, rendent certains résultats difficiles à mettre en perspectives (voir la revue (Pusztai et al. 2006) pour une discussion sur les méthodes de classification dans le cas des cancers du sein). De plus, la détermination de sous-catégories s'est avérée plus délicate, voire impossible, avec d'autres pathologies.

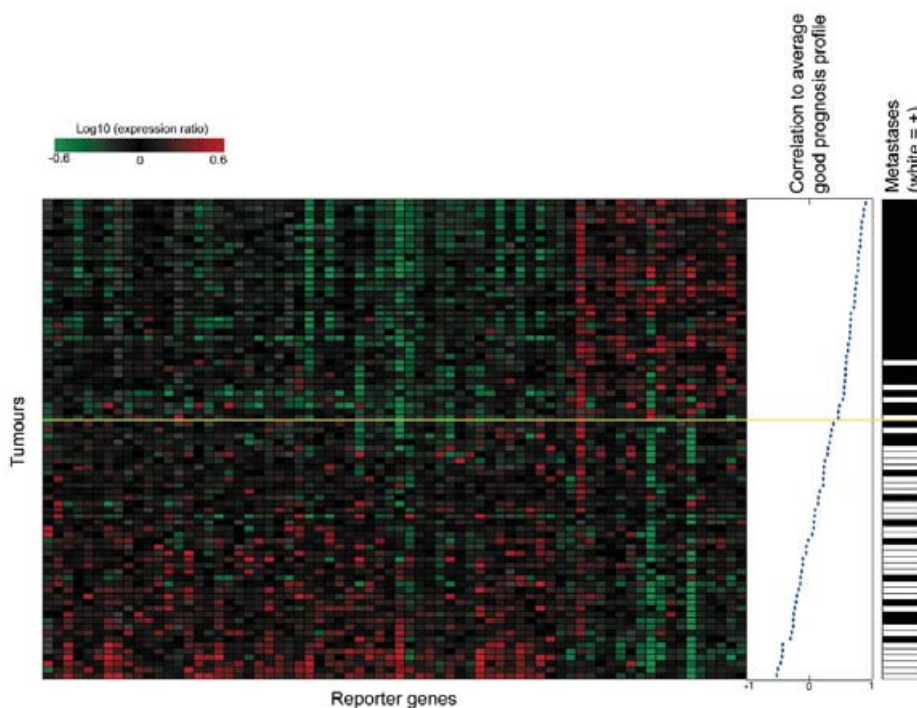


Figure 6. Signatures de tumeurs du sein sur MammaPrint et pronostic (Glas et al. 2006). Profils d'expression pour 70 marqueurs pronostiques. Les ARNm extraits de tumeurs de 78 patients sont analysés et classifiés sur MammaPrint (« Eisengramme », gauche). Les tumeurs sont classées selon leur degré de corrélation avec le profil moyen du groupe ayant un bon pronostic (milieu). Les patients développant des métastases moins de cinq ans après le diagnostic sont indiqués en blanc (droite).

Les analyses d'expression génique par microarray ne se bornent pas non plus aux ARN messagers (codant pour des protéines), mais ont trouvé également leur place dans la course à la recherche de petits ARN non-codants, tels que les microARN, des petits ARN simple brin de 21 à 23 nt issus de la maturation de précurseurs repliés en tiges-boucles et jouant un rôle de répresseur post-

transcriptionnel. Des puces spécialement dédiées à la détection des microARN ont permis de dresser un profil de l'expression de ces nouveaux marqueurs dans plusieurs types de cancers, tels que le cancer du sein (Blenkiron et al. 2007; Yin et al. 2008), du poumon (Yanaiharu et al. 2006) ou de la prostate (Mattie et al. 2006).

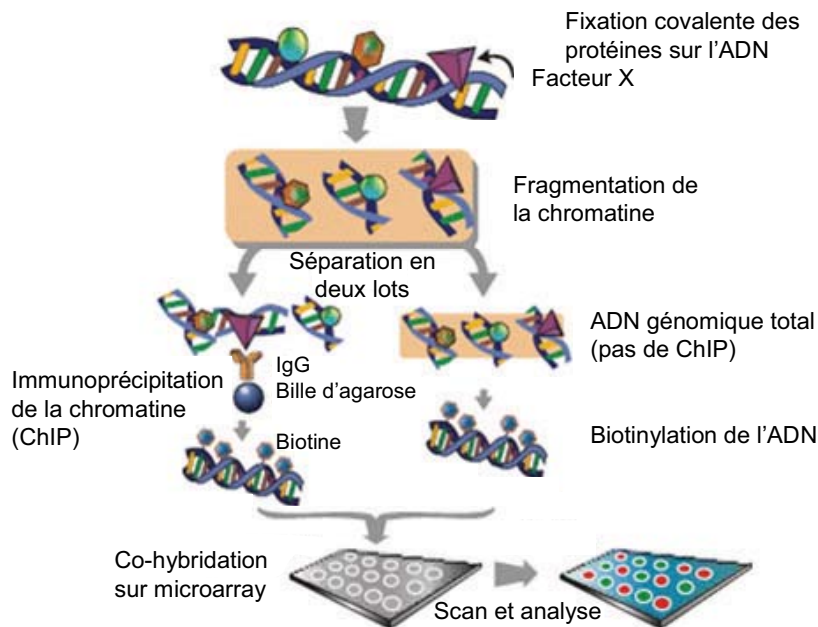


Figure 7. Principe du ChIP-on-chip contre un facteur X.

Toutefois, la quantification de l'expression génique est loin d'être la seule application des puces à ADN. D'autres utilisations ont également bénéficié des améliorations techniques les plus récentes. Ainsi, un grand nombre d'approches à l'échelle du génome ont pu être menées récemment, couplant éventuellement plusieurs techniques à l'emploi de puces. Citons notamment des travaux de génomique fonctionnelle, comme la recherche des cibles directes de facteurs de transcription à l'aide d'immunoprécipitations de chromatine (ChIP) suivie d'une analyse sur puce (ChIP-on-chip ; figure 7), technique notamment employée pour déterminer les cibles du facteur de transcription p63 dans les kératinocytes (Vigano et al. 2006). Le ChIP-on-chip a également permis de montrer que dans des cellules souches embryonnaires humaines, un très grand nombre de gènes codant pour des protéines, y compris ceux supposés inactifs, font l'objet d'une initiation de la transcription, l'élongation étant régulée par certaines modifications des histones (Guenther et al. 2007). Citons aussi des études de génétique, telles la recherche de

mutations ponctuelles (SNP), mais aussi de pertes d'hétérozygotie (LOH) ou de variation du nombre de copies d'un gène (CNV) – micro-délétions ou amplifications – au moyen d'une hybridation génomique comparée sur puces (CGH arrays). Ainsi, près de la moitié des SNP recensés par le projet international HapMap (2003) l'ont été au moyen de puces à ADN, une approche qui s'est avérée aussi fiable que les méthodes traditionnelles par séquençage (Gunderson et al. 2005). Les CGH arrays ont quant à eux permis de découvrir que les variations du nombre de copies concernent près de 12% du génome humain et plusieurs milliers de gènes, bien au-delà de ce qui était supposé précédemment, expliquant probablement une partie des variations phénotypiques normales entre individus (Carter 2007). La combinaison de puces pour l'expression génique et de CGH arrays a également révélé un lien indirect entre des variations de nombre de copies et l'expression de certains gènes dans des carcinomes hépatocellulaires dus au virus de l'hépatite B (Huang et al. 2006), rejoignant des résultats faisant le lien entre amplification ou perte d'hétérozygotie et altérations des profils d'expression génique dans le même type de cancer (Xu et al. 2001). Enfin, la technologie des puces à ADN (ou des versions dérivées sur support microfluidique) commence à être employée pour capturer des séquences génomiques et les séquencer par des techniques de nouvelle génération (Garber 2008).

1.2.1.5. Normalisation et tests statistiques

En raison du grand nombre de types de puces et d'applications, nous ne décrivons ici que l'analyse de données d'expression issues d'hybridations comparatives, c'est à dire du type « deux couleurs » où une condition est marquée en rouge (Cyanine5 ou Cy5) et l'autre en vert (Cy3), les deux étant simultanément cohybridées sur le même support.

Les données issues de puces à ADN sont des données dites « bruitées », c'est à dire qu'une part non négligeable de chaque mesure est soumise à l'influence de paramètres expérimentaux qui peuvent donc la fausser. Les données brutes ne sont donc pas directement analysables et comparables, et nécessitent de s'affranchir de deux types d'erreurs de mesure.

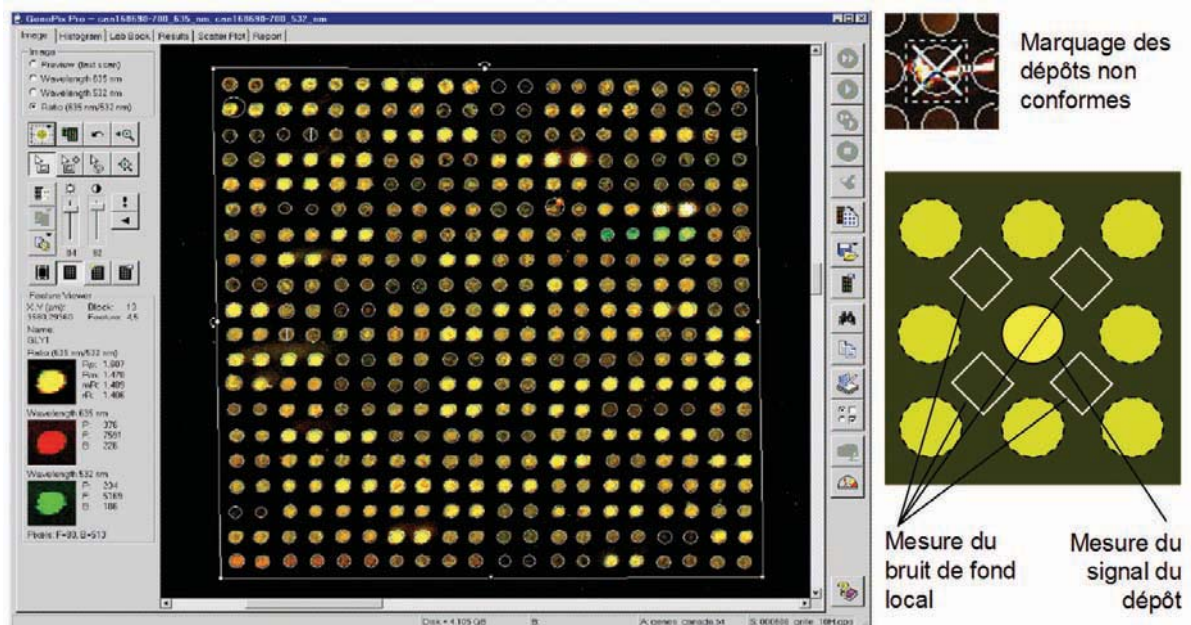


Figure 8. Analyse d'un scan de microarray avec le logiciel Genepix. Les logiciels d'analyse d'images de puces définissent une trame à partir d'un plan de la puce (fichier .gal), puis cette trame est ajustée de façon semi-automatique sur le scan réel. Les dépôts ou hybridations non-conformes (poussière, éraflure...) sont marqués pour être ignorés. Le logiciel quantifie le signal dans l'aire de chaque dépôt (moyenne ou médiane de l'intensité des pixels), mais aussi autour des dépôts pour déterminer le bruit de fond.

Les erreurs systématiques sont dues à des biais se produisant sur l'ensemble des données, avec une amplitude généralement similaire d'une expérience à une autre, et qui surtout est mesurable à partir des données. Elles correspondent notamment à des différences dans les quantités de matériel au départ, dans l'efficacité des différentes étapes d'extraction des ARN, de la transcription inverse, du marquage (différence d'incorporation des différentes cyanines, mais aussi de leur rendement quantique), de l'hybridation, du scan... Ces biais peuvent être corrigés en calibrant les données à comparer, ce qui correspond aux étapes de traitement des données brutes (soustraction du bruit de fond et surtout étape de normalisation).

Les erreurs stochastiques, qui se produisent au niveau d'une partie seulement des cibles à étudier et de façon non reproductible (variabilité inter-individus, dégradation partielle de certains ARN, anomalie locale dans l'hybridation ou dans le dépôt des sondes...) ne peuvent être prises en compte à partir des données d'une seule puce. Ces erreurs stochastiques nécessitent donc de répéter l'expérience un nombre variable de fois selon leur amplitude (et selon la finesse d'analyse souhaitée). Le coefficient de variation d'une même mesure étant difficilement prévisible *a priori*,

il peut être utile de faire quand cela est possible une analyse préliminaire sur quelques échantillons pour estimer le nombre de répétitions nécessaires à l'étude.

Quel que soit le type de puce, la première étape dans le traitement des données est la soustraction du bruit de fond, afin d'éliminer le signal lié à une fixation non spécifique sur le support. La plupart des logiciels d'analyse d'image, gratuits (ex : ScanAlyze, Dapple...) ou payants (GenepixPro pour les scanners Axon, QuantArray, ImaGene), permettent de définir de façon semi-automatique la zone correspondant au dépôt, ainsi qu'une zone autour de chaque dépôt servant comme valeur locale du bruit de fond de l'hybridation (figure 8). Les données (intensités de fluorescence) sont ensuite généralement transformées en données logarithmiques (condition facilitant la normalisation et les tests statistiques ; figure 9).

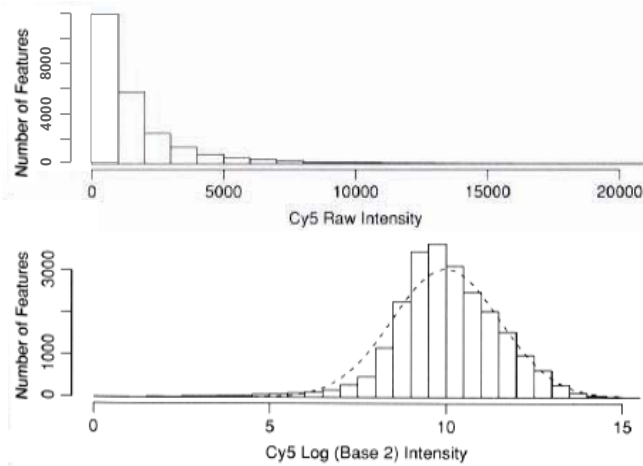


Figure 9. Transformation des données en valeurs logarithmiques. Après la soustraction du bruit de fond, les données ont une distribution non symétrique, la plupart des intensités étant faibles. Le passage en valeurs logarithmiques (\log_2 ou \log_{10}) permet d'obtenir une distribution symétrique par rapport à la médiane des intensités. D'après « Microarray Bioinformatics », Dov Stekel

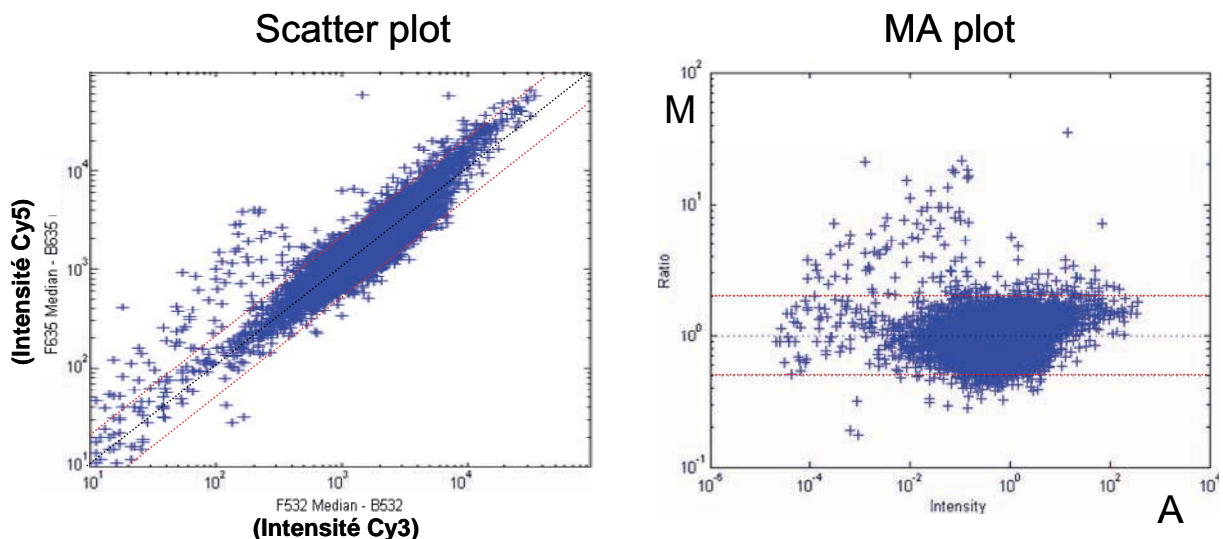


Figure 10 Scatter plot vs. MA plots. Ces deux types de graphes sont couramment employés pour visualiser des données de puces, avant ou après normalisation. Le scatter plot permet de comparer les intensités dans les deux canaux après soustraction du bruit de fond. Le MA plot permet de visualiser le ratio M en fonction de l'intensité moyenne A, soit $M = \log(Cy5/Cy3) = \log Cy5 - \log Cy3$ et $A = (\log Cy3 + \log Cy5) / 2 = \log \sqrt{(Cy3 \times Cy5)}$. D'après <http://www.mathworks.com>

L'étape suivante, la normalisation à proprement parler, dépend fortement du « design » des puces, les différentes stratégies n'étant pas applicables à tous les types de puce. Sur la plupart des puces, le design autorise l'emploi d'étalons internes et/ou externes. Les étalons internes sont des gènes de référence dont l'expression est sensée ne pas varier d'une condition à une autre. L'emploi de cette méthode se heurte à deux écueils : la difficulté de trouver *a priori* de « bons » gènes de ménage, et la nécessité d'en trouver un nombre suffisant afin d'éviter les artefacts de normalisation. En effet, les ratios sur puces étant généralement assez faibles par rapport à d'autres méthodes quantitatives, la moindre variation d'expression des gènes de référence fausse l'ensemble de la normalisation, d'où la nécessité d'identifier au minimum une dizaine de gènes d'expression stable. Les étalons externes sont des ARN exogènes rajoutés en quantité précise dans chaque échantillon et servant à les calibrer entre eux, l'inconvénient majeur de cette méthode étant qu'elle est tributaire des imprécisions de dosage de ces ARN, loin d'être négligeables.

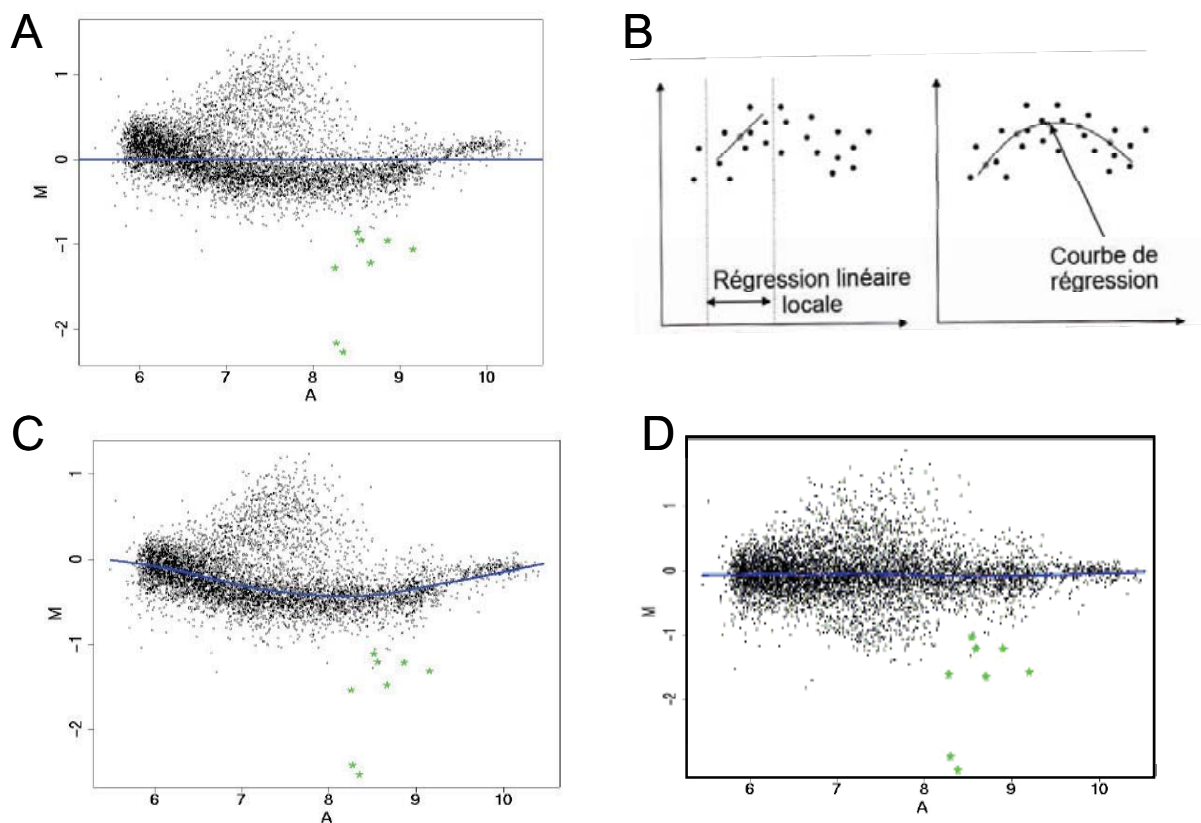
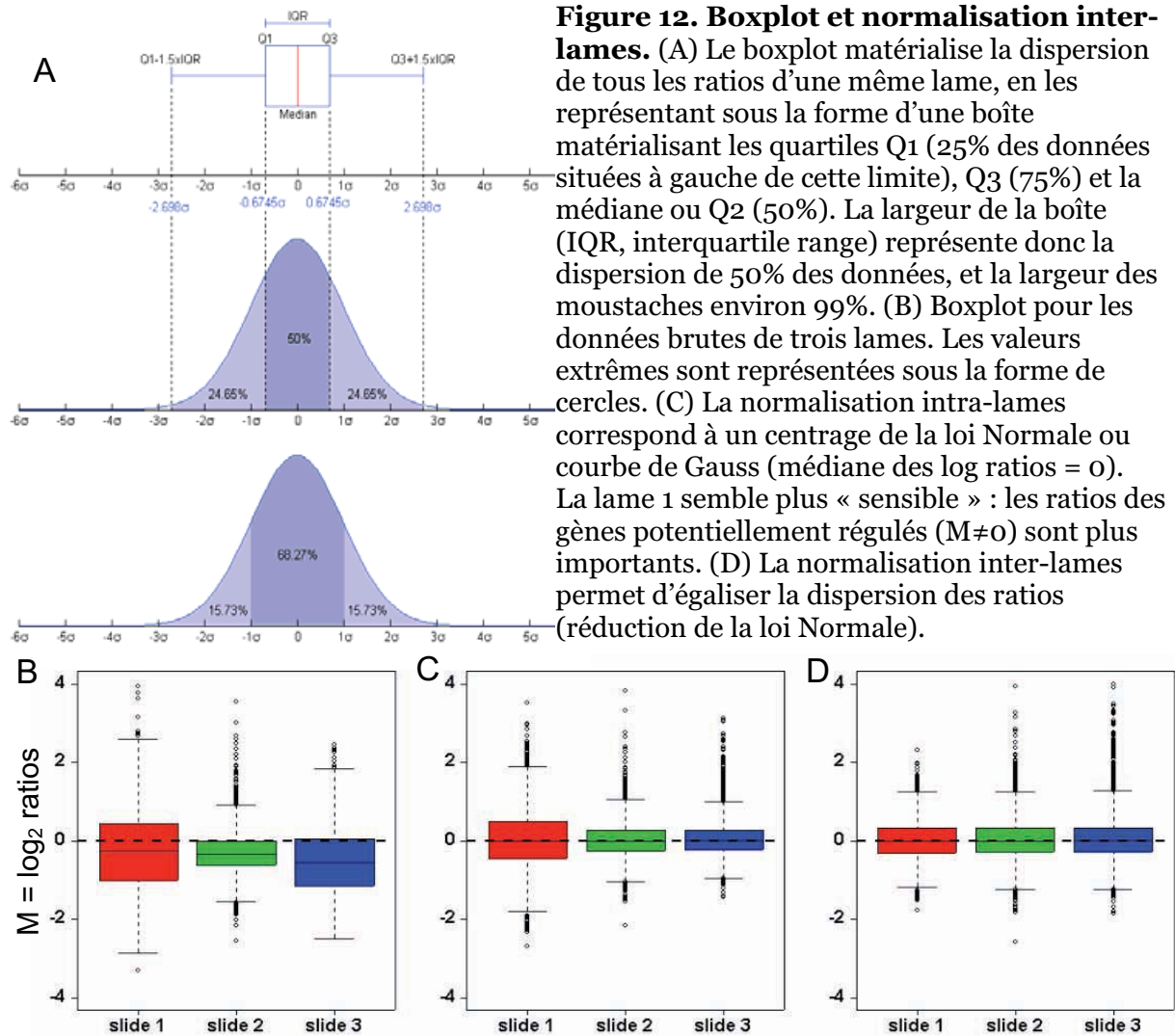


Figure 11. Normalisation des données par la méthode Lowess. (A) Exemple de données de puces avec un fort biais des log ratios M selon l'intensité A. La ligne bleue correspond à la définition de $M = 0$ (ratio = 1) pour une normalisation selon l'intensité moyenne dans chaque canal. (B) La définition d'une courbe de régression par la méthode Lowess ne dépend que de la largeur des fenêtres pour la régression linéaire locale, et de leur degré de chevauchement. (C) Courbe de régression définie pour les données de (A). (D) Données après normalisation. D'après <http://bioinfo.unice.fr>

Le design des puces pangénomiques autorise des méthodes de normalisation s'appuyant sur le grand nombre de gènes, donc moins risquées. Il est notamment possible de se baser sur l'intensité globale de fluorescence de chaque condition (donc sur l'ensemble des dépôts), pour calibrer les marquages entre eux ; sur l'intensité de fluorescence des seuls dépôts considérés comme positifs (dans les deux cas, on suppose que les deux cellules ont une activité transcriptionnelle équivalente quantitativement) ; ou encore sur la supposition que la majorité des gènes sont exprimés au même niveau entre les deux conditions. Deux méthodes sont basées sur cette dernière possibilité : la méthode de la majorité stable, qui recherche le plus grand sous-ensemble de gènes ayant un ratio similaire et se sert de leurs valeurs pour définir un facteur de correction unique, ou la méthode Lowess, initialement développée par W.S. Cleveland et de loin la plus utilisée. La méthode Lowess (LOcally WEighted Scatterplot Smoothing, ou méthode de lissage des valeurs par pondération locale), appliquée aux données de puces à ADN (Yang et al. 2002), permet d'équilibrer les marquages mais aussi de gommer certains artefacts, notamment l'effet « banane », ou non-linéarité des ratios en fonction de l'intensité, un effet particulièrement visible sur les MA plots (figure 10) et qui est responsable d'important biais avec les autres méthodes de normalisation « linéaire ». La méthode consiste à effectuer une régression linéaire locale pour aboutir en définitive à une courbe de régression robuste (figure 11) : pour chaque fenêtre d'intensité, la régression linéaire locale définit une correction adaptée aux données dans cette gamme d'intensité, puis la fenêtre est décalée vers les intensités plus élevées, et ainsi de suite. Cette méthode permet de normaliser de façon adaptée là où un facteur de correction unique s'accommode mal de la nature non-linéaire des données. De nombreux outils de normalisation pour les données de puces « deux couleurs » sont disponibles, citons simplement l'outil Bioconductor et le langage de programmation R (Dudoit et al. 2003), extrêmement employés car libres d'usage et comportant de nombreux modules pour l'analyse d'image, la normalisation, la classification des données... Citons également les portails M@IA (Micro@rray Integrated Applications - <http://maia.genouest.org/>) (Le Behec et al. 2008), basé sur R, et GEPAS (<http://gepas.bioinfo.cipf.es/>) (Herrero et al. 2003), aux fonctionnalités très complètes.

Cette étape de normalisation est faite systématiquement pour chacune des puces (normalisation intra-lames ou centrage). Plus rarement, une étape de

normalisation inter-lames (centrage + réduction) est parfois rajoutée, afin d'équilibrer la dispersion des ratios d'une puce lorsque certaines s'avèrent plus « sensibles » que d'autres (figure 12).



Une fois la normalisation effectuée, la moyenne des ratios d'expression est calculée pour chaque sonde à partir des données normalisées des différentes puces, puis un seuil est choisi, afin de définir les gènes différentiellement exprimés. A cette étape, un test statistique est requis, afin d'évaluer si les conclusions tirées à partir des ratios moyens sont significatives ou susceptibles d'être dues au hasard. Le risque lors d'un test statistique est double : celui de déclarer un gène comme étant différentiellement détecté alors qu'il ne l'est pas (risque de type 1, ou risque de faux positif), et celui de déclarer un gène comme non régulé alors qu'une différence de détection est constatée (risque de type 2, ou faux négatif) (figure 13). Ces deux risques

H_0 = « Pas de différence de détection »
 H_1 = « Différence de détection »

Situation / Décision \	H_0 Vrai	H_0 Faux
H_0 acceptée	$1 - \alpha$	β
H_0 rejetée (H_1 acceptée)	α	$1 - \beta$

α = risque de 1^{ère} espèce (faux positif)
 β = risque de 2^{ème} espèce (faux négatif)
 $1 - \alpha$ = confiance du test
 $1 - \beta$ = puissance du test

Figure 13. Test statistique et risques.

sont liés au choix de deux paramètres α et β . Le seuil α (= 1 – confiance du test), correspondant au seuil d’erreur de type 1 (faux positifs) toléré pour un test, est généralement fixé à 0,05 ou 0,01. Le seuil β (= 1 – puissance du test) correspond au seuil d’erreur de type 2, mais ce risque n’est généralement pas évalué, les faux négatifs n’étant pas considérés par la suite. Les seuils α et β sont liés : plus le seuil α choisi est faible, plus le risque de rejeter une valeur à tort est élevé. Le test statistique le plus employé est le test t de Student ; lorsque

la p-value calculée (probabilité d’observer une différence d’expression au moins égale si en réalité il n’y a pas de différence) est inférieure au seuil α fixé (0,05 ou 0,01), l’expression différentielle est considérée comme significative d’après le test.

Toutefois, ce type de test doit être répété pour chaque ratio, ce qui, même en ne considérant que les ratios de gènes surexprimés ou sous-exprimés représente un grand nombre de tests. En fixant par exemple un seuil α à 0,05, pour le cas d’une analyse où 1000 gènes seraient testés, 50 faux positifs (1000 x 0,05) seraient considérés comme vrais positifs, alors que le nombre réel de vrais positifs (en plus de ces 50) peut être beaucoup plus faible, ce qui explique qu’en raison des tests multiples, la proportion de faux positifs parmi les gènes déclarés positifs peut être très élevée. Une des parades consiste à corriger le test : une des premières corrections inventée est la correction de Bonferroni, qui consiste à diviser le seuil α par le nombre de tests, mais cette correction est trop sévère dans le cas de données de puces, le nombre de tests étant de plusieurs milliers. Une technique récente, qui tend de plus en plus à remplacer le test t de Student, consiste à calculer le taux de fausse découverte ou FDR (False Discovery Rate, méthode décrite initialement par Benjamini et Hochberg en 1995) (Reiner et al. 2003; Strimmer 2008). Le FDR correspond à la proportion attendue de faux positifs parmi les gènes déclarés positifs par le test selon la valeur du seuil α , ce qui permet de déterminer le seuil α adéquat en fonction de la distribution des p-values du test de Student. La méthode de calcul la

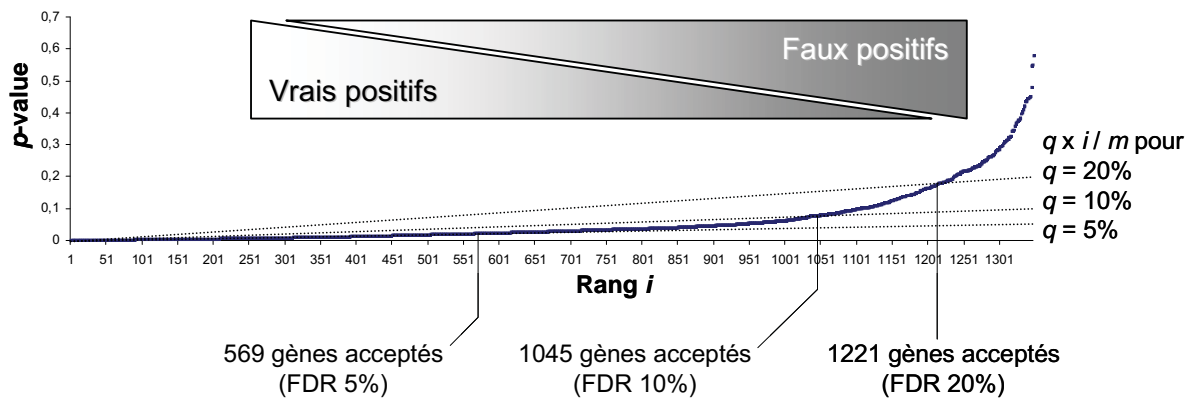


Figure 14. Exemple de la détermination du nombre de gènes pour lesquels H_0 est rejetée en fonction du FDR souhaité (méthode de Benjamini-Hochberg).

Dans cet exemple, 1346 gènes ont un ratio > 2 ou $< 0,5$. Les gènes sont classés par ordre croissant de p-values, déterminant leur rang i . Tant que les p-values restent inférieures à (taux de FP souhaité \times rang i / 1346) (lignes pointillées), H_0 est rejetée et les gènes sont déclarés significativement régulés. Pour un taux de FP fixé à 5%, on attend ainsi environ 28 faux et 541 vrais positifs sur les 569 gènes acceptés.

plus simple pour déterminer le seuil adéquat consiste à classer les m gènes potentiellement régulés par p-value croissante ($p_1 \leq \dots \leq p_i \leq \dots \leq p_m$), et à déterminer le rang i au-delà duquel la p-value dépasse la valeur $q \times (i / m)$ où q est la proportion maximale souhaitée de faux positifs (figure 14).

Si le calcul de la p-value ou du FDR est adapté à la recherche de gènes différentiellement exprimés, la classification non-supervisée des données est une approche alternative fréquemment utilisée en cancérologie pour identifier des signatures spécifiques de certaines tumeurs (cf. partie précédente). La classification non supervisée consiste à déterminer des classes (groupes) de gènes co-exprimés dans les différents échantillons analysés, de sorte que les différences soient minimales au sein d'une classe et maximales entre les classes. La plus couramment utilisée est la classification hiérarchique, qui consiste à générer une suite de classes emboîtées et représentées sous la forme d'un dendrogramme (représentation en arbre). Elle requiert de calculer une « distance » entre les gènes (et échantillons) afin de regrouper les plus similaires dans une classe, et de choisir une règle d'agglomération (distance inter-groupes) pour hiérarchiser ces classes. Les distances de similarité les plus utilisés en biologie sont la distance euclidienne (distance « physique » dans un espace à n dimensions, n étant le nombre d'échantillons ou de gènes, et les coordonnées étant les mesures d'expression elles-mêmes) ou la distance de corrélation de Pearson, tandis que les règles d'agglomération courantes sont le lien moyen (« average linkage », distance moyenne entre les points des deux groupes), le

lien complet (« complete linkage », distance entre les points les plus éloignés) ou le lien simple (« single linkage », distance entre les points les plus proches). Un des premiers logiciels développé spécifiquement pour les données biologiques est le programme Cluster (Eisen et al. 1998), encore couramment employé, ainsi que l'outil Treeview pour construire la représentation visuelle en dendrogramme (ou « Eisengramme », figure 15). D'autres outils, plus récents mais moins utilisés, sont basés sur des alternatives à la classification hiérarchique, tels que la méthode des K-means, la Self-Organizing Map (SOM) et le Self-Organizing Tree Algorithm (SOTA), particulièrement robuste.

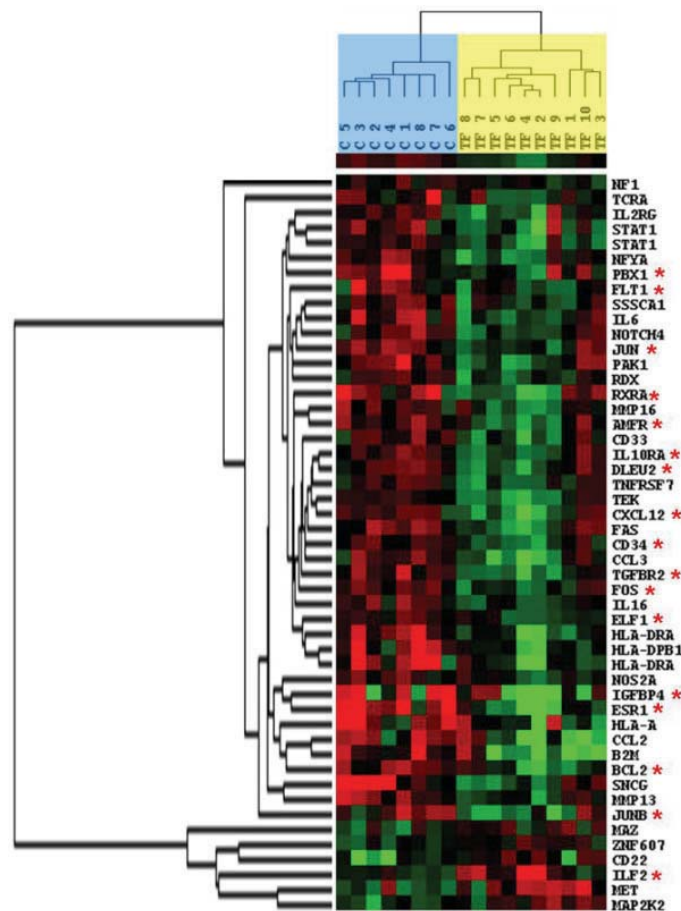


Figure 15. Exemple d'Eisengramme à partir de données d'expression de tumeurs sensibles ou non au tamoxifène (Vendrell et al. 2008).

1.2.2. PCR quantitative en temps réel

1.2.2.1. Origine et principe

Les techniques de PCR utilisent des amorces spécifiques d'un messenger particulier pour détecter son expression parmi l'ensemble des ADNc d'une population cellulaire. Les méthodes classiques n'autorisent pas les comparaisons de niveau d'expression car après quelques dizaines de cycles de PCR, la quantité d'ADN amplifié n'est plus proportionnelle à la quantité de départ. En effet, un fléchissement de la cinétique d'amplification apparaît lorsque l'un des réactifs (amorces, dNTP...) devient limitant, ou que la polymérase n'est plus suffisamment active. Ainsi, le dépôt sur gel d'agarose de produits de PCR obtenus à partir de deux échantillons, peut faire apparaître des bandes d'intensités très voisines, malgré un niveau d'expression du gène d'intérêt très différent.

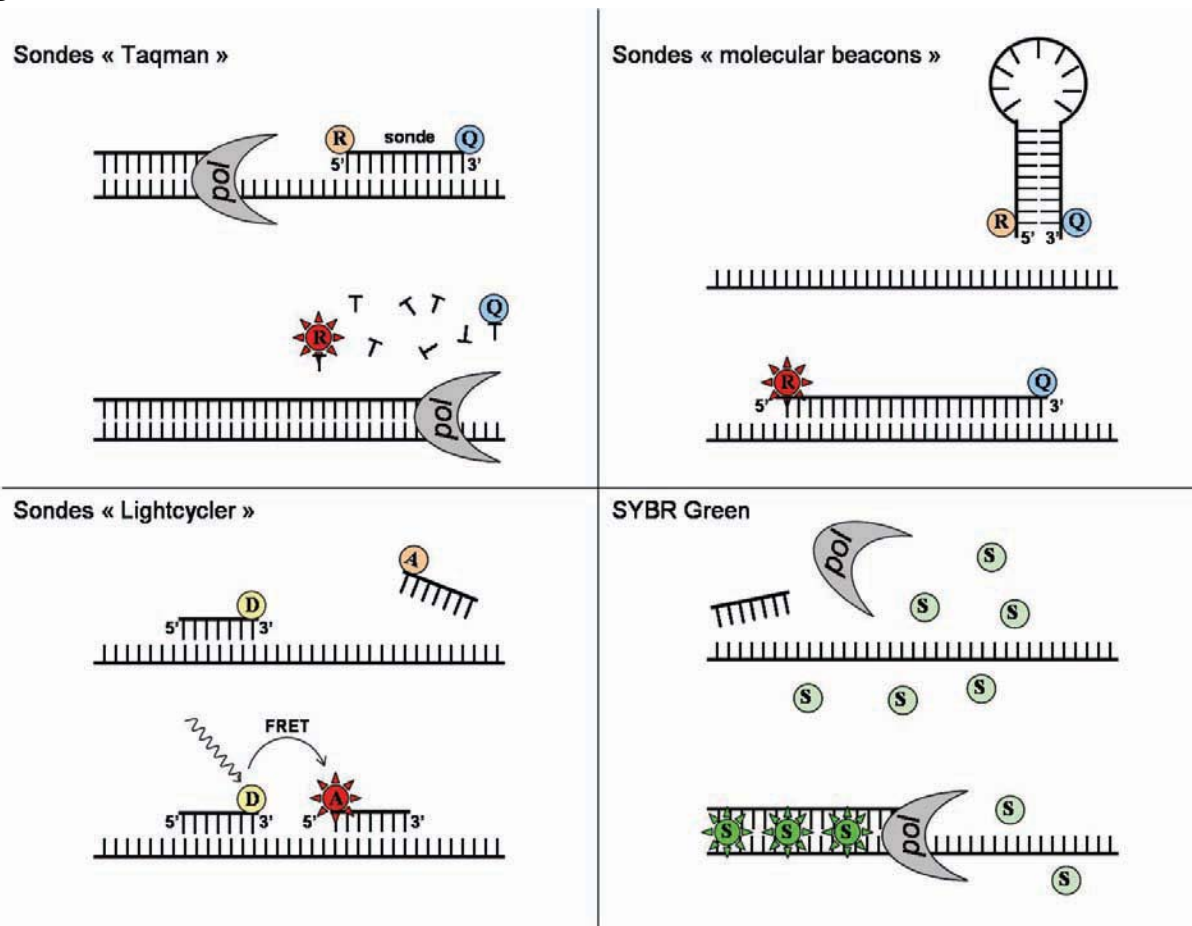


Figure 16. Chimies employées pour la quantification en temps réel par PCR.

Au contraire, la PCR quantitative autorise une mesure de la quantité d'ADN amplifié à chaque cycle et non à la fin de l'expérience. Le principe de base reste toujours inchangé (amorces spécifiques de la cible, Taq polymérase, dNTP, etc.), mais le matériel et la méthode de détection connaissent des variantes, le but étant de quantifier une fluorescence dont l'intensité est directement corrélée à l'amplification de la cible. Ainsi, la méthode de PCR quantitative en Sybr Green (marqueur) consiste à suivre l'apparition en temps réel d'ADN double brin grâce à un marqueur n'interférant pas avec l'activité de la polymérase. D'autres méthodes (Taqman, Lightcycler, Molecular Beacons, Scorpion primers...) sont basées sur un système de sondes couplant généralement un fluorochrome et un « quencheur » qui absorbe la fluorescence du fluorochrome, les deux étant physiquement séparés lors de l'amplification, ce qui entraîne l'apparition d'un signal (figure 16). Ces méthodes sont théoriquement plus spécifiques que la PCR en Sybr green, puisque faisant intervenir une sonde séquence-spécifique pour la détection en plus des deux amorces, mais la synthèse d'amorces marquées est coûteuse. Elles sont donc plus adaptées à la quantification de quelques gènes dans un grand nombre d'échantillons. Au contraire, la PCR quantitative en Sybr Green est plus adaptée à la quantification d'un grand nombre de gènes différents dans un faible nombre d'échantillons.

La PCR quantitative en Sybr green permet soit une quantification absolue, soit relative. La quantification absolue nécessite la réalisation d'une courbe étalon pour chaque gène d'intérêt, alors que la quantification relative consiste simplement à comparer directement les échantillons entre eux. Dans ce dernier cas, la PCR quantitative nécessite de choisir un ou plusieurs gènes de référence afin de réaliser une normalisation, nécessaire pour compenser par exemple les différences de concentration des solutions d'ADNc comparées. Le choix du ou des gènes de référence est critique : en effet, peu des gènes usuellement considérés comme des gènes « de ménage » ont une expression véritablement inchangée entre les conditions comparées, ce qui nécessite, avant toute étude, de déterminer avec soin quels sont les gènes les plus stables. Ainsi, de profonds biais peuvent être introduits dans la quantification lors de la normalisation, si les gènes de références sont inadéquats.

1.2.2.2. Normalisation

Le principe de la PCR quantitative consiste à quantifier l'ADN amplifié à chaque cycle (d'où le terme PCR quantitative *en temps réel*), en se basant sur le fait que la quantité d'ADN en fin de réaction n'est plus corrélée avec la quantité de

départ, mais l'est durant la phase d'amplification exponentielle qui précède la phase de plateau (figure 17). La méthode de détermination du Ct consiste à définir arbitrairement un seuil de fluorescence généralement bas, mais bien supérieur au bruit de fond, de sorte que ce seuil soit dépassé lorsque l'amplification est encore exponentielle. Le Ct ou « Cycle threshold » correspond alors au nombre de cycles requis pour dépasser ce seuil. Or les Ct sont directement corrélés à la quantité d'ADN au départ, les échantillons les plus concentrés au départ étant ceux dont la fluorescence sera détectée en premier. Il est donc possible, en comparant les Ct, de calculer des ratios d'expression. La méthode la plus couramment employée (Livak and Schmittgen 2001) est le calcul du $2^{-\Delta\Delta Ct}$ (les formules exactes et approchées sont détaillées dans la partie Discussion consacrée à la PCR quantitative) : après avoir calculé le Ct du gène d'intérêt, la valeur du Ct du gène de référence est soustraite (ΔCt) afin de rendre les échantillons comparables, puis les Ct ainsi normalisés sont comparés par simple soustraction ($\Delta\Delta Ct$). Si l'on suppose une efficacité de 100%, c'est à dire un doublement de la quantité d'ADN amplifié à chaque cycle (en phase exponentielle), le rapport des niveaux d'expression, ou ratio, correspond à $2^{-\Delta\Delta Ct}$, le signe (-) étant uniquement dû au fait que les gènes les plus exprimés ont les Ct les plus faibles.

1.2.2.3. Applications

En raison de son faible débit, la PCR quantitative est rarement utilisée comme outil de criblage à grande échelle, contrairement aux puces à ADN. Toutefois, sa sensibilité ainsi que sa gamme dynamique (près de 10^4) élevées en font un outil de prédilection pour quantifier de petites variations d'expression. De plus, la reproductibilité et la spécificité élevées de cette méthode en font un outil privilégié pour confirmer des données obtenues par d'autres techniques. Enfin, contrairement aux puces à ADN, la spécificité de la PCR quantitative est aisément vérifiable, d'une part en contrôlant les courbes de dissociation qui témoignent de l'amplification d'une seule séquence d'ADN, mais aussi en réalisant une validation par séquençage des amplicons obtenus au cours de la PCR quantitative. En dehors de son utilisation pour la quantification de l'expression génique ou comme outil de validation, la PCR quantitative en temps réel est également couramment employée pour la détection et quantification de charges virales (Lin et al. 2004), le diagnostic de tumeurs (Rambaldi et al. 2005), le titrage de germes et contaminants (agronomie, sang,

fluides, tissus), la discrimination allélique, l'analyse de SNP, d'amplifications ou de délétions géniques, la quantification du nombre de copies d'un transgène, etc

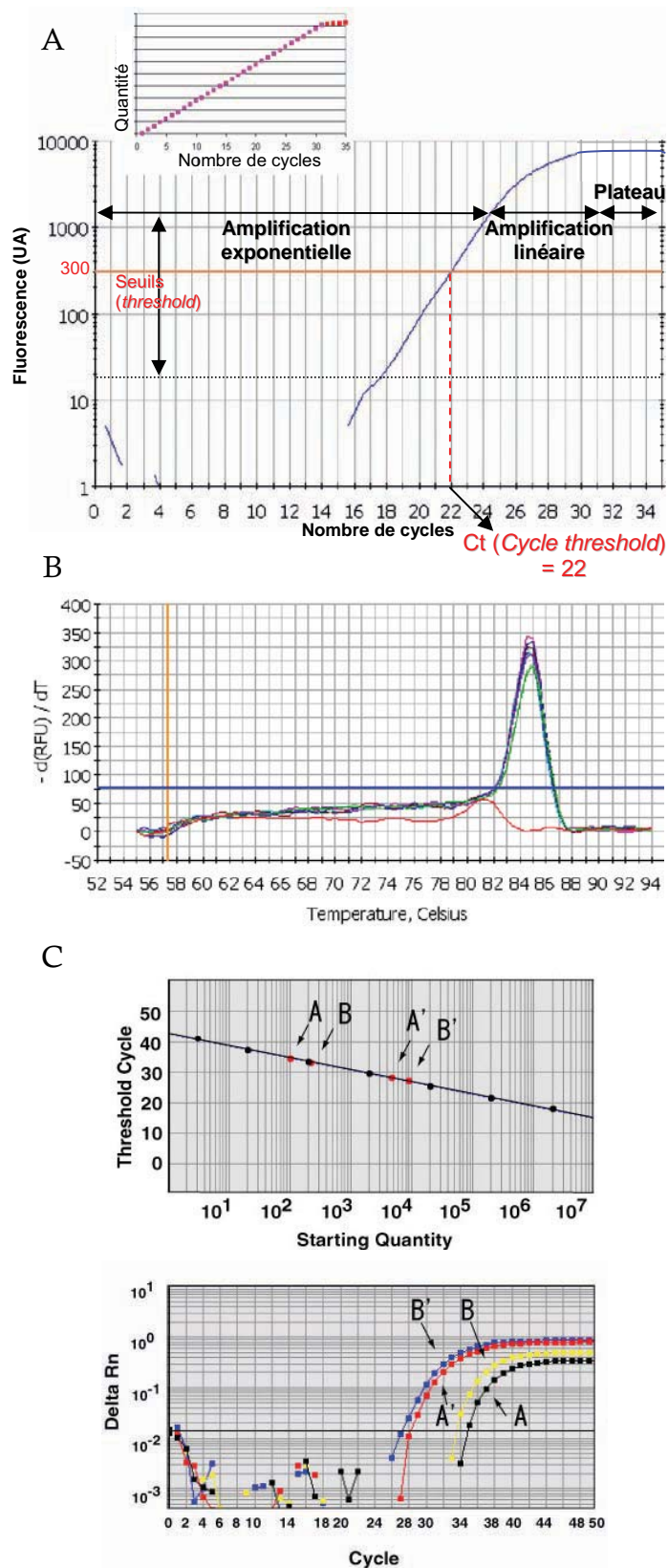


Figure 17. Quantification relative ou absolue par la détermination du Cycle threshold (Ct). (A) Courbe d'amplification et détermination du Ct. (B) Courbe de dissociation. (C) Gamme étalon et quantification absolue du matériel de départ. D'après <http://pathmicro.med.sc.edu/> et <http://pathol2.iwate-med.ac.jp/>

1.2.3. SAGE & Séquençage nouvelle génération

Les méthodes d'analyse du génome et du transcriptome basées sur l'hybridation ont des limites, à la fois en termes de sensibilité et de spécificité. Récemment, le développement de méthodes de séquençage dites de nouvelle génération, ou à haut débit (« deep sequencing »), autorise une conception radicalement différente de l'exploration de l'expression génique. Nous nous concentrerons sur les approches qu'elles permettent de réaliser, la description des techniques elles-mêmes pouvant être trouvée en annexe.

Plusieurs techniques de séquençage à (très) haut débit ont vu le jour ces dernières années et visent à apporter des solutions à plusieurs problèmes posés par la technique de Sanger. Elles permettent notamment de s'affranchir de l'étape préalable

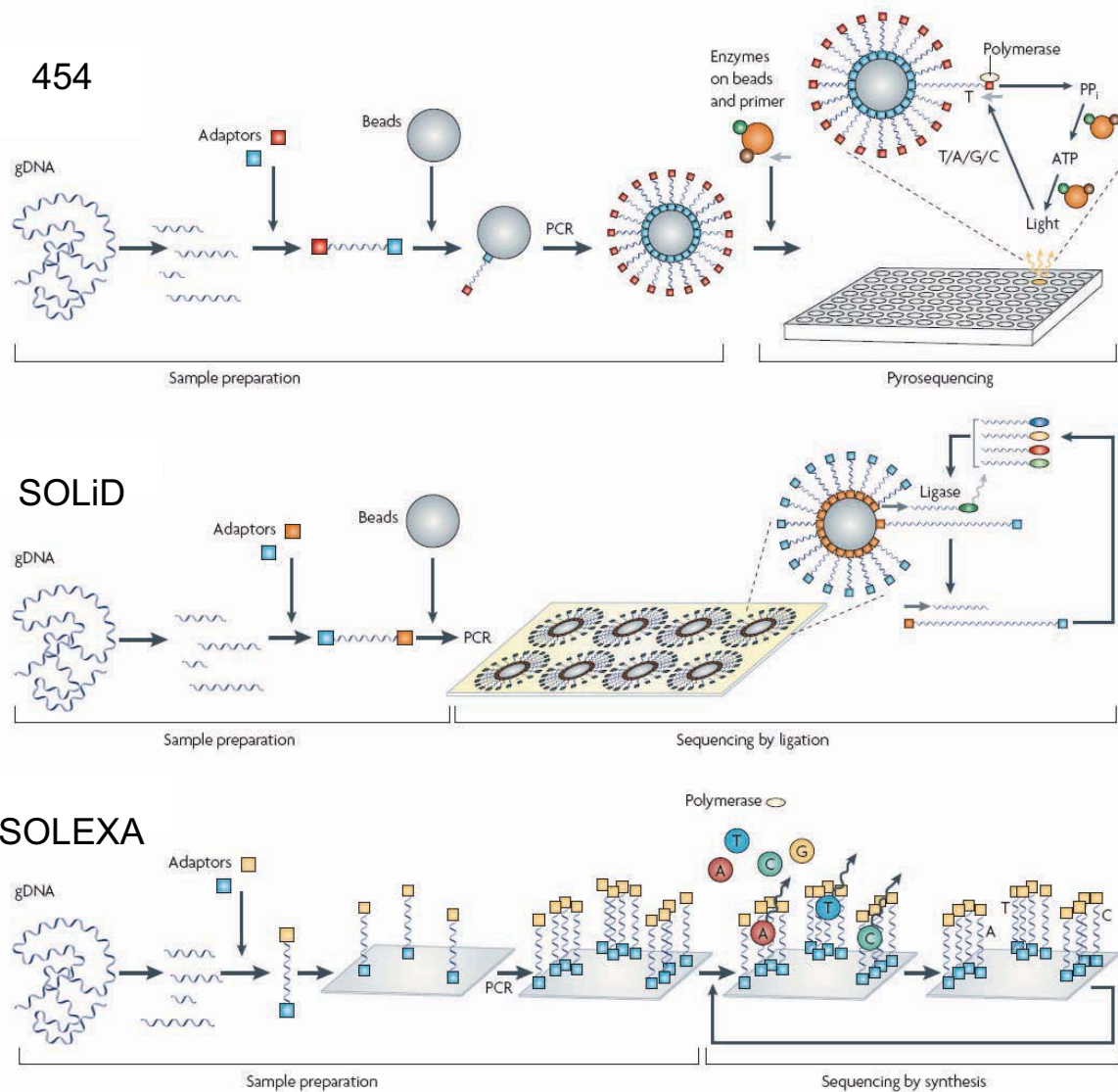


Figure 18. Nouvelles techniques de séquençage à très haut débit (Medini et al. 2008).

de clonage et d'amplification *in vivo* dans des bactéries, susceptible de générer des mutations. Les trois principales approches, développées et commercialisées par des compagnies concurrentes, sont basées sur des chimies distinctes (technologies 454, du Solexa et du SOLiD ; figure 18). Leurs principaux avantages sont la rapidité et surtout la grande quantité de séquences générées (400 à 600 megabases en 10 heures pour le 454, jusqu'à 18 gigabases en 10 jours pour le Solexa et 1 à 3 gigabases en 8 jours pour le SOLiD), ainsi que le faible coût (quelques milliers d'euros), rendant par exemple possible le séquençage complet du génome d'un humain (avec une profondeur x12, soit en moyenne 12 lectures / base) pour 10000 \$ en 2009, chose évidemment impossible avec la technique de Sanger.

Les inconvénients de ces techniques sont néanmoins multiples. En premier lieu, elles sont peu appropriées à de petits projets, la quantité de données générées pouvant être en soi un obstacle pour l'analyse ou même le stockage de ces données – voir chapitre suivant. De plus, la longueur des séquences (environ 35 paires de bases pour deux des trois méthodes), complique fortement l'identification des séquences dans le génome s'il s'agit de portions répétées, et constitue un véritable défi pour l'assemblage s'il s'agit d'un génome jamais séquencé. Pour cette raison, chacune des techniques connaît régulièrement des améliorations visant à la fois à augmenter le débit, donc la quantité de séquences générées, mais surtout leur longueur, avec en moyenne une nouvelle génération d'appareils tous les six mois.

Toutefois, si le grand nombre de données susceptibles d'être générées en un seul run peut compliquer l'analyse des résultats, c'est cette même caractéristique qui rend possible l'exploitation d'un grand nombre de techniques à leur vrai potentiel, c'est à dire à l'échelle de génomes entiers et de façon réellement exhaustive. Nous ne détaillerons ici que les exemples concernant plus spécifiquement l'expression génique (« deep sequencing » et RNA-Seq), pour voir l'impact du séquençage nouvelle génération sur les approches déjà existantes ou sur le développement de nouvelles approches.

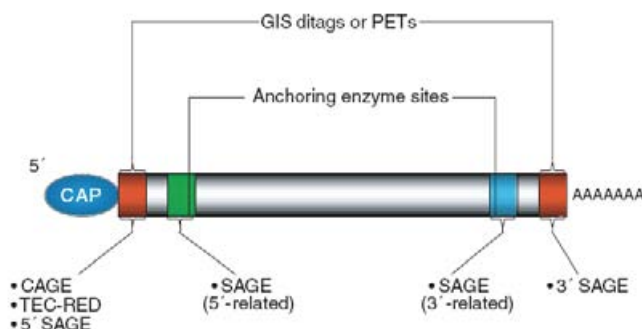


Figure 19. Principales techniques d'étude du transcriptome basées sur la génération d'étiquettes (Harbers and Carninci 2005).

Pour un même messager, les étiquettes ou *tags* peuvent être simples (SAGE, CAGE, TEC-RED...) ou appariées (GIS ditags).

Le deep sequencing consiste à employer le séquençage à très haut débit pour tirer le maximum du potentiel de techniques « anciennes », notamment celles visant à ne séquencer que de courtes portions d'ADNc (étiquettes ou « tags »), ces segments étant généralement concaténés (afin de réduire le nombre de réactions de séquençage de type Sanger initialement employé). Parmi toutes ces techniques (figure 19), citons notamment le SAGE (Serial Analysis of Gene Expression), le CAGE (Cap Analysis of Gene Expression) permettant d'identifier l'extrémité 5' des transcrits, ou encore le GIS (Gene Identification Signature), une stratégie visant à rabouter les extrémités 5' et 3' des ADNc afin d'identifier sans ambiguïté les transcrits, ce qui est difficile lorsque les séquences générées sont trop courtes comme avec le SAGE. Toutes ces techniques basées sur la génération d'étiquettes présentent l'avantage de diminuer fortement le volume de séquences à produire, mais aussi de donner une information à la fois qualitative (quels gènes sont exprimés) et quantitative (le nombre de « tags » pour un gène donné étant théoriquement proportionnel l'abondance de son messenger). Les techniques de séquençage nouvelle génération permettent de se passer de l'étape de concaténation (puisque générant naturellement un très grand nombre de courtes séquences). Elles permettent également d'atteindre une profondeur de lecture suffisante pour identifier les transcrits rares au milieu de transcrits plus abondants, préservant ainsi l'aspect quantitatif de ces méthodes.

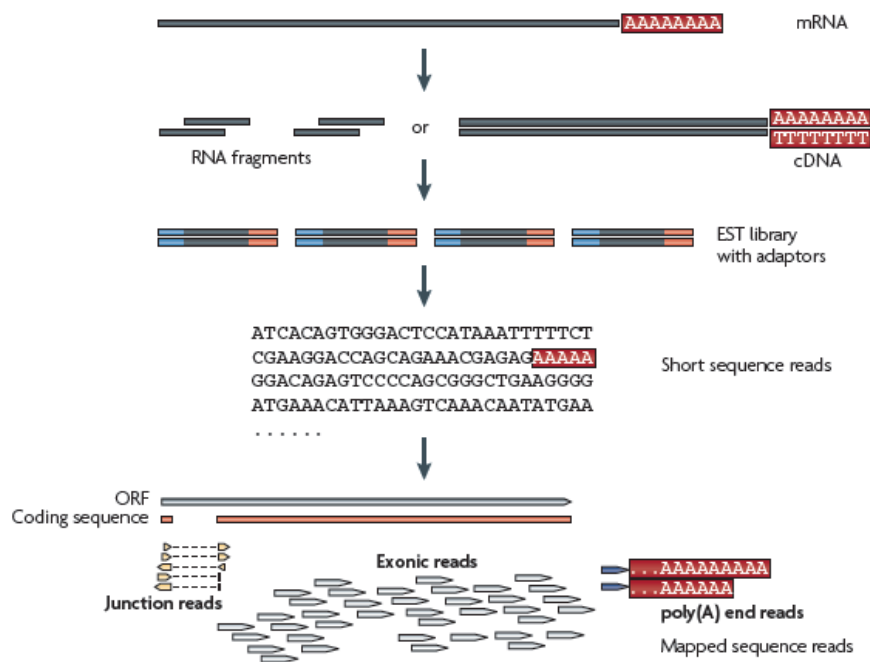


Figure 20. Technique du RNA-Seq (Wang et al. 2009).

Toutefois, en plus des difficultés d'identification, l'usage de courtes séquences livre peu d'information sur la complexité du transcriptome (jonctions exon-exon, épissages alternatifs, etc.)

Or, le débit du séquençage nouvelle génération ne justifie plus les restrictions qu'imposait la méthode de Sanger, comme le fait de se limiter à une portion des ADNc, qu'il s'agisse de tags ou d'ESTs. Ainsi, l'étude de la complexité du transcriptome est plus directement accessible, notamment grâce à la technique récente du RNA-Seq (figure 20). En partant d'une population d'ARN totaux ou purifiés (poly(A)+ par exemple), une librairie de fragment d'ADNc est produite (soit par nébulisation ou hydrolyse des ARN, soit par sonication ou traitement à la DNaseI

Technologie	Tiling microarray	Séquençage (EST/ADNc)	RNA-Seq
Caractéristiques			
Principe	Hybridation	Séquençage de Sanger	Séquençage à très haut débit
Résolution	De quelques bases à 100pb	Une base	Une base
Débit	Haut	Bas	Haut
Dépendance vis-à-vis du séquençage du génome	Oui	Non	Variable selon l'application
Bruit de fond	Elevé	Bas	Bas
Applications			
Cartographie des régions transcrites et mesure de l'expression génique simultanément	Oui	Limitée pour l'expression génique	Oui
Gamme dynamique (expression génique)	Généralement < 100x	Peu applicable	> 8000x
Distinction entre isoformes	Limitée	Oui	Oui
Distinction entre allèles	Limitée	Oui	Oui
Limitations pratiques			
Quantité d'ARN requis	Importante	Importante	Faible
Coût pour des génomes importants	Elevé	Elevé	Moyen

Table 1. Comparaison entre tiling microarray, banques d'ADNc (EST) et RNA-Seq.

des ADNc), ces fragments étant liés à des adaptateurs permettant éventuellement une amplification avant le *deep sequencing*. La possibilité de limiter au maximum les étapes entre l'extraction des ARN et leur séquençage constitue une grande avancée par rapport aux techniques antérieures permettant d'aller aussi loin dans l'étude du transcriptome (tiling microarrays ou séquençage d'ADNc – Table 1). Une récente analyse du transcriptome de tissus murins (cerveau, foie et muscles squelettiques) a été réalisée par RNA-Seq (Mortazavi et al. 2008). Cette approche par deep sequencing a permis de mesurer l'abondance relative de chacun des messagers en fonction du nombre de séquences, mais également de réviser l'annotation du génome. 90% des séquences alignées avec le génome se sont avérées localisées au niveau d'exons connus, les autres séquences définissant de nouveaux sites d'initiation de la transcription, des exons codants alternatifs ou des variants de la région 3' non codante, ainsi que de nouveaux précurseurs de microARN. Deux études menées chez l'homme ont permis de dresser un catalogue de 31 997 (Morin et al. 2008a) et 94 241 (Sultan et al. 2008) jonctions exon-exon, dont 379 et 4096 jamais décrites (respectivement), mettant notamment en valeur la prévalence du saut d'exon sur les autres formes d'épissage alternatifs. Une autre étude ciblant exclusivement les microARN exprimés dans des cellules souches embryonnaires humaines non-différenciées ou différenciées en corps embryoides (amas cellulaire constitué des trois feuillets embryonnaires) a identifié 334 microARN connus et 104 jamais décrits, dont 171 connus et 23 nouveaux différentiellement exprimés entre les deux stades de différenciation (Morin et al. 2008b). Plusieurs autres études employant le RNA-Seq ont été menées chez *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, ainsi que la souris et l'homme (pour revue, voir (Wang et al. 2009)).

Il est important de noter que les applications du séquençage nouvelle génération ne se limitent pas à l'étude du transcriptome, mais concernent également la recherche de sites fixés par des facteurs de transcription (le ChIP-Seq – ImmunoPrécipitation de Chromatine suivie du séquençage à haut débit des fragments de chromatine – remplaçant alors le ChIP-on-chip, les puces à ADN étant moins sensibles et plus contraignantes) ou l'étude de l'état de la chromatine. Le ChIP-Seq a ainsi récemment permis de caractériser à l'échelle du génome les sites de fixation de 13 facteurs de transcription tels que Nanog, Oct4, STAT3, Smad1, Sox2, etc., jouant des rôles-clés dans les cellules souches embryonnaires, définissant ainsi

une première carte du cœur du réseau de régulation transcriptionnelle dans ces cellules (Chen et al. 2008). Le séquençage nouvelle génération facilite également les études de populations comme la recherche de SNP (Van Tassell et al. 2008) ou de CNV en séquençant de nombreux individus, ou même les travaux de métagénomique (identification de nombreuses espèces de microorganismes présentes dans un échantillon complexe, par exemple eau d'étang ou prélèvements de fluides biologiques (Nakamura et al. 2009)), voire de métatranscriptomique (Gilbert et al. 2008).

1.3. Bioinformatique

1.3.1. Banques de données

La première étape dans la gestion et l'analyse des données est avant tout leur pérennisation, qui doit garantir leur disponibilité pour le plus grand nombre. Il existe pour cela de nombreuses bases de données publiques, qui diffèrent notamment par leur « architecture » (la façon dont l'information est collectée, stockée et redistribuée ; ce point ne sera pas traité ici), mais aussi par leur approche, selon qu'il s'agisse d'une banque généraliste ou spécialisée.

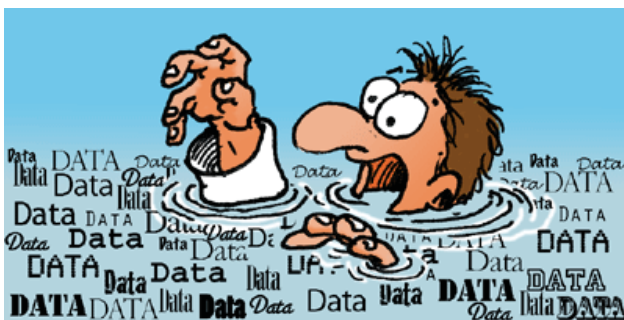
Concernant les banques de séquences nucléotidiques, trois grandes banques internationales existent : GenBank au NCBI (National Center for Biotechnology Information, Bethesda, MD, USA), la Nucleotide Sequence Database (ou EMBL-Bank) de l'EMBL (European Molecular Biology Laboratory, Cambridge, UK), et la DDBJ (DNA Database of Japan, Mishima, Japan), toutes trois interconnectées depuis 1986 grâce aux efforts de l'INSDC (International Nucleotide Sequence Database Collaboration ; <http://www.insdc.org/>). Ce comité est chargé d'unifier les formats de stockage des données de ces trois grandes banques, afin de faciliter par la suite leur annotation ; ainsi, toute séquence soumise à une de ces trois banques sera convertie dans un format commun, et intégrée aux deux autres banques lors d'une mise à jour.

Pour les données de séquençage à haut débit, le volume de stockage requis représente un réel défi pour leur partage et leur exploitation. En effet, un seul « run » de séquençage haut débit peut générer une quantité très importante de données, de 100 mégabases à 3 gigabases suivant la technique. Le volume des données brutes pose de sérieux problèmes de stockage, ce qui explique qu'elles aient été parfois supprimées après leur traitement, d'où des difficultés pour reconstruire la qualité des

séquences. Même en ne considérant que les séquences post-traitement, leur très grand nombre pose problème pour la gestion et l'interrogation des banques de données les hébergeant. Ainsi, les trois banques internationales, devant l'immense afflux de courtes séquences issues des techniques à haut débit, ont créé une section SRA (Short Read Archive) dédiée au dépôt et la consultation de ce type de séquences, en marge des séquences de type Sanger (Trace Archive). A titre d'illustration, lors de la création de cette section, l'EMBL-Bank a reçu en trois mois un volume de séquences correspondant à 1/8^{ème} des données générées en 28 années d'activité. Ainsi, l'architecture de ces banques a dû être adaptée pour permettre le stockage des données de séquençage haut-débit.

Concernant les données d'expression issues de puces à ADN, trois banques coexistent : le Gene Expression Omnibus (GEO) du NCBI (<http://www.ncbi.nlm.nih.gov/geo/>), ArrayExpress de l'EMBL-EBI (<http://www.ebi.ac.uk/microarray-as/ae/>), et le Center for Information Biology gene EXpression database (CIBEX) de la DDBJ (<http://cibex.nig.ac.jp/>), toutes trois imposant le respect du standard MIAME (Minimum Information About Microarray Experiment) créé par la MGED Society, une organisation fondée en 1999 et regroupant des biologistes, bioinformaticiens et bioanalystes dans le but de faciliter le partage, l'interprétation et la reproduction de données issues de méthodes à haut débit. Le standard MIAME impose notamment de spécifier le maximum d'informations sur la totalité du protocole : l'échantillon (nom, date, organisme, tissus/cellules, stade développemental, statut physiologique, génotype...), l'extraction (nom, date, ARN totaux/purifiés, méthode, volume, concentration...), le marquage (nom, date, fluorochrome, méthode, quantité...), les puces employées (plan exact de la puce), le protocole d'hybridation (échantillons, quantité, tampon, température, durée...), le plan expérimental (comparaison de deux types cellulaires, profiling...).

1.3.2. Outils d'annotation : de l'information à la connaissance



Les techniques à haut débit génèrent un tel flux de données d'expression que leur dépôt dans des banques de données n'est pas, de loin, la seule étape posant problème. Ainsi, l'annotation, c'est à dire le retour au

versant biologique de ces approches, s'avère généralement complexe, que ce soit à l'échelle d'un même projet, ou lorsqu'on tente de comparer les données générées par plusieurs équipes ou groupes.

A l'échelle d'un projet, l'étape cruciale est le passage d'une information issue de données analysées, à une connaissance biologique, soit par exemple, le passage de données normalisées de puces à ADN (liste de gènes différentiellement exprimés) à une description des fonctions connues ou supposées de ces gènes. Suivant le volume de données générées, l'analyse manuelle, bien que présentant un niveau de fiabilité élevé, peut s'avérer une solution trop coûteuse en temps, voire impossible, en plus d'avoir le défaut de considérer les gènes un par un, là où il peut être intéressant de mettre en exergue les relations qu'il peut y avoir entre eux.

De nombreux outils d'annotation automatique existent à ce jour, basés généralement sur l'utilisation d'une nomenclature précise décrivant les objets (gènes), leur appartenance à une famille ou un processus (voies), leur fonction, leur relation avec d'autres objets... Ce type de description, employant un vocabulaire structuré précis, ou ontologie, est une nécessité lorsque l'on souhaite avoir une homogénéité entre les annotations de données, et l'exemple le plus connu est Gene Ontology. Le consortium GO, créé en 1998 par les bases de données Flybase (drosophile), Saccharomyces Genome Database (SGD) et Mouse Genome Database (MGD), regroupe à ce jour une vingtaine de bases de données partenaires. Le consortium vise, indépendamment de l'organisme considéré, à donner la description la plus précise possible des produits de gènes, selon trois axes distincts : le composant cellulaire impliqué (ex : réticulum endoplasmique, ribosome, complexe protéique...), le processus biologique associé (voie métabolique, voie de signalisation...), et la fonction moléculaire connue (liaison à un récepteur, catalyse d'une réaction...), chacune de ces trois ontologies étant indépendantes. Un grand nombre d'outils basés totalement ou en partie sur les termes GO sont proposés soit par le consortium lui-même, soit par la communauté GO ; citons AmiGO (sur le site de GO), DAVID (Database for Annotation, Visualization and Integrated Discovery - <http://david.abcc.ncifcrf.gov/>), Ease (Expression Analysis Systematic Explorer - <http://david.abcc.ncifcrf.gov/ease/ease.jsp>), GoMiner (<http://discover.nci.nih.gov/gominer/>)... Toutefois, l'exploitation de ce type d'outil se heurte à deux écueils : la fiabilité des termes GO, parfois erronés ou contradictoires car attribués automatiquement en fonction du degré d'homologie avec une protéine connue et non

vérifiés manuellement ; l'exhaustivité et la précision de la description, qui est loin d'être satisfaisante pour certains produits de gènes exprimés dans un très petit nombre de tissus et/ou peu décrits. Certains outils, tels que Ease ou M@IA, offrent toutefois la possibilité de pallier en partie ce défaut en combinant les informations qui peuvent être trouvées dans différentes bases de données, afin d'obtenir l'annotation la plus complète possible. D'autres outils, tels que Ingenuity Pathway (figure 21), permettent une annotation automatique théoriquement plus sûre car vérifiée manuellement avant d'être rendue disponible ; toutefois leur coût, ainsi que leur non-exhaustivité en raison de l'ampleur de la validation humaine des données, rendent ces outils souvent incomplets.

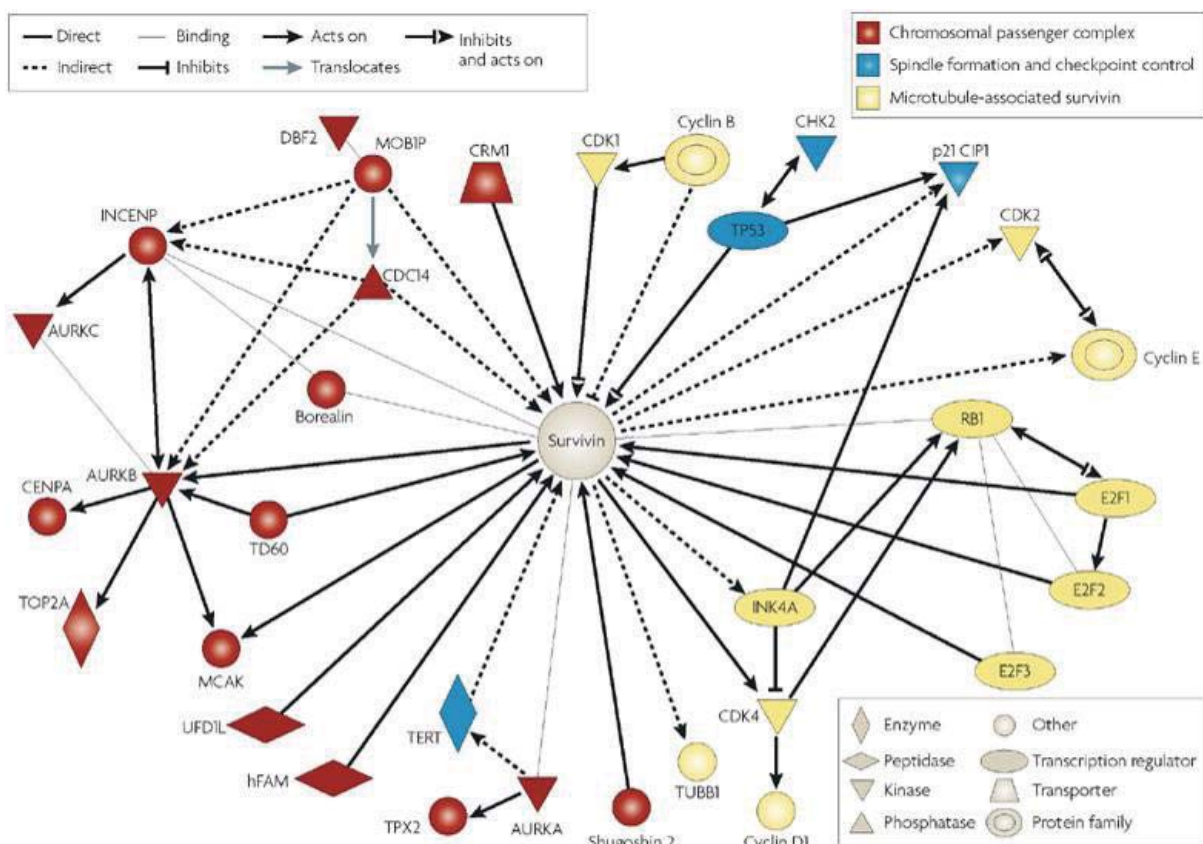


Figure 21. Réseau de protéines associées à la Survivine reconstitué grâce à l'outil d'annotation Ingenuity Pathway (Altieri 2008).

1.3.3. Identification *in silico* de séquences régulatrices

Lorsque la connaissance d'un processus biologique est suffisante, la bioinformatique peut devenir un outil de modélisation et de prédiction permettant de réaliser d'importantes économies de temps en ciblant les recherches à partir des données obtenues *in silico*. Un des exemples les plus frappants est probablement la

recherche de séquences régulatrices au sein des génomes d'organismes séquencés. Plusieurs approches bioinformatiques complémentaires ont été employées, et nous allons détailler ici la recherche de sites de fixation pour facteurs de transcription ou TFBS (Transcription Factor Binding Site), ainsi que la recherche de régions conservées au cours de l'évolution.

1.3.3.1. Recherche de sites de fixation pour facteurs de transcription

Ces dernières années, la connaissance sur les facteurs de transcription (FT) et leur fonctionnement a grandement évolué, en particulier sur la façon dont ils interagissent avec l'ADN. Généralement, le contact FT-ADN se fait par l'intermédiaire de liaisons faibles (électrostatiques, forces de van der Waals) avec les bases nucléotidiques de l'ADN, généralement depuis le grand sillon. Si quelques FT peuvent se fixer sur un grand nombre de régions de façon assez peu spécifique, la plupart reconnaissent spécifiquement plusieurs bases nucléotidiques et donc des sites plus ou moins définis. Cependant, les bases reconnues n'étant pas nécessairement directement adjacentes, de nombreux FT reconnaissent préférentiellement un site de fixation précis, mais aussi des sites plus ou moins dégénérés, avec des affinités variables selon les bases qui diffèrent. Pour cette raison, il n'est pas judicieux de rechercher directement au sein d'une séquence la présence d'un site consensus, en sachant que des sites alternatifs sont susceptibles de ne pas être pris en compte par l'analyse.

La plupart des outils bioinformatiques de recherche de TFBS emploient non pas des séquences consensus, mais des matrices de liaison, établies à partir de sites de fixation déterminés expérimentalement, principalement par la technique de SELEX ou Systematic Evolution of Ligands by EXponential enrichment (Tuerk and Gold 1990; Cui et al. 1995). Pour chaque position, la fréquence observée de A, T, G, ou C est alors calculée, afin de la comparer à la séquence étudiée et de déterminer le degré de ressemblance et la probabilité d'observer un tel motif. Plusieurs bases de données regroupent un grand nombre de matrices : citons les bases de données publiques ooTFD (object-oriented Transcription Factors Database, 457 matrices ; <http://www.ifti.org/oofdf/>) et TransFac (398 matrices ; <http://www.gene-regulation.com/pub/databases.html>), ou encore la base MatBase développée par Genomatix (<http://www.genomatix.de/products/index.html>), très complète (974 matrices) mais payante.

1.3.3.2. Recherche de régions conservées

La recherche de TFBS étant limitée aux facteurs de transcription dont on connaît la spécificité de liaison à l'ADN, une partie des séquences susceptibles de jouer un rôle dans la régulation génique peuvent échapper à cette approche. La recherche de régions conservées au cours de l'évolution part du postulat que parmi les séquences du génome, codantes ou non, les portions contenant une information importante auront tendance à accumuler moins de mutations que les portions contenant peu d'information, parce que ces mutations auront plus de chances d'être délétères. Parmi les séquences non-codantes mais porteuses d'une information cruciale figurent notamment les éléments du cœur des promoteurs (« core elements » tels que la boîte TATA, la séquence initiateur Inr correspond au point +1 de transcription, *etc.* figure 22), ainsi que les sites liés par les facteurs de transcription régulant leur activité, généralement situés dans le promoteur proximal ou éventuellement dans les premiers introns. Les développements de la bioinformatique ont rendu faciles les comparaisons entre les génomes de différentes espèces, et de nombreux logiciels d'alignements multiples (DBA – DNA Block Aligner, PipMaker, zPicture...) permettent d'identifier les régions conservées, les remaniements (inversions, duplications...), *etc.*, et donc de cibler des régions plus susceptibles d'être importantes pour la régulation génique.

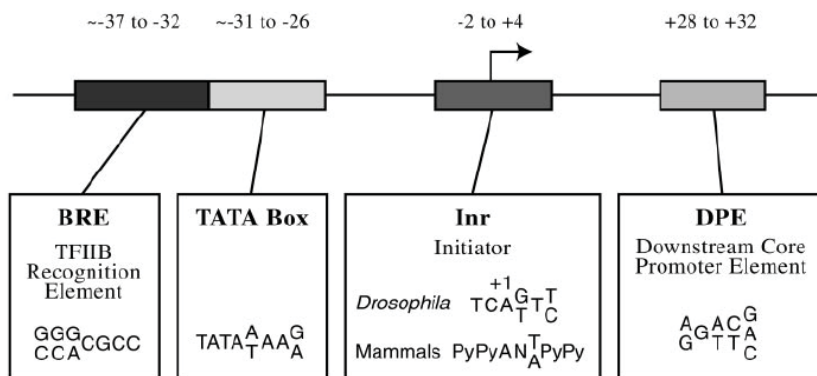


Figure 22. Différents éléments du cœur d'un promoteur eucaryote (Smale and Kadonaga 2003). Chacun de ces éléments n'est présent que dans une partie des promoteurs, et environ 50% des gènes sont associés à des îlots CpG.

Toutefois, l'analyse de la conservation des séquences génomiques ne se limite pas aux promoteurs proximaux et introns. Dès 1993, la comparaison de séquences génomiques d'organismes très éloignés évolutivement, notamment entre les mammifères et le poisson *Takifugu rubripes* ou Fugu, a permis de mettre en évidence l'existence, en dehors des régions codantes, de régions très conservées, et parfois distantes des régions promotrices (Marshall et al. 1994). Grâce (notamment) aux données issues du séquençage de génomes entiers, l'analyse plus poussée de certaines

de ces séquences, appelées CNS ou CNE (Conserved Non-coding Sequences/Elements), a mis en évidence leur association préférentielle avec des gènes cruciaux pour le développement. Ces régions conservées sont notamment très fortement associées avec des « enhancers » connus pour activer à longue distance la transcription de leurs gènes cibles. Toutefois, l'hypothèse selon laquelle le fort degré de conservation de ces enhancers ou CNS serait corrélé avec leur importance dans l'expression de gènes essentiels semble mise à mal par une étude récente, relançant le débat sur le rôle exact de ces séquences (Ahituv et al. 2007).

1.3.3.3. Outils d'analyse des régions régulatrices.

Les deux optiques possibles – recherche de TFBS et de régions conservées – sont des techniques hautement complémentaires. En effet, si la recherche de régions conservées est souvent un bon choix pour cibler de petites régions dans de larges régions génomiques, elles ne livrent aucune autre information en soi.

Inversement, les TFBS étant des séquences nucléotidiques relativement courtes et dégénérées, la probabilité de les rencontrer dans une séquence aléatoire est relativement élevée, et seulement une proportion très faible (<5%) des sites prédits sont réellement fixés par leur facteur, la chromatine n'étant pas nécessairement accessible dans le modèle étudié, ou le facteur de transcription n'étant pas nécessairement présent et actif. Or les facteurs de transcription n'agissent pas de façon isolée, mais en synergie ou en compétition avec d'autres cofacteurs, activateurs ou inhibiteurs, d'où l'intérêt de resserrer l'analyse des TFBS d'un promoteur, en recherchant non pas tous les sites potentiels (plusieurs centaines), mais uniquement des combinaisons de sites de fixation présents dans plusieurs séquences (ex : promoteurs de gènes co-régulés, de gènes orthologues ou paralogues, etc.) En effet, la probabilité que plusieurs séquences aient plusieurs sites en commun, avec plus ou moins le même agencement, est beaucoup plus faible que pour un site isolé.

De nombreux outils, tels que ceux proposés par le Comparative Genomics Center du Lawrence Livermore National Lab (<http://www.dcode.org/>) ou ceux – payants – proposés par la suite logicielle Genomatix, permettent ainsi de cibler des régions conservées et/ou de rechercher des combinaisons de TFBS entre plusieurs séquences homologues (figure 23), qu'il s'agisse de modèles connus (cofacteurs avérés décrits dans un type cellulaire donné) ou non. La présence de tels « modèles » dans les promoteurs de gènes ayant des patrons d'expression comparables permet

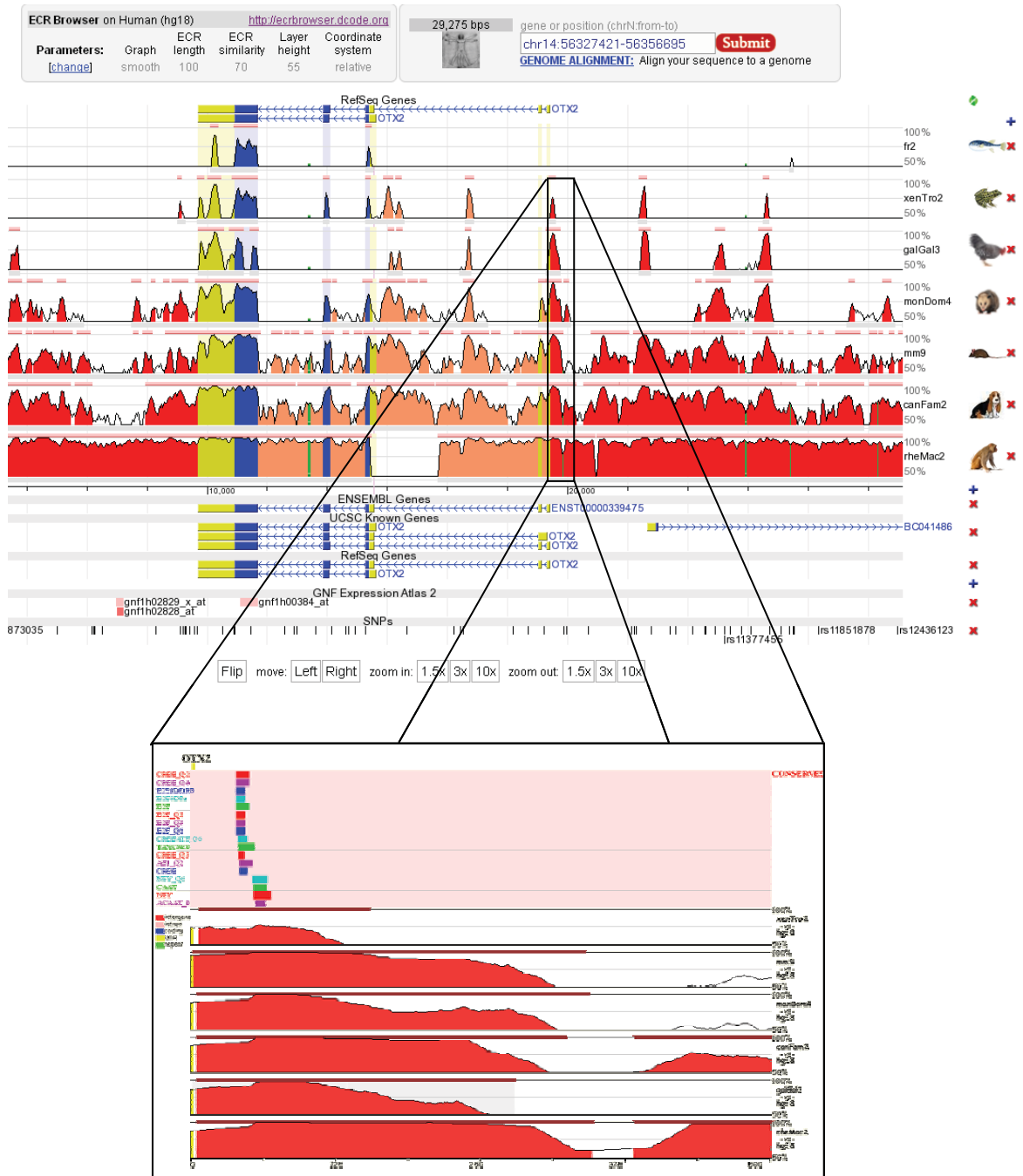


Figure 23. Outils du portail dcode.org. Recherche de régions conservées sur le locus du gène *OTX2*, et identification de TFBS dans le promoteur proximal.

d'identifier plus facilement de des candidats à valider expérimentalement (figure 24). Ces techniques in silico présentent malgré tout plusieurs limites qui sont abordées dans la partie Discussion.

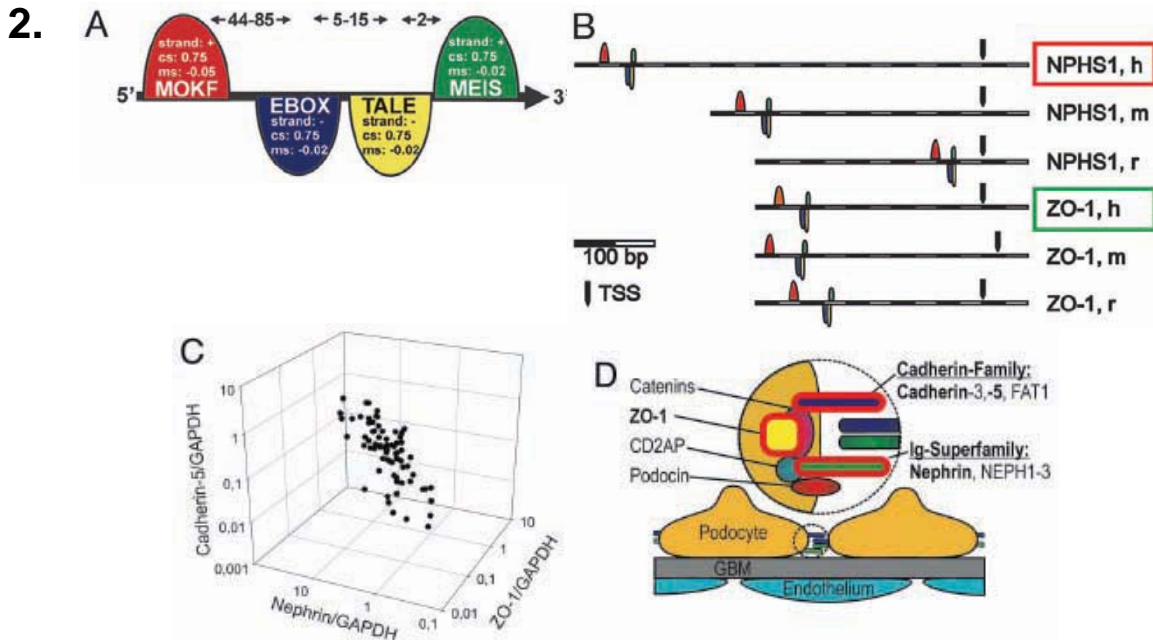


Figure 24. Analyse de promoteur et recherche de modèles de régulation (Cohen et al. 2006). L'analyse par comparaison inter-espèces du promoteur de *NPHS1* (néphrine) met en évidence une combinaison de quatre TFBS (A), retrouvée dans les promoteurs de plusieurs gènes dont *ZO-1* (B) et la cadhérine-5, corégulés d'après des analyses de microdissection de glomérules suivie de PCR quantitative (C), permettant d'identifier de nouvelles protéines de la barrière podocytaire (D).

2. La différenciation épidermique

2.1. Présentation générale

2.1.1. La peau

La peau est à la fois l'organe le plus externe et le plus étendu (en moyenne 1,8 m² de superficie pour un individu adulte). En tant que principale interface, il constitue avant tout une barrière qui protège l'organisme contre les principales agressions du milieu extérieur, telles que l'entrée de pathogènes ou de substances toxiques. Cette barrière s'oppose aussi aux pertes hydriques, potentiellement très importantes dans un environnement aérien. Les autres fonctions incluent notamment le sens du toucher et la thermorégulation. Deux tissus conjonctifs, l'hypoderme et le derme (d'origine mésodermique) et un tissu épithélial de revêtement, l'épiderme (d'origine ectodermique, tout comme les annexes épidermiques – follicules pileux, ongles, glandes sébacées et glandes sudoripares), composent le système tégumentaire (figure 25).

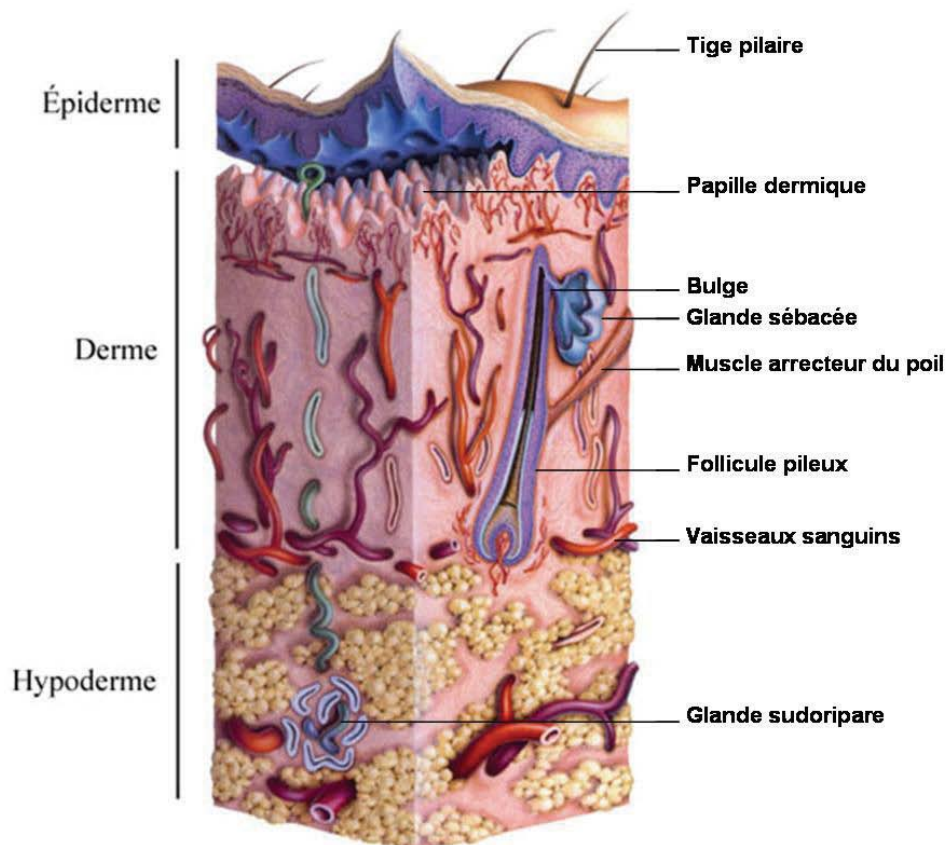


Figure 25. Représentation schématique de la peau.

2.1.2. L'hypoderme et le derme

Egalement nommé tissu adipeux sous-cutané, l'hypoderme est constitué de lobules graisseux séparés par un septum fibreux riche en collagène. Comme l'ensemble du tissu adipeux de l'organisme (dont il représente environ 50%), sa principale fonction est le stockage de lipides sous la forme de triglycérides. Le tissu adipeux est également considéré comme une glande endocrine, en raison de son rôle dans la production de nombreuses hormones (notamment leptine, adiponectine et résistine) ou la conversion des hormones stéroïdiennes. L'épaisseur de l'hypoderme ainsi que sa position périphérique lui confèrent des propriétés plus spécifiques de protection thermique et mécanique.

Le derme est un tissu conjonctif lâche comprenant peu de cellules, majoritairement des fibroblastes (ou fibrocytes) responsables de la synthèse de fibres de collagène (70% de la masse sèche du derme), de fibres élastiques (3%) et de substance fondamentale ou matrice extrafibrillaire (faible masse mais volume important dû aux glycosaminoglycans riches notamment en acide hyaluronique et capables de retenir près de 1000 fois leur volume d'eau). Les différentes fibres confèrent au tissu une grande résistance et élasticité. Le derme a un rôle important de soutien pour l'épiderme et ses annexes : les nombreux vaisseaux fournissent l'oxygène et les nutriments requis pour la survie des cellules épithéliales.

2.1.3. La jonction dermo-épidermique

A l'interface entre deux tissus, la jonction dermo-épidermique ou lame basale est une structure complexe synthétisée à la fois par les kératinocytes de la couche basale de l'épiderme et par les fibroblastes du derme superficiel. Cette structure régule les échanges entre derme et épiderme, et constitue un support pour l'ancrage des cellules de la couche basale de l'épiderme, assuré par les hémidesmosomes. En microscopie électronique, elle apparaît sous la forme de trois couches de densité différente (figure 26) : la *lamina lucida* (ou *lamina rara*), située immédiatement sous la membrane plasmique des cellules épidermiques et dans laquelle on peut distinguer les filaments d'ancrage des hémidesmosomes ; la *lamina densa*, plus profonde ; et la *sub-lamina densa*, correspondant à la partie la plus superficielle du derme dans laquelle sont situées les fibrilles d'ancrage assurant la liaison mécanique entre le derme et la lame basale et principalement constitués de collagène VII.

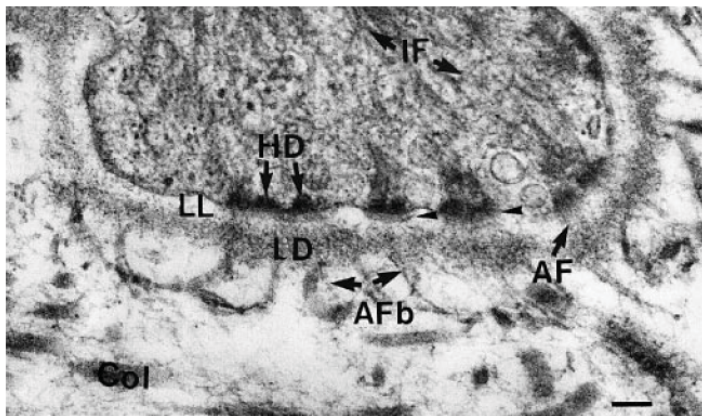
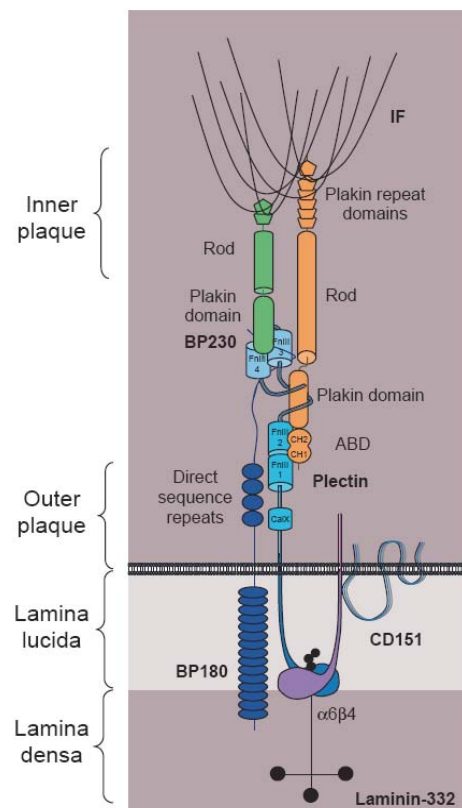


Figure 26. Jonction dermo-épidermique et hémidesmosomes. (Gauche). Microscopie électronique. HD : plaque cytoplasmique des hémidesmosomes ; IF : filaments intermédiaires de kératines ; LL : *lamina lucida* ; LD : *lamina densa* ; AF : filaments d'ancrage ; AFb : fibrilles d'ancrage ; Col : fibres de collagène du derme superficiel. (Borradori and Sonnenberg 1999). (Droite) Représentation schématique (Litjens et al. 2006).



2.1.4. L'épiderme

Les kératinocytes, cellules majoritaires de l'épiderme (~95%), composent un épithélium pluristratifié, pavimenteux et cornifié. En effet, ces cellules, en progressant des couches profondes vers la surface, se différencient progressivement, avant de subir un processus de mort cellulaire programmée, la cornification. Les autres types cellulaires, dont les proportions varient légèrement selon les territoires anatomiques, comprennent : les mélanocytes (~3%), dont le rôle est essentiel dans la protection contre les rayonnements ultraviolets ; les cellules de Langerhans (~2%), cellules présentatrices d'antigènes assurant la fonction de sentinelles du système immunitaire ; les cellules de Merkel (<1%), généralement proches de terminaisons nerveuses et jouant un rôle neurosensoriel.

2.1.5. Les annexes épidermiques

Les annexes épidermiques sont formées au cours du développement par invagination des cellules de l'ectoderme, et sont donc situées dans le derme et l'hypoderme ; elles comprennent les follicules pileux et les ongles (tous deux caractérisés par la présence de cellules cornifiées), ainsi que les glandes sébacées, les glandes sudoripares eccrines, et les glandes apocrines (faussement qualifiées de sudoripares).

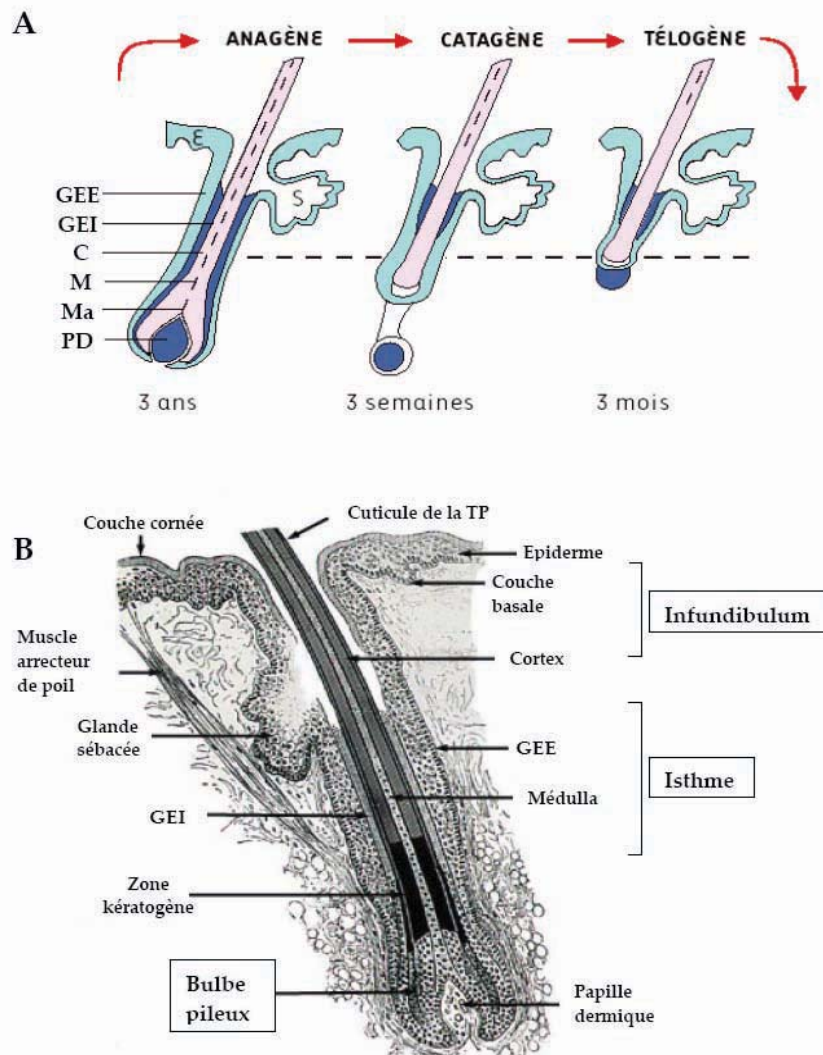


Figure 27. Unité pilo-sébacée et cycle pilaire.

(A) Représentation schématique des trois phases du cycle pilaire. (B) Représentation schématique de l'unité pilo-sébacée avec un follicule en phase anagène. E : épiderme ; GEE : gaine épithéliale externe ; GEI : gaine épithéliale interne ; C : cortex ; M : médulla ; Ma : matrice ; Pd : papille dermique ; S : glande sébacée ; TP : tige pilaire. D'après (Bernard 2002).

Le follicule pileux obéit à un cycle complexe de renouvellement, où une période de croissance (phase anagène) est suivie d'une régression (phase catagène), puis d'une période de latence (phase télogène), avant que la croissance lors du cycle suivant ne provoque l'expulsion du follicule précédant (figure 27). Les follicules pileux sont constitués de trois régions, de la surface vers l'intérieur : l'infundibulum (portion stable du follicule), l'isthme et le bulbe pileux (renouvelés à chaque cycle). Le bulbe est ancré autour de la papille dermique qui joue aussi un rôle de nutrition, et comprend à la fois des mélanocytes responsables de la pigmentation du poil ou du cheveu, et des cellules progénitrices impliquées dans le renouvellement des parties

supérieures. Les portions supérieures du follicule pileux se décomposent en gaine épithéliale externe, qui est en continuité directe avec l'épiderme interfolliculaire, et gaine épithéliale interne, composée de plusieurs feuillets dont les cellules subissent une cornification asynchrone selon les couches. Ce processus est comparable à celui qui a lieu au niveau des kératinocytes granuleux de l'épiderme interfolliculaire, bien que les programmes de différenciation diffèrent par l'expression de marqueurs spécifiques du follicule. Enfin, les cellules de la tige pileuse (cuticule, cortex et medulla) jouent un rôle dans les propriétés mécaniques du poil, et sont cornifiées au niveau de la cuticule où elles donnent naissance au poil proprement dit.

Un phénomène de cornification comparable a également lieu au niveau de la matrice des ongles, où les kératinocytes se différencient pour former la plaque de l'ongle (partie morte).

2.2. La différenciation kératinocytaire

2.2.1. Aspects morphologiques

L'épiderme est un tissu en perpétuel renouvellement : la desquamation des cornéocytes au niveau de la surface de l'épiderme est compensée par la prolifération des kératinocytes de la couche basale, ce qui assure l'homéostasie de l'épiderme. Le

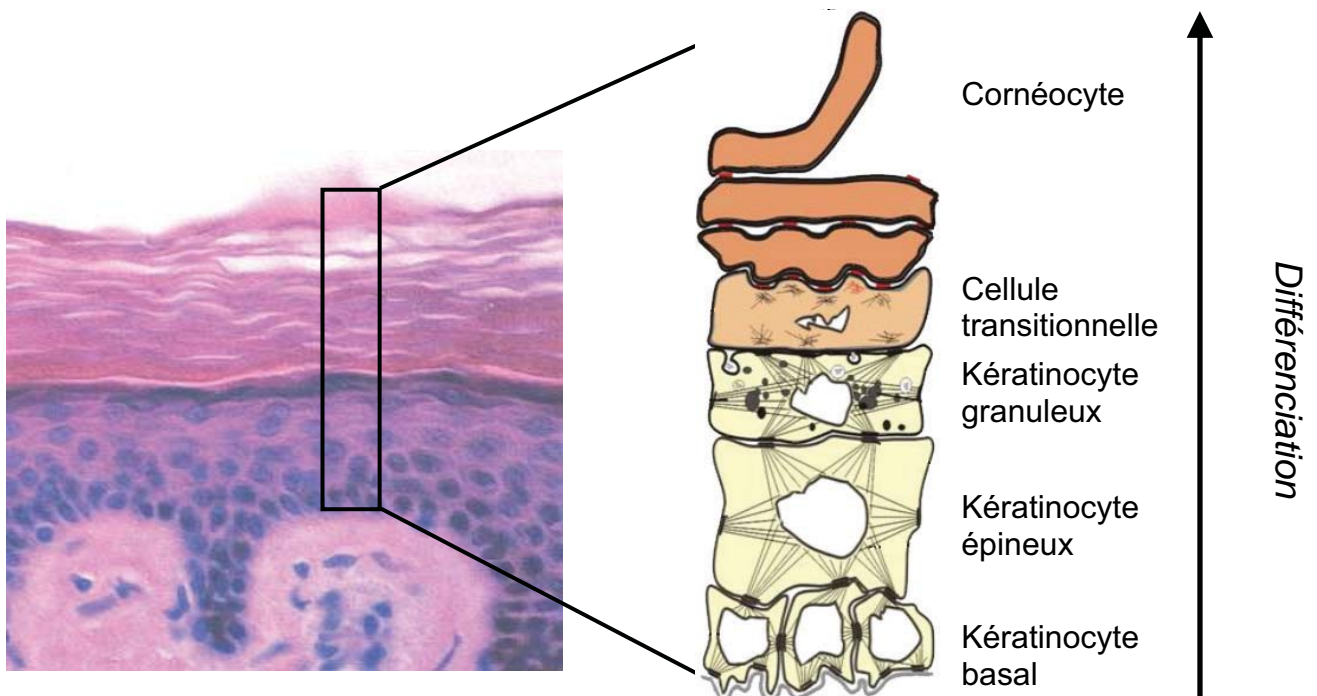


Figure 28. Coupe de peau humaine colorée à l'hématoxyline-éosine et représentation schématique de l'ultrastructure de l'épiderme (d'après un schéma original de Christian Vincent).

renouvellement de l'épiderme se déroule sur une période de 26 à 28 jours, correspondant pour moitié à la durée de l'exécution du programme de différenciation (couches vivantes) et pour moitié à la progression dans les couches mortes, jusqu'à la desquamation. Différents stades dans l'évolution du kératinocyte (prolifération, étapes précoces et tardives de la différenciation, cornification) sont morphologiquement distinguables en microscopie optique et électronique, et permettent de définir les différentes couches de l'épiderme (figure 28).

2.2.1.1. La couche basale (stratum basale)

Seule assise de l'épiderme contenant des kératinocytes prolifératifs, la couche basale de l'épiderme est constituée de cellules cubiques de petite taille, dont le noyau très basophile occupe une large place. Elles sont ancrées à la lame basale par les hémidesmosomes et reliées via les desmosomes. Les kératinocytes basaux ne forment pas en réalité une population homogène au niveau fonctionnel, puisqu'il est possible de distinguer des cellules souches, au fort potentiel prolifératif mais se divisant très rarement, des cellules progénitrices, majoritaires, se divisant activement pour garantir l'homéostasie, et potentiellement des cellules post-mitotiques sur le point de passer dans le compartiment suprabasal. Les mélanocytes, présents également dans la couche basale, produisent de la mélanine (eumélanine et phéomélanine) sous forme de granules, les mélanosomes. Le transfert aux kératinocytes voisins se fait par cytophagocytose de l'extrémité des dendrites contenant les mélanosomes (Singh et al. 2008), aboutissant à la formation d'une protection autour du noyau des kératinocytes.

2.2.1.2. La couche épineuse (stratum spinosum)

Première couche suprabasale de l'épiderme, la couche épineuse est constituée de cinq à dix assises cellulaires. Les caractéristiques cellulaires majeures sont le changement de morphologie (passant de cubique à polyédrique), l'augmentation importante du volume cytoplasmique, ainsi que le très grand nombre de desmosomes, conférant en microscopie optique un aspect épineux à ces cellules après rétraction des membranes lors de la préparation histologique. Au sein du cytoplasme, les desmosomes sont reliés par un très important réseau de filaments intermédiaires de kératines formant une cage périnucléaire particulièrement visible en microscopie électronique. Les desmosomes et filaments intermédiaires sont reliés via la plaque desmosomale formée de protéines adaptatrices, et jouent un rôle crucial dans la

résistance de l'épiderme, assurant la distribution de contraintes locales à l'ensemble du tissu et garantissant sa cohésion (avec le muscle cardiaque, l'épiderme est un des tissus ayant la plus forte densité en desmosomes). Dans les assises supérieures de la couche épineuse, des organites sécrétoires plurimembranaires, nommés corps lamellaires ou kératinosomes, peuvent être repérés en microscopie électronique, au niveau du pôle apical des kératinocytes. Dérivant de l'appareil de Golgi, les kératinosomes présentent un aspect de vésicules membranaires arrondies, d'une taille de 0,2 à 0,3 μm , au contenu riche en bicouches lipidiques formant des saccules aplatis. Toutefois, cet aspect de vésicules isolées et plurimembranaires pourrait être dû à des artefacts de préparation ; en effet, sur des cryocoupes d'épiderme « vitrifié » observées en microscopie cryoélectronique, les corps lamellaires n'apparaissent plus comme des vésicules discrètes, mais comme un réseau tubulaire branché et non lamellaire (Norlen 2001; Norlen et al. 2003).

2.2.1.3. La couche granuleuse (stratum granulosum)

La couche granuleuse comprend 2 à 5 assises de cellules, et celles-ci ayant une morphologie aplatie, la couche granuleuse est bien moins épaisse que la couche épineuse. Les kératinocytes granuleux tirent leur nom de la présence, au sein de leur cytoplasme, de nombreux grains de kératohyaline, correspondant à des agrégats insolubles de deux protéines, la loricrine (L-granules, d'aspect arrondi) et la profilaggrine (F-granules, irréguliers et plus gros), produites en grandes quantités par ces cellules. Les kératinocytes granuleux constituent les dernières cellules vivantes de l'épiderme avant la cornification qui s'accompagne de la destruction du noyau et de la disparition de toute activité transcriptionnelle. Le nombre de kératinosomes (ou la densité du réseau tubulaire branché, selon le modèle considéré) augmente fortement au pôle apical de la dernière assise, et la sécrétion de leur contenu lipidique dans l'espace extracellulaire à l'interface couche granuleuse / couche cornée permet la formation de *lamellae* lipidiques, cruciaux pour l'imperméabilité de la couche cornée, et donc la fonction barrière de l'épiderme.

2.2.1.4. La couche cornée (stratum corneum)

La couche cornée est formée d'un empilement de 15 à 20 assises de kératinocytes morts d'aspect très aplati, qui constituent le stade ultime de la différenciation kératinocytaire, prenant le nom de cornéocytes. Le noyau et les organites principaux étant dégradés au cours de la cornification, les cornéocytes ne

présentent pas d'activité transcriptionnelle ou traductionnelle, voire métabolique. Les cornéocytes sont particulièrement cohésifs dans les assises profondes de la couche cornée (*stratum compactum*), puis de moins en moins jointifs (*stratum disjunctum*), avant de se détacher complètement en surface sous l'effet des contraintes mécaniques.

La cornification se traduit par la mise en place de nouvelles structures bien visibles en microscopie électronique. La membrane plasmique est remplacée par une coque protéique rigide, l'enveloppe cornée, et le cytoplasme est réduit à une matrice fibreuse dense amorphe constituée de kératines réarrangées en macrofibrilles. L'espace intercornéocytaire est rempli de lipides organisés en bicouches, les *lamellae* lipidiques. Les desmosomes prennent le nom de cornéodesmosomes tandis que la plaque cytoplasmique est intégrée à l'enveloppe cornée et le cœur intercellulaire se densifie. Au cours de la progression des cornéocytes vers la surface, les cornéodesmosomes sont progressivement dégradés, les premiers à disparaître étant ceux situés aux pôles apicaux et basaux des cornéocytes ; la dégradation des cornéodesmosomes latéraux permet la libération du cornéocyte, correspondant à la desquamation.

L'épaisseur de la couche cornée est extrêmement variable selon le territoire anatomique et les contraintes mécaniques subies, passant de quelques dizaines de μm pour les paupières, à quelques millimètres pour les régions palmoplantaires ; cet épaissement de la couche cornée dans l'épiderme palmoplantaire s'accompagne d'une persistance prolongée de l'ensemble des cornéodesmosomes jusque dans les dernières assises cornéocytaires. Une autre caractéristique de cet épiderme est l'existence d'une couche claire (*stratum lucidum*) située juste avant la couche cornée et constituée de cellules transitionnelles en cours de cornification, alors que ces mêmes cellules apparaissent dispersées à l'interface couche granuleuse / couche cornée dans les autres régions anatomiques.

2.2.2. Aspects fonctionnels

2.2.2.1. Homéostasie de l'épiderme

L'épiderme est soumis à de nombreuses forces (étirement, cisaillement, abrasion...), et même en absence de traumatisme nécessitant une cicatrisation, un renouvellement permanent du tissu est indispensable pour que la fonction protectrice soit maintenue. Afin que l'homéostasie du tissu soit conservée, il faut que l'auto-renouvellement des kératinocytes dans la couche basale contrebalance exactement le détachement des cornéocytes en surface. A l'heure actuelle, le ou les mécanismes couplant prolifération et desquamation restent obscurs, même si la description de ces deux processus a considérablement progressé ces dernières années.

2.2.2.1.1. Couche basale et auto-renouvellement

La couche basale est le lieu où les kératinocytes assurent l'auto-renouvellement de l'épiderme, les kératinocytes suprabasaux étant des cellules post-mitotiques engagées dans un processus de différenciation terminale irréversible. Les kératinocytes basaux se distinguent des kératinocytes suprabasaux par l'expression d'un certain nombre de marqueurs spécifiques : des protéines impliquées dans la prolifération cellulaire (ex : cycline D1, PCNA), mais aussi des constituants de la lame basale (ex : la laminine-5, ligand des intégrines, sécrétée dans la *lamina lucida*), des hémidesmosomes (intégrines $\alpha 6\beta 4$, le collagène de type XVII $\alpha 1$ COL17A1 aussi nommé BP180 ou Bullous Pemphigoid Antigen 2 BPAG2), du cytosquelette (avec le couple de kératines K5/K14).

La compréhension des mécanismes assurant l'auto-renouvellement de l'épiderme reste sujet à débat et fait de ce tissu un modèle d'étude de premier choix pour les cellules souches. Les méthodes d'identification des cellules souches étant essentiellement fonctionnelles, elles s'avèrent difficilement compatibles avec l'étude *in vivo* chez l'homme. Pour cette raison, l'étude du renouvellement épidermique et des cellules souches s'est surtout effectuée sur des modèles murins.

Les cellules souches épidermiques y ont été initialement identifiées et étudiées grâce à leur très faible taux de division cellulaire comparé aux cellules progénitrices. Ainsi, les cellules souches sont capables de retenir un marquage à la thymidine tritiée ou au BrdU (Bickenbach 1981), ou de conserver plus longtemps (jusqu'à huit mois) que des cellules fortement prolifératives une histone H2B-GFP exprimée transitoirement (Tumbar et al. 2004) (figure 29). L'identification de ces « label-

retaining cells » (LRC) a permis de localiser la niche majeure que constitue le « bulge », un renflement cellulaire en périphérie du follicule pileux, près du site d'attachement des muscles arrecteurs du poil, en dessous des glandes sébacées (Cotsarelis et al. 1990).

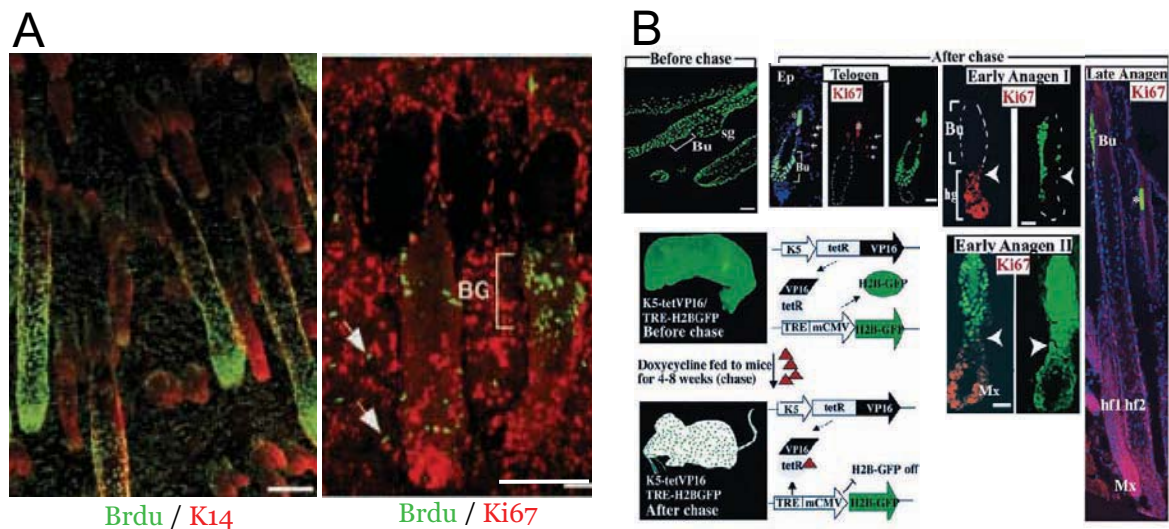


Figure 29. Identification des LRC dans l'épiderme murin in vivo. (A) Des souris reçoivent des injections répétées de BrdU, puis l'épiderme de la queue est prélevé après 2 jours (gauche) ou 70 jours (droite) et marqué avec des anticorps anti-BrdU (vert). Le marquage intense du bulbe ne persiste pas, mais des LRC sont visibles au niveau du bulge (BG) ainsi que dans l'épiderme interfolliculaire (flèches) (Braun et al. 2003). (B) Des souris transgéniques K5-tetVP16 / TRE-H2BGFP sont soumises à un traitement prolongé à la doxycycline pour éteindre l'expression de l'histone couplée à la GFP. Seules les LRC conservent un niveau de fluorescence élevé, confirmant leur présence au niveau du bulge dans toutes les phases du cycle pileux (Tumbar et al. 2004).

Les cellules souches présentes dans le bulge sont en outre capables de se différencier en plusieurs lignages cellulaires : ainsi, une LRC unique isolée du bulge par FACS (H2B-GFP⁺), cultivée puis greffée sur une souris *nude* est capable de



reformer des follicules pileux (Blanpain et al. 2004) (cf. ci-contre). Toutefois, cette pluripotence ne semble pas être

requis dans un contexte physiologique, les cellules souches impliquées dans l'homéostasie de l'épiderme interfolliculaire et celles requises pour la génération des follicules pileux étant distinctes (Levy et al. 2005), les cellules du bulge n'étant mobilisées hors du follicule pileux qu'au cours de la cicatrisation (Ito et al. 2005). Au moyen de souris exprimant au cours de l'embryogenèse la β -galactosidase dans les cellules de la placode donnant naissance au follicule pileux (grâce à l'excision d'une

cassette de terminaison de transcription au moyen d'une recombinaise sous dépendance du promoteur du gène *shh*), Levy et collaborateurs ont ainsi pu montrer que la descendance de ces cellules se retrouvait dans la majorité du follicule pileux, mais jamais dans l'épiderme interfolliculaire en absence de traumatisme (figure 30). Ce résultat confirme des données précédentes montrant que des cellules du bulge suivies grâce à leur expression de la β -galactosidase ne contribuent que très rarement et marginalement au renouvellement de l'épiderme inter-folliculaire (Morris et al. 2004). L'idée que le renouvellement de l'épiderme se fait à partir de cellules souches présentes dans la couche basale de l'épiderme interfolliculaire est en accord avec le fait que l'absence de follicules pileux dans l'épiderme palmoplantaire n'empêche pas le renouvellement du tissu, ainsi que la présence de LRC dans la couche basale de l'épiderme interfolliculaire murin (Braun et al. 2003).

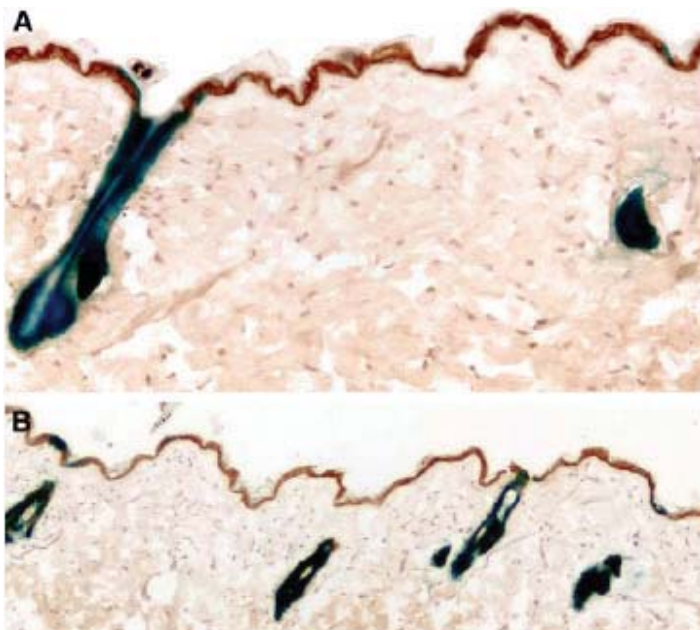


Figure 30 : activité β -galactosidase dans la peau de souris *ShhCre;R26R* (Levy et al. 2005). Seuls les follicules pilo-sébacés sont marqués (cellules bleues). Les couches suprabasales sont marquées en brun avec un anticorps anti-kératine 1. (A) : fort grossissement ; (B) : faible grossissement.

Ces méthodes fonctionnelles ont permis l'identification d'un petit nombre de marqueurs moléculaires. Citons, pour l'homme, une surexpression des kératines 15 et 19 et de l'intégrine β 1 importante pour le maintien à l'état indifférencié (Lyle et al. 1998), de l'intégrine α 6 associée à l'absence du récepteur de la transferrine CD71 (Li et al. 1998), ou encore de la forme longue du facteur de transcription p63 impliquée dans le maintien des cellules souches (Pellegrini et al. 2001) - pour revue (Tiede et al. 2007). De récentes études basées sur le tri de cellules d'épiderme murin selon le marqueur CD34 spécifique du bulge (Trempeus et al. 2007), ou d'épiderme humain avec exclusion du marquage Hoechst 33342 (Larderet et al. 2006), et couplées à l'emploi de puces à ADN, ont permis d'identifier de nouveaux gènes dont l'expression

est augmentée ou diminuée dans les cellules souches épidermiques. C'est le cas notamment des facteurs de transcription Id1 et Id2 ou de la famille des phosphatases Dusp (Larderet et al. 2006), dont la fonction exacte dans les cellules souches reste à encore définir. Bien qu'il n'y ait pas une rupture très marquée entre les profils d'expression des cellules souches et progénitrices, ces descriptions permettent une meilleure compréhension de la régulation du potentiel prolifératif dans les cellules souches.

Dès la fin des années 60, plusieurs travaux ont mis en évidence l'organisation en colonne des cellules de l'épiderme murin, pour aboutir à la démonstration que les cellules s'organisaient en véritables unités prolifératives (Epidermal Proliferative Units - EPU) dans lesquelles l'ensemble des cellules suprabasales et des cornéocytes dérivent de cellules progénitrices situées à la base de la colonne, dans la couche basale (Potten 1974).

Chez l'homme, l'impossibilité de visualiser les LRC ou les EPU *in vivo* a imposé l'étude indirecte du potentiel clonogénique de cellules isolées (Barrandon and Green 1987) ; ceci a notamment permis de confirmer la présence de cellules à fort potentiel prolifératif dans le follicule pileux, mais dans une région située plus bas que le bulge, lequel est généralement impossible à distinguer morphologiquement chez l'homme (Rochat et al. 1994). Concernant l'épiderme interfolliculaire, le modèle le plus couramment admis pour l'organisation de la couche basale des EPU décrit la présence d'une minorité de cellules souches isolées, se divisant très rarement, entourée de cellules d'amplification transitoire (« TA » cells). Les cellules TA, issues de la division asymétrique d'une cellule souche, se divisent un nombre limité de fois (< de 15 cycles) avant de s'engager irrémédiablement dans la différenciation en gagnant la couche épineuse (figure 31A). L'observation des EPU dans l'épiderme humain n'a pu être réalisée qu'assez récemment, au moyen de greffes de biopsies de prépuce sur des souris *nude*, suivies de transfection par des lentivirus contenant le gène de la GFP (Ghazizadeh and Taichman 2005) (figure 32). Après 28 semaines, les cellules exprimant la GFP sont organisées en colonnes typiques. La largeur de ces colonnes semble variable suivant la localisation anatomique (creux ou sommet des papilles dermiques...).

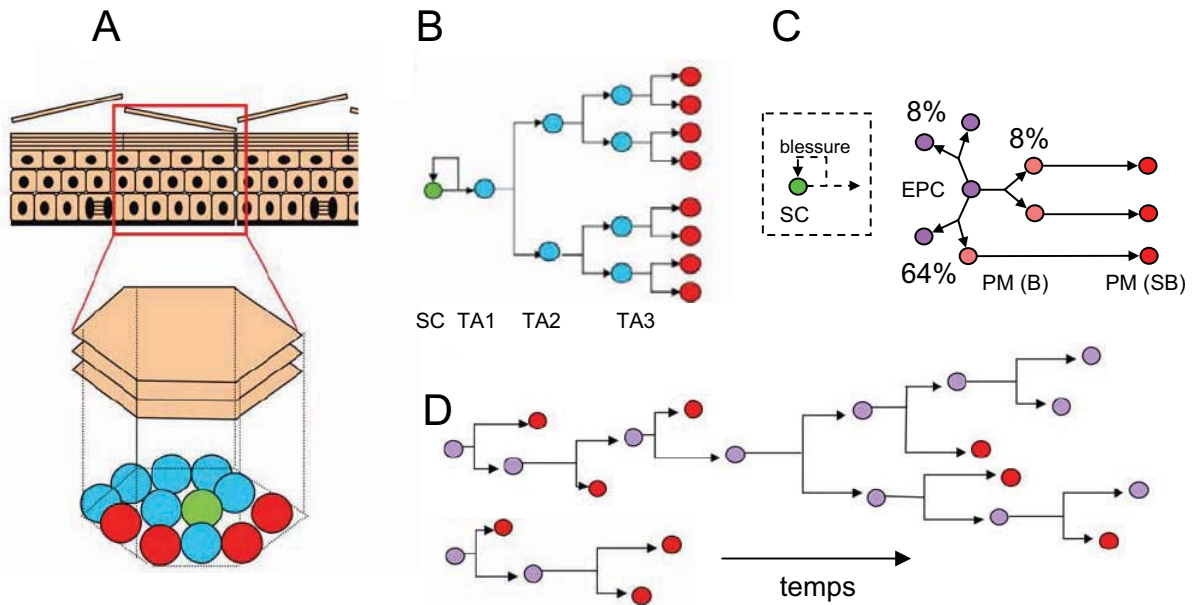


Figure 31. Renouvellement de l'épiderme. (A) Modèle d'organisation des cellules épidermiques en unités (EPU) proposé par Potten. (B) Les cellules à la base des EPU se répartissent en cellules souches (SC) se divisant rarement, en cellules d'amplification transitoire (TA1-3), et en cellules post-mitotiques (PM) passant alors dans le compartiment suprabasal. (C) Dans le modèle proposé par Clayton et collaborateurs, les cellules souches épidermiques n'interviennent pas dans un contexte physiologique ; seules des cellules progénitrices (EPC) au potentiel prolifératif non limité donneraient naissance à zéro, une ou deux cellules post-mitotiques basales (PM (B)) dans des proportions fixes (8-64-8). La vitesse du renouvellement de l'épiderme ne dépendrait alors que du taux de division des cellules progénitrices et de la vitesse du transfert vers le compartiment suprabasal (PM (SB)). (D) Dans ce modèle, les différences de potentiel clonogénique observées *in vitro* s'expliqueraient par la probabilité non nulle d'un arrêt précoce de la prolifération (cas du bas), les cellules progénitrices ne se renouvelant pas dans 8% des divisions.

Toutefois, une étude récente suggère que le modèle en EPU, avec cellules souches et TA, n'explique pas la dynamique réelle du renouvellement de l'épiderme, et propose un nouveau paradigme dans lequel les cellules souches ne joueraient pas de rôle dans des conditions physiologiques. Des cellules progénitrices (au potentiel prolifératif non limité, contrairement au modèle des cellules TA) seraient responsables à elles seules de l'homéostasie, se divisant essentiellement sur un mode asymétrique pour donner une cellule post-mitotique et une progénitrice (84% des divisions), mais aussi plus marginalement deux progénitrices (8%) ou deux cellules post-mitotiques (8%) (Clayton et al. 2007) (figure 31B-C). Ce modèle présente notamment l'intérêt d'expliquer certaines observations, comme la croissance de la taille des EPU au cours de la vie d'une souris.

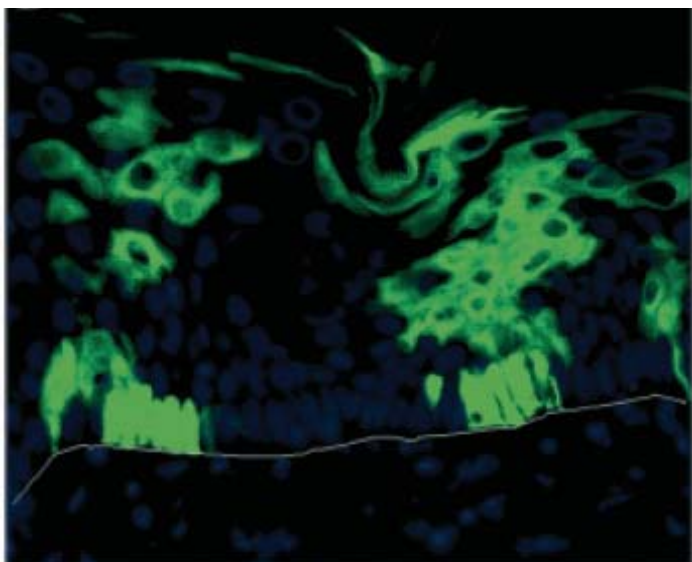


Figure 32. Distribution de la fluorescence dans des unités de prolifération d'épiderme humain greffé chez la souris nude et transfecté par un lentivirus exprimant la GFP (Ghazizadeh and Taichman 2005).

Deux excellentes revues récentes (Jones and Simons 2008; Blanpain and Fuchs 2009) résument bien la complexité du modèle, où plusieurs populations de cellules souches aux potentiels variables occupent des niches distinctes et jouent des rôles probablement différents (voire aucun rôle) suivant le contexte (homéostasie tissulaire, cicatrisation...).

2.2.2.1.2. *Couche cornée et desquamation*

De façon couplée avec le renouvellement des cellules dans la couche basale, les cornéocytes les plus superficiels se détachent régulièrement au cours du processus de la desquamation (du latin *desquamare*, écailler, écorcer, enlever ce qui est autour). Correspondant à l'étape ultime du programme de différenciation, la desquamation se traduit par l'exfoliation d'environ 1000 cornéocytes / cm² / h dans des conditions normales (Roberts and Marks 1980). Les cornéodesmosomes constituent les structures jonctionnelles liant entre eux les cornéocytes, et sont donc à l'origine de la cohésion de la couche cornée. Leur structure dérive de celle des desmosomes mais est modifiée par la cornification. Au cours de leur progression dans cette ultime couche, les cornéocytes voient ces attaches progressivement dégradées, comme en attestent les données de microscopie électronique : la partie extracellulaire des cornéodesmosomes apparaît moins dense et progressivement dégradée, tandis que les lamellae lipidiques entourent progressivement la partie centrale (Fartasch et al. 1993) (figure 33).

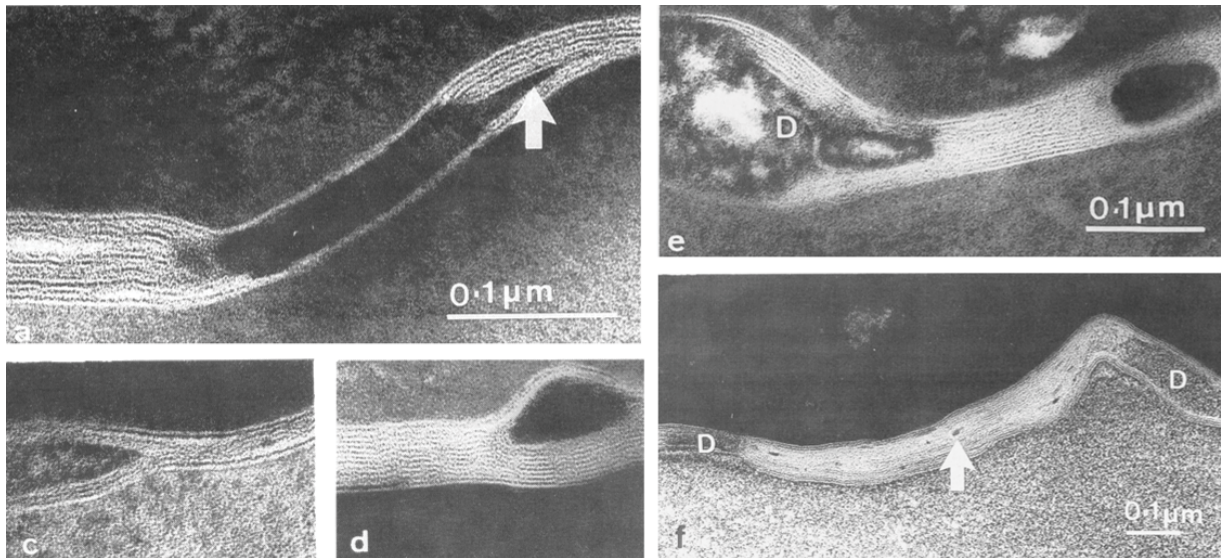


Figure 33. Dégradation progressive des cornéodesmosomes au cours de la desquamation (Fartasch et al. 1993). Le cœur intercornéocytaire, dense aux électrons, est progressivement dégradé, et envahi par les lamellae lipidiques. La densité du cœur diminue fortement, puis seules de rares zones denses isolées persistent au milieu des lamellae.

Ce phénomène de dégradation se produit dans un premier temps au niveau des cornéodesmosomes basaux/apicaux dans la partie inférieure de la couche cornée, et est à l'origine des différences morphologiques entre *stratum compactum* et *stratum disjunctum*. Puis un clivage plus tardif a lieu au niveau des cornéodesmosomes latéraux (Skerrow et al. 1989; Chapman and Walsh 1990), ce qui entraîne alors la libération du cornéocyte. La protéolyse des cornéodesmosomes passe par la dégradation de leur protéines transmembranaires et extracellulaires (cadhérines desmosomales de la famille des desmocollines et desmogléines – cf. partie 2.2.2.2.3 - ainsi que la cornéodesmosine (Lundstrom and Egelrud 1990; Serre et al. 1991; Lundstrom et al. 1994; Guerrin et al. 1998; Simon et al. 2001)). Cette dégradation implique plusieurs protéases présentes dans l'espace intercornéocytaire, leur action étant finement régulée des inhibiteurs de protéases. La cinétique de la desquamation (donc également l'épaisseur de la couche cornée et éventuellement la fonction barrière elle-même) dépend de la balance entre les protéases et les inhibiteurs, tous sécrétés par les kératinocytes granuleux sous forme active ou non.

Une étude en immunofluorescence et en immunomicroscopie électronique montre que plusieurs molécules sécrétées, telles que la kallikréine 7 (KLK7, ou SCCE, Stratum Corneum Chymotryptic Enzyme), la cornéodesmosine (CDSN), la cathepsine

D (CTSD) ou les glucosylcéramides, sont transportées séparément depuis l'appareil de Golgi vers le pôle apical des kératinocytes granuleux par le système sécrétoire spécifique, les kératinosomes (Ishida-Yamamoto et al. 2004). Toutes les cargaisons testées dans cette étude semblent acheminées indépendamment sous la forme d'agrégats vers la membrane plasmique. Dans le cas de KLK7, dont CDSN constitue un substrat au cours de la desquamation, cette séparation pourrait avoir un rôle fonctionnel en empêchant une protéolyse prématurée de la CDSN. Toutefois, certaines enzymes transportées par les kératinosomes, comme les cathepsines ou certaines enzymes du métabolisme des lipides, ont dans d'autres tissus une localisation lysosomiale (Madison et al. 1998). Le mécanisme exact de biogenèse des kératinosomes reste en fait partiellement incompris.

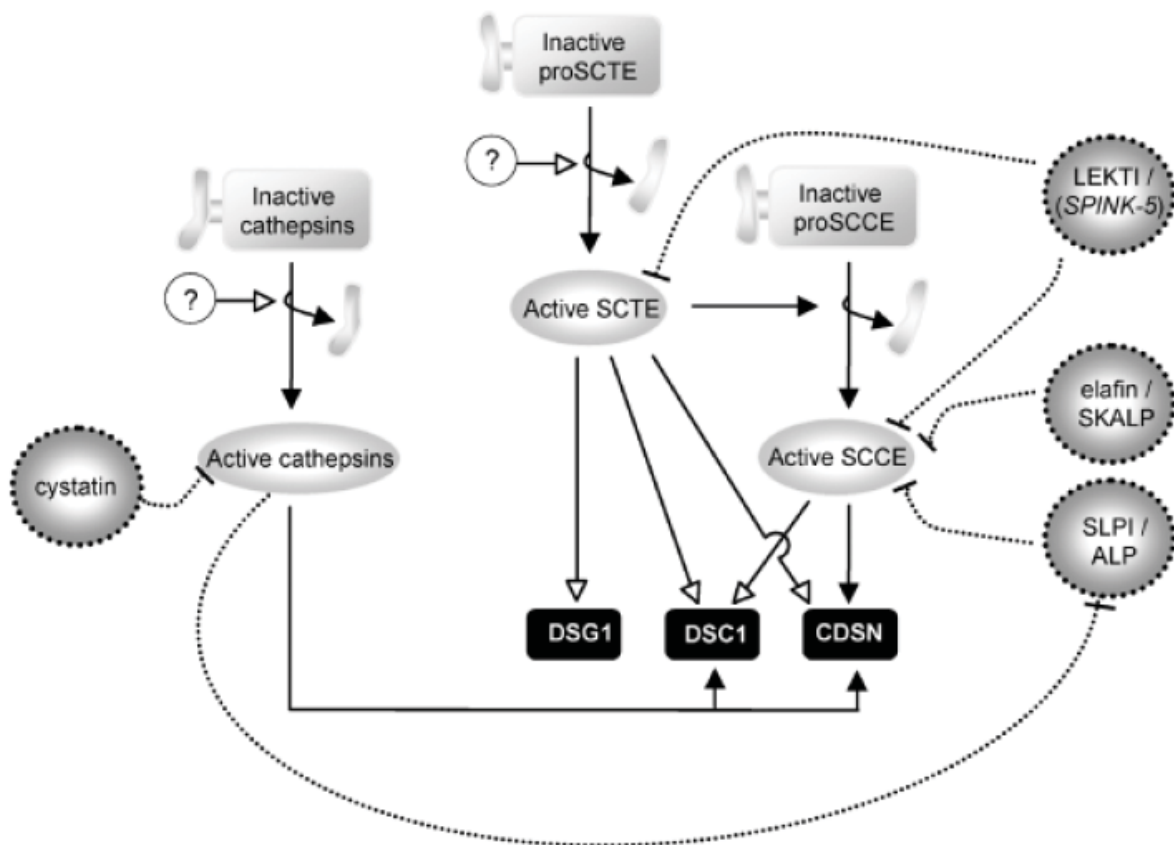


Figure 34. Modèle de régulation de la desquamation (Caubet et al. 2004).

Parmi la dizaine de protéases sécrétées identifiées dans la couche cornée, deux protéases à sérine, la KLK5/SCTE (Ekholm and Egelrud 2000) et la KLK7/SCCE (Hansson et al. 1994) semblent jouer un rôle majeur. *In vitro*, dans des conditions de pH acide proches de celles de la couche cornée, les composants du cornéodesmosome sont des substrats pour la SCTE. La SCCE est quant à elle capable de cliver

efficacement la CDSN et la desmocolline-1 (DSC1) (Caubet et al. 2004). S'appuyant sur ces données et celles de la littérature, une représentation schématique des acteurs principaux intervenant dans la régulation de la desquamation a été proposée (figure 34).

Les cathepsines sont des protéases lysosomiales. Chez l'homme, une mutation de la cathepsine C (CTSC) est responsable du syndrome de Papillon-Lefèvre, une kératodermie palmoplantaire associée à une périodontopathie (OMIM #245000). Peu d'études sont disponibles quant au rôle de cette protéase dans l'épiderme en général et dans la desquamation en particulier. Sa fonction dans les lymphocytes T est toutefois bien documentée. Elle y active les granzymes A et B par protéolyse des proformes. Le granzyme B, sécrété par les kératinocytes, intervient dans la défense antimicrobienne (Berthou et al. 1997). On peut spéculer sur un rôle de CTSC dans l'activation du granzyme B ou celle d'autres protéases jouant un rôle dans la différenciation épidermique (Meade et al. 2006). L'inhibition de la cathepsine D diminue d'un tiers le détachement des cornéocytes à pH acide dans un modèle de desquamation de couche cornée plantaire *in vitro* (Horikoshi et al. 1999). Cette protéase à acide aspartique est détectée en immunomicroscopie électronique dans l'espace intercornéocytaire où elle colocalise avec les cornéodesmosomes (Igarashi et al. 2004). Ces arguments sont en faveur de l'implication de la CTSD dans la desquamation. En plus des cathepsines C et D, les cathepsines V et L semblent impliquées respectivement dans la dégradation des cornéodesmosomes et la maturation des transglutaminases (Zeeuwen et al. 2007).

L'activité protéolytique responsable de la desquamation est régulée par la présence d'un grand nombre d'inhibiteurs de protéases. On peut citer SLPI (Secretory Leukocyte Protease Inhibitor) et SKALP (Skin derived AntiLeukoProteinase, également désigné élafine), deux protéines de petite taille dont les spectres d'inhibition sont complémentaires (Heinzel et al. 1986; Wiedow et al. 1990). SLPI présente une meilleure inhibition vis-à-vis des protéases de la famille de la chymotrypsine telles que SCCE. SKALP est au contraire plus efficace comme inhibiteur de protéases de la famille de la trypsine telles que SCTE. Alors que ces deux inhibiteurs sont très faiblement détectés dans l'épiderme sain, ils sont fortement surexprimés dans les couches différenciées de l'épiderme psoriasique et lors de la cicatrisation (Schalkwijk et al. 1990; Ashcroft et al. 2000). Un autre inhibiteur de protéases, la cystatine M/E, produit du gène *Cst6*, inhibe l'activité de plusieurs

protéases aux cibles et localisations variées, notamment les cathepsines V et L (Zeeuwen et al. 2007). Des mutations nulles de *Cst6* sont responsables chez la souris du phénotype *ichq*, caractérisé par des anomalies de la cornification et de la desquamation et une mortalité néonatale (Zeeuwen et al. 2002). Enfin, LEKTI (LymphoEpithelial Kazal-Type-related Inhibitor), produit du gène *SPINK5* (Serine Protease INhibitor, Kazal-type 5) (Magert et al. 1999), joue un rôle crucial dans la régulation de la desquamation. Ainsi, des mutations du gène codant cet anti-trypsine sont responsables du syndrome de Netherton (Chavanas et al. 2000), génodermatose caractérisée par une ichtyose érythrodermique, une anomalie spécifique des cheveux (cheveux bambous) et une atopie sévère avec un taux d'IgE élevé (OMIM #256500). L'inactivation de ce gène par mutagenèse insertionnelle reproduit le phénotype de la maladie chez la souris (Yang et al. 2004) (figure 35). Les animaux présentent une fragilité desmosomale associée à une protéolyse prématurée de la CDSN. Deux autres études mettent en évidence d'une part l'hyperactivité des protéases SCCE et SCTE et la protéolyse anormale de la desmoplakine et de la desmogléine 1 (Descargues et al. 2005), et d'autre part le rôle du pH (lequel varie dans la couche cornée, passant d'un pH neutre en profondeur à acide en surface) dans la dissociation entre LEKTI et ses cibles (Deraison et al. 2007). Ces résultats font apparaître LEKTI comme un des régulateurs clé du processus de desquamation. LEKTI semble avoir également de façon indirecte un rôle dans la formation des lamellae lipidiques : chez les patients

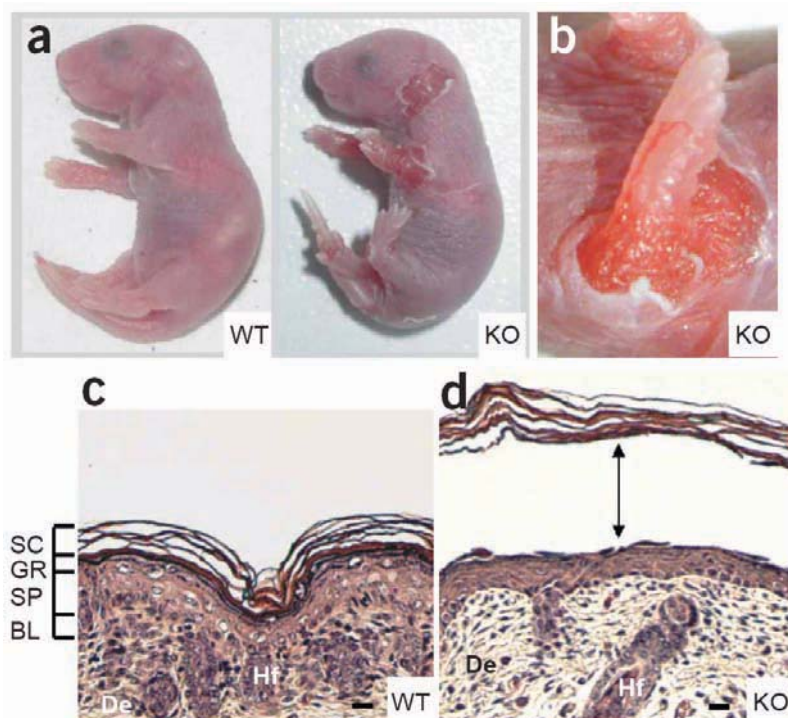


Figure 35. Trouble majeur de la desquamation chez des souris inactivées pour *Spink5* (Descargues et al. 2005). Les nouveaux nés présentent de larges décollements épidermiques (a), notamment aux zones de frottements (b), correspondant à un clivage entre la couche granuleuse (GR) et la couche cornée (SC) (c-d).

souffrant du syndrome de Netherton, les phénotypes les plus sévères associés à une quasi absence de LEKTI présentent une très forte activité des sérines protéases, et une désorganisation massive des lamellae lipidiques qui serait directement liée à la dégradation par ces protéases dérégulées de la β -glucocérébrosidase, une enzyme maturant les lipides inter-cornéocytaires – voir paragraphe 2.2.2.2.1.1. De façon intéressante, en plus de l'expression de LEKTI, un autre inhibiteur de protéase de la même famille, produit du gène *SPINK9* (Brattsand et al. 2009), a récemment été identifié dans l'épiderme palmo-plantaire, et pourrait expliquer en partie les différences observées au niveau de ces territoires où la couche cornée est plus épaisse et les cornéocytes persistent plus longtemps (figure 36).

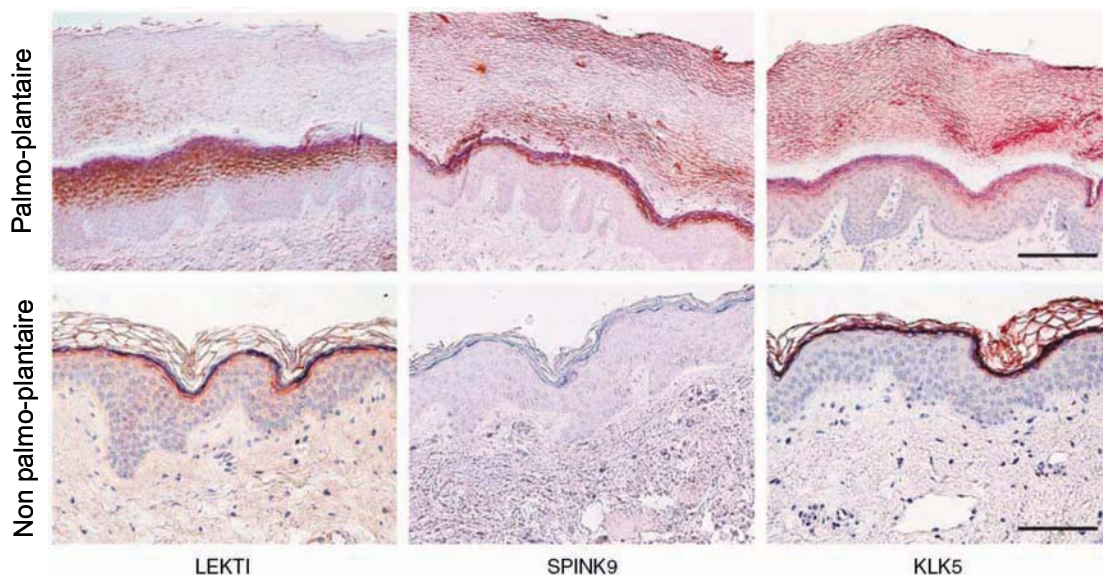


Figure 36. Immunodétection de LEKTI (SPINK5), SPINK9 et KLK5 sur des coupes d'épiderme palmo-plantaire et non palmo-plantaire (Brattsand et al. 2009).

2.2.2.2. Fonction barrière de l'épiderme

La couche granuleuse participe largement à la formation de la barrière épidermique en tant que dernière couche ayant une activité transcriptionnelle et traductionnelle, responsable de la production de la quasi-totalité des acteurs de la cornification (protéines structurales comme enzymes) mais aussi de la desquamation.

La couche cornée, bien que composée de cellules mortes, assure la majorité des aspects de la fonction barrière de l'épiderme, grâce aux remaniements importants de

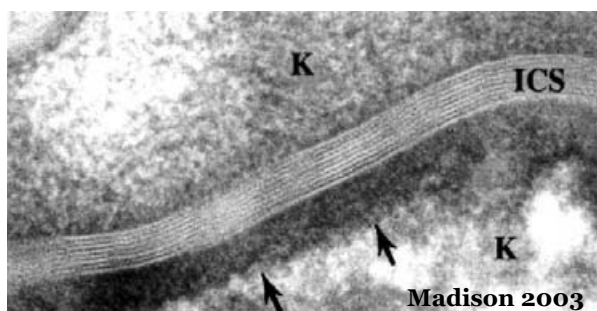
la cornification et aux nouvelles propriétés mécaniques et physico-chimiques qui en émergent. Bien qu'étant un processus de mort cellulaire programmée, la cornification se distingue de l'apoptose par la persistance des cellules mortes, phénomène également constaté au cours d'une autre mort cellulaire programmée, la « dénucléation » des cellules épithéliales du cristallin – cf. revue (Yan et al. 2006), qui là aussi permet l'émergence de nouvelles propriétés – ici optiques – que les cellules vivantes ne peuvent assurer.

Ainsi, la cornification ne se résume pas à la simple disparition du noyau et autres organites, mais permet surtout de former une couche de cellules imperméable, de par la formation des lamellae lipidiques intercornéocytaires, mais aussi extrêmement résistante et cohésive, de par la formation de nouvelles structures (matrice fibreuse et enveloppe cornée) ou via le renforcement de structures existantes (transformation des desmosomes en cornéodesmosomes).

2.2.2.2.1. *Imperméabilité*

L'épiderme constitue une double barrière dans le sens où son imperméabilité concerne à la fois la sortie de substances (eau, ions, etc.), mais aussi l'entrée (substances toxiques, etc.). Deux structures distinctes, les lamellae lipidiques intercornéocytaires et les jonctions serrées de la couche granuleuse, sont impliquées conjointement dans cette double barrière, sans que les liens fonctionnels entre ces deux structures soient clairement établis.

2.2.2.2.1.1. Lipides intercornéocytaires



Les lipides extracellulaires organisés en feuillets lamellaires ou *lamellae* (Swartzendruber et al. 1989) jouent un rôle primordial dans l'imperméabilité de la couche cornée. En microscopie électronique, les lipides présentent une organisation alternant bandes denses aux électrons et bandes claires, qui correspondent, respectivement, aux groupements polaires et aux chaînes carbonées apolaires. L'espace intercornéocytairre contient une quantité variable de triglycérides, d'esters de cholestérol et de squalènes, provenant des glandes sébacées. Cependant, les lipides de la couche cornée sont principalement synthétisés par les kératinocytes. Une analyse en chromatographie en couche mince montre que la

couche cornée contient, selon les sites anatomiques, 20-30% de sphingolipides (majoritairement des céramides), 15-20% de cholestérol, 15-20% d'acides gras libres, 2-6% de sulfate de cholestérol et 3-5% de phospholipides (Lampe et al. 1983). Au niveau palmoplantaire, la couche cornée est plus riche en cholestérol (33%) et en sphingolipides (35%) au détriment des lipides spécifiques des glandes sébacées, lesquelles sont absentes de ce territoire.

Les kératinosomes occupent environ 10% du volume cytoplasmique des kératinocytes granuleux. Ce compartiment représente vraisemblablement un système de stockage des lipides à destinée extracellulaire avant leur sécrétion. Les kératinosomes transportent aussi, en plus de la CDSN et des protéines régulant la desquamation, une partie des enzymes chargées de la transformation des lipides, telles que la stéroïde sulfatase, la beta-glucocérébrosidase, la sphingomyélinase et les phospholipases A2 sécrétées. A ce jour, on peut supposer que toutes les protéines extracellulaires dotées d'un peptide-signal transitent par les kératinosomes.

Les transporteurs de la famille ABC (ATP-Binding Cassette) constituent une superfamille de protéines à plusieurs domaines transmembranaires, évolutivement très conservées, formant un canal qui permet le transport ATP-dépendant de biomolécules à travers les membranes. Les membres de la sous-famille ABCA participent au transfert de lipides à travers les membranes d'un compartiment à l'autre de la cellule (Kaminski et al. 2006), et participent probablement à l'accumulation des lipides dans les corps lamellaires. Ainsi, des mutations inactivant l'expression du transporteur ABCA12 (mutations non-sens et délétions) ont été identifiées chez patients atteints d'ichtyose arlequin (Lefevre et al. 2003; Kelsell et al. 2005). L'ichtyose arlequin (OMIM #242500), génodermatose transmise sur le mode autosomique récessif, congénitale et généralement létale, est caractérisée par une hyperkératose, une compaction de la couche cornée, des fissures cutanées profondes, un ectropion et un eclabium. L'épiderme des individus atteints d'ichtyose arlequin présente des kératinosomes anormaux, dépourvus de contenu lamellaire (Dale et al. 1990). Des analyses en immunomicroscopie électronique montrent que ABCA12 est exclusivement localisé à la membrane des kératinosomes (Akiyama et al. 2005). Des kératinocytes de patients, établis en culture et différenciés *in vitro*, présentent un défaut de sécrétion des glucosylcéramides, ce phénotype étant réversible par transfert du gène *ABCA12* sauvage dans ces cellules. Des analyses en microscopie électronique montrent une accumulation des kératinosomes dans le cytosol près de la membrane

plasmique et des structures lamellaires perturbées. Des substitutions d'acides aminés dans la protéine sont quant à elles associées à l'ichtyose lamellaire de type II (OMIM #601277), pathologie hétérogène transmise sur le mode autosomique récessif. Ces travaux suggèrent ainsi un rôle direct pour ABCA12 dans la sécrétion des lipides (transfert des lipides vers les kératinosomes et/ou extrusion), à rapprocher de la fonction du gène ABCA3 dans la sécrétion du surfactant pulmonaire (principalement de la dipalmitoylphosphatidylcholine) *via* les corps lamellaires des pneumocytes de type II (Shulenin et al. 2004).

– Le sulfate de cholestérol

Le cholestérol sécrété par les kératinosomes provient de la synthèse *de novo* dans les couches vivantes de l'épiderme. Une petite proportion est transformée dans les couches suprabasales en sulfate de cholestérol par la cholestérol sulfotransférase SULT2B1. Ce sulfate de cholestérol stimule la différenciation kératinocytaire par activation des PKC (Protein Kinase C) et par régulation transcriptionnelle de l'involucrine et de la transglutaminase 1 via l'induction de complexes AP-1 contenant Fra1, Fra2 et JunD (Hanley et al. 2001). Le sulfate de cholestérol est également un inhibiteur de serine protéases et retarde la desquamation chez la souris (Sato et al. 1998). Dans la couche cornée, le sulfate de cholestérol est progressivement converti en cholestérol par la stéroïde sulfatase (STS, aussi connue sous le symbole ARSC1) (figure 37). Une déficience de cette enzyme est responsable d'une ichtyose récessive

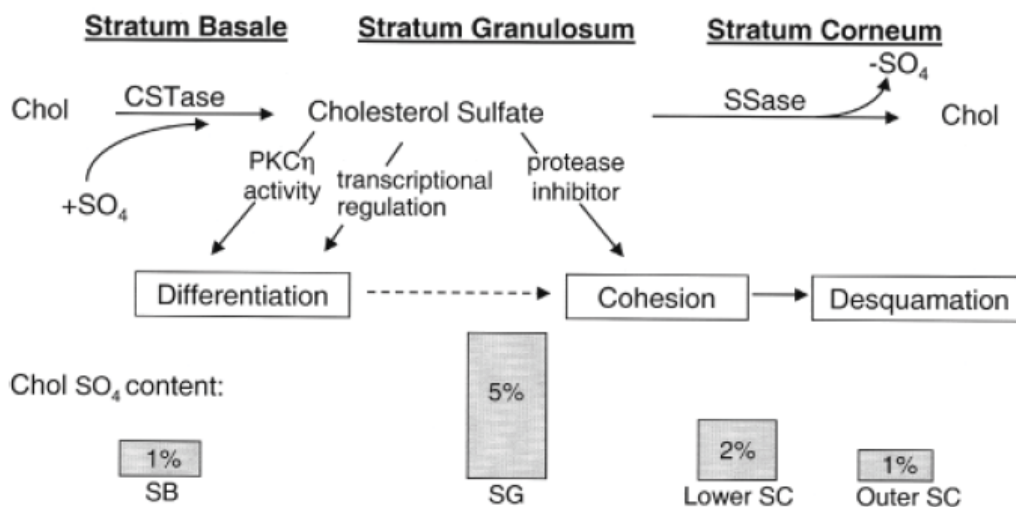


Figure 37. Cycle du sulfate de cholestérol dans l'épiderme normal (Elias et al. 2004).

liée à l'X (OMIM +308100) par accumulation de son substrat, le sulfate de cholestérol, dans la couche cornée de l'épiderme, retardant la desquamation. Les malades présentent une activité réduite des serine protéases, et un retard dans la protéolyse des cornéodesmosomes (Elias et al. 2004).

– Les céramides

Les céramides sont des sphingolipides formés d'un groupement sphingosine lié à un acide gras par une liaison amide. Dans la couche cornée, 9 espèces de céramides sont présentes et diffèrent entre elles par leur groupement sphingosine et par la longueur de leur chaîne carbonée. Parmi eux se distinguent 2 espèces d' ω -hydroxycéramides à très longue chaîne (30-32 C) qui traversent les deux hémicouches de la membrane plasmique et sont liées aux protéines de l'enveloppe cornée au moment de sa formation, en particulier l'involucrine, sous l'action de la transglutaminase 1 (Nemes et al. 1999). Cette couche de céramides fixés de façon covalente autour de l'enveloppe cornée pourrait servir de matrice pour l'orientation des *lamellae* lipidiques (Swartzendruber et al. 1989).

Il existe 2 voies principales de synthèse des céramides, par synthèse *de novo* par la condensation de serine et de palmitoyl-CoA par la serinepalmitoyltransférase et la céramide synthase (Choi and Maibach 2005), et par dégradation du glucosylcéramide par la beta-glucocérébrosidase (GBA) (Hachem et al. 2005; Hachem et al. 2006) ou de la sphingomyéline par la sphingomyélinase (SMPD).

– Les acides gras libres

Ils proviennent quasi-exclusivement du catabolisme des glycérophospholipides, constituants principaux de la membrane plasmique des cellules, par des phospholipases A2. Ces enzymes hydrolysent les liaisons ester des glycérophospholipides, libérant un lysophospholipide et un acide gras. Ainsi, les glycérophospholipides disparaissent progressivement et ne sont plus détectés qu'à l'état de traces au niveau de la couche cornée. L'application topique d'inhibiteurs des phospholipases A2 entraîne chez la souris une perturbation de la fonction barrière, restaurée par l'application d'acides gras libres (Mao-Qiang et al. 1995). Plusieurs phospholipases A2 sécrétées ont été identifiées dans l'épiderme mais seules sPLA2-IB, -IIF, et -X sont spécifiques des kératinocytes granuleux. Leur fonction précise n'a pas encore été clairement établie (Haas et al. 2005). La protéine ABHD5/CGI-58, de

la famille des alpha/bêta-hydrolases, pourrait également être impliquée dans la production d'acides gras en servant de coactivateur des triglycéride lipases – pour revue (Yamaguchi and Osumi 2009). Des mutations dans ce gène sont associées au syndrome de Chanarin-Dorfman (OMIM #275630), caractérisé notamment par l'accumulation de gouttelettes lipidiques dans le cytoplasme de nombreux tissus (Lefevre et al. 2001; Caux et al. 2004; Ben Selma et al. 2007). Enfin, des mutations dans le gène *FATP4*, codant pour un transporteur d'acides gras, ont été récemment identifiées chez des patients atteints d'un syndrome associant ichtyose et prématurité (OMIM #608649) (Klar et al. 2009).

De façon intéressante, en accord avec les données de la littérature sur le métabolisme lipidique des adipocytes et des hépatocytes, les récepteurs nucléaires / facteurs de transcription des familles PPAR (Peroxisome Proliferation Activation Receptor) et LXR (Liver X Receptor), exprimés dans l'épiderme, semblent capables de réguler plusieurs gènes impliqués dans la synthèse ou le transport des lipides requis pour la fonction barrière ; pour une revue récente, voir (Schmuth et al. 2008). Ainsi, le traitement topique par des activateurs des PPAR ou LXR induisent l'expression du transporteur ABCA12 (Jiang et al. 2008), de la cholestérol sulfotransférase SULT2B1 (Jiang et al. 2005), mais aussi accélère la sécrétion des kératinosomes et des phospholipases A2 contenues par ceux-ci (Fluhr et al. 2009).

Toutefois, en plus du rôle crucial des différents lipides de la couche cornée (et des protéines impliquées dans leur synthèse et transport) pour l'établissement d'une barrière imperméable, il a également été mis en évidence ces dernières années un rôle des jonctions serrées localisées au niveau de la couche granuleuse.

2.2.2.2.1.2. Jonctions serrées

Egalement nommées *zonula occludens*, les jonctions serrées forment une ceinture d'adhérence autour du pôle apical des cellules épithéliales (figure 38A) et assurent l'étanchéité de certains épithéliums en empêchant la diffusion de molécules par la voie paracellulaire (certaines sont particulièrement étanches, notamment celles de l'endothélium vasculaire formant la barrière hémato-encéphalique). Plusieurs familles de protéines à quatre segments transmembranaires sont caractéristiques de ces jonctions, telles que les claudines (plus de 24 membres), l'occludine, et la

tricelluline (localisée uniquement aux points de contact entre trois cellules). Du côté cytosolique, plusieurs protéines adaptatrices assurent la liaison au cytosquelette d'actine : les protéines à domaine PDZ (interactions protéine-protéine) ZO(Zonula Occludens)-1, ZO-2 et ZO-3, MUPP1 (multi-PDZ Domain Protein-1), la cinguline, *etc.* Enfin, de nombreuses protéines impliquées dans diverses voies de signalisation (protéines G, PKC...) sont également liées aux jonctions serrées. Une étude alliant protéomique et bioinformatique souligne ainsi la présence, avérée ou potentielle, de plus de 900 protéines différentes (Tang 2006). Les jonctions serrées jouent à la fois un rôle structural, fonctionnel et signalisateur, divisant la membrane plasmique en plusieurs domaines (apical vs baso-latéral) de composition protéique et lipidique variable, et recrutant localement de nombreuses protéines de signalisation ayant notamment une fonction prépondérante dans le maintien de la polarité cellulaire.

Contrairement aux épithéliums monocouches où leur fonction cruciale a été caractérisée depuis de nombreuses années, les jonctions serrées sont longtemps passées inaperçues au sein de l'épiderme, principalement parce que leur localisation est restreinte à quelques assises et leur observation peu aisée (Furuse et al. 2002) (figure 38B). Les jonctions serrées ne sont assemblées et fonctionnelles qu'à partir

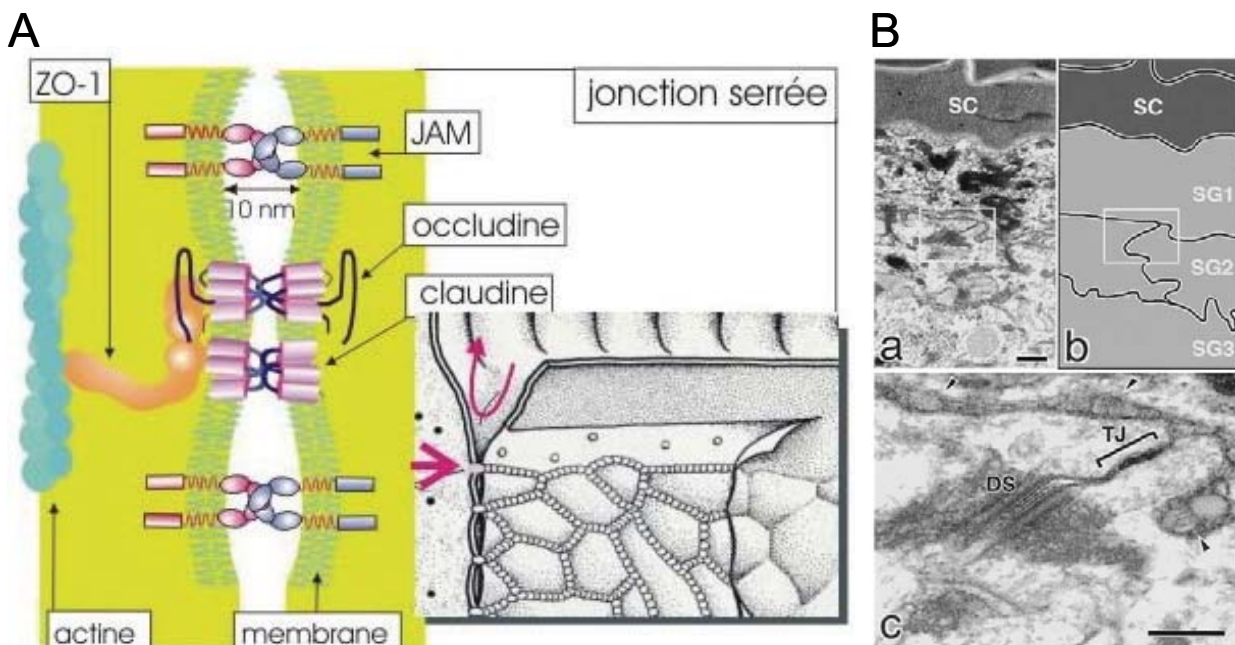


Figure 38. Structure des jonctions serrées. (A) Représentation schématique (http://www.ulyse.u-bordeaux.fr/atelier/ikramer/biocell_diffusion/). (B) Coupe d'épiderme murin observée en microscopie électronique (Furuse et al. 2002). Les jonctions serrées sont visibles dans la couche granuleuse (SG), au pôle apical des kératinocytes. TJ : jonction serrée ; DS : desmosome.

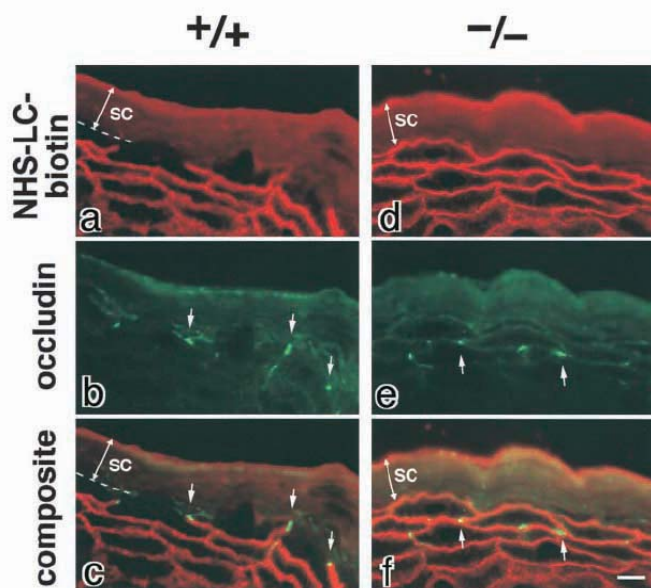


Figure 39. Diffusion paracellulaire d'un marqueur (Furuse et al. 2002). Dans un épiderme de souris sauvage, la diffusion du marqueur est arrêtée au niveau de la couche granuleuse par la présence des jonctions serrées (marquage occludine). Dans l'épiderme de souris inactivées pour la claudine 1, le marqueur diffuse jusqu'à la couche cornée.

des assises centrales de la couche granuleuse, même si certaines protéines sont exprimées plus précocement. Leur fonction dans la barrière épidermique a été mise en évidence dans des souris inactivées pour la claudine 1, celles-ci présentant un trouble important de la barrière épidermique et mourant en quelques heures après la naissance (Furuse et al. 2002). Chez ces souris, la couche cornée apparaît épaissie et compacte en microscopie optique, mais de façon surprenante, possède une apparence normale en microscopie électronique. Par contre, les jonctions serrées ne sont pas fonctionnelles et ne bloquent pas la diffusion paracellulaire d'un marqueur (figure 39). De même, des souris surexprimant la claudine 6 murine sous dépendance du promoteur humain de l'involucrine (*IVL-Cldn6*) ont une désorganisation des jonctions serrées avec perte de la fonction barrière (Turksen and Troy 2002). Chez l'homme, des mutations du gène codant la claudine 1 ont été identifiées chez des patients souffrant du syndrome NISCH (Neonatal Ichthyosis-Sclerosing CHolangitis syndrome, également associé à une alopecie – OMIM #607626) (Hadj-Rabia et al. 2004), touchant le foie, l'épiderme et ses annexes, sites majeurs d'expression de la claudine 1. L'importance des autres protéines des jonctions serrées pour la fonction barrière reste encore à déterminer, l'absence de changement dans l'organisation kératinocytaire ou dans la structure des jonctions serrées dans l'épiderme de souris déficientes en claudine 1 suggérant notamment une possible redondance (mais non compensation) avec les autres claudines présentes dans l'épiderme. Le profil d'expression de plusieurs d'entre elles a été déterminé : en immunofluorescence indirecte, les claudines 1 et 7, JAM1 et MUPP1 sont détectées sur des coupes de peau

humaine dans toutes les couches vivantes de l'épiderme, alors que ZO1 et la claudine 4 sont majoritairement exprimés dans la couche granuleuse, et l'occludine et la cinguline sont présentes exclusivement dans les dernières cellules vivantes – pour revue (Brandner 2008) (figure 40). L'expression de la tricelluline n'a pas été caractérisée sur coupe de peau, toutefois son expression est détectable dans des cultures de kératinocytes dès le début de la stratification (cellules post-confluentes), suggérant que la protéine pourrait être présente dans toutes les couches suprabasales de l'épiderme à l'intersection de trois cellules (Schluter et al. 2007).

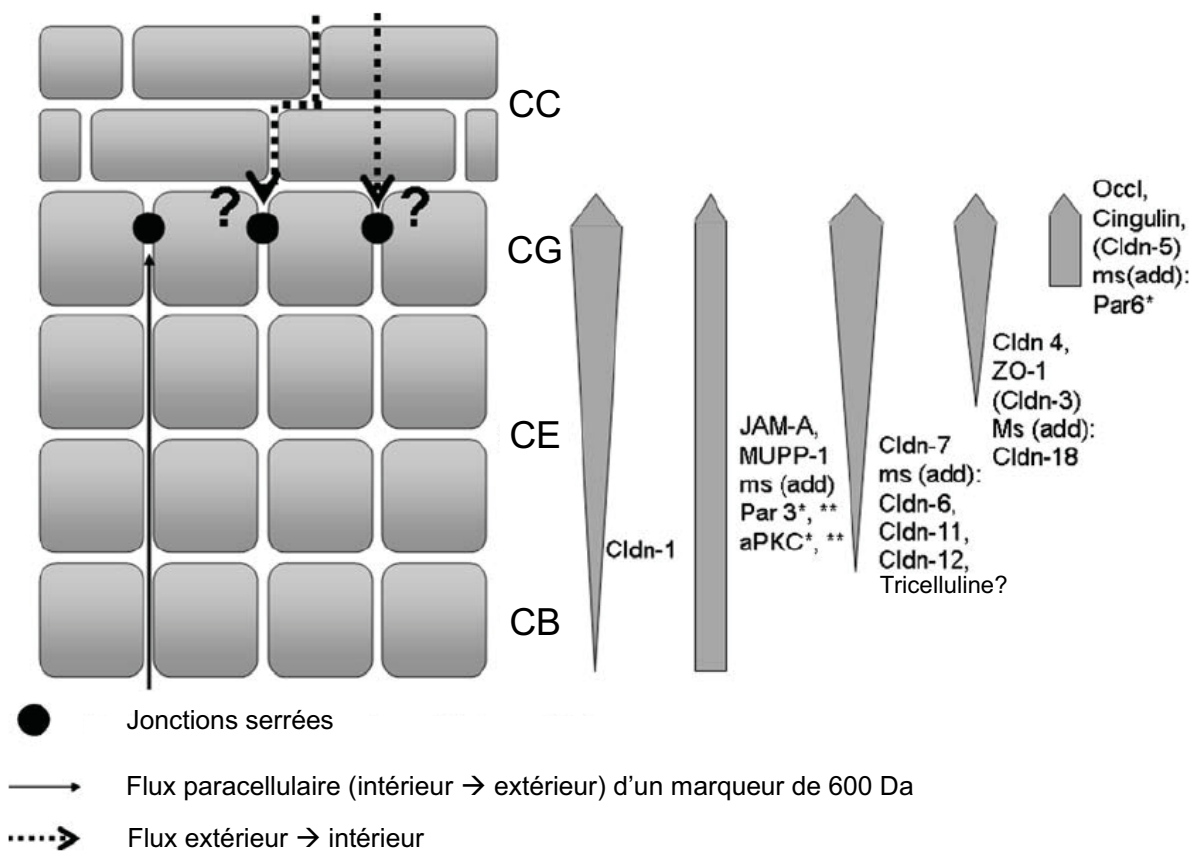


Figure 40. Expression épidermique des protéines des jonctions serrées (d'après Brandner 2008). * Localisation variable selon isoformes. ** localisation membranaire et cytoplasmique. Ms (add) Spécifiques de l'épiderme murin.

Ainsi, il apparaît qu'une double barrière, à la fois lipidique et protéique, est requise pour garantir l'imperméabilité de l'épiderme. Toutefois, il ne s'agit que d'un aspect de la fonction barrière, la résistance de l'épiderme aux fortes contraintes mécaniques auxquelles il est soumis étant également un prérequis pour qu'il assure sa fonction.

2.2.2.2.2. Résistance mécanique

La résistance de la couche cornée fait intervenir trois éléments en continuité physique : la matrice fibreuse composant l'intérieur des cornéocytes, l'enveloppe cornée qui forme une sorte de coque autour de cette matrice et remplace la membrane plasmique, et enfin les cornéodesmosomes liant entre eux les cornéocytes et assurant leur cohésion.

2.2.2.2.2.1. Matrice fibreuse

La compaction des filaments intermédiaires de kératines dans les cornéocytes par la filaggrine est vraisemblablement responsable de l'aplatissement de ces cellules. Le cytoplasme est remplacé par une matrice fibreuse essentiellement formée de kératines agrégées. Les filaments intermédiaires (FI) de kératines sont une composante essentielle du cytosquelette épidermique, permettant la cohésion de ce tissu grâce notamment à leur connexion intercellulaire par l'intermédiaire des desmosomes, et à la lame basale par l'intermédiaire des hémidesmosomes.

Une nouvelle nomenclature a été proposée par plusieurs scientifiques faisant référence dans ce domaine (Schweizer et al. 2006). Les gènes codant pour les

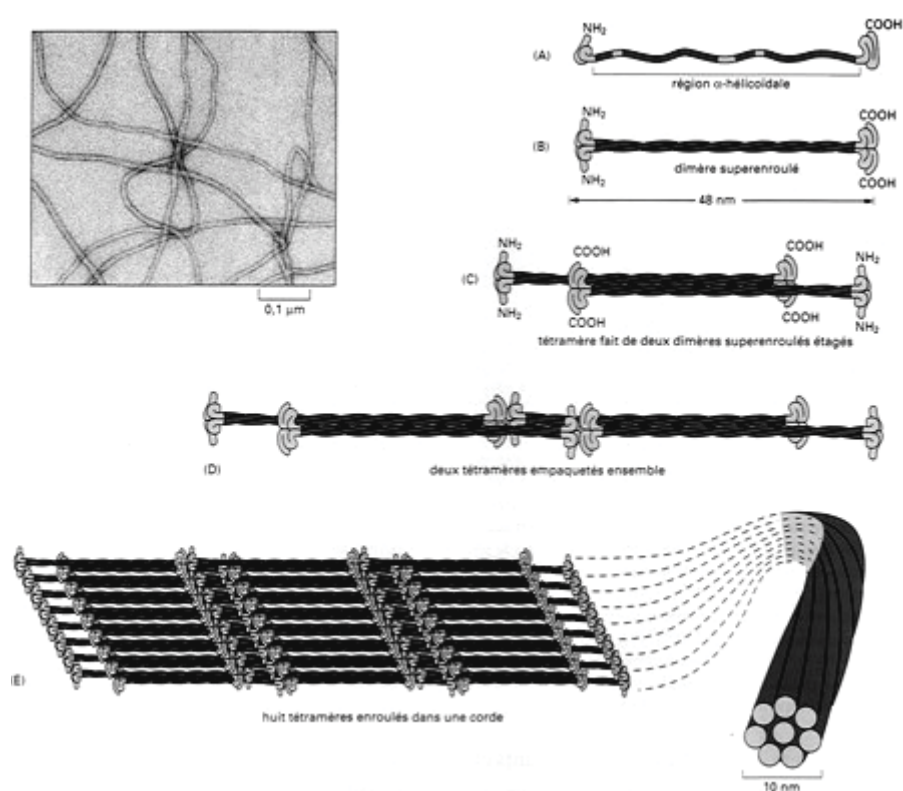


Figure 41. Assemblage des filaments intermédiaires de kératines
(http://www.humans.be/bio_cell_cytosquelette.html).

kératines classiques sont désignés *KRT1-24* et les protéines correspondantes K1-24. Elles sont classées en deux groupes, I (acide, K9 à K23) ou II (basique, K1 à K8). Les kératines acides sont codées par des gènes situés dans un même locus en 17q12, alors que les gènes codant pour les kératines basiques sont situés en 12q13 (Lessin et al. 1988). Toutes les kératines partagent une même organisation structurale avec une partie centrale, constituée d'hélices α reliées par des domaines de liaison, entourée de deux domaines globulaires terminaux. Les kératines s'assemblent en hétérodimères constitués d'une kératine basique et d'une kératine acide puis en tétramères qui forment l'unité de base des protofilaments de 3 nm (figure 41). Huit protofilaments forment finalement le filament intermédiaire de 10 nm de diamètre (Fuchs and Weber 1994).

Dès la première assise de la couche épineuse, les kératinocytes expriment le couple de kératines K1 et K10 qui remplace le réseau préexistant K5 - K14, spécifique des kératinocytes basaux. K2 (anciennement *KRT2A* ou *KRT2e*) est détectée dans les couches suprabasales de l'épiderme, à partir de la troisième ou quatrième assise cellulaire. Des disparités sont à noter dans l'expression des kératines selon les territoires anatomiques : ainsi, K2 est moins présente au niveau des seins, aisselles et verge (Collin et al. 1992), tandis que K9, une kératine très homologue à K10, est exprimée de façon spécifique dans les couches suprabasales de l'épiderme palmoplantaire (Langbein et al. 1993).

De nombreuses génodermatoses sont causées par des mutations des kératines épidermiques, qui ont pour la plupart un effet dominant négatif. Les phénotypes sont plus ou moins sévères selon la couche épidermique au niveau de laquelle sont exprimées les kératines touchées, et selon la position de la mutation sur la protéine (Cheng et al. 1992). Par exemple, une anomalie au niveau des régions globulaires de K5 ou K14 entraîne un grave défaut de cohésion entre les kératinocytes basaux conduisant à la rupture de ces cellules et à la formation de bulles profondes (épidermolyse bulleuse simple (EBS) de type Dowling-Meara, OMIM #131760). Des mutations dans les mêmes régions globulaires des kératines suprabasales K1 ou K10 sont responsables d'une hyperkératose épidermolytique avec rupture au niveau des kératinocytes épineux, dermatose moins sévère (Cheng et al. 1992), ou encore de kératodermies palmoplantaires (Kimonis et al. 1994) ou d'ichtyose hystrix de Curth-Macklin (Sprecher et al. 2001), tandis que des mutations dans le domaine hélicoïdal conduisent à une érythrodermie ichtyosiforme congénitale bulleuse (McLean et al.

1994a). Des mutations de KRT2 sont responsables de l'ichtyose bulleuse de Siemens (McLean et al. 1994b), transmise sur le mode dominant, et caractérisée par une hyperkératose et la formation spontanée de bulles superficielles (OMIM #146800). Enfin, des mutations dans le gène *KRT9* sont responsables de kératodermies palmoplantaires épidermolytiques (OMIM #144200).

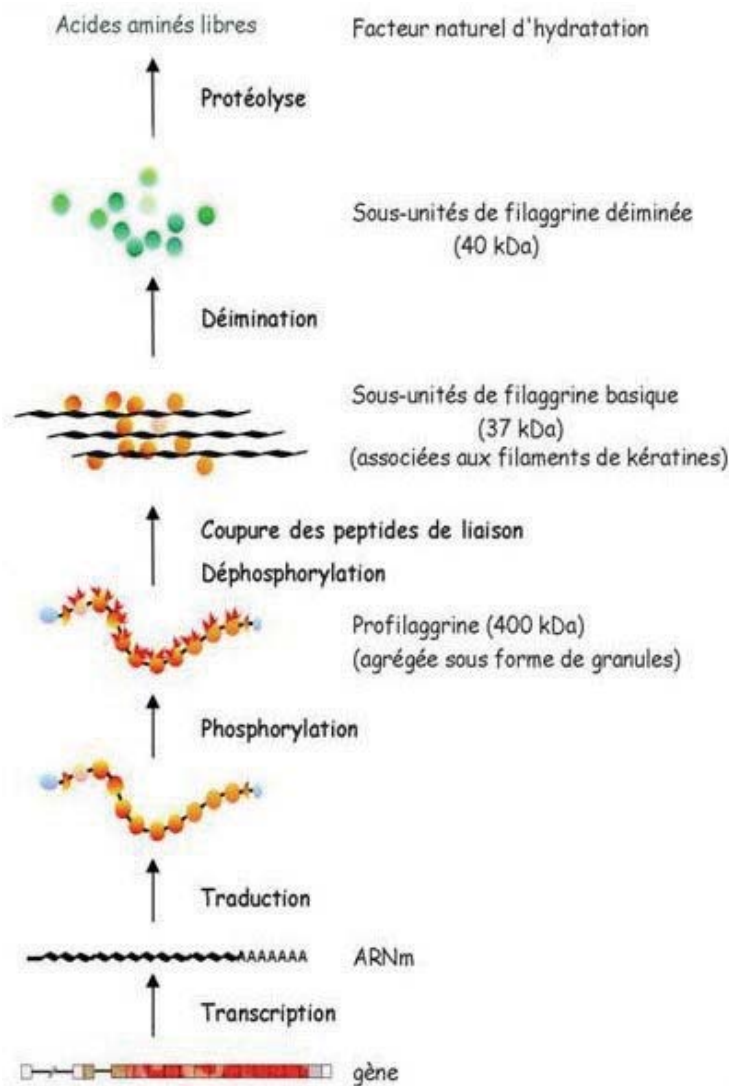


Figure 42. Métabolisme de la profilaggrine (Simon, Méchin et al. 2001).

La profilaggrine, contenue dans les granules de kératohyaline (granules F) caractéristiques du kératinocyte granuleux, appartient à la famille des IFAP (Intermediate Filament Associated Protein). Son gène est localisé chez l'homme dans le complexe de différenciation épidermique en 1q21. Elle est d'abord produite sous la forme d'un précurseur insoluble de 400 kDa environ, phosphorylé, formé d'un domaine amino-terminal de 293 aa, d'une sous-unité tronquée de 173 aa suivie de 10 à 12 sous-unités homologues de 324 aa et enfin d'un domaine carboxy-terminal de 157 aa (Presland et al. 1992). Lors de la cornification, la profilaggrine est déphosphorylée puis clivée en unités de filaggrine (figure 42)

(Gan et al. 1990). Là où les protéases impliquées dans le clivage entre les monomères n'ont à ce jour pas été formellement identifiées, bien que la matriptase (List et al. 2003; Alef et al. 2009) et la caspase 14 (Denecker et al. 2007) puissent être impliquées dans une partie du processus. Les monomères de filaggrine libérés agrègent les filaments intermédiaires en microfibrilles qui formeront la matrice fibreuse dense intracornéocytaire (Steinert et al. 1981). Finalement, la filaggrine est

désiminée dans la couche cornée par des peptidyl-arginine désiminases (PADI) qui catalysent la conversion des résidus arginyl en résidus citrullyl (Nachat et al. 2005) ; la protéine perd son affinité pour les filaments intermédiaires de kératines et est dégradée en acides aminés libres qui participent à la formation du facteur naturel d'hydratation (Scott et al. 1982), possiblement sous l'action de deux protéases, la bléomycine hydrolase et la calpaine-I (Kamata et al. 2009).

L'implication de la profilaggrine dans la physiopathologie de l'ichtyose vulgaire a été longtemps soupçonnée, en particulier en raison de l'absence de granules de kératohyaline dans cette pathologie (Sybert et al. 1985), mais ce n'est que récemment que des mutations du gène ont été identifiées dans 15 familles de patients (Smith et al. 2006). La longueur de la séquence codante (presque 12kb) ainsi que la présence de répétitions ont sans doute été responsable du retard pris dans le séquençage de ce gène chez les malades. Il n'existe d'ailleurs aucun ARNm complet dans les banques de données. Les deux principales mutations entraînant une perte de fonction, R501X ou 2282del4, sont retrouvées chez 9% de la population européenne, et constituent en outre un facteur de risque majeur, à l'état hétérozygote, pour la dermatite atopique (mutations semidominantes) (Palmer et al. 2006) ; pour une revue récente, voir (McGrath 2008).

Enfin, il faut noter que des molécules de kératines suprabasales K1, K2 et K10, ainsi que de filaggrine (Simon et al. 1996), sont liées à l'enveloppe cornée sous l'action des transglutaminases, enzymes qui catalysent la formation de liaisons covalentes, comme le montre la digestion protéolytique d'enveloppes cornées purifiées à partir d'épiderme humain et le séquençage des peptides obtenus (Steinert and Marekov 1995).

2.2.2.2.2. Enveloppe cornée

L'enveloppe cornée est une structure protéique remplaçant la membrane plasmique et formant une coque qui emprisonne la matrice fibreuse. Cette structure est insoluble en présence de détergents et / ou d'agents réducteurs, dans la mesure où les protéines qui la constituent sont liées entre elles par des liaisons covalentes, autres que des ponts disulfure (Rice and Green 1977). L'assemblage de l'enveloppe cornée, qui remplace la membrane plasmique, se déroule de façon séquentielle par incorporation successive de précurseurs de deux sortes : des protéines de la plaque desmosomale, mais aussi des protéines qui viennent se positionner sous la membrane plasmique au fur et à mesure de leur synthèse dans le cytoplasme (figure

43). Les protéines de la plaque desmosomale, notamment la périplakine et l'envoplakine, servent d'échafaudage avec l'involucrine, avant que les autres précurseurs ne soient incorporés. Ces précurseurs sont liés entre eux par une famille d'enzymes calcium-dépendantes, les transglutaminases, dont les isoformes Tgm1, 3 et 5 sont exprimées dans l'épiderme.

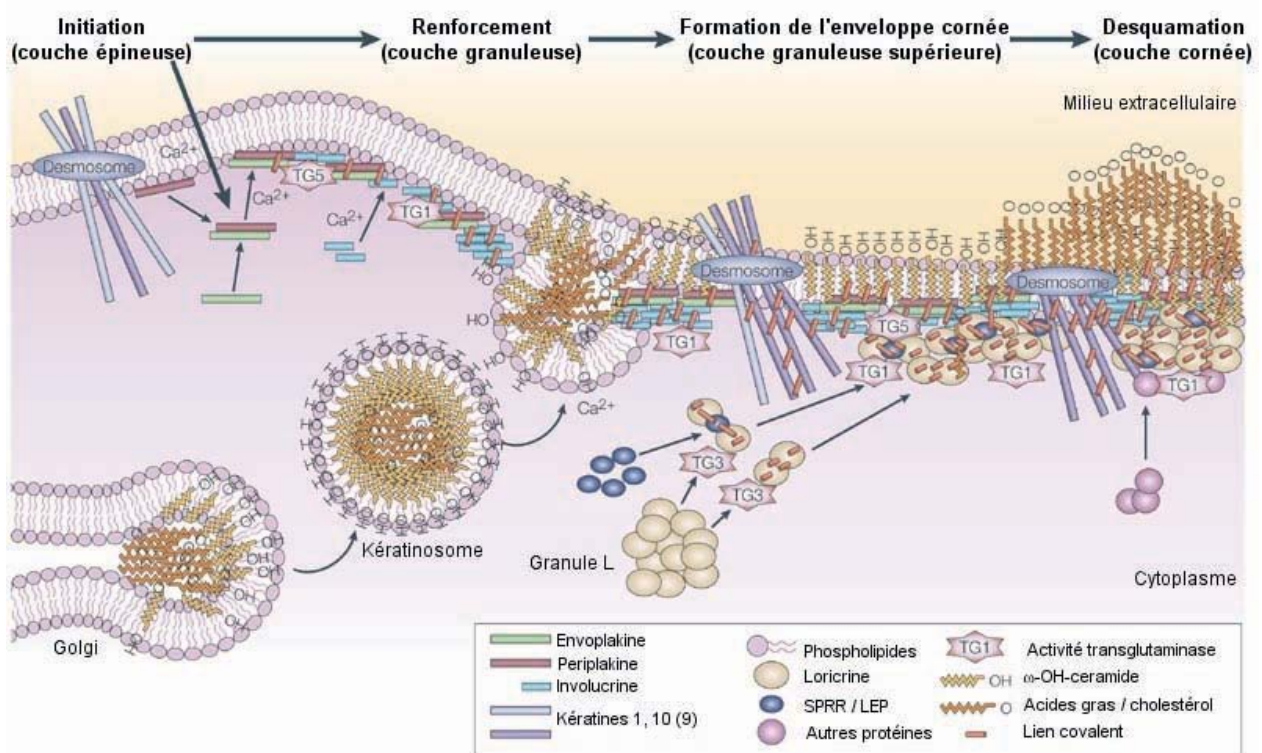


Figure 43. Etapes de l'assemblage de l'enveloppe cornée (Candi et al. 2005).

La plaque desmosomale, dont les constituants sont inclus dans l'enveloppe cornée sous l'action des transglutaminases, comporte des membres de la famille des plakines (desmoplakine, périplakine, envoplakine, épiplakine) ainsi que des protéines à domaine armadillo (plakoglobine et plakophilines). Deux protéines, la périphiline (Kazerounian and Aho 2003) et la kazrine (Groot et al. 2004), sont également associées à la périplakine dans la couche granuleuse, la kazrine jouant de plus un rôle dans le contrôle de la différenciation via la régulation de l'activité de Rho (Sevilla et al. 2008). Plusieurs des protéines de la plaque ont été montrées comme pouvant être impliquées dans diverses maladies génétiques caractérisées par des cardiomyopathies et/ou des anomalies au niveau de l'épiderme et de ses annexes (figure 44), mettant en évidence leur rôle dans la résistance mécanique des tissus soumis à de fortes contraintes. C'est notamment le cas de la desmoplakine, dont différentes mutations

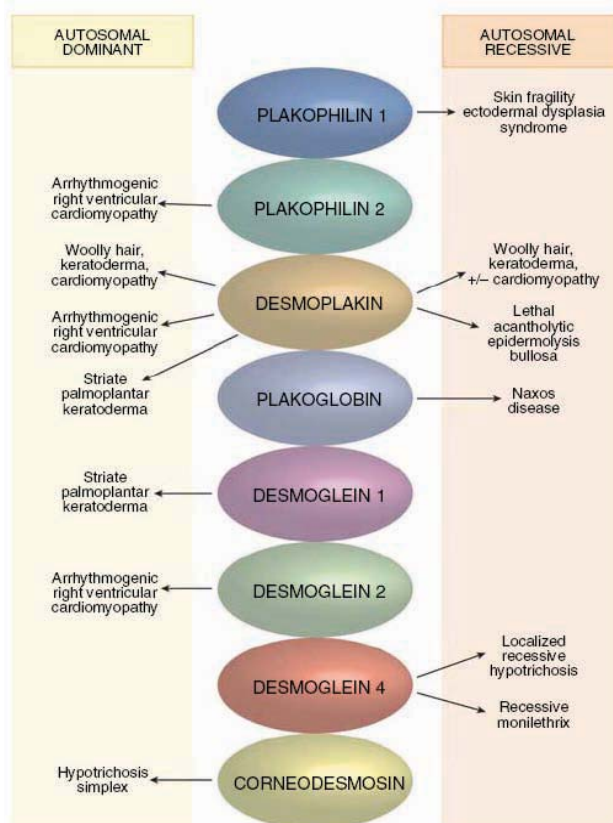


Figure 44. Protéines du (cornéo-) desmosome impliquées dans des maladies humaines (Lai-Cheong et al. 2007)

sont à l'origine de kératodermies palmoplantaires striées de type II, une maladie autosomique dominante due à une haploinsuffisance ; d'autres mutations sont responsables de fragilités épidermiques (pouvant aller jusqu'à une épidermolyse bulleuse acantholytique létale à la naissance) ou de kératodermies généralisées, d'anomalies du follicule pileux et de cardiomyopathies (OMIM +125647). Toutefois, en raison d'une certaine redondance fonctionnelle, l'importance de certaines protéines de la famille des plakines dans l'établissement de la fonction barrière reste très difficile à établir. Ainsi, des souris dont le gène de l'envoplakine (Maatta et al. 2001), de la périplakine

(Aho et al. 2004) ou de l'épiplakine (Spazierer et al. 2006) ont été inactivés séparément ne présentent aucun trouble de la fonction barrière. De plus, une inactivation simultanée de l'involucrine (précurseur de l'enveloppe cornée), de l'envoplakine et de la périplakine ne provoque qu'un retard dans la mise en place de la fonction barrière au cours de l'embryogenèse, accompagné d'une simple anomalie de la desquamation conduisant à une hyperkératose (Sevilla et al. 2007).

En complément de ces protéines de la famille des plakines, l'involucrine participe également au premier échafaudage à l'origine de l'enveloppe cornée. Le gène *IVL*, situé au sein de l'EDC, code chez l'homme pour une protéine de 585 acides aminés composée de courtes répétitions en tandem (39 répétitions de 10 acides aminés) (Eckert and Green 1986). Exprimée précocement, dès la couche épineuse, l'involucrine est incorporée à l'enveloppe cornée comme le montrent des marquages en immunomicroscopie électronique d'enveloppes cornées et le séquençage de peptides obtenus après protéolyse de ces enveloppes (Steinert and Marekov 1997). Sa localisation sous-membranaire lui permet d'être attachée de façon covalente aux

hydroxycéramides qui vont progressivement remplacer la membrane plasmique (Nemes et al. 1999). De manière surprenante, des souris invalidées pour le gène de l'involucrine ne développent aucun phénotype particulier, l'épiderme et les follicules pileux étant histologiquement normaux (Djian et al. 2000). Elles ne présentent pas non plus de différences avec des souris sauvages au cours de la cicatrisation. Enfin, les enveloppes cornées sont morphologiquement identiques à celles de souris sauvages. Cette étude n'a pas montré de mécanisme de compensation par surexpression d'autres protéines de l'enveloppe cornée, néanmoins, seules la loritrine et SPRR1 ont été analysées à cet égard. L'expression de l'involucrine a été étudiée depuis de nombreuses années comme modèle de régulation génique lors de la différenciation kératinocytaire (pour une revue voir (Eckert et al. 2004)). Deux séquences promotrices ont été identifiées : un promoteur distal (-2500) et un promoteur proximal (Welter et al. 1995; Crish et al. 2006). Le promoteur distal est capable de lier les facteurs de transcription AP1 et Sp1, le promoteur proximal lie C/EBP et AP1. L'activation de ce gène est sous la dépendance de la voie des MAPK (Mitogen-Activated Protein Kinases) et des PKC. Le promoteur humain de l'involucrine est également très utilisé pour diriger l'expression d'un transgène dans les kératinocytes suprabasaux de modèles *in vitro* ou *in vivo*, y compris chez la souris (Carroll et al. 1993; Crish et al. 1993). De telles constructions ont notamment été employées pour réaliser des expressions ectopiques ou surexpressions de plusieurs intégrines (Carroll et al. 1995), du gène *nude / foxn1* (Prowse et al. 1999), de la claudine 6 (Turksen and Troy 2002), de la desmogléine 3 (Merritt et al. 2002) et de la connexine 26 (Djalilian et al. 2006).

Après les plakines et l'involucrine, d'autres protéines de structure sont intégrées à l'enveloppe cornée, notamment la loritrine, les SPRR (Small PROline-Rich proteins) et les LCE (Late Cornified Envelope).

Le gène *LOR* codant pour la loritrine est lui aussi situé au sein de l'EDC. La loritrine est spécifiquement détectée dans la couche granuleuse de l'épiderme et à la périphérie des cornéocytes (Mehrel et al. 1990). La comparaison de sa séquence et de la composition des enveloppes cornées suggère qu'elle constitue au moins 70% de leur masse protéique (Hohl et al. 1991; Steinert and Marekov 1997). La loritrine est exprimée tardivement dans les kératinocytes granuleux au sein de « L-granules », plus petits et de forme plus régulière que les « F-granules » de filaggrine (Steven et al. 1990). Très riche en glycine (47% chez l'homme, et plus de 55% chez la souris),

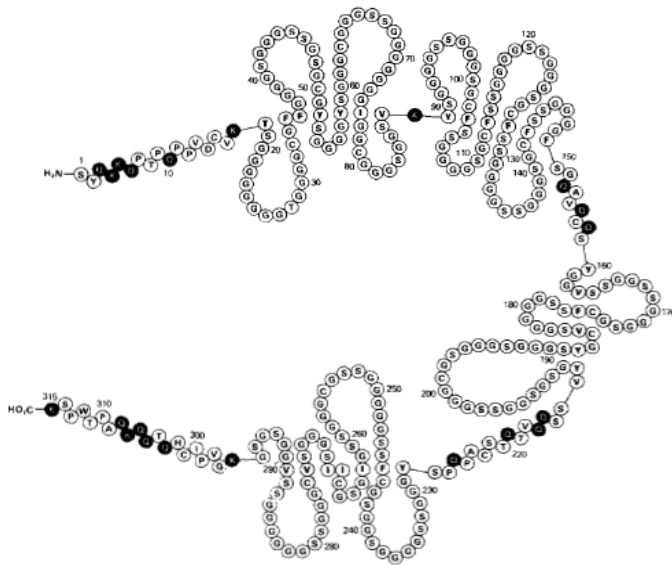


Figure 45. Modèle des « boucles glycine » pour la loricrine (Steinert et al. 1991).

la protéine bien qu'insoluble est supposée extrêmement flexible. Selon le modèle de Peter Steinert (Steinert et al. 1991), les acides aminés aromatiques ou aliphatiques dispersés parmi des régions riches en glycine et serine pourraient interagir entre eux et permettre la formation de « boucles glycine » (figure 45), l'ensemble étant éventuellement stabilisé par des ponts disulfures.

Deux génodermatoses

autosomiques dominantes ont été associées à des mutations de la loricrine, une forme mutilante du syndrome de Vohwinkel (Maestrini et al. 1996) (OMIM #604117) et l'érythrokratodermie progressive symétrique (PSEK, OMIM #602036) (Ishida-Yamamoto et al. 1997). Ces pathologies sont caractérisées par une kératodermie palmoplantaire assortie d'un pseudo-aïnhum (constriction circonférentielle des phalanges pouvant aboutir à leur amputation). Des souris invalidées pour le gène de la loricrine présentent un retard dans la formation de la barrière épidermique à la fin de la vie embryonnaire, une érythrodermie congénitale et une couche cornée fragilisée (Koch et al. 2000). Néanmoins, ce phénotype disparaît au bout de 4 à 5 jours. La surexpression d'autres protéines de l'enveloppe cornée, telles que SPRR2D, SPRR2H et la répétine pourrait témoigner de la mise en place d'un mécanisme de compensation. Produites par la même équipe, des souris transgéniques exprimant une forme mutée de la loricrine proche de celle associée au variant mutilant du syndrome de Vohwinkel et la PSEK présentent un phénotype proche de celui des patients, avec une ichtyose, un défaut de barrière et des constrictions au niveau de la queue (Stern et al. 2006). La sévérité du phénotype est directement corrélée au taux d'expression du transgène. Néanmoins, la loricrine mutée n'interagit pas avec la protéine sauvage et n'est pas incorporée à l'enveloppe cornée mais accumulée dans les noyaux parakératosiques. La forme mutée n'agirait pas comme dominant négatif, mais par un gain de fonction consécutif à l'apparition d'un signal de localisation nucléaire fonctionnel. Une autre étude montre que les mutations caractéristiques de

ces pathologies (insertion d'un nucléotide en position 662, 709 ou 730) entraînent aussi chez l'homme un décalage du cadre de lecture et l'apparition d'un signal de localisation nucléaire (Ishida-Yamamoto 2003). Il est proposé que la forme de loricrine mutée interfère indirectement dans la cornification en empêchant la lyse nucléaire, et/ou en perturbant les fonctions nucléaires requises lors de la différenciation. Cette même étude propose de regrouper la forme variante mutilante du syndrome de Vohwinkel et l'érythrokratodermie progressive symétrique sous le terme de kératodermie impliquant la loricrine (LK, « Loricrin Keratoderma »). Le mécanisme par lequel se forment les constrictions conduisant à des mutilations reste encore totalement incompris.

	N TERMINUS	REPEATS		C TERMINUS	
		number	consensus sequence		
type I	SPRR1A	MNSQQQ--KQPGTPPPQIQQQQVKQFCQPP	} 6X	PKVPEPC.	-----STVTPAFAQQ---KTKQK*
	SPRR1B	MSSQQQ--KQPGTPPPQIQQQQVKQFCQPPP			-----SIVTPAFAQQ---KTKQK*
	SPRR3	MSSYQQ--KQTFTPPPQIQQQQVKQFCQPPP			-----STVTPGPAQQ---KTKQK*
	SPRR4	MSSQQQQRQQCCPPQRAQQQQVKQFCQPPP			----DQKCE-----SAQQA=SKSKQK*
type II	SPRR2A	MSYQQQQCKQPCQPPVVCPT	} 3X	PKCPEPCPP	QQCQQYPPVTPSPPCQSKYPPKSK*
	SPRR2B	MSYQQQQCKQPCQPPVVCPT			QQCQQYPPVTPSPPCQPNYPPKSK*
	SPRR2C	MSYQQ*QCKQPCQPPVVCRL → Ψ			QQCQQYPLVTPSPPCQPKQPPKSK*
	SPRR2D	MSYQQQQCKQPCQPPVVCPT			QQCQQYPPVTPSPPCQPKQPPKSK*
	SPRR2E	MSYQQQQCKQPCQPPVVCPT			QQCQQYPPVTPSPPCQPKQPPKSK*
	SPRR2F	MSYQQQQCKQPCQPPVVCBA			QQCQQYPPVTPSPPCQPKQPPKSK*
	SPRR2G	MSYQQQQCKQPCQPPVVCPT			PFQDCKCPVQYPPCQQYPPKSK*

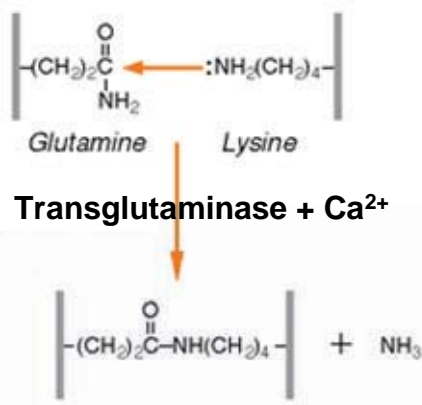
Figure 46. Organisation des protéines SPRR (Cabral et al. 2001b).

Les membres de la famille des SPRR se répartissent en quatre familles, initialement définies en fonction de leur homologie de séquence : SPRR1 (deux membres, SPRR1A et SPRR1B), SPRR2 (six membres chez l'homme, SPRR2A-2G, SPRR2C étant un pseudogène), SPRR3 (un seul membre) (Gibbs et al. 1993), et SPRR4 (un seul membre) (Cabral et al. 2001a). Ces petites protéines (6 à 25 kDa) composées de courtes répétitions sont classées en 2 types selon l'homologie de leurs domaines terminaux et le nombre d'acides aminés de leurs répétitions (figure 46). Les SPRR de type I comprennent des répétitions de 8 aa : 6 pour SPRR1, 16 pour SPRR3 et 4 pour SPRR4, et les SPRR de type II comprennent 3 répétitions de 9 aa (SPRR2A-G) (Cabral et al. 2001b). Cette étude a également montré que les SPRR2G et SPRR4 sont préférentiellement exprimés dans le kératinocyte granuleux, alors que les autres membres sont aussi détectés dans d'autres épithéliums stratifiés. SPRR3 n'est pas détecté dans l'épiderme sain mais seulement dans l'oesophage et le col utérin. Dans des kératinocytes primaires en culture, tous les membres, hormis SPRR2F, voient leur expression induite par le calcium. Ces gènes présentent en outre des réponses différentes à l'irradiation aux UV : SPRR4 et SPRR2G, les plus

spécifiques de l'épiderme, sont aussi les seuls à être fortement induits après irradiation aux UVC in vitro. Ainsi on constate que le patron d'expression des membres du groupe n'est pas corrélé avec leur type ou leur classe. La diversification de la régulation semble dans ce cas plus importante que la diversification de l'organisation structurale.

LCE5A	LEP18, SPRL5A	118 aa	<p>La famille des LCE (Late Cornified Envelope proteins) regroupe 17 gènes anciennement nommés SPRL (Small Proline Rich-like proteins), LEP (Late Envelope Proteins) ou encore xp (skin specific proteins). D'expression très tardive dans les kératinocytes granuleux, ils sont parmi les derniers constituants à être incorporés à l'enveloppe cornée. Les transcrits correspondants sont généralement constitués de 2 exons, et codent pour des protéines de 100 acides aminés en moyenne.</p>
C1orf42	NICE-1	99 aa	
LCE3E	LEP17	92 aa	
LCE3D	LEP16; SPRL6A; SPRL6B	92 aa	
LCE3C	LEP15; SPRL3A	94 aa	
LCE3B	LEP14	95 aa	
LCE3A	LEP13	89 aa	
C1orf46	xp33	-	
LCE2D	LEP12; SPRL1A	110 aa	
LCE2C	LEP11	110 aa	
LCE2B	xp5; LEP10; SPRL1B	110 aa	
LCE2A	LEP9	106 aa	
LCE4A	LEP8; SPRL4A	99 aa	
C1orf68	xp32; LEP7	250 aa	
C1orf45	hkPRP	579 aa	
LCE1F	LEP6	118 aa	
LCE1E	LEP5	118 aa	
LCE1D	LEP4	114 aa	
LCE1C	LEP3	118 aa	
LCE1B	LEP2; SPRL2A	118 aa	
LCE1A	LEP1	110 aa	

Contrairement aux protéines SPRR, les 17 protéines LCE ne comprennent pas de répétitions en tandem identifiables. Elles s'en distinguent aussi par leur contenu élevé en serine, glycine et cystéine. Leur nombre pose la question de la redondance de leur fonction. La comparaison de leur séquence en acides aminés et leur localisation génomique a permis d'établir une classification en 3 familles (LCE1, LCE2, LCE3) et 2 gènes isolés (LCE4A et LCE5A) (Jackson et al. 2005). L'expression de ces gènes chez l'homme a été étudiée par PCR quantitative en temps réel dans différents épithéliums pluristratifiés. Les familles 1 et 2 sont les LCE que l'épiderme exprime majoritairement, en particulier LCE1C et dans une moindre mesure LCE2A et LCE2B. Les gènes de la famille LCE3 sont exprimés dans les épithéliums stratifiés non cornifiés tels que l'oesophage et la langue. Les gènes LCE4A et LCE5A sont très faiblement exprimés dans les tissus et conditions testés dans l'étude. D'autre part, seule l'expression des gènes de la famille LCE2 est inductible par le calcium. Enfin, les gènes LCE1 sont surexprimés lors de l'exposition des kératinocytes aux UVB.



Les transglutaminases, directement responsables de la réticulation précoce des protéines d'assemblage et de la consolidation avec les protéines de structure plus tardives, jouent évidemment un rôle crucial dans la cornification. Ces enzymes, produites sous forme de zymogènes, sont maturées par des protéases encore peu caractérisées, et deviennent probablement actives sous l'effet de l'augmentation du calcium intracellulaire. Ces enzymes catalysent la formation d'une liaison covalente entre le groupement gamma-carboxamide de la chaîne latérale d'un résidu glutamine et le groupement epsilon-amino de la chaîne latérale d'un résidu lysine (EC 2.3.2.13). Bien que les différentes isoformes de transglutaminases présentes dans l'épiderme (TGM1, 3 et 5) aient la même activité catalytique, leur redondance n'est qu'apparente.

La TGM1 est une isoforme liée à la face cytosolique de la membrane plasmique par un acide gras et interviendrait, conjointement avec la TGM5, sur l'assemblage précoce de l'enveloppe cornée en liant entre elles les plakines (essentiellement envoplakine et périplakine) et l'involucrine, selon le modèle proposé par Peter Steinert (figure 45). La TGM1 est également responsable de la liaison sur la face externe de l'enveloppe cornée des céramides à très longues chaînes d'acides gras, les ω -hydroxycéramides (cf. paragraphe 2.2.2.2.1.1) (Nemes et al. 1999). Chez l'homme, des mutations du gène *TGM1* sont responsables d'une génodermatose récessive grave, l'ichtyose lamellaire de type 1 (MIM #242300). Le phénotype de bébé collodion à la naissance, parfois associé à un ectropion et/ou eclabium, conduit généralement à une mort néonatale. L'épiderme est hyperkératosique, le profil de céramides altéré, et l'augmentation des pertes hydriques transcutanées témoigne d'un trouble de la barrière. Comme attendu, des souris invalidées pour ce gène présentent un trouble majeur de la barrière épidermique et meurent quelques heures après la naissance (Matsuki et al. 1998). Les granules de kératohyaline s'accumulent dans le cytoplasme des kératinocytes granuleux, et on observe en microscopie électronique un défaut de formation de l'enveloppe cornée et une anomalie ultrastructurale des lipides intercornéocytaires (Kuramoto et al. 2002). Ces arguments convergent vers un rôle indispensable de TGM1 dans l'attachement des

protéines servant d'échafaudage pour la mise en place de l'enveloppe cornée et la fonction barrière.

Les TGM3 et 5 sont quant à elles cytosoliques et interviendraient dans la consolidation de cette première structure via l'intégration de la LOR, des SPRR et des LCE. L'inactivation de la TGM3 chez la souris provoque une létalité embryonnaire très précoce (Leonard M. Milstone, résultats non publiés), ce qui pourrait expliquer qu'il n'existe chez l'homme aucune pathologie associée à des mutations de ce gène. La TGM5 a quant à elle été récemment associée à un syndrome de peau déciduale acrale (*noninflammatory acral peeling skin syndrome*, OMIM #609796) (Cassidy et al. 2005; Kharfi et al. 2009). Cette pathologie est caractérisée par un détachement indolore de lambeaux de couche cornée successif à la formation de bulles à la transition couche granuleuse / couche cornée.

En plus de l'importance majeure des constituants de l'enveloppe cornée, ceux du cornéodesmosome assurant la forte cohésion entre les cornéocytes sont également requis pour la fonction barrière.

2.2.2.2.3. Cornéodesmosomes

Le cornéodesmosome, en tant que structure dérivant directement des desmosomes des couches vivantes de l'épiderme, est composé des mêmes protéines, à savoir, pour la partie extracellulaire assurant la liaison entre les cornéocytes, des cadhérines desmosomales (Desmocolline 1, Desmogléine 1 et 4), auxquelles s'ajoute la cornéodesmosine, protéine sécrétée par le kératinocyte granuleux et retrouvée au niveau du cœur desmosomal avant la cornification.

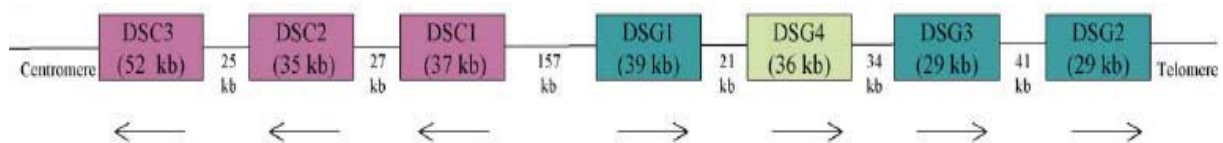


Figure 47. Organisation génomique des cadhérines desmosomales chez l'homme sur le locus 18q12.1 (Kljuic et al. 2003).

Les cadhérines desmosomales, protéines transmembranaires Ca^{2+} -dépendantes assurant la jonction intercellulaire des desmosomes, se divisent en deux familles, les desmocollines (DSC1, DSC2 et DSC3) et les desmogléines (DSG1, DSG2, DSG3 et DSG4) organisées en locus sur le chromosome 18 (figure 47). Le niveau d'expression de ces protéines est lié au stade de différenciation kératinocytaire (figure 48). L'expression de DSG2 et DSC2 est restreinte à la couche basale, DSG3 et DSC3

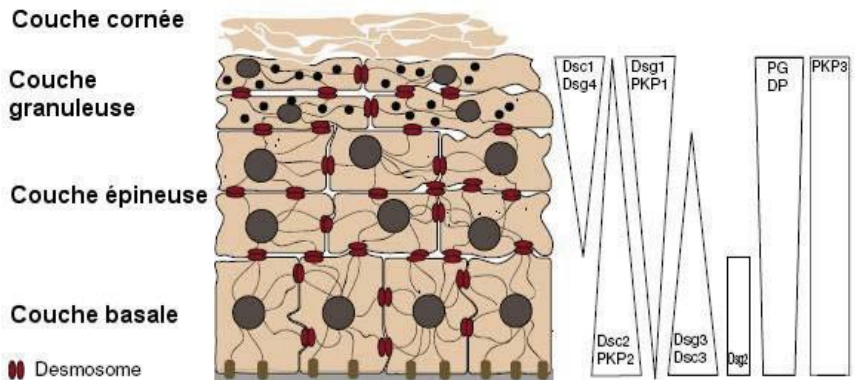


Figure 48. Profil d'expression des composants des desmosomes de l'épiderme (Kottke et al. 2006).

sont présentes dans tout l'épiderme avec un fort gradient vers la couche basale, alors que DSC1 et DSG1 sont au contraire fortement exprimées dans les couches suprabasales. La desmogléine 4, plus récemment identifiée (Kljuic et al. 2003), est exprimée dans la couche granuleuse de l'épiderme interfolliculaire et dans le follicule

pileux (matrice, cellules corticales et gaine épithéliale interne) (Bazzi et al. 2006). Elle est la seule desmogléine présente dans la cuticule de la tige pileuse.

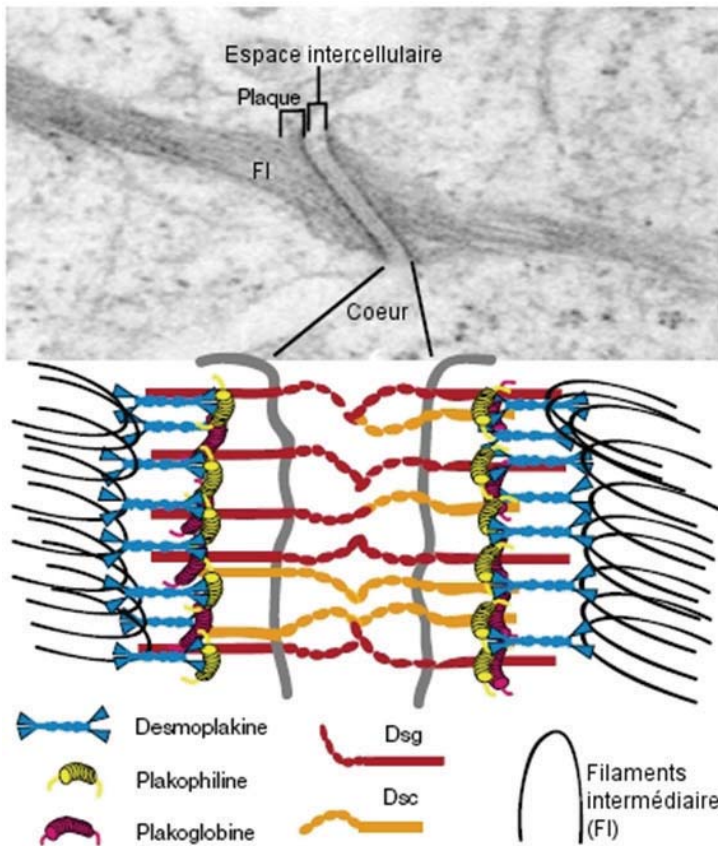


Figure 49. Desmosome de l'épiderme en microscopie électronique et représentation schématique correspondante (Kottke et al. 2006).

Le mode d'interaction entre les cadhérines desmosomales, homophile ou hétérophile, reste encore incertain. Une étude de l'association entre les domaines extracellulaires de DSC2 et DSG2 par la technique du Biacore (résonance plasmonique de surface) montre qu'en présence de Ca^{2+} , DSC2 et DSG2 interagissent préférentiellement en hétérodimères, et seule DSC2 est capable de former des

proposent que desmocollines et

desmogléines pourraient s'associer de façon hétérophile entre 2 cellules adjacentes et formeraient des dimères latéraux (figure 49). De plus, si la fonction de ces protéines transmembranaires met en jeu des interactions entre cellules adjacentes (donc en *trans*), certaines données tendent à montrer que dans de nombreux tissus (dont l'épiderme), leur association est renforcée par des interactions en *cis*, définissant ainsi un état d'hyper-adhérence des desmosomes qui deviennent calcium-indépendants (Garrod et al. 2005; Al-Amoudi et al. 2007). Le passage de l'état « renforcé » à l'état « normal » semble dépendant de l'action de la PKC α sur les protéines de la plaque desmosomale et pourrait jouer un rôle dans la régulation de l'adhérence, notamment au cours de la cicatrisation.

Des souris inactivées pour *Dsc1* présentent un défaut de barrière associé à une acantholyse (Chidgey et al. 2001). Toutefois, la morphologie des desmosomes dans les couches suprabasales est normale comme montré en microscopie électronique, ce qui suggère que la perte de *Dsc1*, seule desmocolline exprimée dans ces cellules, est partiellement compensée par la formation de liaisons homophiles entre desmogléines. Les auteurs n'ont pu mettre en évidence la surexpression d'autres cadhérines desmosomales, toutefois la *Dsg4* n'avait pas encore été identifiée au moment de cette étude. Chez l'homme, des mutations de la *DSG1* sont associées à une kératodermie palmoplantaire de type I, peu sévère (OMIM #148700). Des mutations dans le gène *DSG4* ont été identifiées chez des patients présentant une hypotrichose transmise sur le mode récessif (OMIM #607903), et sont responsables de phénotypes similaires chez le rat et la souris (Bazzi et al. 2005). L'absence de phénotype épidermique suggère une compensation de la perte de *Dsg4*, peut-être liée à la présence de *Dsg1* dans les mêmes couches.

Lors de la transition entre couche granuleuse et couche cornée, les desmosomes changent de morphologie : la plaque est incorporée à l'enveloppe cornée jusqu'à devenir invisible en microscopie électronique et le cœur extracellulaire se densifie. La production au laboratoire d'anticorps monoclonaux de souris immunisées avec des extraits de couche cornée a conduit à l'identification de la cornéodesmosine (CDSN) (Serre et al. 1991; Guerrin et al. 1998). Détectée dans la couche granuleuse où elle est sécrétée via les kératinosomes, la CDSN est localisée au niveau de la partie extracellulaire des desmosomes, au sein desquels elle a vraisemblablement un rôle d'adhérence homophile (Jonca et al. 2002). Sa protéolyse dans la partie supérieure de la couche cornée joue un rôle fondamental lors de la

desquamation (Simon et al. 2001). Récemment, une équipe japonaise a réalisé un KO constitutif du gène *Cdsn* chez la souris (Matsumoto et al. 2008). Les souris homozygotes ne survivent que quelques heures après la naissance et ne possèdent pas de fonction barrière efficace, prouvant ainsi le rôle essentiel de la cornéodesmosine.

Chez l'homme, des mutations non-sens du gène *CDSN* sont responsables de l'Hypotrichose simple du cuir chevelu (HSS, OMIM 146520), génotrichose autosomique dominante caractérisée par la perte progressive et exclusive des cheveux, tous les autres poils du corps étant préservés (Levy-Nissenbaum et al. 2003). Ces mutations induisent la production de formes tronquées de CDSN, correspondant globalement au domaine NH₂-terminal de la protéine qui contient un des domaines riches en boucles glycine. Ces formes tronquées sont retrouvées sous forme d'agrégats dans le derme superficiel du cuir chevelu des patients et à la périphérie des follicules pileux résiduels. L'absence de phénotype observable pour les souris hétérozygotes *Cdsn*^{+/-} (Matsumoto et al. 2008) suggère fortement que le caractère dominant de l'HSS est dû à un gain de fonction de la forme tronquée et non à une haploinsuffisance.

Enfin, le gène *CDSN* est localisé dans le locus *PSORS1* en 6p21.3, locus majeur de susceptibilité au Psoriasis familial (OMIM #177900), maladie inflammatoire chronique dont la physiopathologie est encore mal comprise, mais plusieurs travaux (Helms et al. 2005; Nair et al. 2006) suggèrent fortement que *CDSN* n'est pas le gène responsable de la liaison entre *PSORS1* et la maladie.

2.2.3. Régulation des étapes tardives de la différenciation

2.2.3.1. Voies de signalisation

Si plusieurs voies de signalisation ont été identifiées comme intervenant dans la régulation de la prolifération ou l'initiation de la différenciation, peu de voies ont été caractérisées en ce qui concerne les étapes tardives. Seules les voies de signalisation du calcium, de la vitamine D₃ et des dérivés de l'acide arachidonique, partiellement décrites à ce jour, sont connues pour leur implication directe dans la fonction barrière.

2.2.3.1.1. Voie calcique

L'épiderme présente un gradient de calcium de la profondeur vers la surface des couches vivantes, puis sa concentration diminue dans la couche cornée, comme

démonstré par une analyse ionique de ce tissu aux rayons X (Forslind 1986). *In vitro*, l'augmentation de la concentration de calcium dans le milieu de culture est une méthode très classique pour induire la stratification et la différenciation des kératinocytes (Pillai et al. 1990). La perturbation de la barrière épidermique chez la souris par application topique d'acétone ou par délaminage au ruban adhésif entraîne une disparition locale du gradient calcique épidermique, lequel réapparaît 6h plus tard, parallèlement à la restauration de la fonction de barrière et à la sécrétion de kératinosomes (Menon et al. 1992). Quel que soit la méthode employée pour la perturber, une barrière fonctionnelle est systématiquement requise pour le rétablissement d'un gradient calcique, certaines données suggérant même que le processus serait passif (retour d'un gradient 5h après délaminage au ruban adhésif et pose d'une membrane de Goretex à 4°C, température à laquelle la barrière ne se rétablit que très lentement) (Elias et al. 2002a).

Si le gradient calcique est dépendant de la fonction barrière, l'inverse est aussi vrai si l'on considère que l'expression de plusieurs marqueurs de la différenciation est

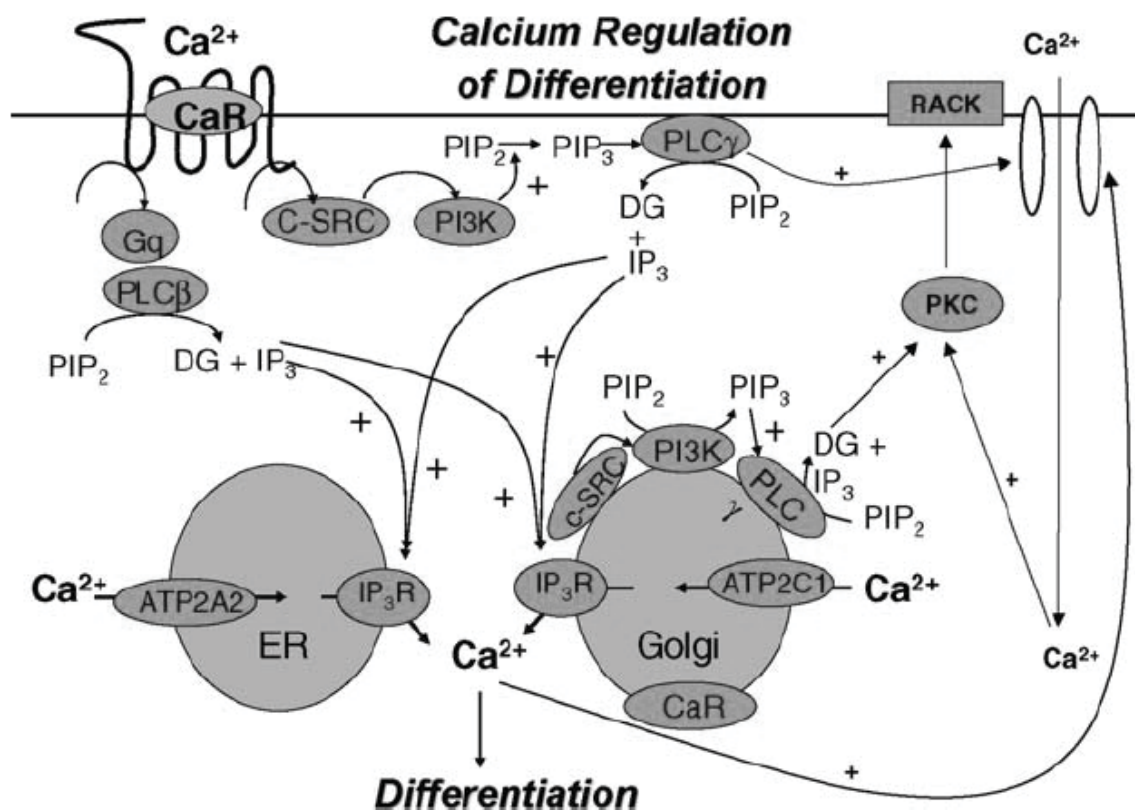


Figure 50. Signalisation calcique médiée par le récepteur membranaire au calcium extracellulaire (Bikle et al. 2005).

dépendante de la concentration calcique, comme démontré sur l'épiderme murin après divers traitements altérant celle-ci (Elias et al. 2002b). De plus, l'activité de nombreuses protéines impliquées dans la différenciation épidermique est directement régulée par la concentration calcique, comme pour les transglutaminases, les cadhérines, les peptidyl-arginine désiminases, la glucosylcéramide synthase, les PKC η et α (via l'activation des PLC), les protéines S100... Le récepteur membranaire au calcium (figure 50), produit du gène *CASR* (Calcium sensing receptor), est requis *in vitro* pour la réponse des kératinocytes au calcium (Tu et al. 2001; Tu et al. 2007). L'inactivation de ce gène chez la souris n'entraîne qu'une altération modérée de la différenciation, en particulier une altération morphologique des granules de kératohyaline probablement liée à la sous-expression de la Flg et Lor, normalement induites par le calcium (Komuves et al. 2002). L'absence de trouble majeur de la barrière permet de supposer l'existence d'un mécanisme alternatif compensant la perte du CaSR. Les ARN pré-messagers du gène *CASR* subissent un épissage alternatif conduisant à la production de deux isoformes, membranaire ou périnucléaire. Ainsi, la fonction du CaSR pourrait également passer par une interaction avec les pompes du RE qui régulent la concentration intracellulaire en calcium (Tu et al. 2007).

De nombreux travaux ont en effet démontré l'importance de ces pompes dans la physiologie épidermique. Les pompes ATP2A2 (ATPase, Ca²⁺-transporting, type 2A, member 2) et ATP2C1 (type 2C, member 1) interviennent dans le maintien d'une concentration calcique très basse dans le cytosol, nécessaire à la signalisation calcique. Des mutations dominantes de ces deux gènes sont responsables respectivement de la maladie de Darier-White (OMIM #124200) et de sa variante de l'acrokératose verruciforme (OMIM #101900), et de la maladie de Hailey-Hailey (OMIM #169600). La maladie de Darier-White, transmise sur le mode autosomique dominant, est liée à des mutations non-sens ou faux-sens du gène *ATP2A2*. Elle se caractérise par la présence de bulles, une ichtyose et une dystrophie unguéale. L'analyse histologique montre un détachement des kératinocytes au niveau de la couche épineuse, accompagné d'une acantholyse. En microscopie électronique, une perte de cohésion au niveau des desmosomes et une désorganisation du réseau de filaments intermédiaires qui s'agrègent à la périphérie du noyau sont mises en évidence. Ces observations suggèrent que la maladie résulte d'un défaut de formation des complexes desmosomes - filaments intermédiaires de kératines. L'acrokératose

verruciforme, d'aspect clinique différent (lésions hyperkératosiques et verruciformes sur la face dorsale des extrémités), est causée par une mutation faux-sens (P602L) du gène *ATP2A2*, ce qui fait de cette maladie un variant allélique de la maladie de Darier. Tandis que les mutations non-sens, entraînant la dégradation de l'ARNm muté par le NMD, plaident en faveur d'un mécanisme d'haploinsuffisance, les différences phénotypiques liées à certaines mutations faux-sens (not. P602L) suggèrent en plus un éventuel gain de fonction des protéines mutées. La maladie de Hailey-Hailey est une affection voisine de la maladie de Darier-White mais où les ongles ne sont pas atteints. Elle est transmise également de façon autosomique dominante ; les nombreuses mutations mises en évidence entraînent généralement une perte de fonction du transporteur, suggérant là aussi un mécanisme d'haploinsuffisance.

Le mécanisme physiopathologique de ces trois affections n'a pas été clairement établi. Des modèles animaux de souris, dont un allèle des gènes *ATP2A2* (Ji et al. 2000) ou *ATP2C1* (Okunade et al. 2007) a été inactivé (seuls les hétérozygotes sont viables), ne reproduisent pas le phénotype épidermique humain. Les études ont mis en évidence une perturbation de l'exocytose Ca^{2+} -dépendante (Zhao et al. 2001) ou une désorganisation du système de Golgi et une accumulation de gouttelettes lipidiques (Okunade et al. 2007), mais pas d'altération des jonctions intercellulaires. Les souris hétérozygotes des deux modèles présentent de plus une prédisposition aux carcinomes épidermoïdes. Les mutations des gènes *ATP2A2* ou *ATP2C1* conduisent donc à des phénotypes très similaires, mais extrêmement différents entre l'homme (troubles de la différenciation) et le modèle murin (carcinomes), sans qu'il soit possible d'expliquer les différences entre les mécanismes mis en jeu.

Enfin, la protéine sécrétée SLURP-1 est également impliquée dans la signalisation calcique (Arredondo et al. 2005). Chez l'homme, des mutations dans le gène *SLURP1* sont ainsi associées au Mal de Meleda (OMIM #248300), une maladie transmise sur le mode autosomique récessif et caractérisée par une kératodermie palmoplantaire inflammatoire débordant sur la face dorsale des mains et des pieds, fréquemment associée à d'autres signes cliniques tels qu'une brachydactylie et des anomalies des ongles (Fischer et al. 2001). Le Mal de Meleda constitue le premier exemple de kératodermie palmoplantaire due à des mutations dans un gène codant une protéine sécrétée. SLURP-1 a une action régulatrice aussi bien sur l'homéostasie épidermique que sur l'inflammation, qui passe par la potentialisation du signal calcique induit par le récepteur nicotinique à l'acétylcholine (ACh) $\alpha 7$ -nAChR. Si

plusieurs types de récepteurs nicotiques (nAChR) et muscariniques (mAChR) sont présents dans l'épiderme, le récepteur $\alpha 7$ -nAChR est essentiellement exprimé dans le haut de la couche épineuse et la couche granuleuse ; or l'acétylcholine produite par les kératinocytes s'accumule particulièrement dans la couche granuleuse - pour revue (Kurzen and Schallreuter 2004). L'activation du récepteur $\alpha 7$ -nAChR favorise l'entrée de calcium dans les cellules, favorisant la différenciation kératinocytaire *in vitro* comme *in vivo* (Arredondo et al. 2002; Chimienti et al. 2003).

Si les relations entre les différents acteurs de la signalisation calcique (voie SLURP-1/ACh/ $\alpha 7$ -nAChR, voie du CaSR et pompes calciques du RE) restent inconnues à ce jour, la vitamine D₃ est connue pour son action synergique.

2.2.3.1.2. Voie de la vitamine D₃

La peau est la source majeure de vitamine D₃ (cholécalférol), produite grâce à une réaction photochimique à partir du 7-déhydroxycholestérol au niveau de l'épiderme. Pour être biologiquement active, la vitamine D₃ doit être convertie, *in situ* ou dans le foie, en 1,25-dihydroxycholécalférol ou 1,25(OH)₂D. Cette molécule lipophile se lie à un récepteur nucléaire, le VDR. Le complexe formé interagit alors avec son partenaire, le récepteur X aux rétinoïdes (RXR), l'hétérodimère VDR-RXR régulant la transcription par liaison à des éléments de réponse spécifiques (VDRE, élément de réponse à la vitamine D) dans les régions promotrices de gènes cibles. Le 1,25(OH)₂D augmente l'expression de l'involucrine, des transglutaminases, des phospholipases C (PLC) ou encore des PAD1 et 3 (Su et al. 1994; Gibson et al. 1996; Palmer et al. 2008). Une étude menée au moyen de puces Affymetrix ciblant 12600 gènes a permis d'identifier à une plus grande échelle les gènes dont l'expression varie suite à un traitement par la vitamine D₃ de kératinocytes immortalisés (lignée KerTr) ou de kératinocytes primaires humains normaux (NHEK) (Lu et al. 2005). Parmi les 82 gènes surexprimés figurent plusieurs membres des PADI (1, 2 et 3), des kallikréines (5, 6, 7, 8, 10, 13) et des serpinines (B1, 6 et 9), mais aussi l'involucrine, la cystatine M/E et les facteurs de transcription c-Fos et KLF4. SPRR1 et TGM1 font partie des 16 gènes réprimés. Ces résultats confirment le caractère pléiotropique de la vitamine D₃, mais ne permettent pas d'identifier les cibles directes du VDR.

L'action du 1,25(OH)₂D dans la différenciation kératinocytaire passe par de nombreuses voies de signalisation communes avec le calcium (figure 51) (Bikle and Pillai 1993). Premièrement, l'expression du CaSR est augmentée par le 1,25(OH)₂D, entraînant une plus grande sensibilité du kératinocyte à la concentration en calcium

extracellulaire (Ratnam et al. 1999). De plus, plusieurs membres de la famille des phospholipases C, également cruciale pour la signalisation calcique dans les kératinocytes, sont induites par le $1,25(\text{OH})_2\text{D}$ (Pillai et al. 1995) aussi bien que par le calcium. Enfin, le calcium et le $1,25(\text{OH})_2\text{D}$ agissent en synergie notamment pour activer la transcription de l'involucrine et des TGM (Su et al. 1994). Dans le promoteur du gène de l'involucrine, des éléments de réponse au calcium (CaRE) et VDRE ont été retrouvés très proches physiquement. Des mutations dans le CaRE bloquent à la fois l'action du calcium et du $1,25(\text{OH})_2\text{D}$, alors qu'une mutation du VDRE bloque uniquement la réponse au $1,25(\text{OH})_2\text{D}$ (Ng et al. 2000; Bikle et al. 2002).

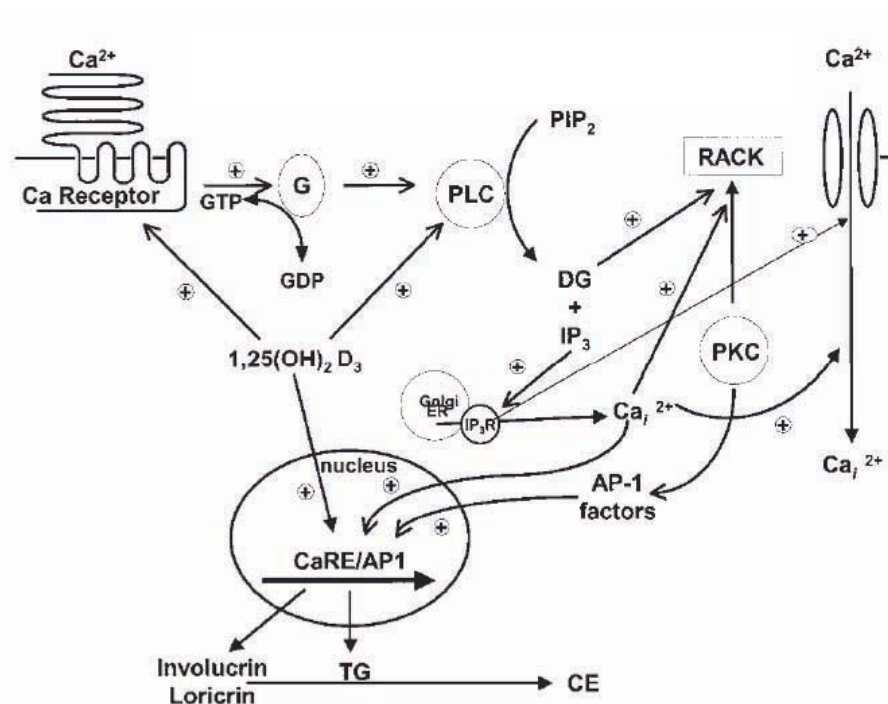


Figure 51. Synergie entre l'action du calcium et de la vitamine D3 (Bikle et al. 2004).

L'ensemble de ces résultats souligne la capacité du $1,25(\text{OH})_2\text{D}$ à renforcer l'action du calcium au cours de la différenciation, bien que la vitamine D_3 ne soit pas strictement requise pour la différenciation épidermique. En effet, des souris déficientes pour le VDR présentent un rachitisme et une alopécie, mais pas de défaut majeur de la différenciation épidermique, malgré une diminution des niveaux d'expression de l'INV, FLG, LOR, ainsi qu'une disparition des granules de kératohyaline (Xie et al. 2002). Ce phénotype rappelle celui, chez l'homme, du rachitisme dépendant de la vitamine D de type II (OMIM #277440), lui aussi

généralement associé à une alopécie, sans autre manifestation épidermique. Toutefois, certains allèles du gène *VDR* semblent être plus rarement trouvés chez les personnes atteintes de psoriasis, et être associés à une meilleure réponse au calcipotriol, un dérivé de la vitamine D₃ (Halsall et al. 2005).

2.2.3.1.3. Voie de l'acide arachidonique

L'acide arachidonique est un acide gras naturellement présent à la membrane plasmique sous la forme de phospholipide, et libéré sous l'action des phospholipases exprimée dans la couche granuleuse de l'épiderme. L'acide arachidonique est progressivement transformé en divers produits de la voie de la 12(R)lipoxygénase (figure 52). Plusieurs enzymes-clés de cette voie sont exprimées dans l'épiderme et requises pour la fonction barrière (voir aussi Table 3, dans le paragraphe 2.3.1) : les gènes *ALOX12B* et *ALOXE3*, responsables de la transformation de l'acide arachidonique en hepoxylin A₃ (HXA₃), sont fréquemment mutés dans des cas d'érythrodermie ichtyosiforme congénitale sèche (Jobard et al. 2002), le lien de cause à effet étant notamment confirmé par un modèle de souris mutantes pour *Alox12b* (Moran et al. 2007) ou délétées pour ce même gène (de Juanes et al. 2009). De plus, la transformation du TXA₃ en 20-COOH-TXA₃ est catalysée par plusieurs enzymes,

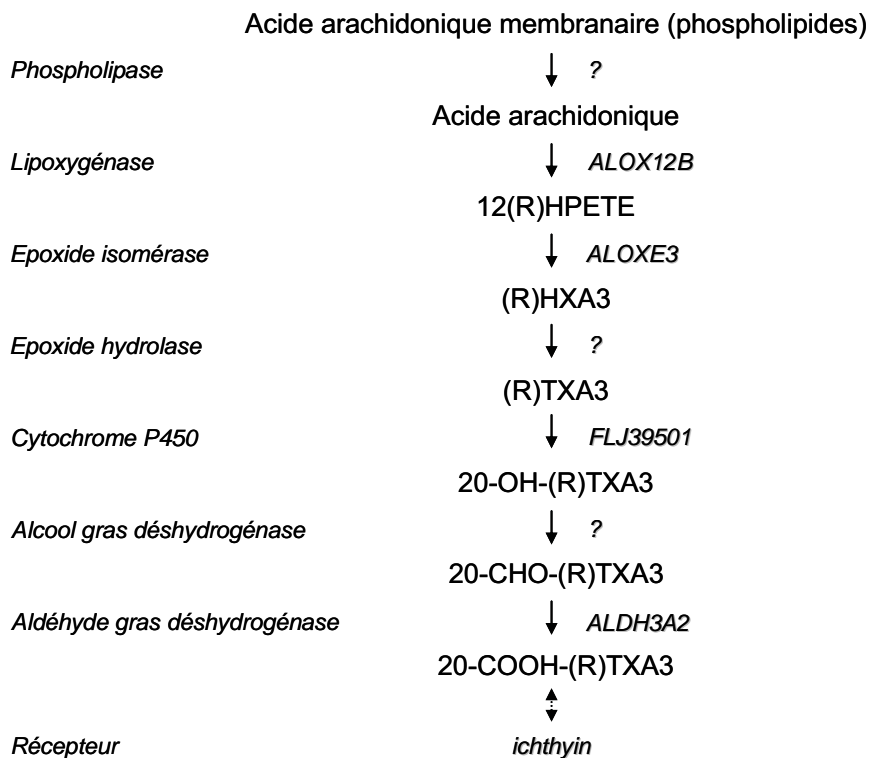


Figure 52. Voie de l'acide arachidonique.
D'après le modèle proposé dans (Lefevre et al. 2006).

notamment une protéine apparentée à la famille des cytochromes P450 (FLJ39501/CYP4F22) et une aldéhyde déshydrogénase (ALDH3A2), dont les gènes sont respectivement mutés dans l'ichtyose lamellaire de type III (Lefevre et al. 2006) et le syndrome de Sjögren-Larsson (De Laurenzi et al. 1996). Toutes ces données suggèrent que le produit final de cette voie métabolique (probablement la 20-COOH-TXA₃) est requis pour la fonction barrière de l'épiderme. L'ichthyin, une protéine à plusieurs domaines transmembranaires codée par le gène NIPAL4, pourrait être le récepteur du 20-COOH-TXA₃, ce gène étant notamment muté dans un sous-type d'ichtyose congénitale autosomique récessive (ARCI1, OMIM#612281) (Lefevre et al. 2004). Les protéines responsables de la transduction du signal de l'ichthyin, potentiellement des protéines G, restent encore à identifier.

2.2.3.2. Facteurs de transcription

2.2.3.2.1. Facteurs ubiquistes

Pendant de nombreuses années, les études menées sur la régulation de l'expression des gènes de la différenciation épidermique n'ont pas conduit à l'identification de facteurs spécifiques, et ont conduit à imaginer un modèle général où l'expression différentielle de facteurs ubiquistes suffirait à expliquer toute la régulation de la différenciation épidermique. De nombreuses études, dont quelques résultats sont cités à titre d'exemple dans la table 2, ont mis en évidence l'implication de plusieurs familles telles que AP-1, AP-2, Sp1, Sp3, CREB-1, etc, ainsi que d'autres facteurs d'expression plus restreinte, tels que les facteurs POU/Oct et ETS, dans la régulation de nombreux gènes spécifiques de la différenciation.

La famille AP-1, composée des facteurs Fos (c-Fos, FosB, Fra-1 et Fra-2), Jun (c-Jun, JunB et JunD) et des cofacteurs Maf et ATF, est impliquée dans la régulation d'un grand nombre de gènes de la différenciation : *FLG* (Jang et al. 2000), *INV* (LaPres and Hudson 1996), *LOR* (Jang and Steinert 2002), *DSG1* (Adams et al. 1998), *TGM1* (Lu et al. 1995), *KRT10* (Maytin et al. 1999)... Son site consensus est cependant présent dans 32,6% des promoteurs de vertébrés, ce qui met en évidence son caractère hautement ubiquiste. Les différents membres de cette famille sont exprimés à des niveaux variables selon les couches de l'épiderme (figure 53). L'étude de la fonction des facteurs AP-1 dans l'épiderme par inactivation chez la souris est limitée par la redondance partielle entre les différents membres de la famille :

l'inactivation de c-Jun par KO au niveau de l'épiderme murin a mis en évidence son rôle dans la réponse à l'EGF, mais sans phénotype spontané (Zenz et al. 2003).

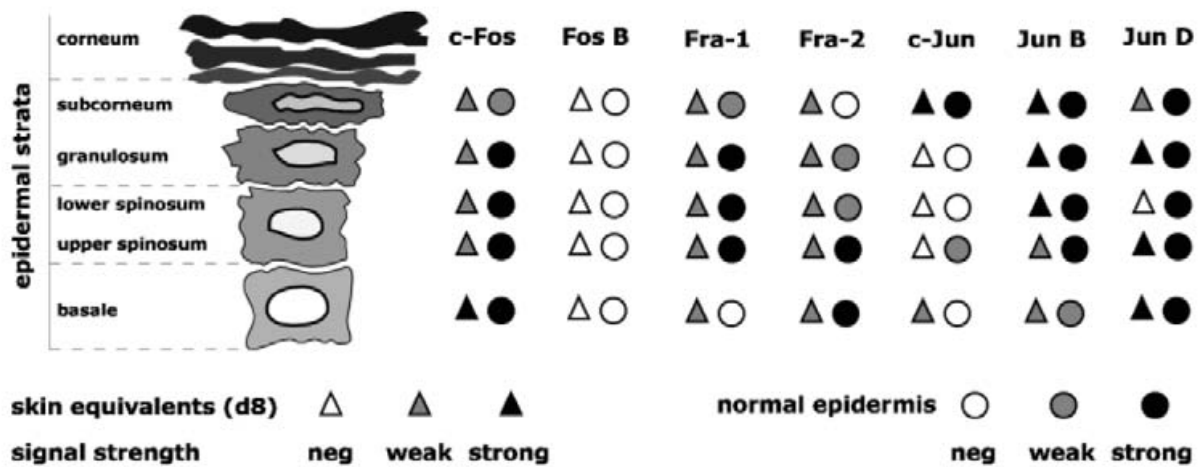


Figure 53. Expression des protéines AP-1 dans l'épiderme (Mehic et al. 2005).

La famille AP-2 comprend cinq membres : AP-2 α , AP-2 β , AP-2 γ , AP-2 δ et AP-2 ϵ (seule isoforme spécifique de l'épiderme) (Tummala et al. 2003). Des sites de fixation pour ces facteurs, fonctionnant en hétéro ou homodimères, ont été mis en évidence aussi bien pour des gènes exprimés dans la couche basale que dans les couches suprabasales (notamment les promoteurs des gènes *KRT14* et *KRT1*), et in vitro selon le facteur AP-2, la construction et le type cellulaire, ces facteurs sont décrits comme activateurs ou comme répresseurs. Les ablations conditionnelles d'AP-2 α (Wang et al. 2006) ou AP-2 γ (Guttormsen et al. 2008) ont essentiellement souligné la redondance entre les facteurs AP-2 ; seule une légère perturbation dans la signalisation de l'EGF et un retard dans l'établissement de la fonction barrière ont été mis en évidence respectivement par ces deux modèles. L'inactivation simultanée de ces deux gènes a livré plus d'information sur leur rôle dans l'épiderme, révélant la fonction des facteurs AP-2 et de la voie Notch dans la transition entre prolifération et différenciation, via la régulation des facteurs CCAAT/enhancer binding protein (Wang et al. 2008).

Ce modèle impliquant divers facteurs ubiquistes a été progressivement complété par la découverte de facteurs plus spécifiques qui pourraient agir en synergie avec les facteurs ubiquistes.

Gènes codant pour protéines de la différenciation épidermique

Gènes	Facteurs de transcription impliqués	Commentaires	Références
<i>Profilaggrine</i>	AP-1: c-Jun, c-Fos, JunB Protéines à domaine POU (Oct6) Ets	La région proximale du promoteur de la profilaggrine lui confère une haute spécificité pour les kératinocytes via l'interaction avec l'hétérodimère c-Jun/c-Fos sur un site AP-1. D'autre part des protéines à domaine POU (Oct1, Skn1/1 et Oct6) agissent en stimulant ou antagonisant l'activité dépendante de c-Jun sur le promoteur. La synthèse de profilaggrine dépend donc d'une fine balance dans les taux et les activités des facteurs AP-1 et des protéines à domaine POU.	Jang et al., 2000
<i>Involucrine</i>	AP-1 : JunD, JunB, Fra-1 Sp1 p300 P/CAF	Le gène de l'involucrine est régulé par un promoteur proximal et par une région distale. Un large complexe multiprotéique incluant les facteurs Fra-1, p300, P/CAF, JunD, JunB, and Sp1 sur le site AP1-5 est nécessaire à l'expression correcte du gène de l'involucrine. Un site AP-1 distal et un site AP-1 proximal sont requis.	Crish et al., 2008
<i>Loricrine</i>	AP-1 : JunB, c-Jun Sp1/Sp3 CREB-1/CREM/ATF-1 p300/CBP	Implication d'un complexe multiprotéique dans la région proximale du promoteur. Il y a compétition entre différents facteurs en fonction du stade de différenciation du kératinocyte. Dans les cellules prolifératives l'expression de la loricrine est inhibée par l'action des protéines JunB, Sp3, CREB-1/CREM/ATF-1 et AP2-like (keratinocyte-specific repressor-1 (KSR-1). Alors que dans les kératinocytes différenciés, les niveaux de ces protéines diminuent, ce qui va permettre la fixation d'un régulateur non identifié, de Sp1, c-Jun et p300/CBP, permettant l'activation de la transcription de la loricrine. Les facteurs CREB et c-jun entre en compétition au niveau du même site AP-1 induisant des effets antagonistes sur la transcription du gène de la loricrine.	Jang et al., 2002
<i>SPRR</i>	SPRR1A : AP-1	Les gènes des SPRR sont organisés en tandem dans une région de 220 kb dans le CDE. 14 gènes chez la souris et 9 chez l'homme. SPRR1A : un promoteur proximal (site AP-1 avec effet enhancer) + région distale (Sark, 1998 #192). Les promoteurs minimums des autres SPRR n'ont pas été étudiés dans les kératinocytes. Plusieurs CNS ont été identifiés et permettraient une régulation coordonnée de ces gènes dans les kératinocytes différenciés.	Martin et al., 2004
<i>TGM 1</i>	AP-1 : c-Jun et JunD Cre Sp1	Promoteur minimal : sites AP1 et CRE, implication des facteurs c-Jun et JunD. Une région distale contenant des sites de liaison pour API et Sp1 sont conservés et responsables d'une activation transcriptionnelle dans des souris transgéniques et des kératinocytes en culture. Comme pour l'involucrine un site AP-1 proximal et un site AP-1 distal sont requis.	Jessen et al., 2000 Phillips et al., 2004
<i>TGM 3</i>	Sp1 Ets-like sites	Promoteur proximal : Sp1 et Ets-like sites	Lee et al., 1996
<i>CAPNS1</i>	AP-1 NRF-1	Promoteur proximal : NRF-1 et AP-1	Asangani et al., 2008
<i>Envoplakine</i>	Sp1 et Sp3		Maata et al., 2000
<i>Caspase 14</i>	AP-1 NFkB		Ballaun et al., 2008

Table 2 : FT impliqués dans la régulation de l'expression de protéines fondamentales de la différenciation épidermique. (thèse V. Adoue 2008)

2.2.3.2.2. Facteurs d'expression tissulaire restreinte

A ce jour, seul un petit nombre de facteurs de transcription non-ubiquistes ont été identifiés pour leur rôle dans les étapes tardives de la différenciation et/ou la mise en place de la fonction barrière de l'épiderme. Il s'agit des facteurs DLX3 (Distal-less homeobox 3), KLF4 (Krüppel-like factor 4), GRHL3 (Grainyhead-like 3), GATA3 (GATA binding protein 3) et BLIMP1/PRDM1 (PR domain containing 1, with ZNF domain).

2.2.3.2.2.1. DLX3

Le facteur DLX3 appartient à la famille des facteurs de transcription à homéodomaine apparentés au gène *distal-less* (*dll*) identifié chez *D. melanogaster*. Cette famille comporte six membres chez les mammifères, répartis en tandem sur trois loci distincts (*DLX1* et *DLX2* en 2q31.1, *DLX3* et *DLX4* en 17q21.33, *DLX5* et *DLX6* en 7q21.3).

Le rôle du facteur de transcription DLX3 dans la morphogenèse et la physiologie de l'épiderme et de ses annexes a été mis en évidence grâce à plusieurs modèles murins. Dans l'épiderme murin, l'expression de *Dlx3* est restreinte aux couches différenciées de l'épiderme, essentiellement la couche granuleuse, ainsi que dans le follicule pileux (figure 54). De fait, des souris exprimant ce facteur de façon ectopique dans la couche basale (souris transgéniques *K5-XDlx3*) présentent un trouble létal de la fonction barrière (Morasso et al. 1996), avec un épiderme atrophique et parakératosique, une diminution du nombre de granules de kératohyaline dans la couche granuleuse, et une expression de plusieurs marqueurs tardifs de la différenciation dans la couche basale (notamment *Flg* dont les auteurs montrent qu'il est une cible directe de *Dlx3*).

Toutefois, en raison de son rôle crucial dans le développement du placenta au cours de l'embryogenèse (Morasso et al. 1999), l'exploration directe du rôle de *Dlx3* par inactivation chez la souris a longtemps été impossible à réaliser. Récemment, la

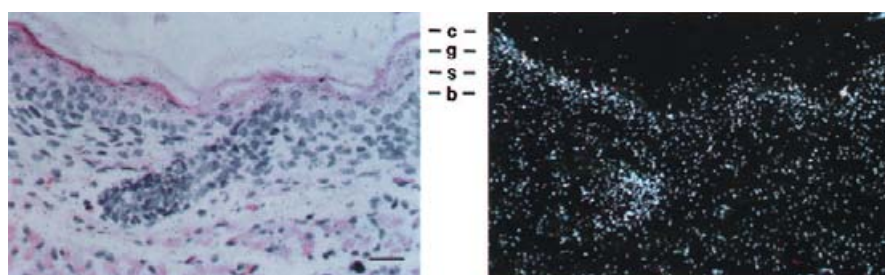


Figure 54. Expression de *Dlx3* dans l'épiderme murin (Morasso et al. 1996).

génération de souris invalidées spécifiquement au niveau de l'épiderme pour *Dlx3* a amené à reconsidérer l'importance de ce facteur dans l'établissement de la fonction barrière (Hwang et al. 2008). En effet, ces souris ne présentent aucun phénotype cutané, mais présentent une alopecie totale démontrant le rôle de *Dlx3* dans la formation et le cycle des follicules pileux. L'absence de phénotype cutané est en accord avec le syndrome tricho-dento-osseux (OMIM #190320) observé chez l'homme, qui est causé par des mutations dominantes situées dans l'exon 3 du gène *Dlx3* (délétions entraînant un décalage du cadre de lecture) et se caractérise par des cheveux cassants, une amélogenèse imparfaite et des malformations des os du crâne, mais aucune anomalie au niveau de l'épiderme. Cette observation suggère une redondance (bien qu'aucun autre facteur DLX n'ait été décrit dans l'épiderme), ou

plus probablement une compensation par d'autres facteurs exprimés dans les kératinocytes différenciés de l'épiderme interfolliculaire.

2.2.3.2.2.2. KLF4

KLF4 est le premier facteur de transcription à avoir été identifié comme spécifiquement requis pour les étapes tardives de la différenciation et jouant un rôle essentiel dans la mise en place de la fonction barrière. Initialement identifié au niveau du colon (Shields et al. 1996), d'où son appellation de « gut-enriched Krüppel-like factor (GKLF) », il est également fortement détecté dans les noyaux des kératinocytes suprabasaux de l'épiderme (figure 55A). Son inactivation par KO chez des souris conduit à une mort rapide

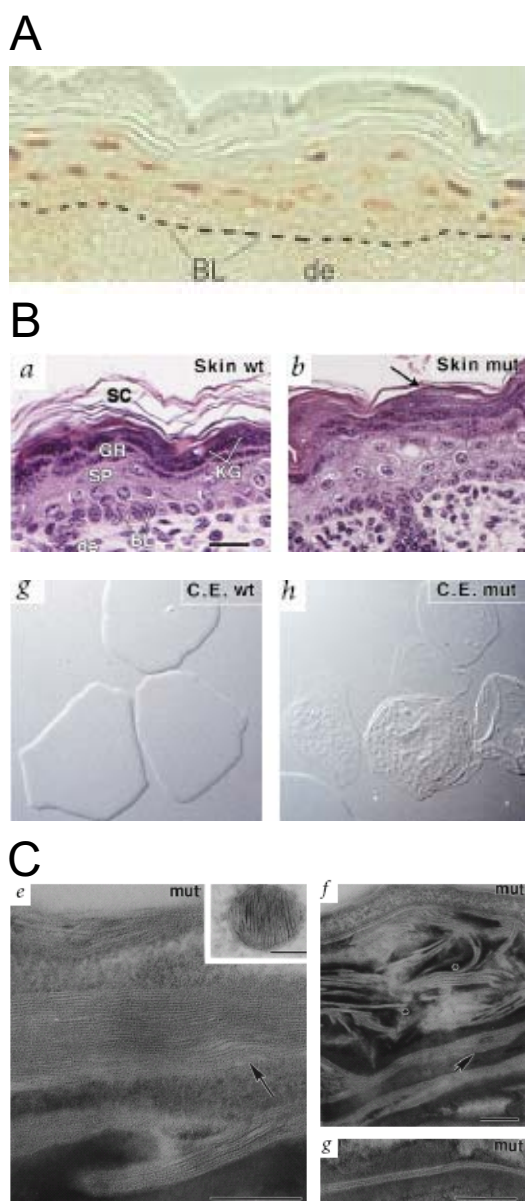


Figure 55. Inactivation de *Klf4* dans l'épiderme murin (Segre et al. 1999). (A) Immunodétection de *Klf4* dans l'épiderme de souris sauvages. (B) Coupes d'épiderme et préparations d'enveloppes cornées de souris sauvages ou *klf4*^{-/-}. (C) Défauts dans la formation des lamellae lipidiques.

dans les heures suivant la naissance, la fonction barrière de l'épiderme étant fortement altérée (Segre et al. 1999). Le tissu présente des anomalies morphologiques strictement limitées aux couches granuleuses et cornées, avec notamment un nombre fortement réduit de granules de kératohyaline, des kératinocytes granuleux moins aplatis que ceux des souris sauvages, et une parakératose (figure 55B). Les enveloppes cornées préparées à partir de peau de souris KO présentent également un aspect irrégulier, fripé et rugueux, alors qu'elles sont normalement polygonales, rigides et lisses. Les auteurs proposent que les anomalies structurales suggérées par l'aspect de ces enveloppes cornées soient à l'origine des altérations observées au niveau des lamellae lipidiques (lamellae désorganisées voire discontinues, ondulées et d'épaisseur variable), les kératinosomes ayant une apparence et un contenu qui semblent normaux (figure 55C). La recherche des cibles potentielles de Klf4 par hybridation soustractive entre les ARNm de ces souris et ceux de souris sauvages a permis de mettre en évidence une surexpression, en absence de Klf4, de trois protéines intégrées à l'enveloppe cornée : SPRR2A, PAI2 (Plasminogen Activator Inhibitor 2, aussi nommé SERPINB2, SERine Protease INhibitor B2) et la répétine. SPRR2A ayant un site de fixation pour Klf4 dans son promoteur, les auteurs suggèrent que Klf4, à l'image de certains autres membres des Klf, pourrait avoir une action de répresseur transcriptionnel. Lorsque Klf4 est exprimé de façon ectopique dans la couche basale de souris transgéniques (Krt5-tTA : TRE-*Klf4*), la mise en place de la fonction barrière *in utero* est accélérée, passant de E16.5-E17.5 à E15.5-E16.5 (Jaubert et al. 2003) (voir l'Annexe 1 pour les différents stades de la formation de l'épiderme au cours du développement murin). L'expression des marqueurs principaux de la différenciation (K1, Lor, Flg) est également plus précoce.

Un second travail d'identification des gènes régulés directement ou non par Klf4, par comparaison des profils d'expression entre épiderme *Klf4*^{+/+} et *Klf4*^{-/-} sur puce à ADN, a permis d'identifier non seulement *Spr1b* et *Spr2a* comme cibles réprimées par Klf4, mais aussi les connexines 26 et 30 (Djalilian et al. 2006) (figure 56). De plus, la régulation de la connexine 26 par Klf4 semble directe, celui-ci liant sa région promotrice et réprimant son activité *in vitro*. Les auteurs suggèrent que la répression de *Cx26* par Klf4 pourrait avoir un rôle physiologique afin d'éviter un relargage trop important d'ATP dans l'espace extracellulaire via les connexons non-appariés des kératinocytes, ce qui entraînerait l'activation de récepteurs

A

Comparison of *Klf4*^{-/-} and psoriatic skin expression profiles

Gene symbol	Gene name	Psoriasis	<i>Klf4</i> ^{-/-}
ACPP	Acid phosphatase prostate	3.0	3.9
CTSHPC	Cathepsin C	2.4	1.9
CLCA2	Chloride channel calcium activated 2	2.5	3.3
DSC2	Desmocollin 2	7.0	2.5
EPRG	Epiregulin	2.0	11.4
ERO1-like	ERO1-like	2.6	2.5
EVA1	Epithelial V-like antigen 1	4.0	1.8
GJB2(CX26)	Gap junction protein, β 2	>10	3.4
GJB6(CX30)	Gap junction protein, β 6	5.2	3.4
GK	Glycerol kinase	5.9	3.9
HK2	Hexokinase 2	2.6	3.4
HIF1a	Hypoxia-inducible factor 1 α	2.3	1.9
IFITM1	Interferon-induced transmembrane 1	2.2	1.8
IL1R1N	Interleukin 1 receptor antagonist	2.4	2.4
MAD	MAX dimerization protein	3.2	2.7
SLC6A14	Solute carrier family 6, member 14	>10	3.3
SPRR1B	Small proline-rich protein 1B	4.6	>10
SPRR2A	Small proline-rich protein 2A	5.9	5.0

Levels of gene upregulation in psoriatic and *Klf4*^{-/-} skin. *Klf4*^{-/-} expression levels were determined by microarray hybridization, and psoriatic expression profiling data are published results from Bowcock laboratory (2).

B

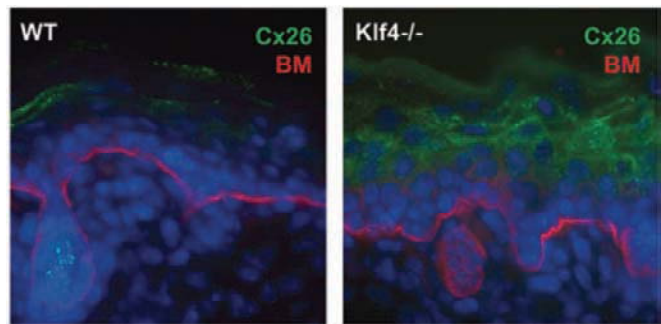
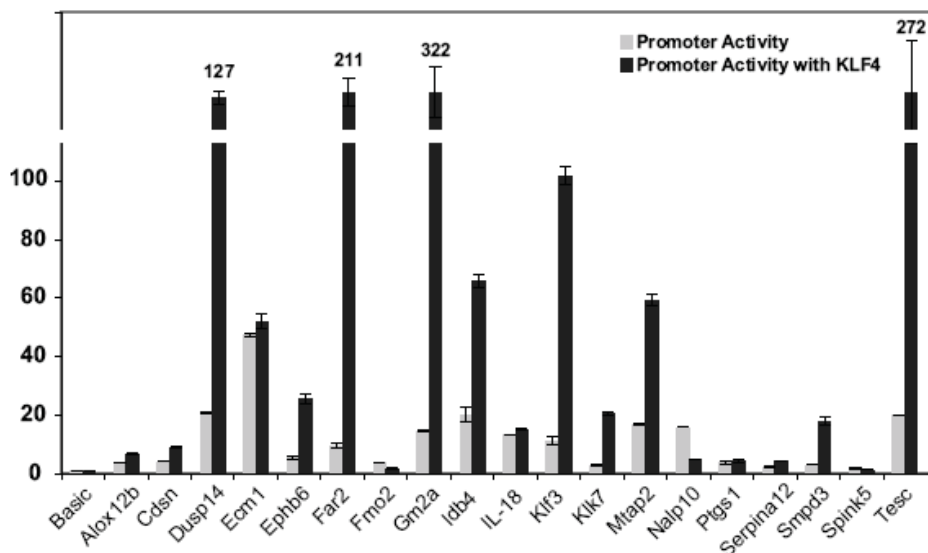


Figure 56. (A) Gènes surexprimés dans l'épiderme de souris *Klf4*^{-/-} ou dans l'épiderme humain psoriasique, comparé à un épiderme normal. (B) Immunofluorescence contre la connexine 26 sur coupe d'épiderme de souris sauvage ou mutée. (Djalilian et al. 2006)

purinergiques et la perturbation de la régulation du calcium intracellulaire, hypothèse corroborée par un modèle murin surexprimant *Cx26* sous dépendance du promoteur humain de l'involucrine. Une autre étude, comparant sur puces à ADN et par PCR quantitative les profils d'expression génique d'épiderme de souris sauvages ou *Klf4*^{-/-}, a permis également d'identifier plusieurs cibles potentielles de *Klf4* (Patel et al. 2006), notamment des gènes connus pour leur fonction cruciale dans

A



B

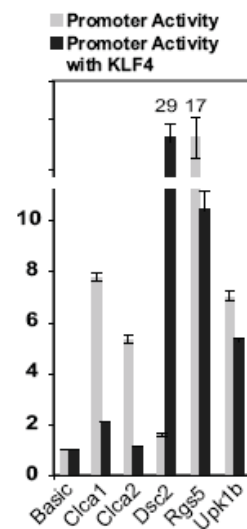


Figure 57. Régulation de l'activité du promoteur proximal de plusieurs gènes d'expression tardive par *Klf4* (Patel et al. 2006). (A) Gènes sous-exprimés dans l'épiderme des souris *kfl4*^{-/-}. Le test de transactivation in vitro indique une activation directe du promoteur proximal de 12 gènes sur 19. (B) Gènes surexprimés dans l'épiderme des souris *kfl4*^{-/-}. Pour les gènes *clca1* et *clca2*, la répression par *Klf4* semble directe.

l'épiderme tels que *Cdsn*, *Klk7*, *Serpina12* et *Spink5*, ainsi que *Alox12b* et *Far2*, tous deux impliqués dans le métabolisme des lipides, la transcription de la plupart de ces cibles semblant être directement activée par la fixation de Klf4 sur leur promoteur proximal (figure 57), contrairement aux travaux précédents qui attribuaient un rôle de répresseur à Klf4.

Au final, KLF4 semble donc réguler plusieurs gènes impliqués dans divers aspects de la différenciation kératinocytaire (signalisation, métabolisme lipidique, enveloppe cornée, desquamation...). De plus, son rôle d'activateur ou de répresseur transcriptionnel suivant les cibles suggère que son action est probablement modulée par d'autres facteurs de transcription, comme cela a été démontré pour GRHL3 et LMO4 par exemple.

2.2.3.2.2.3. GRHL3

Grainyhead-like 3, aussi nommé *Get-1*, est membre d'une famille de trois gènes, orthologues de *grainyhead* identifié chez *D. melanogaster*. Initialement décrit chez les mammifères comme un gène essentiel pour la fermeture du tube neural pendant l'embryogenèse (Ting et al. 2003), *Grhl3* joue aussi un rôle essentiel dans la mise en place de la fonction barrière de l'épiderme. Le gène est majoritairement exprimé au cours du développement murin dans l'ectoderme de surface dès E10.5, puis persiste dans l'épiderme en formation, dans les follicules pileux et vibrisses (cortex) ainsi que plusieurs autres épithélia de recouvrement tels que l'épithélium buccal ou nasal (Kudryavtseva et al. 2003) ; à partir de E15.5 ainsi que chez le nouveau-né et l'adulte, l'expression est limitée aux couches suprabasales de l'épiderme interfolliculaire. Des souris dont le gène *Ghrl3* a été constitutivement inactivé meurent en période périnatale, avec un phénotype caractéristique (*spina bifida* et *curly tail*) dû aux anomalies du tube neural, et ne présentent pas de barrière fonctionnelle (Ting et al. 2005). L'épiderme montre une hyperacanthose, une hyperkératose ainsi qu'une parakératose (figure 58A). Au niveau ultrastructural, les enveloppes cornées sont plus fines, et les lamellae lipidiques sont quasi absents, ce qui s'expliquerait par le fait que les kératinosomes de la couche granuleuse semblent dépourvus de contenu (figure 58B). Certaines de ces altérations ressemblant au phénotype observé pour des souris invalidées pour *Klf4* ou *Tgm1*, l'expression de ces deux marqueurs a été quantifiée, et si le niveau d'expression de *Klf4* est inchangé, l'ARNm codant pour la transglutaminase 1 est cinq fois moins abondant

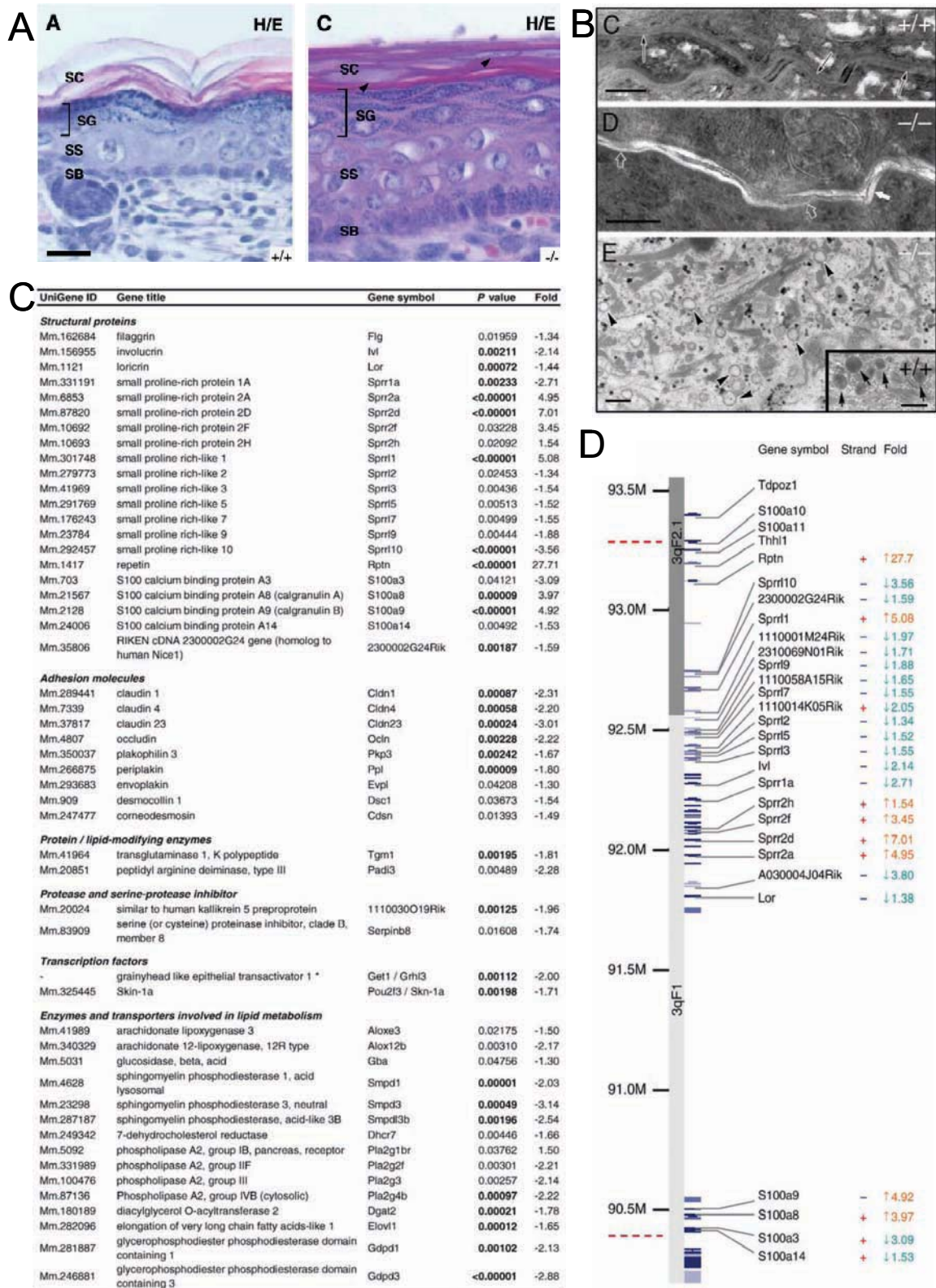


Figure 58. Inactivation du gène *Grhl3* (Yu et al. 2006). (A) Coupe histologique d'épiderme de souris sauvage ou *grhl3*^{-/-}. (B) Observation des défauts structurels des kératinosomes en microscopie électronique. (C) Gènes différentiellement exprimés dans l'épiderme entre souris sauvage et *grhl3*^{-/-}. (D) Gènes régulés appartenant au Complexe de Différenciation Epidermique.

dans l'épiderme des souris *Grhl3*^{-/-}. Les auteurs ont caractérisé le site consensus lié par Grhl3 (A/T A A A C C G G T T T/A T/A T/A), et pu montrer qu'un site de liaison est présent dans le promoteur proximal du gène *Tgm1*. Une étude ultérieure avec une nouvelle lignée de souris invalidées pour Grhl3 a confirmé le phénotype cutané, et, par comparaison d'épiderme de souris sauvages ou KO sur puce Affymetrix, permis d'identifier de nombreuses cibles directes ou indirectes de Grhl3, et en particulier un grand nombre de gènes du locus de différenciation épidermique (Yu et al. 2006) (figure 58C-D). La recherche de sites de fixation pour Grhl3 dans les promoteurs de ces gènes sous- ou sur-exprimés a désigné une dizaine de cibles directes, parmi lesquelles *Tgm1*, mais aussi *Sult2b1* ou le gène de la répétine. Enfin, les auteurs ont montré une interaction physique *in vitro* et fonctionnelle *in vivo* entre Grhl3 et Lmo4, un facteur de transcription lui aussi exprimé dans l'épiderme à E14.5, puis restreint à la couche basale de l'épiderme à partir de E17.5. L'inactivation de ce gène n'entraîne pas à elle seule de phénotype épidermique, mais provoque un trouble majeur de la différenciation dans des souris déjà déficientes pour Grhl3, avec une quasi-absence de couche cornée à E18.5.

Il est intéressant de noter que les deux autres facteurs de la famille, Grhl1 et Grhl2, sont également exprimés dans l'épiderme murin au cours du développement (dès E10.5 et E8.5 respectivement). L'expression de Grhl1 persiste chez l'adulte dans les couches suprabasales de l'épiderme interfolliculaire ainsi que dans les follicules pileux (particulièrement la gaine épithéliale interne). Des souris invalidées pour Grhl1 ne présentent aucun trouble de la barrière, mais une alopecie ainsi qu'une kératodermie palmoplantaire associée à une expression réduite de la desmogléine 1, les gènes *Dsg1α*, *Dsg1β* et *Dsg1γ* étant apparemment des cibles directes de Grhl1 (Wilanowski et al. 2008). Grhl2 semble réguler directement l'expression de la E-cadhérine dans l'épiderme (Cunningham and Jane, données non publiées). De façon surprenante, certains de leurs gènes cibles sont conservés chez d'autres espèces, parfois très éloignées des mammifères. Ainsi, un membre de la famille Grainyhead identifié chez *Caenorhabditis elegans*, Grh-1, régule lui aussi plusieurs gènes de la famille des cadhérines, de même que GRH chez *D. melanogaster*. De plus, chez la drosophile, GRH régule aussi l'expression de deux enzymes, dont la Dopa décarboxylase, impliquées dans la formation de liens covalents entre les constituants de la cuticule qui assure les fonctions de barrière et d'exosquelette de l'insecte. Bien que les enzymes impliquées dans ce phénomène n'aient aucune homologie avec les

transglutaminases exprimées dans l'épiderme de mammifères et régulées par Grhl3, dans les deux systèmes, la formation d'une couche protectrice passe par la création de structures résistantes et hydrophobes.

Le rôle des facteurs Grainyhead dans la régulation des jonctions cellulaires et l'établissement d'une fonction barrière paraît donc conservé dans le règne animal, bien que les cibles directes diffèrent.

2.2.3.2.2.4. GATA3

GATA3 appartient à la famille des GATA-binding factors qui comprend 6 membres chez l'homme et est impliquée notamment dans l'hématopoïèse. GATA3 est

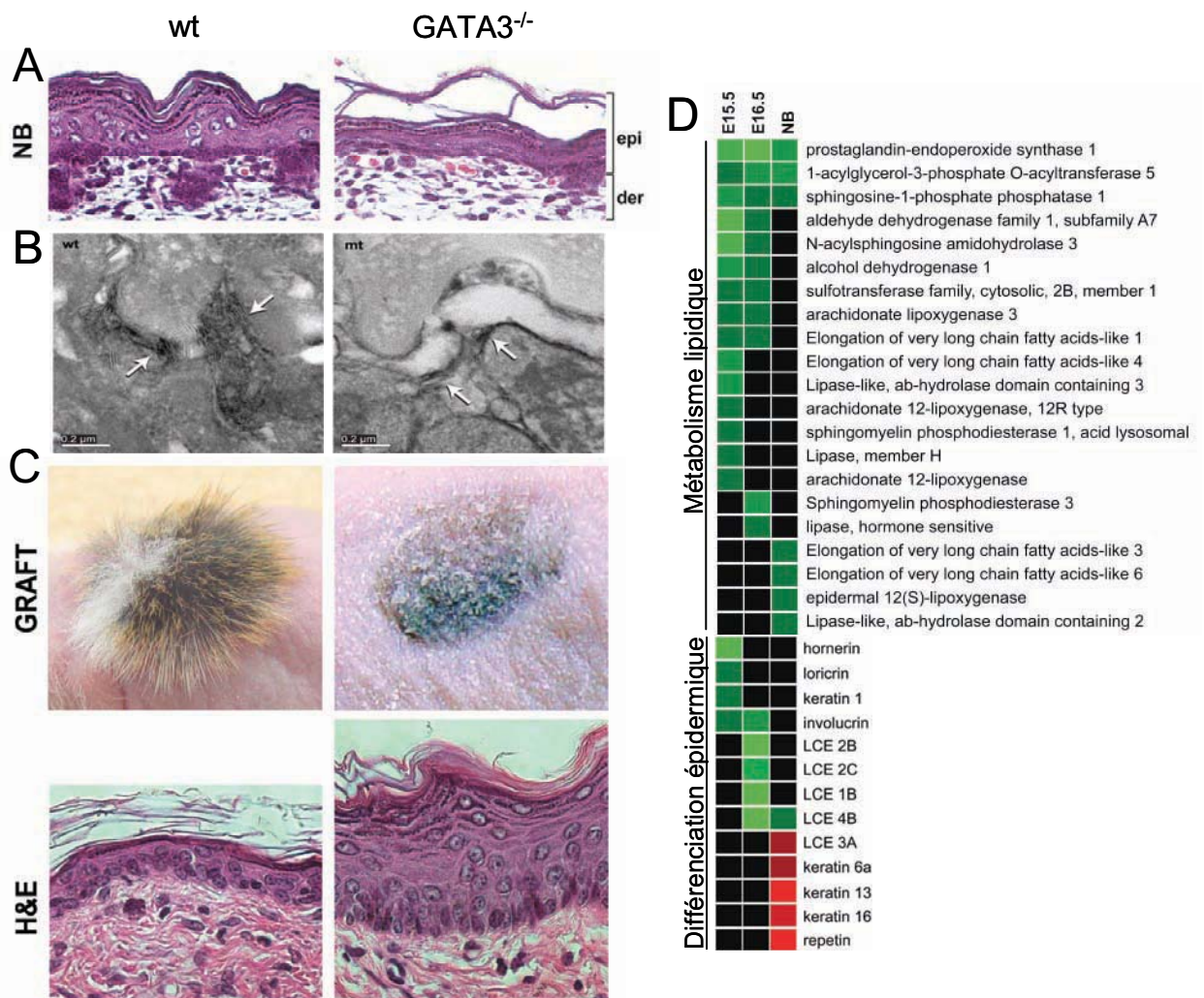


Figure 59. Inactivation du gène *Gata3* (de Guzman Strong et al. 2006).

(A) Coupe histologique d'épiderme de souris sauvages ou mutantes (nouveaux-nés).

(B) Défaut structurels des kératinosomes. (C) Greffe d'épiderme sur souris *nude*.

(D) Gènes d'expression épidermique différenciellement exprimés chez les souris mutantes [vert : sous-expression ($R < 0,5$), rouge : surexpression ($R > 2$)].

fortement exprimé dès les premières couches suprabasales de l'épiderme interfolliculaire, mais aussi au niveau des follicules pileux. Des souris homozygotes *Gata3* *-/-* présentent un trouble majeur de la barrière et meurent de déshydratation quelques heures après la naissance (de Guzman Strong et al. 2006). L'analyse histologique montre un défaut de différenciation, notamment des lamellae lipidiques moins nombreux et désorganisés dans l'espace intercornéocytaire. Ces manifestations s'accompagnent d'une hyperprolifération compensatoire lorsque la peau des nouveaux nés est greffée sur des souris *nude* (figure 59). La comparaison par puce à ADN des profils d'expression génique entre peau de souris inactivées pour *gata3* ou sauvages met en évidence une sous-expression de nombreux marqueurs connus de la différenciation, mais également de plusieurs gènes impliqués dans le métabolisme lipidique. Il apparaît en particulier que le gène *agpat5*, codant pour une enzyme clé de la synthèse des glycérolipides, est une cible directe de *Gata3*, avec quatre sites de fixation prédits dans le promoteur proximal et confirmés par immunoprécipitation de chromatine. Il est intéressant de noter que le niveau d'expression de *Klf4* n'est pas altéré dans l'épiderme *Gata3* *-/-*, et vice versa, suggérant que ces deux facteurs régulent deux ensembles indépendants du programme de différenciation.

2.2.3.2.2.5. PRDM1 (BLIMP-1)

Le facteur de transcription BLIMP-1 (B Lymphocyte-Induced Maturation Protein-1), produit du gène *PRDM1* (PR domain containing 1) et initialement décrit comme nécessaire et suffisant pour la différenciation des cellules B en plasmocytes, est également exprimé dans d'autres tissus au cours du développement comme chez l'adulte. Il est notamment exprimé, dans l'épiderme murin, au niveau des progéniteurs des glandes sébacées (puis les sébocytes matures), de la papille dermique et de la GEI des follicules pileux en développement (puis dans la GEI le cortex et la medulla des poils matures), et dans les couches les plus différenciées de l'épiderme interfolliculaire (Horsley et al. 2006). Ainsi, il représente à ce jour le facteur le plus tardivement exprimé au cours de la différenciation : son expression, qui débute entre E15.5 et E17 (formation du *stratum granulosum*), est restreinte aux dernières cellules vivantes de la couche granuleuse, y compris chez l'adulte (figure 60). Bien qu'il ne soit pas strictement requis pour la fonction barrière, il s'agit pourtant d'un régulateur de la cornification (Magnusdottir et al. 2007).

Des souris déficientes en Blimp-1 dans l'épiderme ont un retard dans l'apparition de la fonction barrière (quelques heures *post partum*, au lieu de E17.5), mais survivent à la naissance et ne développent qu'un phénotype modéré et transitoire (épiderme fripé et écailleux quatre à huit jours *post partum*, léger retard dans l'apparition du pelage). En histologie, des glandes sébacées anormalement larges et une hyperkératose des infudibula (« hyperkeratinization » ou *keratosis pilaris*) entraînant la formation de kystes sont visibles entre J+15 et J+29. L'épiderme est hyperplasique et hyperkératosique dès la première semaine, la couche granuleuse comportant un nombre plus important d'assises cellulaires et les cornéocytes apparaissant anormalement épais. Dans l'épiderme de nouveau-né

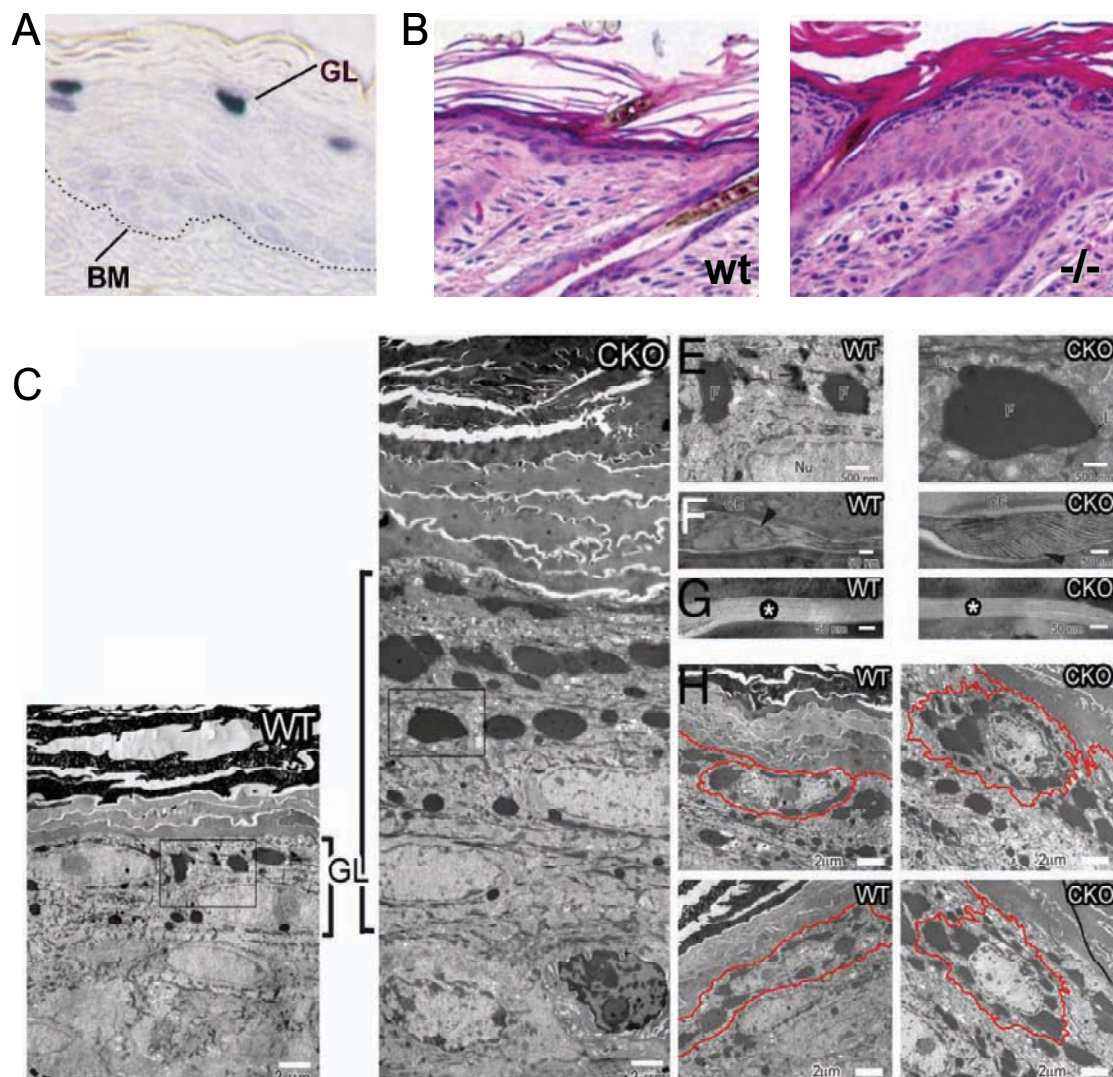


Figure 60. Inactivation du gène *Prdm1* (Magnusdottir et al. 2007).

(A) Expression du gène *Prdm1* dans la couche granuleuse de l'épiderme murin.

(B) Coupes histologiques d'épiderme de souris sauvages ou mutées.

(C) Ultrastructure de l'épiderme : couche granuleuse épaissie, grains de profilaggrine plus gros, kératinosomes et lamellae lipidiques normaux et augmentation du volume cellulaire.

Prdm1^{-/-} observé en microscopie électronique, les kératinocytes granuleux apparaissent moins aplatis que dans un épiderme normal, et le nombre ainsi que le volume des grains de kératohyaline est augmenté, mais aucune anomalie dans la sécrétion des kératinosomes et la formation des lamellae lipidiques n'est visible.

La recherche des cibles Blimp-1 a permis d'identifier 250 gènes différentiellement régulés, pour la plupart surexprimés dans l'épiderme des souris KO, suggérant que Blimp-1 agit plutôt comme un répresseur transcriptionnel. Parmi ces cibles, il est intéressant de noter que plusieurs gènes, codant notamment pour le transporteur d'osmolytes SLC6A12, l'aquaporine AQP9, ou encore le facteur de transcription NFAT5 (cible directe confirmée par ChIP), sont normalement impliqués dans la régulation de l'osmolarité cellulaire. Les auteurs suggèrent que le retard dans l'établissement de la barrière, l'hyperkératose des infundibula et surtout l'accumulation de kératinocytes granuleux et de cornéocytes d'un volume anormalement élevés, soient liés à l'absence de répression de ces gènes de l'équilibre osmotique. Cette répression favoriserait dans un épiderme normal la perte très importante de volume cellulaire constatée entre les couches épineuses et cornées. Parmi les autres cibles directement inhibées par Blimp-1, les auteurs ont identifié plusieurs gènes codant pour des facteurs de transcription, notamment *Fosl1* (*Fra1*) et *Prdm1* lui-même, ainsi qu'une phosphatase régulant la voie des MAPK, *Dusp16*.

Au final, il apparaît que la grande majorité des gènes identifiés à ce jour sont des effecteurs de la différenciation kératinocytaire, qu'il s'agisse de protéines structurales, d'enzymes ou d'inhibiteurs. Au contraire, les voies de signalisation et le « pool » de facteurs de transcription ne sont que très partiellement décrits, en particulier pour les étapes les plus tardives pourtant cruciales pour la mise en place et le maintien de la fonction barrière. Ils constituent pourtant des éléments essentiels pour comprendre comment un programme aussi complexe peut être exécuté correctement, donc également pour comprendre certaines pathologies liées à des altérations de ce programme.

2.3. Pathologies associées au programme de différenciation épidermique

2.3.1. Les génodermatoses

De nombreuses maladies monogéniques sont causées par des mutations des gènes exprimés au cours de la différenciation épidermique. Une grande partie des génodermatoses sont caractérisées par un trouble de la cornification associé à une perturbation de la différenciation kératinocytaire, et conduisent à des pathologies ichtyosiformes généralisées (sécheresse cutanée importante et formation de larges squames) ou localisées (kératodermies palmoplantaires). Voici un tableau récapitulatif des principales génodermatoses liées à des mutations de gènes exprimés au niveau des couches suprabasales de l'épiderme. Les gènes et la plupart de ces pathologies étant cités dans les chapitres précédents, le mécanisme moléculaire (quand il est connu) n'est pas indiqué.

Légende : A transmission autosomique ; X transmission liée au chromosome X ; R récessive ; D/SD dominante/semidominante

Pathologie	Transmission	Gène affecté	Manifestations cliniques/ anatomopathologiques
Ichtyose lamellaire de type I OMIM #242300	A R	TGM1	Bébé collodion, grandes squames brunes et adhérentes, prurit, ectropion / eclabium ; Hyperkératose massive
Ichtyose lamellaire de type II OMIM #601277	A R	ABCA12 (perte de fonction partielle)	Cf. IL type I.
Ichtyose arlequin OMIM #242500	A R	ABCA12 (perte de fonction totale)	Généralement létale, fissures cutanées profondes, ectropion / eclabium ; Hyperkératose, compaction de la couche cornée
Ichtyose lamellaire de type III OMIM #604777	A R	FLJ39501/ CYP4F22	Cf. IL type I.

Ichtyose congénitale autosomique récessive liée à l'ichthyin (ARCII) OMIM #612281	A R	NIPAL4 (Ichthyin)	Bébé collodion, kératodermie palmoplantaire, signes cliniques proches de l'IL ou NCIE (trouble massif et généralisé de la desquamation, érythème)
Erythrodermie ichtyosiforme congénitale sèche (NCIE) OMIM #242100	A R	ALOXE3, ALOX12B	Bébé collodion, érythème, ectropion / eclabium ; Kératose hyperproliférative
Erythrodermie ichtyosiforme congénitale bulleuse (BCIE) OMIM #113800	A R	KRT1, KRT10	Hyperkératose, fragilité cutanée ; bulles intraépidermiques et vacuolisation des couches suprabasales
Maladie de Naxos OMIM #601214	A R	JUP (plakoglobine)	Cardiomyopathie, kératodermie palmoplantaire, cheveux laineux
Acral peeling skin syndrome OMIM #609796	A R	TGM5	Atteinte acrale, érythème indolore, détachement de l'épiderme au niveau de la couche cornée
Ichtyose autosomique récessive avec hypotrichose (ARIH) OMIM #610765	A R	ST14 (Matriptase)	Hyperkératose diffuse, hypotrichose généralisée et diffuse, prurit. Syndrome similaire (OMIM #602400), avec également une atrophodermie folliculaire et une hypohidrose, dû à des mutations du même gène (Alef et al. 2009).
Syndrome de Papillon-Lefèvre OMIM #245000	A R	CTSC	Hyperkératose palmoplantaire, périodontite conduisant à la perte de la dentition
Syndrome de Netherton OMIM #256500	A R	SPINK5	Erythrodermie ichtyosiforme, cheveux bambous, atopie dans ¾ des cas

Maladie de Gaucher OMIM #230800	A R	GBA (gluco- cérébrosidase)	Déficit de la fonction barrière (TEWL augmenté), ichtyose, mort à l'âge de 18 mois due à des troubles neurologiques
Syndrome de Chanarin-Dorfman OMIM #275630	A R	CGI-58/ABHD5	Erythrodermie ichtyosiforme, hépatosplénomégalie, accumulation de gouttelettes lipidiques (nbx tissus)
Syndrome de Sjögren- Larsson OMIM #270200	A R	ALDH3A2	Ichtyose, troubles moteurs et retard mental, dégénérescence rétinienne
Ichtyose liée à l'X OMIM +308100	X R	STS	Hyperkératose, opacification asymptomatique de la cornée
Ichtyose vulgaire OMIM #146700	A SD	FLG	Peau sèche ; Hyperkératose modérée, déficit de la fonction barrière
Kératodermie palmoplantaire striée de type I OMIM #148700	A D	DSG1	Hyperkératose palmoplantaire
Kératodermie palmoplantaire striée de type II OMIM +125647	A D	DSP	Hyperkératose palmoplantaire, cheveux laineux, cardiomyopathie
Ichtyose bulleuse de Siemens OMIM #146800	A D	KRT2A	Hyperkératose et érosion superficielle de l'épiderme avec acantholyse.
Syndrome de Vohwinkel OMIM #124500	A D	GJB2 (Cx26)	Ichtyose, kératodermie palmoplantaire, pseudoainhum, surdité
Variant du syndrome de Vohwinkel OMIM #604117	A D	LOR	Ichtyose, kératodermie palmoplantaire, pseudoainhum
Maladie de Darier OMIM #124200	A D	ATP2A2	Défaut de cohésion kératinocytaire, acantholyse
Maladie de Hailey- Hailey OMIM #169600	A D	ATP2C1	Phénotype proche de celui de la maladie de Darier, atteinte des ongles

Table 3 : principales génodermatoses liées à des mutations de gènes exprimés au niveau des couches suprabasales de l'épiderme.

Il est important de noter qu'à ce jour la nature causative de certaines mutations n'est pas parfaitement établie et que les mécanismes moléculaires impliqués sont parfois peu décrits. De plus, même lorsque des mutations sont identifiées, la corrélation entre génotype et phénotype reste parfois difficile à établir, avec parfois une hétérogénéité phénotypique au sein d'une même famille, voire d'un même individu. C'est notamment le cas des ichtyoses congénitales autosomiques récessives (ARCI), un groupe à la fois génétiquement et phénotypiquement hétérogène constitué des ichtyoses lamellaires, de l'ARCI liée à l'ichthyin et de l'érythrodermie ichtyosiforme congénitale sèche (Fischer 2009). Le manque de précision dans la définition des signes cliniques, ainsi que les difficultés pour établir la nature causative des mutations identifiées dans les gènes *TGM1*, *ABCA12*, *CYP4F22*, *NIPAL4*, *ALOXE3* et *ALOX12B* (sans oublier que dans 22% des cas aucune mutation n'est identifiée), compliquent également l'étude de ces maladies.

2.3.2. Le psoriasis

Cette dermatose inflammatoire qui touche environ 2% de la population caucasienne, est caractérisée par l'apparition de plaques érythémato-squameuses de forme, de taille et de localisation très variables. Au niveau des lésions, l'épiderme est hyperplasique. Cette hyperprolifération kératinocytaire est associée à une disparition de la couche granuleuse, une hyperkératose et une parakératose. Les papilles dermiques sont plus allongées et des micro-abcès à polynucléaires sont souvent présents. L'infiltration de l'épiderme et du derme respectivement par des lymphocytes T CD8+ et CD4+ aboutit à la production de nombreuses cytokines proinflammatoires, telles que l'IL1, l'IL8, l'interféron- γ et le TNF α (figure 61).

L'origine de cette pathologie, kératinocytaire ou immunitaire, n'a pas été définitivement tranchée. Concernant les kératinocytes, des mesures du TEWL (pertes en eau trans-épidermiques) montrent que la fonction barrière est défectueuse au niveau des lésions psoriasiques (Serup and Blichmann 1987). L'absence de couche granuleuse est liée à une forte diminution de l'expression de la profilaggrine et de la loricrine, en plus de la diminution d'autres marqueurs (transglutaminase 3, caspase 14, DNase1L2...). A l'inverse, d'autres marqueurs de la différenciation, tels que l'involucrine, la cystatine A, SLPI, SKALP, la transglutaminase 1, ou encore de nombreuses protéines à activité antimicrobienne faiblement exprimées dans l'épiderme normal (hBD2, hBD3, LL-37, S100A7), sont surexprimés ou exprimés plus précocement dans cette pathologie – pour revue (Iizuka et al. 2004; Tschachler

2007). D'autres éléments sont plutôt en faveur d'une origine immunologique de la maladie. L'amélioration spectaculaire des lésions psoriasiques après traitement par la cyclosporine, une molécule immunosuppressive spécifique des lymphocytes T, suggère qu'ils ont un rôle important dans la maladie. De plus, l'injection de lymphocytes de patients dans le derme de souris *nude* porteuses de greffes d'épiderme humain normal induit l'apparition de lésions de type psoriasiques (Wrone-Smith and Nickoloff 1996). Que l'altération de la barrière au niveau des lésions soit à l'origine de la pathologie ou une conséquence de l'inflammation, les anomalies de la différenciation jouent probablement un rôle important dans le maintien de la maladie et sa chronicité, comme en attestent la disparition de l'inflammation et le rétablissement d'une différenciation normale lorsque les plaques psoriasiques sont recouvertes d'une protection inerte (Hwang et al. 2001).

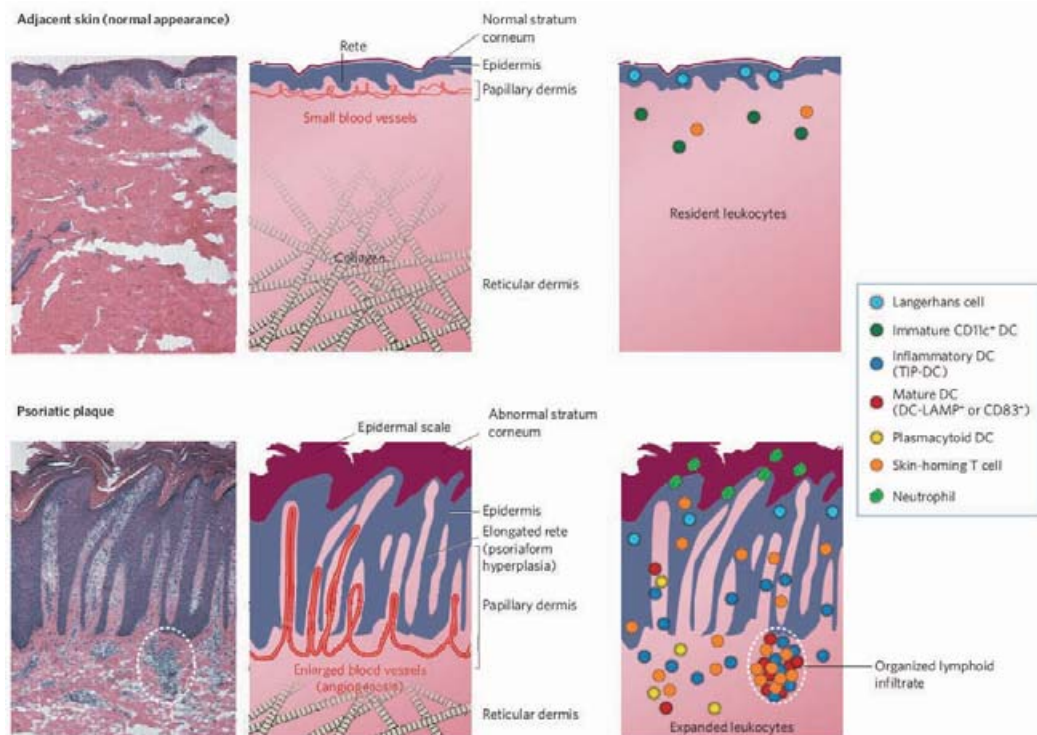


Figure 61. Comparaison entre peau normale et psoriasique (Lowe et al. 2007).

Selon la classification proposée en 1985, on distingue le psoriasis de type I, dit familial, d'apparition précoce (habituellement avant l'âge de 40 ans) et qui constitue 30% des cas, et le psoriasis de type II, sporadique et d'apparition plus tardive (Henseler and Christophers 1985). Le psoriasis de type I est une maladie multigénique et multifactorielle, de grandes disparités étant constatées selon les différentes populations considérées. De nombreuses études de liaison ont défini une

vingtaine de loci de susceptibilité pour le psoriasis familial, dont les plus documentés sont nommés *PSORS* : *PSORS1* (6p21.3), *PSORS2* (17q25), *PSORS3* (4q), *PSORS4* (1q21), *PSORS5* (3q21), *PSORS6* (19p13), *PSORS7* (1p), *PSORS8* (16q), *PSORS9* (4q31-q34), *PSORS10* (18p11.23), et *PSORS11* (5q31.1-q33.1).

PSORS1 contribuerait jusqu'à 65% dans la forme familiale du psoriasis (Gudjonsson et al. 2006), et contient les gènes codant pour le HLA-C et la CDSN. Si des travaux génétiques ont initialement suggéré une liaison entre certains allèles du gène *CDSN* et la susceptibilité à cette maladie (Capon et al. 2003; Capon et al. 2004), la contribution de ce gène a été écartée par deux études plus récentes dans des populations différentes, faisant du gène *HLA-Cw6* (allèle *Cw*0602*) le candidat le plus probable au sein du locus *PSORS1* (Helms et al. 2005; Nair et al. 2006).

La possible contribution du locus *PSORS4* est plus évidente : situé en 1q21, il correspond à une partie de l'EDC, le complexe de différenciation épidermique comportant un grand nombre de gènes spécifiques du kératinocyte granuleux. Dans une récente étude, plusieurs gènes faisant partie du locus de susceptibilité au psoriasis *PSORS4*, et notamment les gènes de la famille *SPRR*, sont ainsi apparus associés au psoriasis dans une population finlandaise (Kainu et al. 2009). De façon surprenante, trois études récentes menées sur des populations différentes ont également montré une association entre les gènes codant pour les LCE et le psoriasis (Liu et al. 2008; de Cid et al. 2009; Zhang et al. 2009), alors que ces gènes, bien que situés dans l'EDC, se trouvent en dehors des bornes du locus *PSORS4* tel qu'il a été redéfini (Capon et al. 2001).

Peu d'études de liaison avec les gènes des autres loci de susceptibilité ont été publiées, et se heurtent souvent au problème des contributions faibles et variables d'une population à l'autre. Les données sur les différents loci identifiés, gènes candidats et populations d'étude sont résumées dans la Table 4 (Jullien & Barker 2006). En outre, les évolutions récentes des techniques de recherche d'associations à l'échelle du génome ont permis d'obtenir des résultats plus précis qu'avec des études « classiques » de liaison et d'agrandir le nombre de gènes susceptibles de contribuer à la physiopathologie du psoriasis. Elles ont notamment souligné le rôle potentiel de plusieurs interleukines (IL-4, IL-12B, IL-13, IL-15 et IL-23A dont les gènes sont situés en dehors des loci de susceptibilité décrits auparavant, mais aussi le récepteur IL23R situé dans le locus *PSORS7* – pour revue : (Duffin and Krueger 2009)), la voie de

signalisation TNFalpha/NF-kB (Nair et al. 2009) ou encore la protéase transmembranaire ADAM33 (Lesueur et al. 2007).

Locus	Region	Associated gene	Study population	
PSORS7	1p	ND	British	
PSORS4	1q	Many within epidermal differentiation complex	Italian and American	
	2p	ND	American and British	
	2q	ND	British	
	3p	ND	Swedish	
	3q	SLC12A8	Swedish	
PSORS5	4q13	ND	American and Swedish	
	4q31	ND	Chinese, American, Northern Europe	
PSORS9	4q31	ND	Chinese, American, Northern Europe	
PSORS3	4q34	IRF-2	Irish	
PSORS1	6p21.3	HLA-C, HERV-K, HCG27, CDSN, PSORS1C3, POU5F1, TCF19, CCHCR1, LMP, SEEK1, SPR1.	Many	
	6q	ND	American and German	
	7	ND	American	
	8q	ND	British	
	10q	ND	American and German	
	14q	ND	American and British	
	15p	ND	Swedish	
	16q		American, German, Icelandic	
	PSORS2	17q25	SLC9A3R1, NAT9, RAPTOR	American and British
		18p11	ND	Finnish
PSORS6	19p13	JunB	German and British	
	20p	ND	German, British and American	

Table 4. Liste des loci putatifs de susceptibilité au psoriasis familial et de quelques gènes candidats (Jullien and Barker 2006).

Afin d'étudier plus facilement les mécanismes impliqués dans le psoriasis, un grand nombre de modèles murins ont été développés, qu'il s'agisse de modèles spontanés, de xéno- ou allogreffes, de modèles transgéniques ou inactivés pour un gène. Toutefois, certaines observations sont limitées par la densité très élevée de follicules pileux dans l'épiderme murin, sans compter que la plupart des modèles présentent d'autres signes phénotypiques non observés dans le psoriasis humain ; pour un résumé des principaux modèles, voir l'excellente revue publiée récemment (Gudjonsson et al. 2007).

3. Etudes à grande échelle du transcriptome épidermique

Le nombre d'EST produits à partir d'épiderme humain et/ou de kératinocytes en culture, est environ 300 fois plus faible que le nombre de ceux obtenus à partir d'organes tels que le cerveau (environ 1 600 contre 510 500 avant les premiers travaux à grande échelle de notre équipe, pour un total de 8,2 millions d'EST humains), ce qui montre le retard pris dans l'étude de ce tissu. Grâce à la miniaturisation des outils et au changement d'échelle opéré par les techniques les plus récentes, l'annotation du génome et la compréhension de mécanismes cellulaires complexes s'appuient de plus en plus sur des données obtenues à partir de prélèvements humains plutôt que de cellules en culture, et à partir de populations homogènes de cellules plutôt que d'organes entiers. Malgré cela, seul un petit nombre d'études se sont intéressées à l'étude du transcriptome de l'épiderme. De nombreuses études ont été menées dans un contexte pathologique, et les études menées dans un contexte physiologique ont principalement eu pour but d'évaluer l'intérêt des équivalents cutanés, donc *in vitro*. Seuls quelques rares travaux ont été menés à partir d'épiderme (murin ou humain) *ex vivo*.

3.1. Etudes du transcriptome lors de pathologies épidermiques

3.1.1. Transcriptome et psoriasis

Une des premières études à grande échelle consacrée au psoriasis a été menée au moyen de puces Affymetrix U95A (12 000 gènes humains) afin d'analyser des échantillons de peau de 15 patients (peau lésionnelle et non-lésionnelle pour 11 d'entre eux) et 6 individus témoins (Bowcock et al. 2001). 7129 transcrits ont été détectés et un clustering hiérarchique a permis de mettre en évidence 177 gènes spécifiquement régulés dans la peau lésionnelle, dont seulement 16 réprimés. Parmi les gènes surexprimés, plusieurs sont connus pour leur rôle dans la différenciation épidermique, dont certains situés sur l'EDC (*IVL*, *SPRRs*, plusieurs protéines de la famille *S100A*). On trouve également les gènes *TGM1* et *3*, *SKALP*, *ALOX12B*, les *SERPINB3* et *B4*, le marqueur membranaire *CD24*, les facteurs de transcription *STAT1* et *3*, et les kératines spécifiques des kératinocytes hyperprolifératifs, K6a, K16 et K17. Ces résultats semblent bien corrélés avec ceux obtenus ultérieurement par Haider et collaborateurs sur des lésions psoriasiques et des carcinomes spinocellulaires (Haider et al. 2006), ceux-ci ayant identifiés dans les deux

pathologies des gènes impliqués dans la protéolyse et sa régulation (ADAM23, KLK6, 10, 13, SERPINB3, B4 et B8), l'inflammation (interleukines, S100A, CD24), et dans la différenciation kératinocytaire (SPRR, DSG3, TGM1, IVL, ALOX12B...). L'étude de Bowcock et collaborateurs a également mis en évidence de nombreux gènes surexprimés codant pour des protéines du protéasome, qui pourraient intervenir dans la présentation antigénique via le CMH. Cette donnée est à rapprocher avec l'association entre susceptibilité au psoriasis familial et certains haplotypes du complexe HLA. Le transcrite présentant la plus forte induction (93 fois) code pour la transcobalamine I, une protéine de liaison à la vitamine B12 spécifique des cellules inflammatoires. 5 gènes ont montré une surexpression à la fois dans la peau lésionnelle et non-lésionnelle par rapport aux individus contrôle : CD47, l'interleukine 8, SPRR2C, ECGF1 (Endothelial Cell Growth Factor 1) et STAF50 (Stimulated Trans-Acting Factor 50kDa). Une étude récente employant des puces Affymetrix contre environ 14 500 gènes a permis d'établir le profil d'expression génique de kératinocytes traités à l'interleukine-1, et de mettre en évidence de nombreux points communs avec l'expression génique dans le psoriasis, soulignant le rôle important de l'IL-1 et de l'inflammation dans les altérations de la différenciation kératinocytaire observées dans cette maladie (Yano et al. 2008).

Malgré ces résultats, l'infiltrat de nombreuses cellules immunitaires au niveau des lésions psoriasiques, ainsi que l'activation de nombreux gènes impliqués dans l'inflammation mais non spécifiques de la maladie, rendent les données de transcriptome « global » particulièrement difficile à décrypter dans le cas du psoriasis, celles-ci n'ayant livré que peu d'information sur les mécanismes impliqués.

3.1.2. Transcriptome et inflammation

Le TNF α étant un des principaux médiateurs de l'inflammation dans l'épiderme (avec l'IL-1), le traitement de kératinocytes en culture par cette cytokine est susceptible d'induire une différenciation aberrante qui ressemble à celle observée dans un contexte d'inflammation cutanée. La technique du SAGE a permis d'identifier environ 10 000 gènes exprimés par les kératinocytes différenciés *in vitro*, parmi lesquels 90 subissent une régulation (surexpression pour 47 d'entre eux, sous-expression pour les 43 autres) en réponse au traitement par le TNF α (Jansen et al. 2001). Les auteurs relèvent notamment un niveau d'expression relativement stable pour les gènes impliqués dans la formation de plusieurs structures essentielles de l'épiderme (*KRTs*, *SPRRs*, *IVL*, *TGMs*), même si certains gènes suprabasaux tendent

à être légèrement réprimés tandis que *K14* est surexprimé. Parmi les gènes fortement exprimés figurent également des gènes codant pour des protéines sécrétées à activité antimicrobienne, avec notamment une surexpression de *S100A8* et *S100A9* codants pour les calgranulines A et B. De façon intéressante, ces deux protéines sont surexprimées dans l'épiderme hyperprolifératif ou psoriasique, et il a récemment été montré qu'elles favorisent la production de nombreuses cytokines proinflammatoires (dont le TNF α) et sont induites en retour, la boucle favorisant le maintien de l'inflammation et une augmentation de la prolifération kératinocytaire (Nukui et al. 2008). Enfin, malgré l'effet pro-apoptotique du TNF α dans certains contextes, aucune différence significative dans l'expression de gènes pro- ou anti-apoptotiques n'est relevée. Il est important de noter que plusieurs des marqueurs de la différenciation (LOR, FLG, KRT2) ne sont pas détectés par cette étude dans les kératinocytes différenciés *in vitro* non traités, suggérant un problème de différenciation très incomplète (en accord avec la morphologie des cultures obtenues), plutôt qu'un problème d'exhaustivité.

3.1.3. Transcriptome et carcinogenèse

La même équipe a également cherché à mettre en évidence par la technique du SAGE des gènes spécifiquement surexprimés dans des lésions épidermiques précancéreuses (kératose actinique) par rapport à la peau non-lésionnelle du même individu (van Ruissen et al. 2002a). Parmi les 12000 gènes identifiés, un ensemble de gènes spécifiques de l'épiderme précancéreux (donc de potentiels marqueurs moléculaires de la cancérogenèse) a été mis en évidence, dont *POLQ* codant pour l'ADN polymérase θ , *S100A7*, *S100A8* et *S100A9*, *CSTA* (cystatine A), *SFN* (stratifine) ou encore TRAF5 codant pour un récepteur au TNF α . Certains gènes étant également surexprimés dans le psoriasis, les auteurs proposent que la progression vers le cancer passe par une activation non réversible des gènes impliqués dans la prolifération, là où celle-ci est réversible dans le psoriasis. La technique utilisée pose cependant un problème de sensibilité, puisque de nombreux gènes représentés par une seule étiquette SAGE n'ont pas pu être inclus dans une analyse statistique quantitative.

De nombreuses autres études se sont ciblées sur l'expression génique des carcinomes épidermoïdes. Les carcinomes épidermoïdes (ou carcinomes malpighiens) sont des tumeurs malignes qui reproduisent de façon plus ou moins élaborée la structure d'un épithélium pluristratifié cornifié. La différenciation au sein de la tumeur se fait généralement de la périphérie vers le centre pouvant aboutir à la

formation de « perles cornées ». On parle de carcinome spinocellulaire pour les carcinomes épidermoïdes d'origine épidermique. Les carcinomes épidermoïdes développés aux dépens d'un épithélium non malpighien prennent le nom de carcinomes épidermoïdes métaplasiques. Les plus fréquents siègent au niveau des bronches, de la sphère ORL, de l'oesophage (HNSCC, head and neck squamous cell carcinoma) ou du col utérin. Les carcinomes épidermoïdes ayant fait l'objet de nombreux séquençages systématiques à partir de biopsies, ils sont fortement représentés dans les banques de données. De ce fait, la majorité des EST correspondant à des gènes de la différenciation épidermique proviennent de carcinomes épidermoïdes. Ceux-ci expriment de nombreux marqueurs de la différenciation épidermique (Haider et al. 2006), même si à ce jour aucun n'a été identifié comme constituant un facteur pronostique. De plus, très peu de données sont disponibles sur les facteurs de transcription et autres régulateurs impliqués dans la différenciation observable dans ces carcinomes, malgré l'intérêt de comparer ce programme avec la différenciation kératinocytaire normale. A ce jour, les rares données disponibles suggèrent l'implication de gènes par ailleurs connus pour leur rôle dans d'autres types de cancer, tels que p53, Bcl2, la TERT ou plusieurs cyclines. Plusieurs travaux démontrent également l'activation constitutive de STAT3 dans la plupart des HNSCC (Sriuranpong et al. 2003; Neiva et al. 2009). NFκB, AP1, c-Myc et la β-caténine semblent également impliqués, notamment dans les carcinomes du nasopharynx – pour revue : (Allen et al. 2007; Chou et al. 2008).

Dans le but de mieux comprendre les gènes impliqués dans l'invasivité des cellules tumorales, des carcinomes spinocellulaires ont été comparés à des lésions psoriasiques, ces deux pathologies étant notamment caractérisées par une hyperprolifération kératinocytaire mais seul le carcinome présentant un pouvoir invasif (Haider et al. 2006). Des puces Affymetrix U95 contenant des sondes pour environ 12 000 gènes ont été utilisées pour comparer chaque échantillon pathologique (8 carcinomes et 8 lésions psoriasiques) à de la peau saine provenant du même individu. Dans un premier temps, 1048 gènes ont été identifiés comme étant surexprimés et 870 sous-exprimés dans les carcinomes par rapport à la peau saine. Pour faire la part des gènes dérégulés dans le cadre de l'hyperprolifération bénigne et celle des gènes réellement impliqués dans la cancérogenèse, cette liste a été confrontée à celle des gènes surexprimés dans les lésions psoriasiques par rapport à la peau non lésionnelle. Parmi les gènes spécifiquement surexprimés dans les

carcinomes et pas dans le psoriasis, plusieurs membres de la famille des métalloprotéases (MMP1, 10, 13) sont présents, ainsi que des gènes impliqués dans la prolifération, mais aussi la KLK7 (SCCE), la cathepsine L2, le facteur de transcription STAT3, des récepteurs de la voie de signalisation WNT. Au contraire, les facteurs JUNB, JUND et FOS, ainsi que les marqueurs membranaires CD8 (lymphocytes T) et CD83 (cellules dendritiques), sont moins fortement exprimés dans les SCC que dans la peau normale ou le psoriasis, respectivement. Pour une partie de ces gènes, les résultats ont été confirmés par PCR quantitative en temps réel, et en immunohistochimie pour les MMP, CD8 et CD83. La principale limite de cette étude, probablement liée à l'étude directe de maladies complexes impliquant de nombreux types cellulaires, est le très grand nombre de gènes identifiés comme dérégulés dans les tumeurs (582 gènes surexprimés dans les tumeurs et non régulés dans le psoriasis ; 2520 gènes au total sont différentiellement régulés dans les 2 pathologies).

Parmi les autres études à grande échelle menées sur les carcinomes épidermoïdes, on peut citer par exemple les travaux de Roepman et al. qui ont cherché à améliorer le diagnostic clinique en recherchant des marqueurs spécifiques exprimés par les tumeurs primaires et susceptibles d'indiquer la présence de métastases non détectées (Roepman et al. 2005). Des échantillons provenant de 82 patients atteints de tumeurs de la sphère ORL, parmi lesquels 45 présentaient des métastases ganglionnaires, les autres n'ayant pas développé de métastases 3 ans après le traitement chirurgical, ont été utilisés. L'analyse par puces pangénomiques à oligonucléotides (Qiagen, 21 329 gènes humains) montre que 1986 gènes présentent une surexpression dans au moins 30 des 82 échantillons tumoraux. Une classification des profils d'expression des tumeurs ayant formé des métastases ou non a permis la mise en évidence de 102 gènes prédictifs de leur potentiel invasif. Il est intéressant de noter que 2/3 d'entre eux sont sous-exprimés dans les tumeurs à l'origine de métastases. On retrouve en particulier des gènes intervenant au niveau de l'adhérence cellulaire (plakines, desmogléine 3), la différenciation épidermique (transglutaminase 3, involucrine, loricrine, dermokine, SLPI, SKALP, SPINK5, ALOX12B, ABCA12) et la mort cellulaire. De nombreux gènes codant pour des protéines de la matrice extracellulaire et de son remodelage (métalloprotéases et inhibiteurs) sont présents, certains étant surexprimés et d'autres sous-exprimés. L'analyse de ces 102 gènes prédictifs du potentiel métastatique dans 22 nouveaux échantillons a montré une bonne corrélation pour 19 échantillons, sans aucun faux

négatif. La mise en place de ces nouvelles méthodes d'analyse des tumeurs primaires devrait permettre de mieux adapter les traitements aux patients.

3.2. Etudes du transcriptome dans un contexte physiologique

3.2.1. Transcriptome et cicatrisation

Dans le but d'étudier les modifications précoces de l'expression génique en réponse à une blessure, Cole et collaborateurs ont mis au point un protocole original pour l'étude de la cicatrisation *in vivo* chez l'homme. Dans le cadre d'une mammoplastie, 4 blessures adjacentes sont réalisées par biopsie « punch » de 8mm dans une région destinée à l'exérèse. Le chirurgien commence par l'autre sein, et organise l'opération de telle sorte que la moitié de la région blessée soit excisée au bout de 30 min, et l'autre 1h après la blessure (Cole et al. 2001). Le transcriptome de ces échantillons de peau est comparé à celui de peau non blessée du même individu à l'aide de filtres de nylon contenant des sondes pour 4400 gènes. Au bout de 30 min, 124 gènes sont surexprimés au niveau de la blessure et aucun n'est sous-exprimé. Il s'agit notamment d'activateurs de transcription (*BCL3*, *TCF3*, *RFXAP*, *MLL...*) et de gènes impliqués dans la signalisation et l'inflammation (*SOCS-1*, codant pour une protéine capable d'inhiber les voies de signalisation de plusieurs cytokines ; *MS-1*, codant pour une protéine sécrétée pro-inflammatoire...). Après 1h, 46 gènes sont surexprimés et 264 sont sous-exprimés. L'ensemble de ces résultats suggère que la première phase de réponse à une blessure est caractérisée uniquement par l'activation transcriptionnelle de certains gènes, alors qu'après 1h une importante proportion de ces gènes est réprimée. On peut toutefois reprocher à cette étude la trop petite proportion de gènes étudiés (4000), probablement assez peu représentatifs du transcriptome, ainsi que les prélèvements utilisés, composés de nombreux types cellulaires qui pourraient se trouver en proportions variables dans les différents prélèvements.

Dans le but de définir les gènes impliqués dans la cicatrisation indépendamment de l'inflammation induite par la blessure, des souris dont le gène *PU.1* (*sp11*) a été inactivé ont été comparées à des souris sauvages (Cooper et al. 2005). Le facteur de transcription PU.1 est en effet impliqué dans la différenciation de plusieurs lignées hématopoïétiques, et les souris invalidées pour ce gène sont dépourvues de macrophages et de neutrophiles fonctionnels. La peau de souris P2 (2 jours) est incisée au niveau dorsal et les animaux sont traités avec un

antibiotique jusqu'au sacrifice. Une réépithélialisation complète des lésions des souris sauvages comme des souris mutées est généralement observée au bout de 24h. Les prélèvements de peau sont effectués 30 min, 6h, 12h et 24h après la blessure. Les échantillons ont été analysés avec des puces Affymetrix ciblant environ 12 000 gènes, en comparaison à la peau non blessée provenant du même animal. 1001 gènes sont régulés au cours de la cicatrisation avec un ratio supérieur ou égal à 2 par rapport à la peau saine. Ces gènes ont ensuite été regroupés selon leur cinétique d'expression (taux d'ARNm en unités arbitraires ; figure 62) :

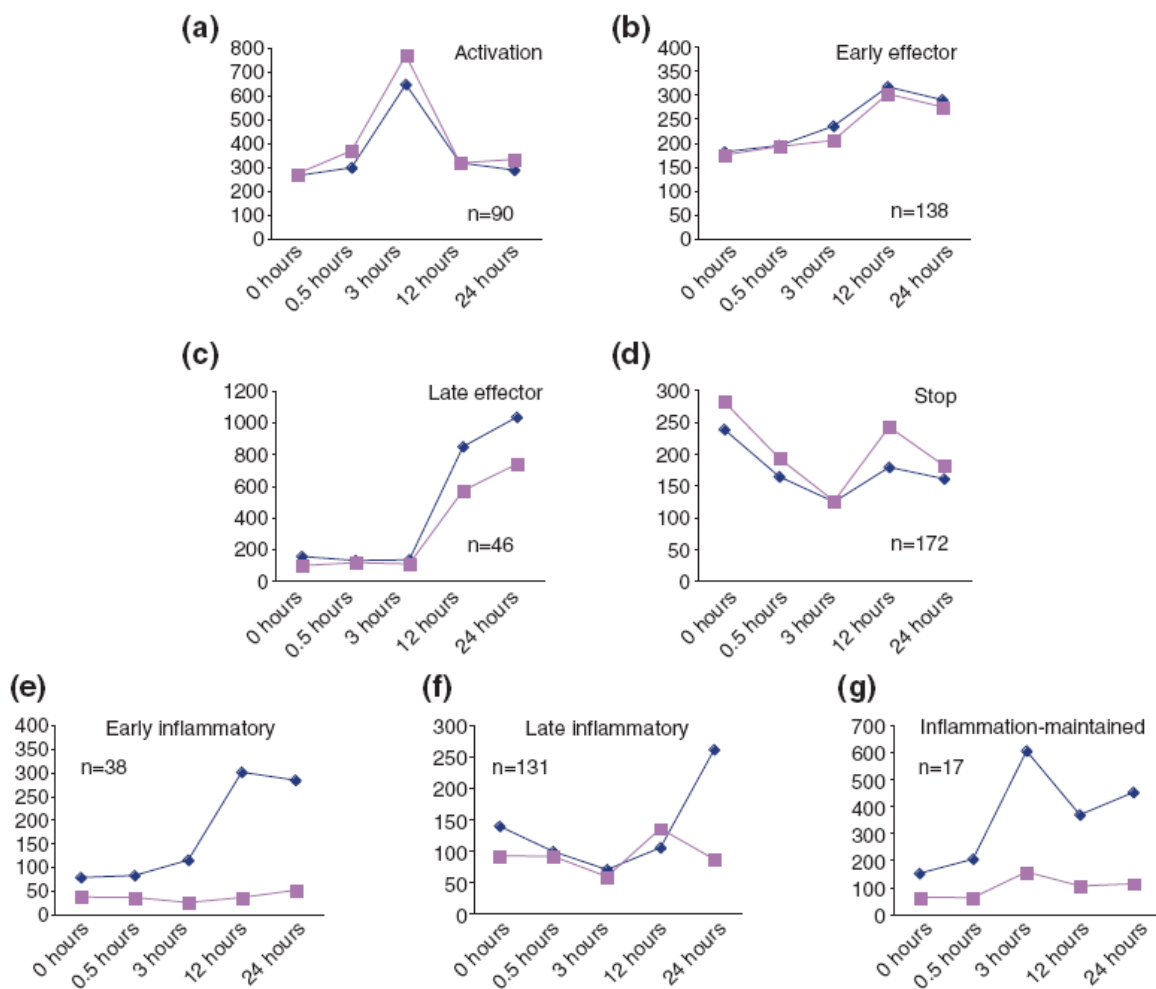


Figure 62. Cinétique d'expression des gènes régulés pendant la cicatrisation (Cooper et al. 2005). Les données obtenues avec des souris sauvages sont représentées en bleu, celles des souris PU1^{-/-} en rose.

Environ 450 gènes, considérés comme des gènes impliqués dans la réparation tissulaire, présentent une régulation au cours de la cicatrisation identique chez les souris sauvages et mutées, et donc indépendante de la présence de cellules de l'immunité innée (groupes a, b, c et d). Au contraire, 200 gènes voient leur expression

modulée par la présence de macrophages (groupes e, f et g). Le groupe a, contenant les gènes de réponse précoce, est principalement constitué de facteurs de transcription (Egr1, Fosb, Fra1, Junb, Klf9, Klf10, Myc...). Parmi les « effecteurs » (groupes b et c) sont largement représentés des gènes codant pour des protéines de la matrice extracellulaire, des protéines de structure ou de signalisation. La kératine 6, connue pour être induite au cours de la cicatrisation, fait partie comme attendu des effecteurs d'expression tardive. Enfin, les gènes du groupe d pourraient être impliqués dans l'inhibition de contact qui intervient lors de la fermeture de la blessure comme les éphrines et leurs récepteurs, le récepteur NOTCH ou DSC3.

3.2.2. Etudes de la différenciation épidermique *in vitro*

Plusieurs équipes ont décrit le transcriptome de différents modèles d'épiderme reconstruit *in vitro* et de kératinocytes cultivés en monocouche (Bernard et al. 2002; Gazel et al. 2003; Lemaitre et al. 2004; Mehul et al. 2004). Ces études s'appuient toutes sur la technologie des puces ou filtres à ADN pour la comparaison du transcriptome dans les différents modèles étudiés.

L'étude de Bernard et collaborateurs a comparé des échantillons d'épiderme humain normal (détachement à froid à partir de peau de sein), d'épiderme reconstitué produit par le laboratoire "SkinEthic" (kératinocytes pluristratifiés à l'interface air-liquide) et de kératinocytes primaires en culture monocouche. Les ARN radiomarqués ont été hybridés sur filtres de nylon contenant 475 sondes correspondant à 5 groupes de gènes : matrice extracellulaire/différenciation/adhésion, signalisation/cytokines/facteurs de croissance, apoptose et radicaux libres, structure, et enfin, métabolisme. 25 à 30% des gènes du filtre sont détectés. Les ARNm de la kératine K6b et des intégrines alpha-3, alpha-6 et beta-1 sont détectés dans les kératinocytes *in vitro* (épiderme reconstitué et culture monocouche), mais de manière surprenante pas dans l'épiderme normal, ce qui suggère une faible sensibilité de la méthode utilisée. Au contraire, des marqueurs de la différenciation épidermique tels que K1 et K10, la filaggrine, la loricrine et la cornéodesmosine sont détectés dans l'épiderme normal et reconstitué, et pas dans les cultures monocouche. Les auteurs ont de ce fait choisi l'épiderme reconstitué pour analyser les effets de l'acide rétinoïque sur l'expression génique. Même si ce travail a le mérite de tenter d'analyser à grande échelle la différence d'expression génique entre l'épiderme *in vivo* et les modèles *in vitro*, le faible nombre de gènes testés et le manque de sensibilité de la méthode en diminuent la portée.

L'étude de Gazel et collaborateurs utilise des puces Affymetrix capables de détecter 12000 gènes pour comparer le même type d'échantillons (épiderme humain normal disséqué, épiderme reconstitué "SkinEthic", et kératinocytes primaires prélevés 24h après confluence). Au total, 3240 gènes présentent une différence d'expression avec un ratio supérieur ou égal à 2 entre au moins deux des échantillons. Ce travail montre là aussi, mais cette fois-ci réellement à grande échelle, que les modèles d'épiderme reconstitué expriment un programme génique plus proche de l'épiderme normal que les kératinocytes en culture monocouche. On peut noter que l'épiderme reconstitué, bien qu'ayant une expression génique plus proche de celle de l'épiderme *ex vivo*, exprime plus faiblement certains gènes codant pour des protéines de signalisation sécrétées et leurs récepteurs. Les auteurs proposent que cette différence soit largement due à l'absence des types cellulaires autres que les kératinocytes. De façon étonnante, plusieurs gènes mitochondriaux sont sous-exprimés dans l'épiderme normal par rapport à l'épiderme reconstitué, ce qui pourrait refléter une activité métabolique moins intense *in vivo*.

L'étude de Mehul et collaborateurs est basée sur l'utilisation de puces dédiées, obtenues par dépôt de produits PCR sur membrane de nylon. Les 504 gènes étudiés ont été choisis pour leur rôle connu ou supposé dans la biologie cutanée. Trois

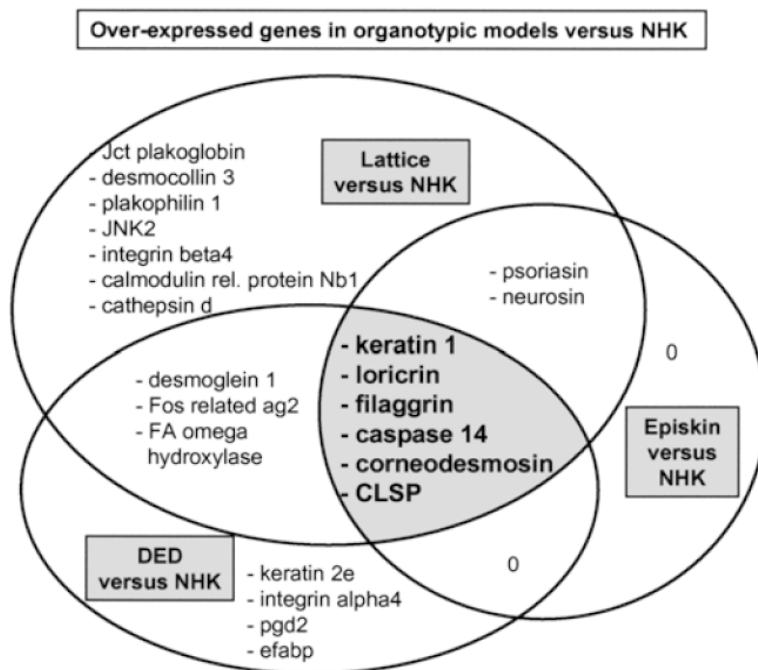


Figure 63. Gènes différentiellement exprimés selon les modèles d'épiderme reconstitué (Mehul et al. 2004)

modèles d'épiderme reconstitué à partir de kératinocytes primaires différenciés à l'interface air-liquide, différant par le support utilisé, matrice de derme irradié, mélange de collagène bovin et de fibroblastes humains, ou membrane recouverte de collagène I et IV ("Episkin"), ont été analysés. La grande majorité des gènes (86%) est détectée dans les 3 modèles et leurs profils

d'expression sont globalement proches. Chacun des modèles a été comparé à des kératinocytes primaires cultivés en monocouche. Tous les épidermes reconstitués reproduisent l'expression de plusieurs gènes associés à la différenciation terminale comme la cornéodesmosine, la loricrine, la caspase 14 ou la filaggrine, même si chacun montre quelques spécificités (figure 63).

Afin d'identifier plus directement les gènes spécifiquement exprimés au cours de la stratification, des kératinocytes primaires ont été mis en culture sur du derme irradié et maintenus submergés ou ont été différenciés à l'interface air/liquide (Koria and Andreadis 2006). Les ARNm sont extraits à partir des cultures submergées ou à l'interface air/liquide après 3 ou 7 jours. Leur transcriptome, analysé à l'aide de puces pangénomiques Affymetrix, montre que la plupart des gènes surexprimés dans les cultures à l'interface air/liquide interviennent dans le métabolisme des carbohydrates, la biosynthèse des lipides ou la phosphorylation oxydative. De façon étonnante, très peu de gènes tardifs de la différenciation épidermique apparaissent exprimés de manière différentielle entre des cultures submergées et des cultures différenciées à l'interface air/liquide. D'autre part, l'analyse morphologique des cultures submergées montre un épiderme beaucoup plus fin avec une à deux assises cellulaires, mais qui développe une couche cornée. Les auteurs proposent que le manque de nutriments, en particulier de glycogène, dans le milieu de culture pourrait affecter la stratification. Malgré le manque de couches suprabasales intermédiaires, un programme transcriptionnel conduisant à la cornification des cellules est mis en place.

Une étude menée la même année sur des kératinocytes humains immortalisés (HaCaT), cultivés en conditions prolifératives (J3 après mise en culture) ou différenciantes (post-confluence, à J10), combine une approche de transcriptomique utilisant des puces à ADN contre 2500 gènes et une approche protéomique par électrophorèse 2D suivie de spectrométrie de masse (Lemaitre et al. 2004). Cette double approche a permis d'identifier 15 protéines différentiellement détectées sur gel, 4 surexprimées et 11 sous-exprimées lors de la différenciation, et 171 gènes régulés parmi 1370 détectés (89 positivement et 82 négativement). L'analyse des transcrits et protéines différentiellement exprimés souligne l'implication de plusieurs fonctions cellulaires au cours de la différenciation, telles que le contrôle du cycle, le métabolisme des carbohydrates, la présentation antigénique, l'adhérence cellulaire ou encore les voies de signalisations liées aux cytokines ou le contrôle de la transcription.

Au delà de ces processus, l'étude a surtout permis d'identifier de nouveaux gènes candidats dans la régulation de la différenciation, avec 7 protéines et 33 transcrits jamais décrits comme étant régulés dans l'épiderme. Parmi les nouveaux gènes dont la régulation a été confirmée par immunohistochimie sur coupe de peau, la stratifine apparaît surexprimée, tandis que la stathmine est réprimée. La stratifine (SFN ou 14-3-3 σ) est une protéine de liaison impliquée dans le contrôle du cycle cellulaire via la séquestration du complexe MPF (Chan et al. 1999) ; une mutation dominante identifiée dans ce gène est responsable des défauts de l'épiderme et des follicules pileux chez les souris de la lignée *Er* et favorise également l'apparition de carcinomes épidermoïdes (Li et al. 2005). La stathmine, une phosphoprotéine impliquée dans la formation du fuseau mitotique, est surexprimée dans plusieurs types de cancers, notamment les cancers spinocellulaires de la cavité orale (Kouzu et al. 2006).

Un point commun aux différentes études concernant les modèles *in vitro* est la différence importante entre les modèles en monocouche et les épidermes reconstruits, ces derniers exprimant un plus grand nombre de gènes associés aux étapes tardives de la différenciation épidermique. Malgré cela, les épidermes reconstruits restent des systèmes simplifiés peu adaptés pour l'identification de gènes critiques pour la différenciation épidermique *in vivo*.

3.2.3. Etudes de la différenciation épidermique *in vivo*

Une étude menée en 2006 sur des souris sauvages ou mutées pour le récepteur à la mélanocortine (Mcr1), un marqueur des mélanocytes, a permis d'identifier de nombreux gènes potentiellement impliqués dans la différenciation kératinocytaire au moyen de puces à ADN pangénomiques de souris (April and Barsh 2006). Les auteurs ont analysé le transcriptome des fibroblastes du derme, des cellules de la couche basale (où sont logés les mélanocytes) ou de l'épiderme suprabasal de souris sauvages ou mutées. Ces 3 populations cellulaires ont été comparées 2 à 2, permettant de dégager une « signature » transcriptionnelle spécifique de chacune des fractions cellulaires (indépendante du génotype des souris ; figure 64). Au total, sur 17267 transcrits détectés (toutes fractions confondues), 552, 2097 et 883 transcrits se sont avérés spécifiques du derme, de la couche basale et des couches suprabasales de l'épiderme, respectivement. Parmi ces 883 transcrits spécifiques des kératinocytes différenciés, on peut noter de nombreux marqueurs déjà décrits dans la littérature ,

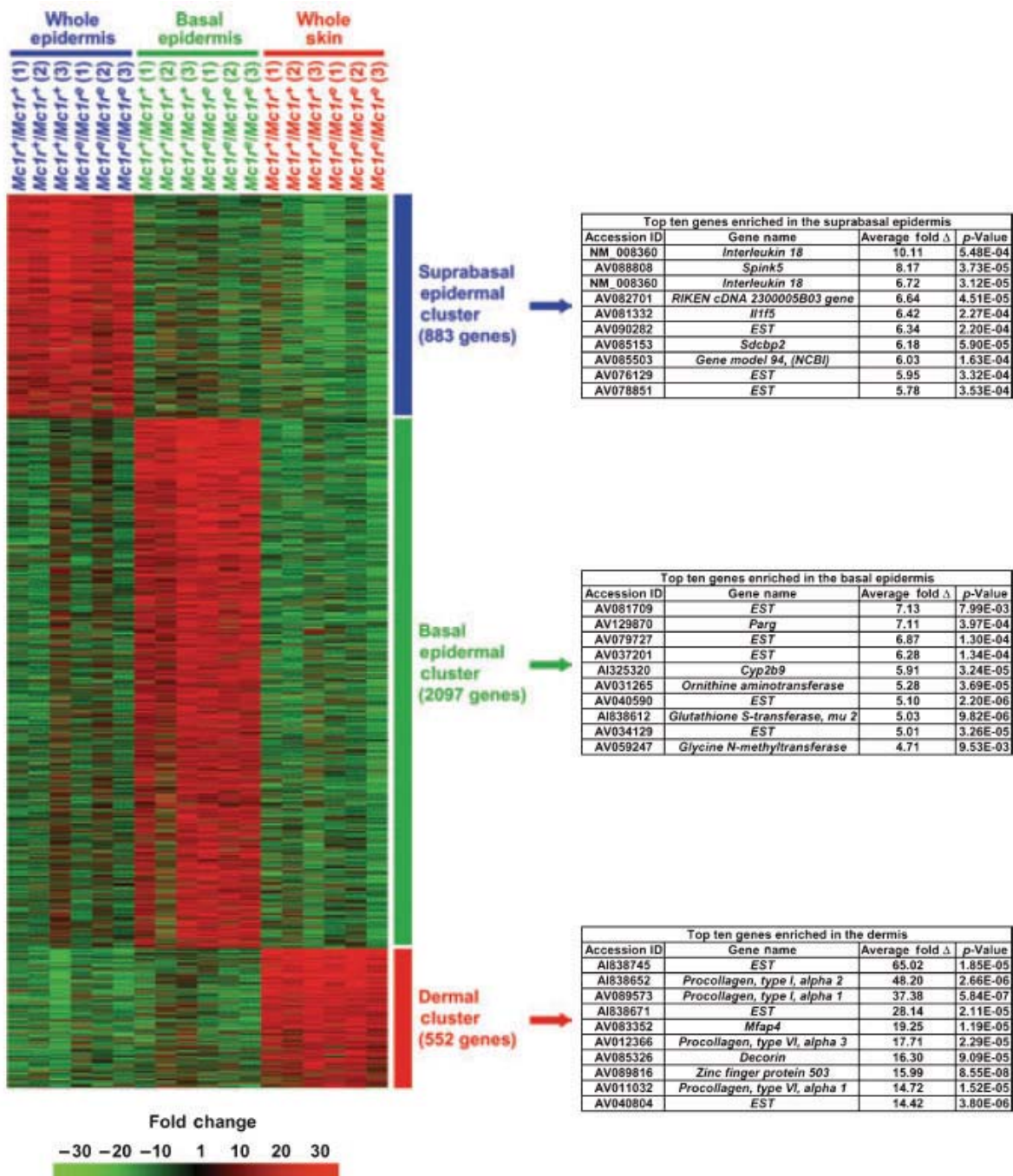


Figure 64. Signature transcriptionnelle des différentes couches de la peau murine (April and Barsh 2006).

(Krt10, périplakine, Gjb5 (connexine 31.1), Pkc epsilon, Ovo-like 1, Ror alpha, répétine, suprabasine, corniféline, Tgm1, Sprr (1b, 2j), Lce (1a1, 1a2, 1h, 1m et 3c), Csta, Aloxe3, Fabp5...). Ce travail a également mis en évidence des gènes peu ou pas caractérisés dans l'épiderme, impliqués dans le métabolisme lipidique (Acsbg1 (lipidosine), Pla2 groupes IVB & VI, Pla2 activating protein, Pld2, Galc...), diverses

voies de signalisation (Il1b, Il4, Il1rn (antagoniste IL1R), Chrna1 (récepteur cholinergique à la nicotine), Stat1 & 4, Calmoduline 1, Camkk1...), le cycle cellulaire (Stathmin-like 3), la défense antimicrobienne (Camp, Defa22...), la régulation de la protéolyse (Spink2)... Toutefois, de nombreux marqueurs connus sont absents (Flg, Ivl, Tgm3, Padi...), tandis que Lor est détectée dans plusieurs fractions alors qu'elle n'est attendue que dans la fraction suprabasale, peut être à cause d'une contamination de la fraction basale lors de la dissociation par récurage manuel des cellules en bain de trypsine.

La première étude partielle du transcriptome de l'épiderme humain entier obtenu par clivage enzymatique a utilisé la méthode SAGE, peu adaptée à la caractérisation de nouveaux gènes par manque de spécificité (van Ruissen et al. 2002b). Au total, 15 131 étiquettes SAGE, dont 7645 uniques, ont été collectées et comparées aux bases de données. Parmi les 20 étiquettes les plus représentées, 6 n'ont pu être identifiées et 10 correspondent à des protéines ribosomiques. La galectine 7, une lectine épidermique exprimée dans toutes les couches, et le gène hypothétique *KIAA0375*, identifié depuis comme codant pour RUSC2 (RUN and SH3 domain containing 2, ou Iporine, un partenaire de Rab1 exprimé de façon ubiquiste), font partie des gènes les plus fortement détectés (Cartharius et al. 2005). Enfin, le fait que seulement 2 étiquettes correspondent à des gènes de la différenciation épidermique, *KRT10* et *KDAP* (*Keratinocyte Differentiation Associated Protein*), souligne les limites de la méthode SAGE dans l'étude de transcriptomes peu caractérisés.

De récents travaux ont comparé le profil d'expression de différentes populations cellulaires épidermiques (Radoja et al. 2006). L'épiderme est séparé du derme après incubation d'une nuit à 4°C avec la dispase, puis les cellules épidermiques sont dissociées par traitement à la trypsine à 37°C. Les kératinocytes basaux sont ensuite isolés à l'aide de billes magnétiques couplées à des anticorps dirigés contre l'intégrine β_4 , spécifiquement exprimée par ces cellules. Des puces Affymetrix HU95 (10 000 gènes) ont été utilisées pour comparer les cellules β_4+ aux cellules β_4- , et à l'épiderme entier (à partir de 3 donneurs sains). Comme attendu, les gènes plus fortement exprimés dans la population de kératinocytes basaux β_4+ codent pour des protéines de l'hémidesmosome, du cycle cellulaire et de la réplication de l'ADN. Les gènes plus fortement exprimés dans la population de cellules β_4-

correspondent aux marqueurs de la différenciation épidermique et en particulier aux gènes situés dans l'EDC (*LOR*, *IVL*, *FLG*, *SPRRs*...). Les 2 populations cellulaires expriment aussi de façon différentielle les gènes codant pour des protéines impliquées dans l'adhésion, la matrice extracellulaire, le métabolisme et la signalisation. D'autre part, 49 gènes surexprimés dans les cellules β_4^- sont spécifiques des mélanocytes. En fait, cette dernière population est très hétérogène puisque composée de kératinocytes suprabasaux ainsi que des autres types cellulaires épidermiques, mélanocytes, cellules de Langerhans et cellules de Merkel. Une autre limitation dans cette étude vient du fait que les kératinocytes granuleux sont plus résistants à la trypsine que les autres kératinocytes et demeurent accrochés à la couche cornée ; de ce fait, ils sont majoritairement absents de la fraction cellulaire β_4^- , ce qui limite l'identification des gènes les plus tardivement exprimés. Ceux-ci ne sont identifiés que par déduction, puisque détectés dans l'épiderme entier et absents des fractions β_4^+ et β_4^- (ratio épiderme/K. épineux > 1,7). Cette approche est donc moins susceptible d'identifier de nouveaux gènes d'expression tardive qu'une comparaison directe des kératinocytes granuleux avec une référence.

L'état des connaissances concernant le transcriptome des différentes couches de l'épiderme apparaît donc très fragmentaire, en particulier *in vivo*. Pour cette raison, notre équipe s'est attachée à décrire le transcriptome des kératinocytes granuleux à partir d'épiderme humain normal, dans le but d'identifier des gènes spécifiquement exprimés dans ces cellules et impliqués dans la régulation du programme de différenciation, dans la cornification ou dans la desquamation.

Les résultats expérimentaux qui suivent décrivent notamment une procédure de purification par incubations itératives en bain de trypsine permettant un enrichissement progressif en kératinocytes granuleux. Cette technique a permis dans un premier temps d'identifier près de 3400 gènes spécifiquement exprimés dans les kératinocytes granuleux, par séquençage d'une banque d'ESTs produits par la technique des ORESTEs (Toulza et al. 2007). Cette méthode autorise la détection de l'expression, même très faible, d'un grand nombre de gènes, mais ne permet pas de distinguer lesquels sont spécifiques de la couche granuleuse, ou exprimés dans tout l'épiderme. Pour cette raison, plusieurs autres techniques complémentaires à grande échelle ont été employées pour rechercher spécifiquement des gènes d'expression

tardive, plus susceptibles d'être impliqués dans la fonction barrière de l'épiderme : exploration de la banque d'ESTs, puces à ADN pangénomiques et analyses bioinformatiques de promoteurs pour l'identification de candidats, PCR quantitative en temps réel et séquençage pour l'identification des gènes régulés (Mattiuzzo et al. 2009).

Enfin, la technique de PCR quantitative en temps réel, utilisée pour ces deux études d'expression génique, a également été employée pour établir le profil d'expression de plusieurs marqueurs de différenciation dans un modèle de souris inactivées pour le gène de la cornéodesmosine (Leclerc et al. 2009). Cette étude étant distincte des approches à grandes échelles, l'article correspondant est donné en Annexes.

II. RESULTATS EXPERIMENTAUX

GENOME BIOLOGY

8:R107 (2007)

Large-scale identification of human genes implicated in epidermal barrier function

Eve Toulza, **Nicolas R. Mattiuzzo**, Marie-Florence Galliano, Nathalie Jonca, Carole Dossat, Daniel Jacob, Antoine de Daruvar, Patrick Wincker, Guy Serre and Marina Guerrin.

Résumé

Au cours de la différenciation épidermique, les kératinocytes quittent le compartiment basal et cessent de proliférer. Les cellules progressant dans les couches suprabasales subissent des remaniements biochimiques complexes et finement régulés conduisant à la cornification, puis la desquamation. Ce processus de mort cellulaire programmée aboutit à la formation des cornéocytes qui assurent la fonction de barrière protectrice. Les kératinocytes granuleux, dernières cellules vivantes, produisent la quasi-totalité des protéines et lipides requis pour la fonction barrière ; pour cette raison, une meilleure compréhension des processus mis en jeu passe par l'étude à grande échelle de leur transcriptome.

Dans cet objectif, des kératinocytes granuleux ont été purifiés à partir d'épiderme humain par incubations successives en présence de trypsine. Le niveau d'enrichissement en kératinocyte granuleux a été contrôlé par RT-PCR quantitative en temps réel (qRT-PCR). Plus de 22 500 EST ont été produits par la méthode des ORESTES puis séquencés. Parmi eux, 16 406 EST correspondent à 3 387 gènes différents, dont 14 rétrogènes et 5 gènes non-codants. Le niveau de normalisation de la méthode s'avère satisfaisant (4,6 EST par gène), bien que certains transcrits abondants soient plus représentés. Environ 330 gènes exprimés sont représentés par moins de 100 EST dans la banque de données UniGene, et sont donc susceptibles d'être spécifiques des kératinocytes granuleux. Parmi eux, 73 codant pour des protéines potentiellement impliquées dans la fonction barrière ont été testés par RT-PCR quantitative (qRT-PCR) ; la moitié correspond à de nouveaux marqueurs tardifs de la différenciation, parmi lesquels des protéases, des inhibiteurs de protéases et des protéines impliquées dans le métabolisme et le transport des lipides. Deux membres du Complexe de Différenciation Epidermique (EDC), l'Ifapsoriasine et une nouvelle LCE, ainsi qu'un rétrogène de la Prosaposine, sont décrits pour la première fois.

Au final, l'emploi couplé d'une base de données de séquençage et de la qRT-PCR s'avère efficace pour identifier les gènes exprimés tardivement, qui pourraient permettre une meilleure compréhension de la physiologie et des pathologies épidermiques.

Large-scale identification of human genes implicated in epidermal barrier function

Eve Toulza^{*}, Nicolas R Mattiuzzo^{*}, Marie-Florence Galliano^{*},
Nathalie Jonca^{*}, Carole Dossat[†], Daniel Jacob[‡], Antoine de Daruvar[‡],
Patrick Wincker[†], Guy Serre^{*} and Marina Guerrin^{*}

Addresses: ^{*}UMR 5165 "Epidermis Differentiation and Rheumatoid Autoimmunity", CNRS - Toulouse III University (IFR 30, INSERM - CNRS - Toulouse III University - CHU), allées Jules Guesde, 31073 Toulouse, France. [†]Genoscope and CNRS UMR 8030, rue Gaston Crémieux, 91057 Evry, France. [‡]Centre de Bioinformatique Bordeaux, Université V. Segalen Bordeaux 2, rue Léo Saignat, 33076 Bordeaux Cedex, France.

Correspondence: Marina Guerrin. Email: mweber@udear.cnrs.fr

Published: 11 June 2007

Genome Biology 2007, 8:R107 (doi:10.1186/gb-2007-8-6-r107)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/6/R107>

Received: 1 March 2007

Revised: 24 May 2007

Accepted: 11 June 2007

© 2007 Toulza et al.; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: During epidermal differentiation, keratinocytes progressing through the suprabasal layers undergo complex and tightly regulated biochemical modifications leading to cornification and desquamation. The last living cells, the granular keratinocytes (GKs), produce almost all of the proteins and lipids required for the protective barrier function before their programmed cell death gives rise to corneocytes. We present here the first analysis of the transcriptome of human GKs, purified from healthy epidermis by an original approach.

Results: Using the ORESTES method, 22,585 expressed sequence tags (ESTs) were produced that matched 3,387 genes. Despite normalization provided by this method (mean 4.6 ORESTES per gene), some highly transcribed genes, including that encoding dermokine, were overrepresented. About 330 expressed genes displayed less than 100 ESTs in UniGene clusters and are most likely to be specific for GKs and potentially involved in barrier function. This hypothesis was tested by comparing the relative expression of 73 genes in the basal and granular layers of epidermis by quantitative RT-PCR. Among these, 33 were identified as new, highly specific markers of GKs, including those encoding a protease, protease inhibitors and proteins involved in lipid metabolism and transport. We identified filaggrin 2 (also called ifapsoriasin), a poorly characterized member of the epidermal differentiation complex, as well as three new lipase genes clustered with paralogous genes on chromosome 10q23.31. A new gene of unknown function, *C1orf81*, is specifically disrupted in the human genome by a frameshift mutation.

Conclusion: These data increase the present knowledge of genes responsible for the formation of the skin barrier and suggest new candidates for genodermatoses of unknown origin.

Background

High-throughput genomic projects focusing on the identification of cell- and tissue-specific transcriptomes are expected to uncover fundamental insights into biological processes. Particularly intriguing are genes in sequenced genomes that remain hypothetical and/or poorly represented in expressed sequence databases, and whose functions in health and disease remain unknown. Some of these are most probably implicated in organ-specific functions. Their characterization is essential to complete the annotation of sequenced genomes and is expected to contribute to advances in physiology and pathology. In order to achieve such goals, transcriptome studies on tissues rather than cultured cells, and eventually on a single cell type at a precise differentiation step are more likely to provide new information.

The epidermis is a highly specialized tissue mainly dedicated to the establishment of a barrier that restricts both water loss from the body and ingress of pathogens. The barrier function of the epidermis is known to involve the expression of numerous tissue-specific genes, most of which are specifically expressed in the late steps of keratinocyte differentiation. In order to establish and constantly maintain this barrier, keratinocytes undergo a complex, highly organized and tightly controlled differentiation program leading to cornification and finally to desquamation. During this process, cells migrate from the basal, proliferative layer to the surface, where they form the cornified layer (stratum corneum). According to the current model of skin epithelial maintenance, basal keratinocytes encompass a heterogeneous cell population that includes slow-cycling stem cells [1]. These stem cells give rise to transiently amplifying keratinocytes that constitute most of the basal layer. They divide only a few times and finally move upward while differentiating to form the spinous layer. The proliferating compartment is characterized by the specific expression of cell cycle regulators and integrin family members responsible for the attachment of the epidermis to the basement membrane. Growth arrested keratinocytes undergo differentiation, mainly characterized by a shift in cytokeratin expression from KRT5 (keratin 5) and KRT14 in the basal layer to KRT1 and KRT10 in suprabasal layers. As differentiation progresses, keratinocytes from the spinous layers progressively express a small number of specific differentiation markers, like involucrin. However, the differentiation program culminates in the granular layer, where keratinocytes express more than 30 epidermis-specific proteins, including proteins that are stored in cytosolic granules characteristic of granular keratinocytes (GKs). These proteins include well known components of the cornified layer, like loricrin and elafin, but also recently identified ones, such as keratinocyte differentiation associated protein (KDAP), hornerin, suprabasin, keratinocyte proline rich protein (hKPRP), and so on [2-5].

GKs undergo a special programmed cell death, called cornification, which gives rise to corneocytes that no longer exhibit

transcriptional or translational activity and are devoid of organelles. Rather, their intracellular content consists of a homogeneous matrix composed mainly of covalently linked keratins. The cornified envelope, a highly specialized insoluble structure, encapsulates corneocytes in place of their plasma membrane (see Kalinin *et al.* [6] for a recent review). The lipid-enriched extracellular matrix, which subserves the barrier, is produced by a highly active lipid factory mainly operative in the granular layer and comprises secretory organelles named the epidermal lamellar bodies [7]. In addition to the provision of lipids for the barrier, lamellar bodies deliver a large number of proteins, including lipid-processing enzymes, proteases and anti-proteases that regulate desquamation, antimicrobial peptides and corneodesmosin, an adhesive protein secondarily located in the external face of the desmosomes, as they turn into corneodesmosomes [8]. Therefore, the components of the stratum corneum, responsible for most of the protective cutaneous functions, are produced by GKs.

Transcriptome studies of selected cell types of the human epidermis are expected to contribute to the elucidation of the mechanisms responsible for barrier function. They will also shed further light on the causes of monogenic genodermatoses and the pathomechanisms of common complex skin disorders like psoriasis. However, present knowledge on the gene repertoire expressed by keratinocytes remains largely fragmentary. Among the approximately eight million human expressed sequence tags (ESTs) from the dbEST division of the GenBank database, only 1,210 are annotated as originating from the epidermis, although these are, in fact, derived from cultured keratinocytes, which do not fully recapitulate the complex *in vivo* differentiation program. In this article, we describe the results of a large-scale cDNA sequencing project on GKs of healthy human skin, purified by a new method. In order to characterize genes expressed at a low level and to avoid the repetitive sequencing of highly expressed ones, we used the ORESTES (open reading frame EST) method to prepare a large series of small size cDNA libraries using arbitrarily chosen primers for reverse transcription (RT) and PCR amplification [9]. The sequencing of about 25,000 clones has produced a list of 3,387 genes expressed by GKs. Some of them, analyzed by quantitative RT-PCR, were shown to be expressed in a cell-specific manner. This effort resulted in a large number of novel candidate genes of importance for the epidermis barrier function and the etiology of genodermatoses.

Results

Purification of human granular keratinocytes

As a first step in this transcriptome project we devised a method to purify GKs. Iterative incubations of pieces of human epidermis with trypsin were performed to give three suspended cell fractions (hereafter named T1-T3) and finally to isolate cells attached to the stratum corneum (T4 fraction).

Table 1**Expression ratios for *KRT14* and *KLK7* as measured by real-time PCR from four independent samples**

Expression ratio	Sample no.			
	1	2	3	4
<i>KRT14</i> (T1/T4)	7.5	5.9	25	13.6
<i>SCCE/KLK7</i> (T4/T1)	164	189	120	54

Morphological analyses revealed that after three treatments, residual epidermal fragments were mostly composed of corneocytes and GKs (Figure 1). Quantitative real-time PCR was performed to quantify the enrichment in GKs. To first select a reference gene for normalization, the relative expression of eight housekeeping genes (*GAPDH*, *SOD1*, *ACTB*, *B2M*, *HPRT1*, *HMBS*, *TBP* and *UBC*) in each cell fraction (T1-T4) was analyzed using GeNorm [10]. In agreement with previous data [11], beta-2-microglobulin (*B2M*) appeared to be stably expressed during epidermis differentiation, and was thus chosen for normalization. In addition, we used the lectin Galectin-7 (*LSGAL7*), which was previously shown by *in situ* hybridization to be equally expressed in all epidermal layers [12]. *BPAG2* (bullous pemphigoid antigen 2) or *KRT14*, and *KLK7* (kallikrein 7, also called stratum corneum chymotryptic enzyme (*SCCE*)) were selected as specific for the basal layer or the GKs, respectively [13,14]. For four cell fractionations from different individuals, the mean T1/T4 expression ratio of *KRT14* was 13, whereas the mean T4/T1 expression ratio of *SCCE* was approximately 130 (Table 1). The *KRT14* ratio might be indicative of a slight contamination of the T4 fraction with basal keratinocytes. Nevertheless, the large *SCCE* ratio indicates that very few, if any, GKs were present in the T1 fraction. From this, we concluded that the T4 fraction was highly enriched in GKs and thus suitable for a large-scale study of their transcriptome.

An ORESTES dataset from human granular keratinocytes

PolyA⁺ RNA was extracted from the T4 fraction from individual 3 (Table 1) and used to generate cDNA mini-libraries using the ORESTES method [9]. This sample was chosen as it presents the highest T1/T4 expression ratio for the *KRT14* gene, suggesting a low contamination of the T4 fraction by basal keratinocytes. This method uses arbitrarily chosen primers for reverse transcription and PCR amplification. The successful amplification of a mRNA thus depends primarily on partial sequence homology with the primer, rather than on its abundance. This, and the elimination of cDNA preparations that display prominent bands on gels (indicative of the selective amplification of particular mRNAs), results in a normalization process and allows the detection of rare transcripts. We constructed 150 cDNA libraries with different primers, the analysis of 100-200 clones from each leading to the production of 22,585 sequences (Figure 2a). Among

these, 1,453 (approximately 6%) corresponded to empty plasmids or uninformative sequences, 377 (1.7%) were of bacterial origin, and 2,303 (10%) matched the human mitochondrial genome. Despite two rounds of polyA⁺ RNA purification, 1,859 sequences (8.2%) arose from ribosomal RNA. In addition, 187 sequences corresponded to unspliced intergenic DNA and may reflect spurious transcriptional activity. The remaining 16,591 sequences (73%) matched known or predicted transcribed regions, of which 62% aligned with the human genome in several blocks, and thus corresponded to spliced transcripts. After clustering, we observed the transcription of 3,387 genes by GKs. Additionally, 23 sequences matched overlapping exons belonging to two genes transcribed in opposite orientations and thus could not be attributed to a single gene.

The normalization ability of the ORESTES method was examined by classifying genes according to the number of matching sequences in the dataset (Figure 2b). Half of the genes were represented by a unique sequence and 76.3% by three or less sequences, thus showing an acceptable level of normalization, with a mean of 4.6 ORESTES per gene. However, the ORESTES method only partially compensates for transcript abundance, as several genes were represented by a large ORESTES number. In these cases, we examined the number of sequences in the corresponding UniGene clusters, a rough measure of gene expression level. This revealed two situations: first, the gene is strongly expressed in many cell types including GKs (a high number of both ORESTES and UniGene entries); and second, the gene is particularly expressed in GKs (a high number of ORESTES, but low number of UniGene entries). The first category mainly includes housekeeping genes from the translation machinery (for example, *RPS8*, *EEF1A1*, *RPL3*, *RPL7A*, *RPL28*; Table 2). The second category contains genes previously described as implicated in epidermis barrier function (for example, *KRT1*, *DMKN*, *LEP7*, *FLG*, *KRT2A*, *SPRR2E*, *CASP14*, *CDSN*, *hKRP*, *SBSN*) and, interestingly, new candidates for this function (*TSPAN5*, *DUOX2*, *TMEM14C*, *SERPINA12*, *SLC22A5*, *FLG2*, *C7orf24*). Dermokine (*DMKN*), represented by 217 ORESTES, was shown to be selectively transcribed in mouse GKs by high-throughput *in situ* hybridization [15] and signal sequence trap [16] screens. The present ORESTES dataset allowed us to describe 13 novel human *DMKN* splicing isoforms with distinct subcellular locations and expression patterns [17].

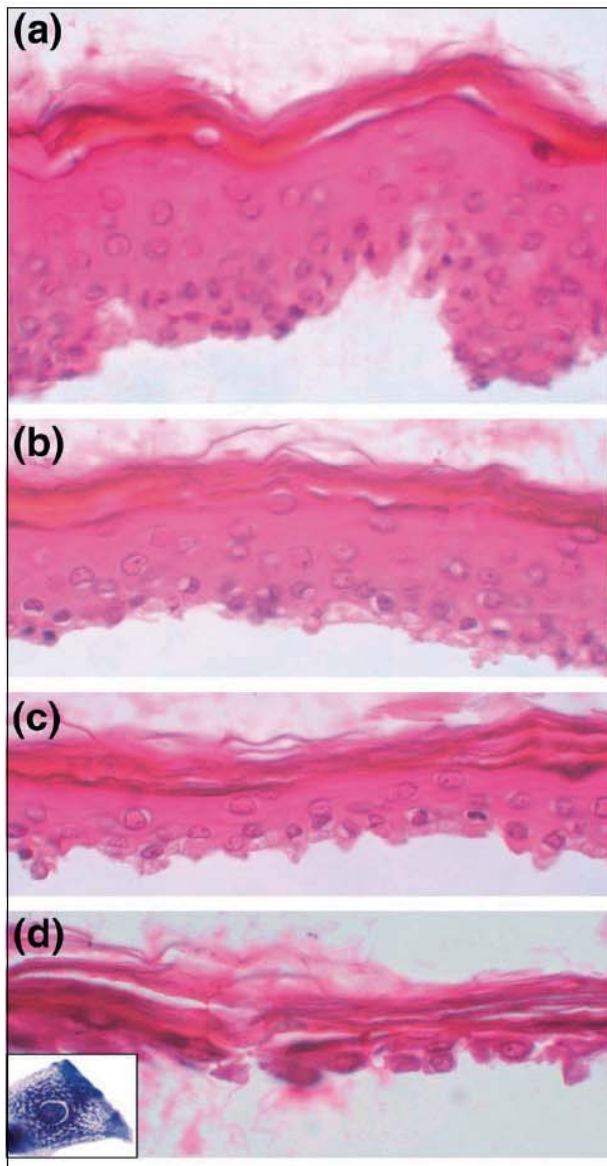


Figure 1
 Histological analysis of epidermis samples. (a) Hematoxylin-eosin stained sections of entire epidermis after thermolysin incubation and removal of the dermis. (b,c,d) Epidermis fragments remaining after the first, second, and third trypsin incubation, respectively. Fragments shown in (d) are mainly composed of GKs attached to the cornified layer and constitute the T4 fraction. Inset: higher magnification showing the characteristic cytological aspect of a GK with cytoplasmic keratohyalin granules.

The ORESTES dataset was aligned with the human genome using BLAT [18]. The BLAT results were used to write a custom track that allows the visualization of the position of a particular ORESTE relative to other annotations such as RefSeq genes, vertebrate orthologues, single nucleotide polymorphisms, microarray expression data, and so on, and is freely available online [19]. A screen copy of a UCSC Genome

Browser window showing the ORESTES obtained for the *C1orf81* gene is presented as an example (Additional data file 1). Indeed, this gene was characterized and a cDNA (DQ983818) was cloned for the first time in this study (see below). Our dataset includes the 16,591 ESTs matching known or predicted transcribed regions. These sequences have also been deposited in public databases (GenBank:EL593304-EL595248, GenBank:CU442764-CU457374).

Poorly represented genes in expressed sequence databases

As few sequencing projects from human epidermis have been performed so far (relative to other organs), genes expressed during the late steps of epidermis differentiation are poorly represented in sequence databases. Among the 3,375 genes from our set, 330 (10%) corresponded to UniGene clusters containing less than 100 mRNA/EST sequences, and were thus good candidates for epidermis late-expressed genes. These were subdivided into five classes. The first one contains all the genes (50) already known to be specifically expressed in the suprabasal layers (Table 3). This confirms that late-expressed genes are poorly represented in EST databases. The second class consisted of 31 genes with known or inferred functions that were previously known as mainly expressed in a specific tissue different from epidermis (Table 4). We suggest that some of them might play a specific role in epidermal differentiation. This could be the case for *SERPINA12*, *DUOX2*, and, to a lesser extent, *CASZ1*, which are represented by a large ORESTES number. We also suspect that *CLDN23* might play an important role in GKs, since claudin-based tight junctions in the granular layer contribute to barrier function of the epidermis [20]. Accordingly, claudin-1-deficient mice display a lethal defect in skin permeability [21]. The third class gathered 32 uncharacterized paralogues of known genes (Table 5). The fourth class was composed of 105 genes that remain hypothetical and about which nothing is known regarding their normal function or disease relevance (Table 6). The fifth class contained genes that are expressed, most probably at low levels, in numerous tissues, but whose epidermal expression is, to the best of our knowledge, described here for the first time (Additional data file 2). Several genes from these five classes were selected to quantify their expression in the course of epidermal differentiation by real-time PCR (see below).

Expressed retrogenes and pseudogenes

Pseudogenes generally correspond to retrocopies with many disruptions in their open reading frame (ORF). However, it is now recognized that a large number of retrocopies are transcribed and can encode functional proteins [22]. Among the top 50 transcribed retrocopies reported by these authors, 11 were detected in GKs by the ORESTES method. Among these, calmodulin-like 3 (*CALML3*) was previously shown to be specific for keratinocyte terminal differentiation [23]. We identified two other expressed retrogenes corresponding to the

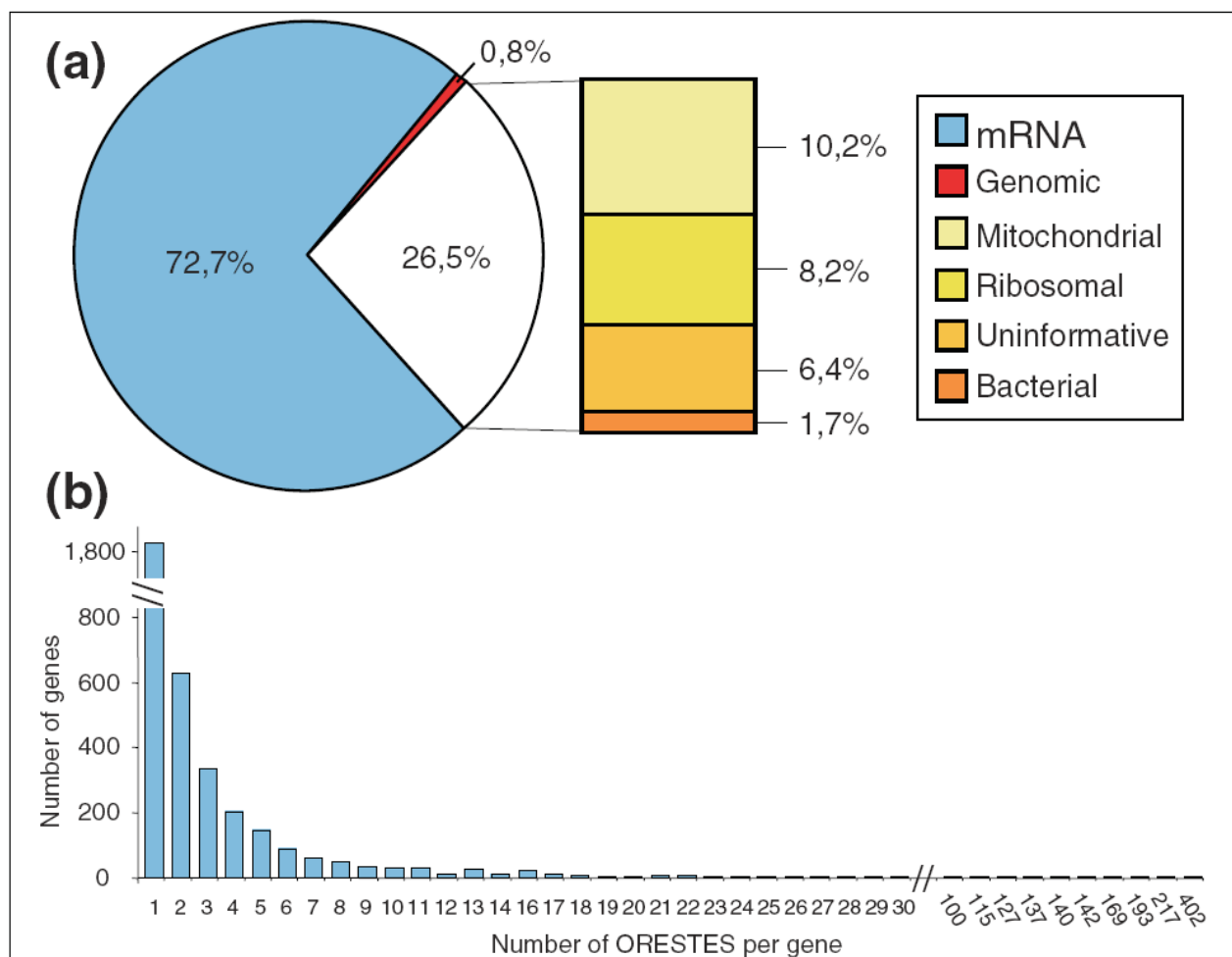


Figure 2
 Analysis of the ORESTES dataset from GKs. **(a)** Pie graph of the 22,585 sequences obtained from the T4 fraction enriched in GKs. The treatment of the mRNA samples with DNase resulted in minimal contamination with genomic sequences. Despite two rounds of polyA+ mRNA purification, rRNA sequences still represent approximately 8% of the dataset. **(b)** Histogram showing the number of ORESTES at each level of redundancy. The vast majority of genes are represented by less than five ORESTES, illustrating the normalization capability of that method. However, a small number of genes are represented by a large number of ORESTES (up to 402).

retrotransposition of the cutaneous T-cell lymphoma associated antigen 5 (*CTAGE5*), and CCR4-NOT transcription complex, subunit 6-like (*CNOT6L*). These genes can be considered as 'intact', that is, they show no disablements such as premature stop codons or frameshift mutations when compared to the ORF of their parental genes. Of note, the *CNOT6L* retrogene is specific for hominoids (Additional data file 3), while the *CTAGE5* retrogene is specific for primates (data not shown).

Moreover, six unspliced ORESTES correspond to a part of intron 8 of the *PPP2R5A* gene, and include the small nucleolar RNA (snoRNA) U98b sequence. The snoRNAs are non-protein-coding RNAs that guide the 2'O-ribose methylation (C/D box snoRNAs) or the pseudouridylation (H/ACA box snoRNAs) of ribosomal RNAs, and are generally processed

from introns of RNA polymerase II transcripts [24]. Interestingly, the U98b snoRNA is a primate-specific retroposon of the *ACA16* snoRNA hosted by the *PNAS-123* gene [25]. We thus suggest that the ORESTES from the *PPP2R5A* gene correspond to a precursor form of the U98b snoRNA, and that snoRNA retroposons can indeed be expressed when located in an intron of a new host gene in the sense orientation. Therefore, our ORESTES dataset included transcripts from retrogenes, originating either from spliced pre-mRNAs or from an intron-encoded snoRNA gene.

Non-protein-coding genes

We obtained two long spliced ORESTES highly similar to the BC070486 mRNA form of the *GAS5* gene, a non-protein-coding gene that belongs to the 'growth arrest specific' family but is disrupted in its ORF by a premature stop codon. The *GAS5*

Table 2**Representative sample of genes with the highest number of ORESTES**

No. of ORESTES	Gene symbol	No. of UniGene ESTs	Full name (alias)
Ubiquitously expressed genes with a high number of UniGene ESTs			
142	<i>RPS8</i>	3,382	Ribosomal protein S8
115	<i>EEF1A1</i>	29,374	Eukaryotic translation elongation factor I alpha I
77	<i>HLA-B</i>	4,536	Major histocompatibility complex, class I, B
71	<i>RPL3</i>	11,561	Ribosomal protein L3
62	<i>NCL</i>	2,970	Nucleolin
55	<i>RPL28</i>	2,394	Ribosomal protein L28
55	<i>RPL7A</i>	5,864	Ribosomal protein l7a
51	<i>RPSA</i>	5,623	Ribosomal protein SA
50	<i>PABPC1</i>	4,385	Poly(A) binding protein, cytoplasmic I
34	<i>RPS18</i>	2,292	Ribosomal protein S18
Known epidermis specific genes			
402	<i>KRT1</i>	134	Keratin I
217	<i>DMKN</i>	275	Dermokine
140	<i>LEP7</i>	5	Late envelope protein 7 (xp32)
100	<i>FLG</i>	5	Filaggrin
71	<i>KRT2A</i>	12	Keratin 2A
62	<i>SPRR2E</i>	36	Small proline-rich protein 2E
61	<i>CASP14</i>	19	Caspase 14
59	<i>CDSN</i>	91	Corneodesmosin
56	<i>hKPRP</i>	7	Human keratinocyte proline rich protein
54	<i>PKP1</i>	263	Plakophilin I
32	<i>SBSN</i>	49	Suprabasin
30	<i>DSG1</i>	61	Desmoglein I
Genes with unknown function			
193	<i>TSPAN5</i>	526	Tetraspanin 5
127	<i>DUOX2</i>	64	Dual oxidase 2
99	<i>TMEM14C</i>	476	Transmembrane protein 14C
99	<i>SERPINA12</i>	11	Serpin peptidase inhibitor, clade A, member 12
66	<i>SLC22A5</i>	142	Solute carrier family 22, member 5
56	<i>FLG2</i>	10	Filaggrin 2 (ifapsoriasis)
41	<i>C7orf24</i>	309	Chromosome 7 open reading frame 24

gene is the host gene for 10 C/D box snoRNAs [26]. Other snoRNA host genes included in our ORESTES dataset are *RPS11*, *RPS12*, *RPL10* and *EIF4A1*. In certain cases, ORESTES contain the snoRNA sequence (U39B in *RPS11*, mgU6-77 in *EIF4A1*, U70 in *RPL10*), and probably correspond to alternative splicing forms of the host gene mRNA, with intron retention.

We furthermore obtained sequences for long, non-protein-coding transcripts. Metastasis associated lung adenocarcinoma transcript 1, (*MALAT-1*, 22 ORESTES) is a conserved long non-protein-coding RNA (>8,000 nucleotides (nt)) of unknown function that is highly expressed in numerous healthy organs and overexpressed in metastatic non-small cell lung carcinomas [27]. Close to *MALAT-1* on 11q13.1, trophoblast-derived noncoding RNA (*TncRNA*, 44

ORESTES) is a 481 nucleotide (nt), non-protein-coding RNA involved in trophoblastic major histocompatibility complex suppression by inhibiting class II transactivator (*CIITA*) transcription [28]. *H19* is a non-protein-coding, maternally imprinted mRNA (two spliced ORESTES) [29] that is highly transcribed in extraembryonic and fetal tissues, as well as in adult skeletal muscle. It has been shown that *H19* is involved in the genomic imprinting of the insulin-like growth factor 2 (*IGF2*) gene [30]. Moreover, *IGF2* is expressed throughout the epidermis [31] and its overexpression increases the thickness of the epidermis and the proportion of dividing cells in the basal layer [32]. We suggest that *H19* could participate in the regulation of *IGF2* transcription by maintaining the genomic imprinting of its promoter in adult epidermis. In addition to numerous protein-coding genes, we thus detected several non-protein-coding RNAs whose expression in the

Table 3

Genes with less than 100 UniGene ESTs encoding known GK expressed proteins

No. of ORESTES	Gene symbol	No. of UniGene ESTs	Full name (alias)
2	<i>LCE1F</i>	1	Late cornified envelope 1F
4	<i>LCE2C</i>	1	Late cornified envelope 2C
1	<i>C1orf46</i>	2	Chromosome 1 open reading frame 46 (xp 33)
1	<i>LCE2A</i>	3	Late cornified envelope 2A
6	<i>LCE5A</i>	3	Late cornified envelope 5A
11	<i>LCE1A</i>	3	Late cornified envelope 1A
2	<i>PGLYRP3</i>	5	Peptidoglycan recognition protein 3
5	<i>LCE1C</i>	5	Late cornified envelope 1C
100	<i>FLG</i>	5	Filaggrin
140	<i>LEP7</i>	5	Late envelope protein 7
2	<i>RPTN</i>	6	Repetin
1	<i>LCE2B</i>	7	Late cornified envelope 2B
56	<i>hKPRP</i>	7	Human keratinocyte proline rich protein
9	<i>LOR</i>	11	Loricrin
71	<i>KRT2A</i>	12	Keratin 2A
2	<i>C1orf42</i>	15	Chromosome 1 open reading frame 42 (NICE-1)
5	<i>TGM5</i>	15	Transglutaminase 5
13	<i>DSC1</i>	17	Desmocollin 1
16	<i>KRT1B</i>	18	Keratin 1B
61	<i>CASP14</i>	19	Caspase 14
1	<i>CNFN</i>	20	Cornifelin
4	<i>CALML5</i>	25	Calmodulin-like 5
1	<i>ALOXE3</i>	28	Arachidonate lipoxygenase 3
8	<i>ALOX12B</i>	30	Arachidonate 12-lipoxygenase, 12R type
62	<i>SPRR2E</i>	36	Small proline-rich protein 2E
17	<i>IVL</i>	38	Involucrin
2	<i>EPPK1</i>	42	Epiplakin 1
5	<i>POU2F3</i>	45	POU domain, class 2, transcription factor 3 (oct-11)
4	<i>ICHTHYIN</i>	48	Ichthyin
32	<i>SBSN</i>	49	Suprabasin
2	<i>KLK8</i>	53	Kallikrein 8 (neuropsin/ovasin)
4	<i>TGM3</i>	54	Transglutaminase 3
1	<i>ABCA12</i>	55	ATP-binding cassette, sub-family A (ABC1), member 12
3	<i>PADI1</i>	56	Peptidylarginine deiminase, type 1
30	<i>DSG1</i>	61	Desmoglein 1
2	<i>GJB3</i>	65	Gap junction protein, beta 3 (connexin 31)
1	<i>CALML3</i>	68	Calmodulin-like 3
13	<i>SASPase</i>	69	Skin aspartic protease
15	<i>KLK7/SCCE</i>	69	Kallikrein 7 (Stratum corneum chymotrypticenzyme)
6	<i>A2ML1</i>	76	Alpha-2-macroglobulin-like 1
1	<i>CST6</i>	78	Cystatin E/M
1	<i>SULT2B1</i>	80	Sulfotransferase family, cytosolic, 2B, member 1
2	<i>KLK11</i>	83	Kallikrein 11
3	<i>HAL</i>	86	Histidine ammonia-lyase (histidase)
14	<i>EVPL</i>	91	Envoplakin
59	<i>CDSN</i>	91	Corneodesmosin
3	<i>PDZK1IP1</i>	92	PDZK1 interacting protein 1
4	<i>TGM1</i>	92	Transglutaminase 1
2	<i>SERPINB8</i>	99	Serpin peptidase inhibitor, clade B, member 8
20	<i>SCEL</i>	99	Sciellin

Table 4**Genes with 100 or less UniGene ESTs, known as mainly expressed in a specific tissue different from epidermis**

No. of ORESTES	Gene symbol	No. of UniGene ESTs	Full name	Main specificity
99	<i>SERPINA12</i>	11	Serpin peptidase inhibitor, clade A, member 12	Adipocytes
1	<i>BSND</i>	12	Bartter syndrome, infantile, with sensorineural deafness	Kidney and inner ear
1	<i>OPN1LW</i>	16	Opsin 1, long-wave-sensitive	Eye
5	<i>GRIN2</i>	16	G-protein-regulated inducer of neurite outgrowth	Brain
2	<i>IL1RL2</i>	24	Interleukin 1 receptor-like 2	Neurons
13	<i>LCTL</i>	25	Lactase-like	Kidney
1	<i>PPEF2</i>	31	Protein phosphatase, EF-hand calcium binding domain 2	Retina
2	<i>SLC6A3</i>	34	Solute carrier family 6, member 3	Neuron
3	<i>CDC42BPG</i>	41	CDC42 binding protein kinase gamma	Heart and skeletal muscle
4	<i>GPR75</i>	41	G protein-coupled receptor 75	Retina
1	<i>OTX1</i>	45	Orthodenticle homolog 1	Neurons
1	<i>K5B</i>	46	Keratin 5b	Tongue
1	<i>TBX15</i>	46	T-box 15	Embryo
3	<i>TMPRSS5</i>	53	Transmembrane protease, serine 5 (spinesin)	Spinal chord
3	<i>LEAP-2</i>	53	Liver-expressed antimicrobial peptide 2	Liver
1	<i>BMP8B</i>	56	Bone morphogenetic protein 8B	Embryo
1	<i>PTGFR</i>	59	Prostaglandin F receptor	Uterus
2	<i>TEC</i>	59	Tec protein tyrosine kinase	Hematopoietic cells
3	<i>SLC5A1</i>	60	Solute carrier family 5, member 1	Intestine and kidney
20	<i>CASZ1</i>	61	Castor homolog 1, zinc finger	Mesenchyme
2	<i>KCNJ12</i>	62	Potassium inwardly rectifying channel, subfamily J, 12	Heart
1	<i>P11</i>	64	26 serine protease	Placenta
14	<i>SERPINB7</i>	64	Serpin peptidase inhibitor, clade B, member 7	Mesangial cells
127	<i>DUOX2</i>	64	Dual oxidase 2	Thyroid
3	<i>GDPD2</i>	68	Glycerophosphodiester phosphodiesterase containing 2	Osteoblasts
3	<i>PDE11A</i>	75	Phosphodiesterase 11A	Testis
2	<i>CLDN23</i>	76	Claudin 23	Placenta
1	<i>PLCL4</i>	78	Phospholipase C-like 4	Neurons
1	<i>EYA4</i>	81	Eyes absent homolog 4	Heart and cochlea
1	<i>LIPH</i>	85	Lipase, member H	Intestine
1	<i>RBP3</i>	88	Retinol binding protein 3	Retina

epidermis had not been previously assessed, evoking the possibility that they might play a specific role in this tissue.

Real-time PCR expression profiling of selected genes

Genes involved in the establishment of the skin barrier are expected to be specifically overexpressed by granular keratinocytes. To compare the expression levels of candidate genes between the basal layer and GKs, quantitative real-time PCR experiments were performed with the T4 and T1 cell fractions. Based on predicted domains and homologies, 73 genes represented by less than 100 ESTs were selected (Table 7). The relative T4/T1 ratio could not be calculated for 20 of them due to very low expression levels. Ten genes were equally expressed in the two layers, and nine were overexpressed in the basal layer, even if expressed at a low level in the granular layer. Interestingly, 33 were overexpressed in the granular layer with T4/T1 ratios ranging from 6 to 800. For several genes, the T4/T1 expression ratio was thus much larger than that observed for the *KLK7* gene, used as a specific marker of the GKs in our cell purification experiments (Table

1). Therefore, these data emphasize the high degree of purity of the GKs we have purified from healthy human skin. They also provide one with new, highly specific markers for this cell type.

Identification of new genes**FLG2**

The epidermal differentiation complex (EDC) spans 1.62 megabases on 1q21.3 and contains approximately 50 genes specifically involved in the barrier function, such as those encoding involucrin, loricrin, filaggrin, small proline rich proteins (SPRR1-4) or late cornified envelope proteins (LCE1-5) (Figure 3a). We cloned many sequences corresponding to known genes of this locus (Figure 3b), but also a large number of sequences for a previously poorly characterized transcript encoding filaggrin 2 (FLG2; also called ifapsoriasis (IFPS); (GenBank:AY827490)). FLG2 displays features of the fused-family genes (encoding filaggrin, trichohyalin, or repetin), with three exons and a large predicted protein sequence (2,391 amino acids) containing two

Table 5

Genes with 100 or less UniGene ESTs, corresponding to uncharacterized paralogues of known genes

No. of ORESTES	Gene symbol	No. of UniGene ESTs	Full name
13	ASAH3	4	N-acylsphingosine amidohydrolase 3
1	LIPL2 (LIPK)	4	Lipase-like, ab-hydrolase domain containing 2
11	CLEC2A	7	C-type lectin domain family 2, member A
2	IGFL3	12	Insulin growth factor-like family member 3
3	LYG2	14	Lysozyme-like
1	PNPLA1	15	Patatin-like phospholipase domain containing 1
9	GSDM1	16	Gasdermin 1
1	GRID2IP	17	Glutamate receptor, ionotropic, delta 2 interacting protein
2	IL1F7	17	Interleukin 1 family, member 7
2	FCRL6	17	Fc receptor-like 6
3	AADA2L2	20	Arylacetamide deacetylase-like 2
1	LIPL3 (LIPM)	20	Lipase-like, ab-hydrolase domain containing 3
1	LAMB4	26	Laminin, beta 4
10	THEM5	26	Thioesterase superfamily member 5
1	FLJ90165	27	Gamma-glutamyltransferase 6 homolog
3	FLJ45651	28	Phospholipase A2, group IVE
1	LPIN3	31	Lipin 3
1	SLC25A34	36	Solute carrier family 25, member 34
1	GPRI15	37	G protein-coupled receptor 115
1	LRP5L	41	Low density lipoprotein receptor-related protein 5-like
1	HSPC105	45	NAD(P) dependent steroid dehydrogenase-like
1	QPCTL	52	Glutamyl-peptide cyclotransferase-like
3	PLA2G4F	56	Phospholipase A2, group IVF
1	KIAA0605	58	ADAMTS-like 2
1	CTGLF1	60	Centaurin, gamma-like family, member 1
2	UGT3A2	65	UDP glycosyltransferase 3 family, polypeptide A2
1	GALNT17	74	Polypeptide N-acetylgalactosaminyltransferase 17
1	BAIAP2L2	76	BAI1-associated protein 2-like 2
1	FLJ43692	80	ARHGEF5-like
1	VILL	86	Villin-like
1	LOC203427	87	Similar to solute carrier family 25, member 16
1	IL17RE	100	Interleukin 17 receptor E

calcium binding EF-hand domains and a large domain made of repeated segments of about 25 amino acids. The amino acid composition of FLG2 is very similar to that of filaggrin, with a high content of serine (22%), glycine (20%), histidine (10%) and glutamine (10%). The expression of this gene is likely restricted to the epidermis, as shown by PCR on a panel of cDNAs from 16 healthy human tissues and organs (Figure 4). Real-time PCR also showed a strong overexpression of the FLG2 gene in GKs, with a T4/T1 ratio of 800 (Table 7). These results thus suggest that this gene is a new functional member of the EDC complex, in agreement with its similarity to the filaggrin gene, whose function in the epidermal barrier is well established.

Lipase-like genes

Two ORESTES were identified as the human orthologues of the murine lipases *Lipl2* (NM_172837) and *Lipl3* (BC031933), previously identified by large-scale mouse cDNA sequencing by the Riken Institute [33] and the Mammalian Gene Collection program [34], respectively. The corresponding human genes *LIPL2* and *LIPL3* were clustered in a 665 kb interval on chromosome 10q23.31 with genes encoding two experimentally characterized lipases, *LIPA* (lysosomal acid lipase, MIM +27,8000) and *LIPF* (gastric lipase, MIM #601980) and two hypothetical lipase-like proteins, *LIPL1* and *LIPL4* (Figure 5a). Therefore, our study contributed to the elucidation of a specialized human genomic locus that includes six lipase genes and four other genes (*ANKRD22*, *STAMBPL1*, *ACT2* and *FAS*) of apparently unrelated function (Figure 5a). In accordance with the Hugo Gene

Table 6**Unknown genes with 100 or less UniGene ESTs**

No. of ORESTES	Gene symbol	No. of UniGene ESTs	Full name
1	<i>FLJ43861</i>	3	Fij43861
1	<i>LOC389791</i>	3	Hypothetical gene supported by AK094537
1	<i>LOC285435</i>	4	Hypothetical LOC285435
2	<i>LOC387846</i>	6	Hypothetical LOC387846
4	<i>LOC401062</i>	6	Hypothetical gene supported by AK092973
1	<i>IMAGE:5260914</i>	7	Image:5260914
5	<i>LOC338667</i>	7	Hypothetical protein LOC338667
5	<i>PSORS1C2</i>	8	Psoriasis susceptibility 1 candidate 2
1	<i>DKFZp779B1540</i>	9	Hypothetical protein dkfzp779b1540
5	<i>C14orf72</i>	9	Chromosome 14 open reading frame 72
1	<i>FLJ37989</i>	10	Fij37989
56	<i>FLG2</i>	10	Filaggrin 2 (ifapsoriasis)
10	<i>WFDC5</i>	11	WAP four-disulfide core domain 5
1	<i>LOC402110</i>	12	Hypothetical LOC402110
1	<i>PLEKHN1</i>	12	Pleckstrin homology domain containing, family N member 1
1	<i>LOC441240</i>	13	Hypothetical protein LOC441240
4	<i>FLJ38159</i>	14	Hypothetical protein FLJ38159
1	<i>C1orf177</i>	15	Chromosome 1 open reading frame 177
1	<i>HMCN2</i>	16	Hemicentin 2
2	<i>MGC23985</i>	16	Similar to AVLV472
1	<i>OFCC1</i>	17	Orofacial cleft 1 candidate 1
1	<i>LOC441860</i>	17	Novel KRAB box containing C2H2 type zinc finger protein
5	<i>AMIGO3</i>	18	Adhesion molecule with Ig-like domain 3
1	<i>LOC441257</i>	20	Hypothetical protein LOC441257
1	<i>LOC285484</i>	20	Hypothetical protein LOC285484
1	<i>C20orf91</i>	20	Chromosome 20 open reading frame 91
1	<i>LOC202460</i>	21	Hypothetical protein LOC202460
2	<i>FLJ25664</i>	21	Fij25664
8	<i>FLJ41623</i>	21	Fij41623
10	<i>LOC342897</i>	21	Similar to F-box only protein 2
1	<i>LOC339237</i>	23	Similar to Envoplakin
13	<i>LOC126248</i>	24	Hypothetical protein LOC126248
1	<i>LOC389142</i>	27	Hypothetical LOC389142
3	<i>C20orf95</i>	28	Chromosome 20 open reading frame 95
1	<i>FKBP9L</i>	31	FK506 binding protein 9-like
1	<i>FNDC8</i>	31	Fibronectin type III domain containing 8
3	<i>FLJ46311</i>	31	FLJ46311 protein
1	<i>C3orf47</i>	33	Chromosome 3 open reading frame 47
1	<i>LOC283143</i>	35	Hypothetical protein LOC283143
1	<i>LOC388727</i>	35	Hypothetical LOC388727
2	<i>FLJ44317</i>	35	Fij44317
1	<i>FLJ31184</i>	36	Fij31184
1	<i>LOC125893</i>	39	Hypothetical protein LOC125893
1	<i>ZNF311</i>	40	Zinc finger protein 311
1	<i>BC041923</i>	40	Image:5300199
2	<i>ZNF600</i>	43	Zinc finger protein 600
3	<i>MCMDC1</i>	43	Minichromosome maintenance deficient domain containing 1
3	<i>FLJ13646</i>	46	Hypothetical protein FLJ13646
1	<i>C14orf121</i>	48	Chromosome 14 open reading frame 121
1	<i>FAM83F</i>	49	Family with sequence similarity 83, member F
3	<i>ABHD9</i>	51	Abhydrolase domain containing 9
1	<i>LOC134466</i>	52	Hypothetical protein LOC134466
1	<i>CXorf33</i>	52	Chromosome X open reading frame 33

Table 6 (Continued)

Unknown genes with 100 or less UniGene ESTs

2	<i>FLJ25006</i>	52	Hypothetical protein FLJ25006
2	<i>DKFZp434N062</i>	53	Hypothetical protein dkfzp434n062
9	<i>LASS3</i>	53	LAG1 longevity assurance homolog 3
1	<i>Cl4orf21</i>	54	Chromosome 14 open reading frame 21
1	<i>Cl7orf67</i>	56	Chromosome 17 open reading frame 67
1	<i>FAM62C</i>	58	Family with sequence similarity 62, member C
2	<i>Cl4orf29</i>	60	Chromosome 14 open reading frame 29
1	<i>FLJ21736</i>	61	Esterase 31
3	<i>LOC349114</i>	61	Hypothetical protein LOC349114
1	<i>MGC26885</i>	62	Hypothetical protein MGC26885
1	<i>SMA3</i>	62	Sma3
8	<i>FAM46B</i>	62	Family with sequence similarity 46, member B
13	<i>ELMOD1</i>	62	ELMO/CED-12 domain containing 1
3	<i>DENND2C</i>	63	DENN/MADD domain containing 2C
13	<i>ANKRD35</i>	64	Ankyrin repeat domain 35
5	<i>LOC401553</i>	66	Hypothetical gene supported by BC019073
1	<i>LOC390927</i>	67	Similar to zinc finger protein 569
1	<i>ZNF696</i>	67	Zinc finger protein 696
2	<i>CCDC9</i>	69	Coiled-coil domain containing 9
6	<i>Cl5orf40</i>	70	Chromosome 15 open reading frame 40
1	<i>LOC148137</i>	73	Hypothetical protein BC017947
1	<i>ZC3H12C</i>	74	Zinc finger CCCH-type containing 12C
1	<i>APXL2</i>	74	Apical protein 2
1	<i>ZMYND19</i>	75	Zinc finger, MYND-type containing 19
1	<i>LRRC37B</i>	77	Leucine rich repeat containing 37B
2	<i>FLJ32356</i>	77	Family with sequence similarity 109, member A
3	<i>DQX1</i>	77	DEA Q box polypeptide 1
2	<i>C9orf9</i>	79	Chromosome 9 open reading frame 9
4	<i>FNDC6</i>	79	Fibronectin type III domain containing 6
1	<i>MSTP9</i>	82	Macrophage stimulating, pseudogene 9
1	<i>HES2</i>	83	Hairy and enhancer of split 2
1	<i>FLJ37464</i>	84	Hypothetical protein FLJ37464
1	<i>KIAA1862</i>	84	KIAA1862 protein
1	<i>LOC196264</i>	86	Hypothetical protein LOC196264
1	<i>Clorf51</i>	86	Chromosome 1 open reading frame 51
2	<i>ANKRD5</i>	86	Ankyrin repeat domain 5
1	<i>CXorf23</i>	87	Chromosome X open reading frame 23
2	<i>SMCR8</i>	88	Smith-Magenis syndrome chromosome region, candidate 8
2	<i>DKFZp686L1814</i>	88	Hypothetical protein dkfzp686l1814
2	<i>MGC34647</i>	88	Hypothetical protein MGC34647
1	<i>C6orf105</i>	89	Chromosome 6 open reading frame 105
1	<i>FLJ23186</i>	90	Chromosome 3 open reading frame 52
1	<i>SAMD10</i>	91	Sterile alpha motif domain containing 10
2	<i>KIAA1287</i>	91	Kiaa1287
2	<i>Cl9orf36</i>	91	Chromosome 19 open reading frame 36
1	<i>ZNF662</i>	93	Zinc finger protein 662
2	<i>ZNF429</i>	97	Zinc finger protein 429
12	<i>SMPD3</i>	98	Sphingomyelin phosphodiesterase 3
1	<i>PCGF1</i>	100	Polycomb group ring finger 1
1	<i>Cl7orf61</i>	100	Chromosome 17 open reading frame 61
1	<i>KIAA1853</i>	100	Kiaa1853
1	<i>C22orf23</i>	100	Chromosome 22 open reading frame 23

Table 7**Comparison of gene expression in the T4 and T1 cell fractions by real-time PCR**

No. of ORESTES	Gene symbol	Full name	No. of UniGene ESTs	T4/T1 expression ratio
56	<i>FLG2</i>	Filaggrin 2 (ifapsoriasis)	10	800
9	<i>GSDM1</i>	Gasdermin 1	16	800
13	<i>ASAH3</i>	N-acylsphingosine amidohydrolase 3	4	300
2	<i>IL1F7</i>	Interleukin 1 family, member 7	17	200
13	<i>ELMOD1</i>	ELMO/CED-12 domain containing 1	62	160
0	<i>LIPL4 (LIPN)</i>	Lipase-like, ab-hydrolase domain containing 4	9	150
1	<i>LIPL3 (LIPM)</i>	Lipase-like, ab-hydrolase domain containing 3	20	130
3	<i>C2orf95</i>	Chromosome 20 open reading frame 95	28	120
1	<i>WFDC12</i>	WAP four-disulfide core domain 12 (WAP2)	3	110
1	<i>LIPL2 (LIPK)</i>	Lipase-like, ab-hydrolase domain containing 2	4	100
12	<i>SMPD3</i>	Sphingomyelin phosphodiesterase 3	98	70
10	<i>LOC440449</i>	Similar to WDNM1 homolog (LOC645638)	-	50
99	<i>SERPINA12</i>	Serpin peptidase inhibitor, clade A, member 12	11	50
1	<i>P11</i>	26 serine protease	64	35
5	<i>PSORS1C2</i>	Psoriasis susceptibility 1 candidate 2	8	30
1	<i>C3orf52</i>	Chromosome 3 open reading frame 52	90	30
2	<i>CLDN23</i>	Claudin 23	76	25
1	<i>PNPLA1</i>	Patatin-like phospholipase domain containing 1	15	20
10	<i>THEM5</i>	Thioesterase superfamily member 5	26	20
3	<i>ABHD9</i>	Abhydrolase domain containing 9	51	20
2	<i>TMEM16H</i>	Transmembrane protein 16H	78	16
6	<i>SERPINB12</i>	Serpin peptidase inhibitor, clade B, member 12	6	15
1	<i>PLEKHN1</i>	Pleckstrin homology domain containing, family N member 1	12	12
1	<i>FAM83F</i>	Family with sequence similarity 83, member F	49	12
3	<i>AADACL2</i>	Arylacetamide deacetylase-like 2	20	12
10	<i>LOC342897</i>	Similar to F-box only protein 2	21	10
9	<i>LASS3</i>	LAG1 longevity assurance homolog 3	53	10
20	<i>CASZ1</i>	Castor homolog 1, zinc finger	61	10
8	<i>FAM46B</i>	Family with sequence similarity 46, member B	62	10
14	<i>SERPINB7</i>	Serpin peptidase inhibitor, clade B, member 7	64	10
1	<i>CARD14</i>	Caspase recruitment domain family, member 14	86	8
1	<i>GGT6</i>	Gamma-glutamyltransferase 6 homolog	27	6
1	<i>CXorf33</i>	Chromosome X open reading frame 33	52	6
10	<i>WFDC5</i>	WAP four-disulfide core domain 5 (WAP1)	11	1
5	<i>KIAA0514</i>	Kiaa0514	16	1
5	<i>AMIGO3</i>	Adhesion molecule with Ig-like domain 3	18	1
3	<i>PLA2G4E</i>	Phospholipase A2, group IVE	28	1
1	<i>HSPC105</i>	NAD(P) dependent steroid dehydrogenase-like	45	1
3	<i>PLA2G4F</i>	Phospholipase A2, group IVF	56	1
127	<i>DUOX2</i>	Dual oxidase 2	64	1
1	<i>APXL2</i>	Apical protein 2	74	1
3	<i>RAB38</i>	RAB38, member RAS oncogene family	79	1
1	<i>AFMID</i>	Arylformamidase	87	1
2	<i>ANKRD5</i>	Ankyrin repeat domain 5	86	0.46
3	<i>MCMD1</i>	Minichromosome maintenance deficient domain containing 1	43	0.4
1	<i>C14orf21</i>	Chromosome 14 open reading frame 21	54	0.38
3	<i>LEAP-2</i>	Liver-expressed antimicrobial peptide 2	53	0.36
1	<i>LRP5L</i>	Low density lipoprotein receptor-related protein 5-like	41	0.35
4	<i>FNDC6</i>	Fibronectin type III domain containing 6	79	0.3
2	<i>KIAA1287</i>	Kiaa1287	91	0.28

Not determined: *FLJ21736*, *HES2*, *ZNF696*, *LYG2*, *MGC34829*, *KIAA0963*, *ALKBH*, *ADAMTSL2*, *c6orf105*, *AW367250*, *KCNJ12*, *LOC338667*, *TMPRSS5*, *CLEC2A*, *FLJ21736*, *c14orf29*, *c14orf72*, *KRBA1*, *FLJ37464*, *LOC134466*.

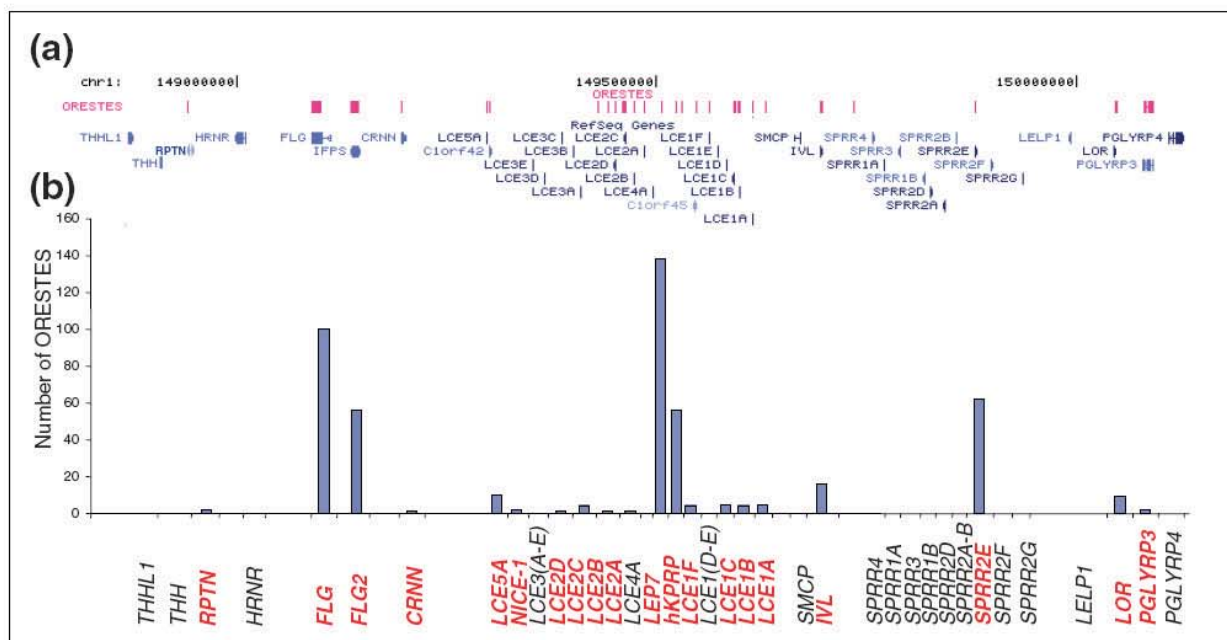


Figure 3 Genes of the EDC present in the ORESTES dataset. (a) Screen copy of a UCSC Genome Browser window (chr1:150,300,000-151,590,000; hg17, May 2004) showing the RefSeq genes from the EDC, and the ORESTE custom track. (b) Number of ORESTES for each gene of the locus. The genes for which at least one ORESTE was sequenced are in red bold characters.

Nomenclature Committee (HGNC), these new hypothetical lipase genes, *LIPL1*, *LIPL2*, *LIPL3* and *LIPL4*, have been renamed *LIPJ*, *LIPK*, *LIPM*, and *LIPN*, respectively.

Rapid amplification of cDNA ends (RACE) cloning experiments of the human *LIPK* mRNA localized the cap site 165 nt upstream of the conceptual ATG initiation codon. For the *LIPM* and *LIPN* mRNAs, the entire coding sequences have been cloned using primers deduced from alignments with the mouse cDNAs (GenBank:EF426482, GenBank:EF426484 and GenBank:EF426483).

The conceptual translation of *LIPJ*, *LIPK*, *LIPM*, and *LIPN* ORFs led to proteins of 366, 399, 423 and 398 amino acids, respectively. As is the case for *LIPA* and *LIPF*, these proteins contain an abhydrolase associated lipase region (PFAM PF04083) with an alpha/beta hydrolase fold domain (PFAM PF00561). More precisely, they display the characteristic Ser-Asp-His catalytic triad (Figure 5b) also present in various hydrolases, such as serine proteases [35,36]. Surprisingly, *LIPJ* (NM_001010939) appeared as the only member in this family to lack a cleavable signal peptide (Figure 5b) and might thus play a distinct role in human physiology.

The clustering of six lipase genes on the human 10q23.31 locus led us to examine in detail the syntenic mouse locus (19qC1), where the orthologous genes except for *LIPN* and *LIPJ* had already been identified or predicted. Sequence

analysis showed that the AK154333 cDNA clearly corresponds to the *LIPN* orthologue (89% homology). As in the human genome, this gene resides between the *Lipl2* and *Lipl3* genes. Intriguingly, we could not precisely identify a single orthologue of *LIPJ*. Instead, we found upstream of the *Lipf* gene a cluster of five predicted genes encoding secreted alpha/beta hydrolases as well as at least three pseudogenes containing premature stop codons (Figure 5a). A syntenic cluster, albeit of smaller size, resides in the rat genome. These five hypothetical functional mouse lipases form by themselves a separate group in the phylogenetic tree (Figure 5c; Additional data file 4). Their genomic localization suggests that they could have arisen from a large expansion by tandem duplications well after the separation of the rodents from the primate ancestors, or from a lipase gene lost in the human genome. These five new genes have thus been renamed *Lipdc1-5* for lipase domain containing1-5 in accordance with the Mouse Gene Nomenclature Committee (MGNC) (Additional data file 5).

The human *LIPK*, *LIPM* and *LIPN* genes appear to be exclusively expressed in the epidermis, as shown by PCR on a panel of 16 cDNAs (Figure 4). Their expression was highly specific for GKs, with real-time PCR T4/T1 ratios of 100-150 (Table 7). *LIPJ* expression was also detected in the epidermis by PCR, but was too weak to allow real-time PCR experiments to be performed. These results strongly suggest that the *LIPK*, *LIPM* and *LIPN* genes play a highly specific role in the last

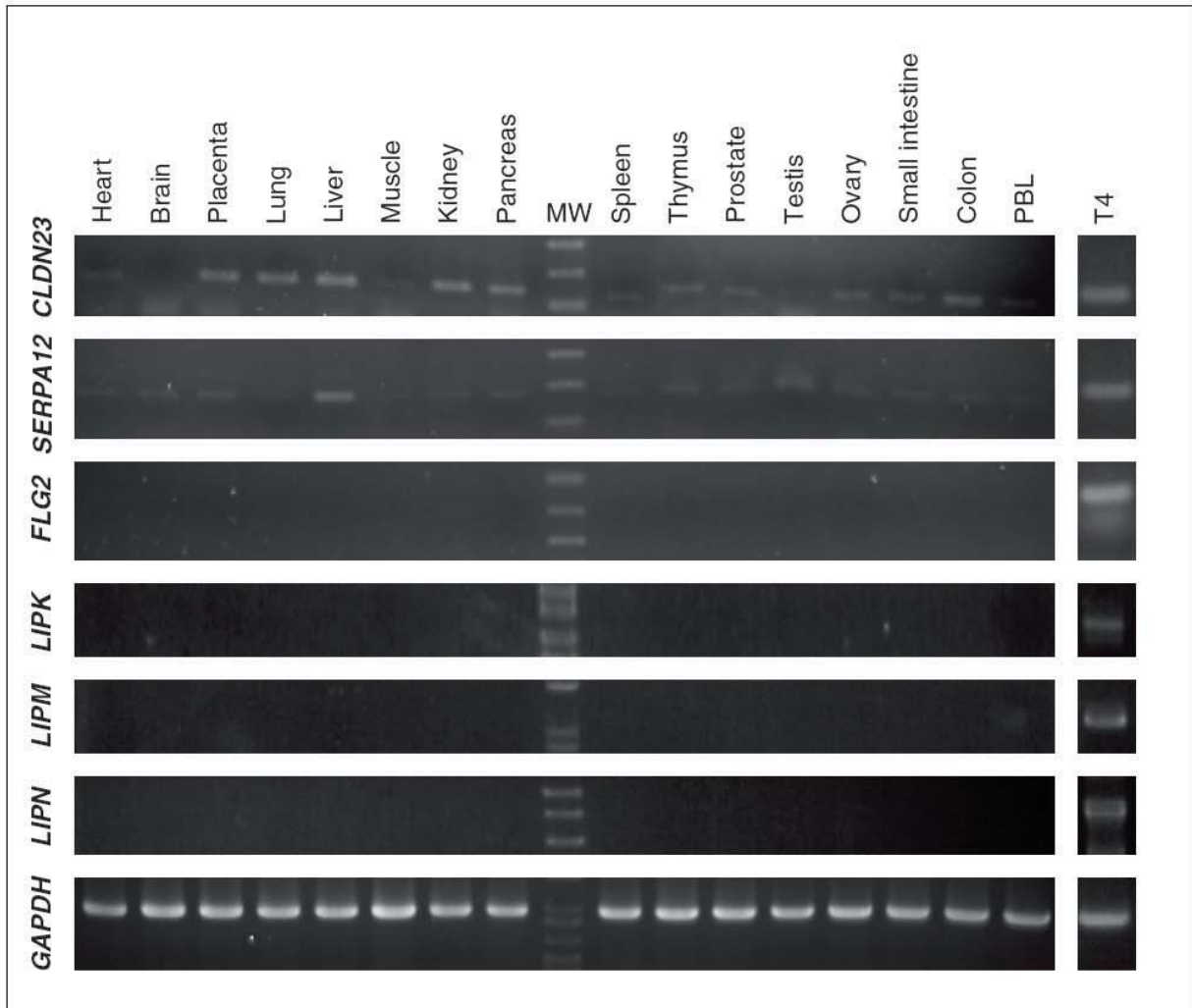


Figure 4
 Expression profile of newly identified genes. PCR experiments were performed with a commercial panel of cDNAs from 16 human tissues (PBL, peripheral blood leukocytes) and with cDNAs prepared from the T4 fraction enriched in GKs. For each gene, PCR primers were chosen to amplify a cDNA fragment encompassing at least two exons. Note the highly specific expression pattern of *FLG2*, *LIPK*, *M*, *N*, and, to a lesser extent, *SERPINA12* genes. The apparent size variation of the *CLDN23* fragment results from an artifactual gel distortion. Expression of *GAPDH*, assessed with the primers provided by the manufacturer, was used as a control.

step of keratinocyte differentiation. Although highly related, the LIPJ protein might play a different role, as previously suggested by its lack of a signal peptide.

C1orf81

Six spliced ORESTES corresponded to a totally unknown gene without mRNA sequences in the databases. We performed RACE experiments and identified 16 exons and 15 introns with canonical splice sites and a consensus polyadenylation signal (GenBank: [DQ983818](http://www.ncbi.nlm.nih.gov/nuccore/DQ983818)). In agreement with the HGNC, we named this gene *C1orf81*. To date, only 16 ESTs are present in the databases, one from testis and the

others from pooled tissues (Additional data file 1). PCR on a panel of cDNAs from 16 healthy human tissues and organs showed that spliced transcripts for this gene are detectable in most samples (Figure 6a). From its conceptual translation, the *C1orf81* mRNA might encode a 373 amino acid protein. However, a second ORF was present in another frame that overlapped the first one by 87 amino acids. This suggested that a longer protein might be produced by a ribosomal frameshift mechanism. We explored the *C1orf81* gene orthologues in vertebrate genomes, and particularly the region corresponding to the ORF overlap (exons 7-9). Compared with the *C1orf81* gene from 10 other mammals, exon 8 of the

human gene contained a 1 base-pair insertion that creates the frameshift (Figure 6b). The possibility of a sequence error in the present human genome assembly was excluded by sequencing the exon 8 region from two individuals. Accordingly, the human *C1orf81* gene product would be 373 amino acids long, but 762 and 714 amino acids long in the chimpanzee and rhesus monkey, respectively. In all cases, the analysis of the peptide sequence did not reveal known domains or signal peptide. Further studies are needed to uncover the role of this new gene in mammalian physiology, and evaluate the consequences of its possible inactivation in human.

Discussion

We have described here the first large-scale study of the transcriptome of human epidermal cells. As we are interested in genes that participate in the barrier function of the skin, we focused on GKs. They correspond to the ultimate step in the course of keratinocyte differentiation, and are the last epidermal cells to display gene expression activity before undergoing a particular programmed cell death leading to cornification. Because these cells represent less than 10% of the epidermis population, a preliminary step was to design an efficient purification method starting from healthy human skin fragments. After unsuccessful attempts using size filtration, Ficoll gradients, and fluorescence-activated cell sorting, we used successive short-term enzyme incubations to progressively detach cells from the deep layers, and purify the cells that remain attached to the cornified layer. These incubations were performed at 4°C to stop cellular metabolic activity and preserve the mRNA pool from degradation. This point is highly relevant as many growth factors, cell cycle regulators and transcription factors are encoded by short-lived mRNA. Quantitative PCR experiments were used to assess the relative expression levels of several genes in the successive cell fractions.

ORESTE technique

In the present transcriptome project, the ORESTE methodology was selected because it produces sequences that are distributed predominantly within the central part of the corresponding transcripts and is biased towards less-abundant mRNAs [9]. When using arbitrarily chosen primers for reverse transcription and PCR, the amplification of a given transcript is proportional to its length and to the probability for the primer to anneal at low stringency (37°C). This conceptual normalization was secondarily strengthened by removing cDNA libraries with highly abundant amplification products. After the elimination of irrelevant sequences, we produced 16,591 ORESTES representing 3,387 genes. The distribution of the ORESTE number per gene (Figure 2) and the low number of ESTs in the corresponding UniGene entries fully confirmed that the transcript normalization obtained with this method is at least comparable to that of

hybridization-subtraction techniques (for an example, see [37]).

Overrepresented genes

Nevertheless, a few genes were represented by more than 100 sequences obtained from different mini-libraries. These include ubiquitous, highly expressed genes like *RPS8* and *EEF1A1*, and genes already known to be highly transcribed in GKs (those encoding keratin 1, late envelope protein 7 or profilaggrin). Other members of this class, *DMKN* (formerly *ZD52F10*), *SERPINA12* and *FLG2* were not previously known to be overexpressed in the skin, and were thus studied in more detail. Two isoforms of transcripts for the dermokine gene were previously reported [15,16] and we describe 13 novel mRNA isoforms that are either ubiquitous or epidermis-specific depending upon alternative promoter usage. The epidermis-specific forms encode secreted proteins of still unclear function, and are abundantly transcribed in this tissue [17].

SERPINA12, also named *Vaspin* (visceral adipose tissue-derived serine protease inhibitor), is an extracellular serine protease inhibitor that displays insulin-sensitizing effects [38]. In addition to the previously documented expression in the liver, we observed that the *SERPINA12* gene is highly expressed in the human epidermis (Figure 4). Moreover, real-time PCR showed that it is specifically expressed by GKs (T4/T1 ratio of 50). The Serpin A12 protein might thus play a role in the regulation of the complex balance between various proteases and their inhibitors operative in the desquamation process.

The filaggrin 2 (*FLG2*; or ifapsoriasis (*IFPS*)) gene is a new member of the EDC (chromosome 1q21.3), a cluster of approximately 50 genes involved in cornification. Our real-time RT-PCR experiments revealed that *FLG2* expression displays strong specificity for GK. Interestingly, its composition is very close to that of filaggrin. The degradation of filaggrin is considered to be at the origin of the free amino acid pool of the natural moisturizing factor, which is capable of attracting and retaining water in the cornified layer to achieve skin softness and flexibility. We suggest that *FLG2* might have a similar function.

C1orf81 might have lost its function during hominization

A new gene characterized in this study, *C1orf81*, is particularly intriguing as the corresponding transcript (2,284 nt) displays an ORF disruption by a human-specific, single nucleotide insertion in the eighth exon (Figure 6b). As a result, a premature stop codon resides in exon 9, 103 nt upstream of its 3' end. Consequently, the human *C1orf81* mRNA might be degraded by the nonsense mediated mRNA decay (NMD) pathway [39]. This appears unlikely, however, as its expression in various tissues was detected by RT-PCR (Figure 6a). However, its expression in epidermis was too weak to

comment

reviews

reports

deposited research

referenced research

interactions

information

perform real-time PCR experiments. Our analysis suggests that the human *C1orf81* gene encodes a truncated protein relative to other mammalian species, including the rhesus monkey. Nevertheless, we do not exclude the possibility that the translation of the human *C1orf81* mRNA might produce a full-length protein by a +1 ribosomal frameshift mechanism. To our knowledge, such a phenomenon has been uniquely described in human for the ornithine decarboxylase antizyme gene family [40]. This would thus constitute a further example of a (partial) loss of gene function during hominoid evolution. The systematic comparison of the human and chimpanzee genomes has revealed nine other human-specific frameshift mutations (pseudogenization) leading to the mRNA decay or carboxy-terminal protein truncation [41]. Gene losses that occurred after human-chimpanzee divergence could play a role in adaptive evolution, as shown for caspase 12 inactivation during hominization [42]. In this framework, it is intriguing that a member of the hair keratin gene complex, *KRTHAP1*, and a serine protease inhibitor gene, *SERPINA13*, are also among the few genes specifically inactivated in human by a mutation [43,44].

In addition, we sequenced ORESTES for *CTAGE5* and *CNOT6L* functional retrogenes, which are specific for primate and hominoid lineages, respectively. Moreover, the *CTAGE5* retrogene is included in a larger, human-specific insertion (data not shown). It is indeed known that new genes have emerged after a burst of retroposition in primates [45]. Not surprisingly, this suggests that modifications of genes expressed in the skin participate in obvious differences between human and chimpanzee.

Real-time PCR experiments

Real-time PCR experiments were used to measure the relative expression ratio of selected genes in the T4 (mainly GKs) and T1 (basal layer) fractions. In initial experiments using the *SCCE/CLK7* gene as a GK specific marker, the T4/T1 ratios (54-189; Table 1) indicated that the T4 fraction was indeed highly enriched in GKs. However, subsequent experiments led to even higher ratios (for example, 800 for *FLG2* and *GSDM*). Our approach thus led to the discovery of new, exquisitely specific gene markers for GKs that constitute valuable tools for detailed studies of epidermis architecture by histochemistry and *in situ* hybridization methods. Moreover, such highly specific expression strongly suggests that the corresponding genes play key roles in barrier function. The real-

time PCR studies also revealed equal expression of several genes in the T4 and T1 fractions, suggesting that they are transcribed during all steps of keratinocyte differentiation. In addition to providing one with a highly purified GK fraction, our human skin fractionation method thus constitutes a new tool for the characterization of genes involved in the successive steps of terminal differentiation in the epidermis.

GK-specific candidate genes

To further characterize genes poorly represented in databases, the 330 genes with the lowest EST number in the UniGene database (≤ 100 ESTs) were analyzed in more detail. Among these, the known specific genes involved in keratinocyte terminal differentiation account for only 15% of the panel, whereas 42% (139) encode hypothetical proteins. This shows that genes expressed specifically in the uppermost layers of the epidermis are poorly represented in the sequence databases, and suggests that some genes encoding hypothetical proteins may play a functional role in the late steps of epidermal differentiation. We specially focused on genes potentially involved in desquamation regulation as well as lipid metabolism and transport, considering their importance for barrier function establishment. Thus, 73 candidate genes were chosen for a quantitative study of their expression in the course of epidermal differentiation by real-time RT-PCR. The T4/T1 ratio for 20 of them could not be calculated by quantitative RT-PCR using the SYBR Green method, due to a very low expression level. This underlines the sensitivity of the ORESTES method in detecting rare transcripts. Among the 52 remaining genes, real-time RT-PCR experiments revealed that 33 are upregulated during late epidermis differentiation, and hypotheses on the function of some of them are presented below. Moreover, the expression pattern of 6 of these genes is mostly restricted to epidermis as shown by PCR on a panel of 16 human cDNAs prepared from various organs (Figure 4). Among the 330 genes poorly represented in EST databases that we identified, a significant proportion of granular keratinocyte specific genes are thus suspected to be present.

Expression of proteases and protease inhibitors in the human epidermis

The balance between proteases and protease inhibitors is essential to desquamation [46]. The inactivation of the protease inhibitor cystatin M/E gene in mice causes lethality and defects in epidermal cornification [47]. Mutations in the *SPINK5* gene (encoding another protease inhibitor) are

Figure 5 (see following page)

Analysis of new human lipase genes. (a) Schematic representation of the lipase gene cluster on chromosome 10q23.31. Six lipase genes, including the newly described *LPLI* (*LIPJ*), *LPL2* (*LIPK*), *LPL3* (*LIPM*) and *LPL4* (*LIPN*), form a cluster also containing four unrelated genes. (b) Alignment of the protein sequences of the six human lipases from the chromosome 10q23.31 cluster. The amino acids of the catalytic triad are boxed. The alignment was generated with Multalin software [71]. (c) Hierarchical clustering of human and mouse abhydro-lipase gene family members. The human *LIPA*, *LIPF*, *LIPK*, *LIPM* and *LIPN*, but not *LIPJ*, proteins have clear mouse orthologues (lower-case gene names). The six hypothetical mouse genes found in place of the *LIPJ* gene (*Lipdcl-5*) form a separate branch of the phylogenetic tree. This tree was generated with the Tree Top software [72]. Bootstrap values are indicated in red.

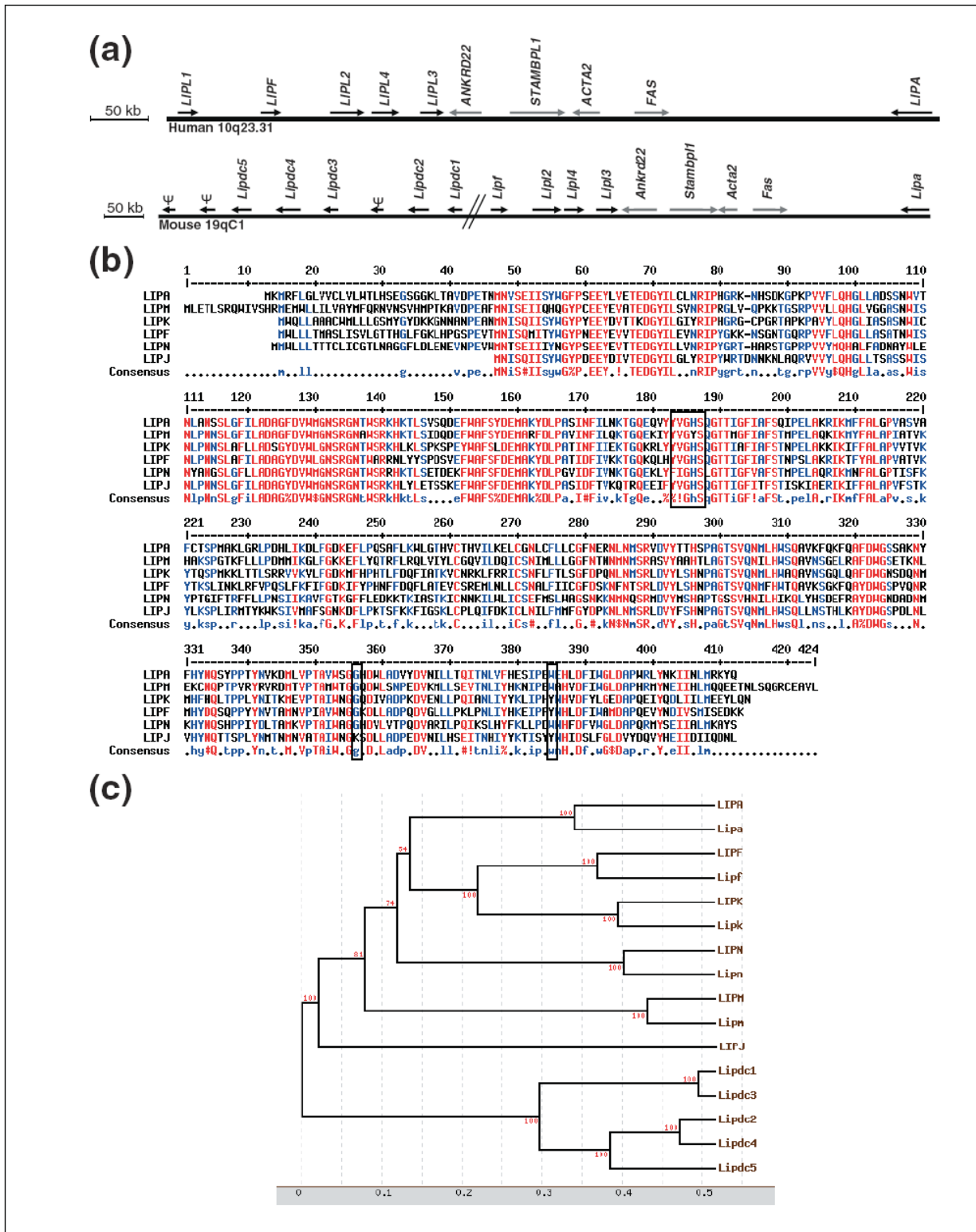


Figure 5 (see legend on previous page)

comment
reviews
reports
deposited research
refereed research
interactions
information

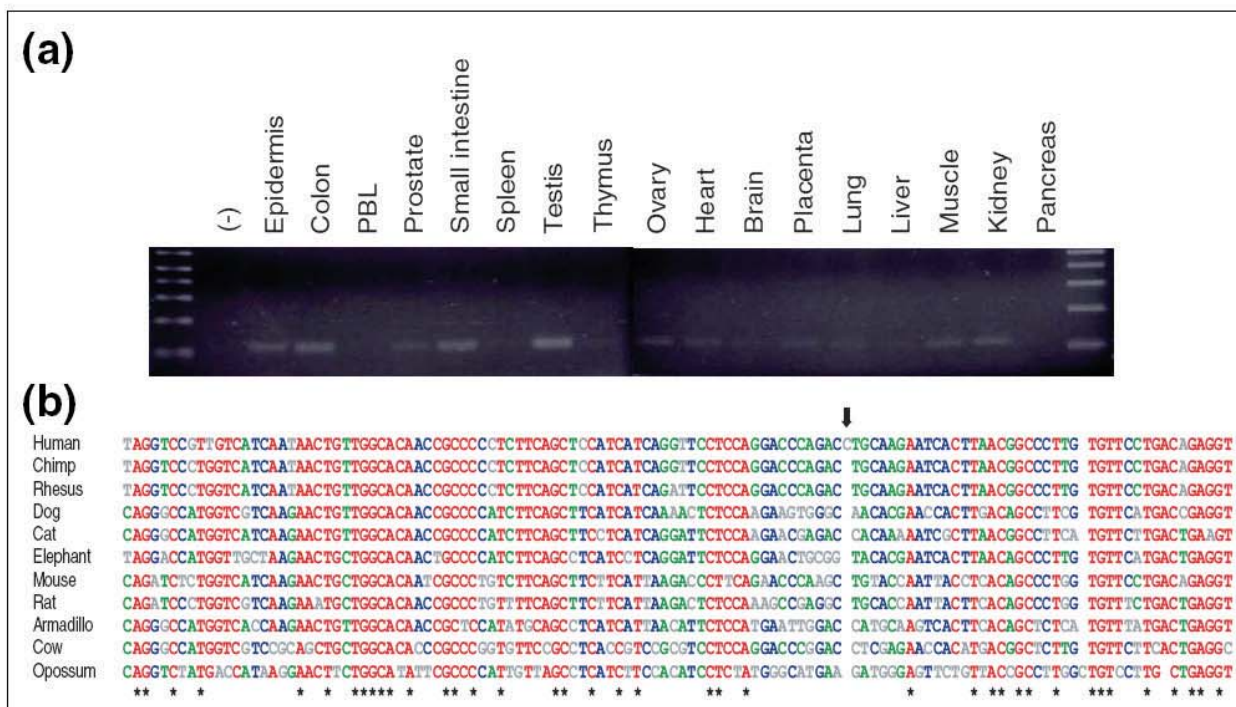


Figure 6
Clorf81 mRNA expression and conservation of the eighth exon among mammals. **(a)** Expression pattern of the *Clorf81* gene. PCR was performed with a commercial panel of cDNAs from 16 human tissues (PBL, peripheral blood leukocytes) and with cDNA prepared from epidermis. The amplified fragment (120 nt) encompasses exons 13-14. **(b)** Sequence alignment of the eighth exon of the *Clorf81* gene from 11 mammals. The sequences were retrieved from the multiz17way table of the UCSC Genome Browser [73], and from a BLAST search of the cat genome. The consensus splicing signals are boxed. The black arrow indicates the single nucleotide insertion in the human gene. The alignment was created with Multalin software [71]. Asterisks indicate the positions conserved in all the sequences. The colors correspond to various levels of consensus, with red for high consensus and grey for low consensus.

responsible for Netherton syndrome (OMIM #256500), characterized by ichthyosiform erythroderma, bamboo hair and atopic dermatitis. In this framework, we identified three protease inhibitors potentially involved in desquamation.

In addition to the serine protease inhibitor *SERPINA12* discussed above, we identified two other members of the serpin superfamily, *SERPINB7* and *SERPINB12*, whose expression in the epidermis is reported here for the first time. Moreover, our real-time PCR experiments clearly show that these two protease inhibitors (as well as *SERPINA12*) are overexpressed in the uppermost epidermal layers. *SERPINB7*, also known as Megsin, is deposited in the extracellular matrix by kidney mesangial cells [48], but its targets have not been identified to date. *SERPINB12* is expressed in many tissues, and displays inhibitor activity against trypsin-like serine proteases [49]. To understand the roles of these protease inhibitors in desquamation, it is of key interest to determine their molecular targets. Proteases expressed in the skin and potentially involved in desquamation are interesting candidates. Our ORESTES data set includes the serine protease kallikrein 7 (*SCCE*), which plays a key function in desquamation by cleaving two corneodesmosome components, desmo-

collin 1 and corneodesmosin [50]. Surprisingly, we did not detect kallikrein 5 (*SCTE*), which cleaves another corneodesmosome component, desmoglein 1. However, we detected for the first time the transcription of the *P11* gene in the epidermis. This gene encodes a secreted serine protease previously shown to be expressed in the human placenta and various neoplasms of the breast, ovary, testis, and stomach [51]. Its specific targets remain unknown. In epidermis, *P11* is upregulated in the granular layer with a T4/T1 ratio of 35 (Table 7). We suggest that *P11* could act in the course of desquamation, either by cleaving corneodesmosome components, or by activating other proteases.

Protease inhibitors might also play a documented role in protecting the body from infection. In addition to the liver-expressed antimicrobial peptide *LEAP-2*, we detected the expression of the *WAP1* and *WAP2* genes, which encode serine protease inhibitors with antimicrobial activity in mouse tongue and kidney [52]. *WAP1* appears to be expressed in all epidermal layers, whereas *WAP2* is overexpressed in GKs and could thus play an antimicrobial role in uppermost epidermal layers. Therefore, our study contributes to enlarge the panel of proteases and protease inhibitors potentially involved in

Table 8

Monogenic diseases due to mutations of genes involved in lipid metabolism and displaying an epidermal phenotype

Gene	Function	MIM/reference	Pathology
<i>ABCA12</i>	ABC lipid transporter	#242500	Harlequin ichthyosis
<i>STS</i>	Steroid sulfatase	+38100	X-linked ichthyosis
<i>GBA</i>	Glucocerebrosidase	#230800	Gaucher disease
<i>ALOXE3/ALOXB12</i>	Arachidonate lipoxygenases	#242100	Non-bullous congenital ichthyosiform erythroderma
<i>CGI58 (ABHD5)</i>	Putative triglyceride lipase	[62]	Chanarin-Dorman syndrome
<i>LIPH</i>	Phospholipase A I	[63]	Hair growth defect

barrier function, regulation of desquamation and defense against microorganisms.

Genes involved in lipid metabolism

Mutations of genes involved in various aspects of lipid metabolism are at the origin of several human genodermatoses (Table 8), underlying the key interest in the identification of new, lipid-processing genes expressed in the skin. We identified three new human genes, *LIPK*, *LIPM* and *LIPN*, which encode proteins containing two characteristic domains, the α/β hydrolase fold and the abhydrolase associated lipase region. They furthermore contain the consensus pattern of the active domain [53], suggesting that they are *bona fide* lipase genes (Figure 5b; Additional data file 4). The *LIPK*, *LIPM* and *LIPN* genes are strongly specific for the epidermis (Figure 4), and real-time RT-PCR experiments revealed a highly specific expression in GKs, with T4/T1 ratios >100.

Phylogenetic studies showed that the *LIPJ*, *LIPK*, *LIPM* and *LIPN* proteins are very close to well-characterized members of the family *LIPA* and *LIPF*, both also encoded in the 10q23.31 locus. *LIPA* is a ubiquitous lysosomal cholesterol ester hydrolase (EC 3.1.1.13) [54,55], while *LIPF* is a secreted triglyceride lipase (EC 3.1.1.3) [56]. Interestingly, both *LIPA* and *LIPF* have a low pH optimum, in agreement with the acidic pH of the extracellular space in the stratum corneum [57,58].

The extracellular hydrolysis of triglycerides in free fatty acids and glycerol is essential to stratum corneum hydration [59]. Furthermore, long chain free fatty acids represent 16% of extracellular lipids, cholesterol esters 15%, and free cholesterol 32% [60]. Triglyceride lipase enzymatic activity has been detected in the intercellular space of the human stratum corneum [61], and little is known regarding the metabolism and extracellular modifications of cholesterol esters. The *LIPK*, *LIPM* and *LIPN* proteins are most probably secreted and could thus participate in the establishment of the barrier function by catalyzing the maturation of extracellular lipids. The weakly expressed, non-secreted *LIPJ* protein might play a distinct role.

The *LIPK*, *LIPM* and *LIPN*, but not *LIPJ*, gene orthologues could be identified in several vertebrate species, including mouse and rat. Surprisingly, a cluster of tandem duplicated genes encoding new lipases resides in the mouse and rat genomes, which could eventually increase the lipase repertoire of these species.

The new *LIPK*, *LIPM* and *LIPN* lipase hydrolase genes may play an essential function in lipid metabolism of the most differentiated epidermal layers, and are thus interesting gene candidates for genodermatoses of unknown origin. Accordingly, mutations in the *CGI58/ABHD5* gene, which encodes a putative triacylglycerol lipase, are responsible for the Chanarin-Dorman syndrome, a neutral-lipid storage disease with ichthyosis [62]. Mutations in the *LIPH* gene are responsible for a hair growth defect [63]. Both *CGI58/ABHD5* and *LIPH* proteins resemble the lipases encoded by genes from the 10q23.31 locus as they include an α/β hydrolase fold, but they lack the abhydrolase associated lipase region.

In addition to the lipases from the 10q23.31 locus, nine additional genes, predicted to be involved in lipid metabolism, were overexpressed in GKs (Table 8). Their homology with known genes suggests that they might act on fatty acid (*PNPLA1*, *THEM5*, *ADHD9*, *FAM83F*) or ceramide (*LASS3*, *ASAH3*, *SMPD3*) metabolism. Compared to other tissues, lipid metabolism in the epidermis presents many distinct characteristics, as it mainly occurs in the extracellular space. Extracellular lipids play key roles in the barrier function, particularly in hydrophobicity of the skin surface. Our study thus unraveled new actors in this particularly important process, and might shed new light on the etiology of genodermatoses.

Genes of miscellaneous function

In addition to genes involved in protein degradation and lipid metabolism, we characterized new genes that might be of key importance in skin function. *CASZ1* is a transcription factor induced during embryogenesis in the course of mesenchyme differentiation [64]. As its gene targets are presently unknown, its relatively high expression level in the human skin (20 ORESTES) calls for detailed functional studies in this tissue. Finally, eight genes encoding hypothetical proteins without known domains, such as *C20orf95*, *CXorf33*, or

LOC440449, also displayed GK-specific expression. Their roles in epidermal differentiation remain completely elusive.

Conclusion

We have described an original and efficient method for purifying GKs from healthy human epidermis. It is now of key interest to adapt this technique to biopsy fragments from patients suffering from various genodermatoses. The GK is the last cell type in the skin to display transcription activity before cornification, and we describe the expression of 3,387 genes, a proportion of which are expressed in GKs in a highly specific manner. We presume that many of them are important for the establishment of the barrier function, and as such deserve detailed functional studies. Moreover, we provide the scientific community with a list of gene candidates for genodermatoses of unknown origin. In particular, the understanding of complex diseases associated with defects in barrier function, such as psoriasis or atopic dermatitis, might benefit from identification of new epidermis-expressed genes located in associated loci. Among the genes described, some fit with, and even improve, our present knowledge of the barrier function, in particular concerning the fine-tuning of protein degradation and lipid metabolism. The precise function of the corresponding proteins will be assessed using mouse models and immunochemistry methods starting from healthy and pathological human skin. However, many of the genes described herein, often specifically expressed in GKs at high levels, encode putative proteins whose functions are totally obscure but that might well participate in the establishment of the skin barrier. Incidentally, we characterized a new gene, *C1orf81*, which is specifically inactivated or truncated in humans. Whether this gene loss participated in the establishment of the human species, and thus fits the 'less-is-more' hypothesis [65] remains a fascinating question. Our study of the human GK transcriptome thus opens new avenues for future research in many fields, including the normal functioning of the epidermis, the origin of genodermatoses, and even the emergence of the human species.

Materials and methods

Skin samples and RNA extraction

Normal human skin was obtained from patients undergoing abdominal plastic surgery (kindly provided by Professor JP Chavoïn, "Service de Chirurgie Plastique et des Brûlés", Centre Hospitalier Universitaire Rangueil, Toulouse, France) after informed consent and in accordance with Helsinki principles. Subcutaneous fat was promptly removed and strips of skin were incubated, epidermis side up, for 1 h at 4°C in phosphate-buffered saline (PBS) containing 0.5 mg/ml thermolysin (T7902, Sigma, St Louis, MO, USA). The epidermis was dissected free of dermis tissue with forceps and rinsed in cold PBS. Epidermal fragments were either immediately frozen for total RNA extraction or incubated in 1× trypsin-EDTA solution (25300-054, Invitrogen, Carlsbad, CA, USA) at 4°C

under gentle agitation for 15 minutes. The remaining epidermal fragments were rinsed in cold PBS and incubated in another trypsin-EDTA solution, while fetal calf serum (10270098, Invitrogen) was added to the suspended cells (10% final concentration). After centrifugation, the cells were frozen as dry pellets. The procedure was repeated twice, leading to three successive fractions of dissociated cells named T1, T2 and T3. The residual fragments (T4 fraction) were drained on a gauze compress, frozen and ground to a powder under liquid nitrogen. Total RNA was extracted from the various cell batches using the RNeasy extraction kit (Qiagen, Hilden, Germany). Purification of poly(A)+ RNA was always performed from the T4 fraction of individual patients, using oligo(dT)25-tagged magnetic beads according to the manufacturer's instructions (Dynal, Oslo, Norway). Two rounds of hybridization to the beads were performed. The mRNAs were treated with DNase I (Invitrogen) and the absence of genomic DNA was confirmed by PCR using primers for the corneodesmosin gene (GenBank: [AF03130](#)).

Morphological analysis of epidermis samples

After each trypsin incubation, an aliquot of epidermal fragments was fixed in Bouin's solution, embedded in paraffin, and sections (10 µm) were stained with hematoxylin-eosin.

Production and analysis of ORESTES

ORESTES production was essentially performed as described [66]. Purified mRNA (20 ng) was heated 10 minutes at 65°C, reverse transcribed at 37°C for 1 h with 200 U of Moloney murine leukemia virus reverse transcriptase (Promega, Madison, WI, USA) and 10 pmol of an arbitrary selected primer (18-25 nt) in a final volume of 20 µl. The reaction products (1 µl) were amplified by PCR using either the primer used for the reverse transcription, or a single, alternative arbitrary chosen primer. The hybridization step of the first PCR cycle was set at 37°C, while the 35 remaining cycles were performed in standard conditions with a hybridization temperature complying with the length of the primer (typically 55°C). After gel electrophoresis, products with predominant bands reflecting the amplification of highly abundant sequences were not further processed. Smear-like reaction products were gel-purified with a 500 bp cut-off. Mini-libraries were then produced by T/A cloning of the purified PCR fragments (TOPO-TA cloning kit, Invitrogen).

Plasmid purification and sequencing

Sequencing was performed by standard procedures (ABI Prism Dye Terminator Cycle Sequencing kit, Applied Biosystems, Foster City, CA, USA) after either plasmid purification (Wizard miniprep kit, Promega) or rolling circle amplification of the plasmids.

Sequence analysis

An automated protocol for the sequence analysis was used to verify sequence quality. Sequences were then analyzed using a stepwise approach. Starting with RepeatMasker [67],

ORESTES were clustered with the PHRAP algorithm [68]. Consensus sequences or singletons were annotated using BLAST searches against human databases (best-hit among successively RefSeq, Uniprot and EST_Human databases [69]). Sequences were also aligned on the human genome (May 2004 assembly) using BLAT [18] and inserted as a custom track into the UCSC Genome Browser.

Analysis of gene expression

For quantitative real-time RT-PCR experiments, all primer pairs (available upon request) were chosen to generate amplicons of 100-250 bp encompassing different exons, thus avoiding the amplification of potential contaminating genomic DNA. The primer sequences were designed using Primer3 software [70] and BLAST analysis [69] ensured the absence of similarity to any other human sequence. Reverse transcription was performed by standard procedures, starting from 100 ng of total RNA of each cell batch and using a mixture of oligo(dT) and random hexamers. Amplification assays were performed with the ABI prism 7000 Sequence Detection System and analyzed with the corresponding software (Applied Biosystems, Foster City, CA, USA) using the qPCR ROX- β GO Green mix (MP Biomedicals, Irvine, CA, USA). Fluorescence was quantified as Ct (threshold cycle) values. Samples were analyzed in triplicate, with differences between the three Ct values lower than 0.3. Expression levels were calibrated using *galectin 7* (*LGALS7*), or *beta-2-microglobulin* (*B2M*) mRNA as internal controls. The differences between the mean Ct values of the various amplicons and the reference genes are denoted (Δ Ct). The difference between Δ Ct obtained with the indicated cell samples are labeled $\Delta\Delta$ Ct; $2^{\Delta\Delta$ Ct gave the relative level of gene expression between the T1 and T4 fractions. Control wells containing the SYBR Green PCR master mix and primers without template cDNA emitted no significant fluorescence after 40 cycles.

Human Multiple Tissue cDNA panels I and II obtained from Clontech (Palo Alto, CA, USA) were used as templates for PCR analysis. A control reaction with T4 cDNAs was carried out in parallel. The reactions were conducted for 35 cycles in standard conditions. The PCR products were separated on 1.5% agarose-TAE gels.

RACE-PCR experiments

We performed 5' RNA ligase-mediated (RLM)-RACE using the FirstChoice RLM-RACE kit (Ambion, Austin, TX, USA). Briefly, total RNA was dephosphorylated with calf intestine phosphatase then decapped using tobacco acid pyrophosphatase to target full-length mRNA. An adapter was then ligated to mRNA and reverse transcription was performed using random decamers. PCR was performed to amplify the resulting cDNA using the Outer 5' RLM-RACE primer and a gene specific lower primer. Nested PCR was then performed with the Inner 5' RLM-RACE primer. The RACE nested PCR products were cloned into the pCRII-TOPO vector using a TOPO T/A cloning kit (Invitrogen) and sequenced.

NCBI gene ID references

GAPDH, 2597; *SOD1*, 6647; *ACTB*, 60; *B2M*, 567; *HPRT1*, 3251; *HMBS*, 3145; *TBP*, 6908; *UBC*, 7316; *PNAS-123*, 85028; *PPP2R5A*, 5525; *RPS11*, 6205; *RPS12*, 6206; *RPL10*, 6134; *EIF4A1*, 1973; *KRTHAP1*, 8686.

Additional data files

The following additional data are available with the online version of this paper. Additional data file 1 is a screen copy of the ORESTES custom track of the human UCSC Genome Browser that shows the genomic localization of the newly described *C1orf81* gene with the corresponding ORESTES (chr1:199,164,695-199,202,419; hg18, March 2006). Additional data file 2 is a table of genes (with 100 or less UniGene ESTs) expressed in various tissues but for which epidermal expression had never been assessed. Additional data file 3 is a colored text file containing the alignment of human and rhesus genomic sequences showing the location of the insertion point and tandem site duplication for the hominoid-specific CNOT6L processed retrogene. Additional data file 4 is the alignment of protein sequences of the ten mouse lipases, including Lipdc1-5. The alignment was generated with Multalin software [71]; the amino acids of the catalytic triad are boxed. Additional data file 5 is a table providing gene nomenclature and IDs for five new mouse lipase genes.

Acknowledgements

Our special thanks go to M-T Ribouchon for excellent technical assistance. We are grateful to H Brun and C Offer from the "Service Commun de Séquençage - IFR30" for sequencing the clones EL593304 to EL595248. We thank D Bobermian for her contribution to the production of minilibraries and M Fraisse for histochemistry. This work was supported in part by grants from the "Centre National de la Recherche Scientifique", the Toulouse III Paul Sabatier University, the "Société Française de Dermatologie" and from the "Société de Recherche Dermatologique". Sequencing was funded by the "Consortium National de Recherche en Génomique".

References

1. Watt FM, Celso CL, Silva-Vargas V: **Epidermal stem cells: an update.** *Curr Opin Genet Dev* 2006, **16**:518-524.
2. Park GT, Lim SE, Jang SI, Morasso MI: **Suprabasin, a novel epidermal differentiation marker and potential cornified envelope precursor.** *J Biol Chem* 2002, **277**:45195-45202.
3. Oomizu S, Sahuc F, Asahina K, Inamatsu M, Matsuzaki T, Sasaki M, Obara M, Yoshizato K: **Kdap, a novel gene associated with the stratification of the epithelium.** *Gene* 2000, **256**:19-27.
4. Makino T, Takaishi M, Morohashi M, Huh NH: **Hornerin, a novel profilaggrin-like protein and differentiation-specific marker isolated from mouse skin.** *J Biol Chem* 2001, **276**:47445-47452.
5. Lee WH, Jang S, Lee JS, Lee Y, Seo EY, You KH, Lee SC, Nam KI, Kim JM, Kee SH, et al.: **Molecular cloning and expression of human keratinocyte proline-rich protein (hKPRP), an epidermal marker isolated from calcium-induced differentiating keratinocytes.** *J Invest Dermatol* 2005, **125**:995-1000.
6. Kalinin AE, Kajava AV, Steinert PM: **Epithelial barrier function: assembly and structural features of the cornified cell envelope.** *Bioessays* 2002, **24**:789-800.
7. Elias PM, Cullander C, Mauro T, Rassner U, Komuves L, Brown BE, Menon GK: **The secretory granular cell: the outermost granular cell as a specialized secretory cell.** *J Invest Dermatol Symp Proc* 1998, **3**:87-100.
8. Guerrin M, Simon M, Montezin M, Haftek M, Vincent C, Serre G:

- Expression cloning of human corneodesmosin proves its identity with the product of the **S** gene and allows improved characterization of its processing during keratinocyte differentiation. *J Biol Chem* 1998, **273**:22640-22647.
9. Dias Neto E, Correa RG, Verjovski-Almeida S, Briones MR, Nagai MA, da Silva W Jr, Zago MA, Bordin S, Costa FF, Goldman GH, et al.: **Shotgun sequencing of the human transcriptome with ORF expressed sequence tags.** *Proc Natl Acad Sci USA* 2000, **97**:3491-3496.
 10. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.** *Genome Biol* 2002, **3**:RESEARCH0034.
 11. Bonnet-Duquennoy M, Abaibou H, Tailhardat M, Lazou K, Bosset S, Le Varlet B, Cleuziat P, Kurfurst R: **Study of housekeeping gene expression in human keratinocytes using OLISA trade mark, a long-oligonucleotide microarray and qRT-PCR.** *Eur J Dermatol* 2006, **16**:136-140.
 12. Magnaldo T, Fowles D, Darmon M: **Galectin-7, a marker of all types of stratified epithelia.** *Differentiation* 1998, **63**:159-168.
 13. Fisher C, Angus B, Rees J: **In-situ hybridization using digoxigenin-labelled probes in human skin.** *Br J Dermatol* 1991, **125**:516-520.
 14. Sondell B, Jonsson M, Dyberg P, Egelrud T: **In situ evidence that the population of Langerhans cells in normal human epidermis may be heterogeneous.** *Br J Dermatol* 1997, **136**:687-693.
 15. Matsui T, Hayashi-Kisumi F, Kinoshita Y, Katahira S, Morita K, Miyachi Y, Ono Y, Imai T, Tanigawa Y, Komiya T, et al.: **Identification of novel keratinocyte-secreted peptides dermokine-alpha/beta and a new stratified epithelium-secreted protein gene complex on human chromosome 19q13.1.** *Genomics* 2004, **84**:384-397.
 16. Moffatt P, Salois P, St-Amant N, Gaumont MH, Lanctot C: **Identification of a conserved cluster of skin-specific genes encoding secreted proteins.** *Gene* 2004, **334**:123-131.
 17. Toulza E, Galliano MF, Jonca N, Galliano H, Mechin MC, Ishida-Yamamoto A, Serre G, Guerrin M: **The human dermokine gene: description of novel isoforms with different tissue-specific expression and subcellular location.** *J Invest Dermatol* 2006, **126**:503-506.
 18. Kent WJ: **BLAT - the BLAST-like alignment tool.** *Genome Res* 2002, **12**:656-664.
 19. **UCSC Custom Track: ORESTES from Granular Keratinocytes of Human Epidermis** [<http://genome.ucsc.edu/cgi-bin/hgTracks?org=human&hgt.customText=http://udear.cnrs.free.fr/orestes.txt>]
 20. Brandner JM, Kief S, Wladykowski E, Houdek P, Moll I: **Tight junction proteins in the skin.** *Skin Pharmacol Physiol* 2006, **19**:71-77.
 21. Furuse M, Hata M, Furuse K, Yoshida Y, Haratake A, Sugitani Y, Noda T, Kubo A, Tsukita S: **Claudin-based tight junctions are crucial for the mammalian epidermal barrier: a lesson from claudin-1-deficient mice.** *J Cell Biol* 2002, **156**:1099-1111.
 22. Vinckenbosch N, Dupanloup I, Kaessmann H: **Evolutionary fate of retroposed gene copies in the human genome.** *Proc Natl Acad Sci USA* 2006, **103**:3220-3225.
 23. Rogers MS, Kobayashi T, Pittelkow MR, Strehler EE: **Human calmodulin-like protein is an epithelial-specific protein regulated during keratinocyte differentiation.** *Exp Cell Res* 2001, **267**:216-224.
 24. **The snoRNABase** [<http://www-snoRNA.biotoul.fr/>]
 25. Weber Mj: **Mammalian small nucleolar RNAs are mobile genetic elements.** *PLoS Genet* 2006, **2**:e205.
 26. Smith CM, Steitz JA: **Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes.** *Mol Cell Biol* 1998, **18**:6897-6909.
 27. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, et al.: **MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer.** *Oncogene* 2003, **22**:8031-8041.
 28. Geirsson A, Paliwal I, Lynch RJ, Bothwell AL, Hammond GL: **Class II transactivator promoter activity is suppressed through regulation by a trophoblast noncoding RNA.** *Transplantation* 2003, **76**:387-394.
 29. Zhang Y, Tycko B: **Monoallelic expression of the human H19 gene.** *Nat Genet* 1992, **1**:40-44.
 30. Thorvaldsen JL, Duran KL, Bartolomei MS: **Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2.** *Genes Dev* 1998, **12**:3693-3702.
 31. Blakytyn R, Jude EB, Martin Gibson J, Boulton AJ, Ferguson MW: **Lack of insulin-like growth factor I (IGF1) in the basal keratinocyte layer of diabetic skin and diabetic foot ulcers.** *J Pathol* 2000, **190**:589-594.
 32. Bennett WR, Crew TE, Slack JM, Ward A: **Structural-proliferative units and organ growth: effects of insulin-like growth factor 2 on the growth of colon and skin.** *Development* 2003, **130**:1079-1088.
 33. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J, et al.: **Antisense transcription in the mammalian transcriptome.** *Science* 2005, **309**:1564-1566.
 34. Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, et al.: **Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.** *Proc Natl Acad Sci USA* 2002, **99**:16899-16903.
 35. Kraut J: **Serine proteases: structure and mechanism of catalysis.** *Annu Rev Biochem* 1977, **46**:331-358.
 36. Brzozowski AM, Derewenda U, Derewenda ZS, Dodson GG, Lawson DM, Turkenburg JP, Bjorkling F, Huges-Jensen B, Patkar SA, Thim L: **A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex.** *Nature* 1991, **351**:491-494.
 37. Smith TP, Grosse WM, Freking BA, Roberts AJ, Stone RT, Casas E, Wray JE, White J, Cho J, Fahrenkrug SC, et al.: **Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle.** *Genome Res* 2001, **11**:626-630.
 38. Hida K, Wada J, Eguchi J, Zhang H, Baba M, Seida A, Hashimoto I, Okada T, Yasuhara A, Nakatsuka A, et al.: **Visceral adipose tissue-derived serine protease inhibitor: a unique insulin-sensitizing adipocytokine in obesity.** *Proc Natl Acad Sci USA* 2005, **102**:10610-10615.
 39. Singh G, Lykke-Andersen J: **New insights into the formation of active nonsense-mediated decay complexes.** *Trends Biochem Sci* 2003, **28**:464-466.
 40. Matsufoji S, Matsufoji T, Miyazaki Y, Murakami Y, Atkins JF, Gesteland RF, Hayashi S: **Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme.** *Cell* 1995, **80**:51-60.
 41. Hahn Y, Lee B: **Identification of nine human-specific frameshift mutations by comparative analysis of the human and the chimpanzee genome sequences.** *Bioinformatics* 2005, **21**(Suppl 1):186-194.
 42. Wang X, Grus WE, Zhang J: **Gene losses during human origins.** *PLoS Biol* 2006, **4**:e52.
 43. Winter H, Langbein L, Krawczak M, Cooper DN, Jave-Suarez LF, Rogers MA, Praetzel S, Heidt PJ, Schweizer J: **Human type I hair keratin pseudogene pihH1A has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the Pan-Homo divergence.** *Hum Genet* 2001, **108**:37-42.
 44. Hahn Y, Lee B: **Human-specific nonsense mutations identified by genome sequence comparisons.** *Hum Genet* 2006, **119**:169-178.
 45. Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H: **Emergence of young human genes after a burst of retroposition in primates.** *PLoS Biol* 2005, **3**:e357.
 46. Zeeuwen PL: **Epidermal differentiation: the role of proteases and their inhibitors.** *Eur J Cell Biol* 2004, **83**:761-773.
 47. Zeeuwen PL, van Vlijmen-Willems IM, Hendriks W, Merlx GF, Schalkwijk J: **A null mutation in the cystatin M/E gene of ichg mice causes juvenile lethality and defects in epidermal cornification.** *Hum Mol Genet* 2002, **11**:2867-2875.
 48. Miyata T, Inagi R, Nangaku M, Imasawa T, Sato M, Izuhara Y, Suzuki D, Yoshino A, Onogi H, Kimura M, et al.: **Overexpression of the serpin meginin induces progressive mesangial cell proliferation and expansion.** *J Clin Invest* 2002, **109**:585-593.
 49. Askew YS, Pak SC, Luke CJ, Askew DJ, Cataltepe S, Mills DR, Kato H, Lehoczyk J, Dewar K, Birren B, et al.: **SERPINB12 is a novel member of the human ov-serpin family that is widely expressed and inhibits trypsin-like serine proteinases.** *J Biol Chem* 2001, **276**:49320-49330.
 50. Caubet C, Jonca N, Lopez F, Esteve JP, Simon M, Serre G: **Homo-oligomerization of human corneodesmosin is mediated by its N-terminal glycine loop domain.** *J Invest Dermatol* 2004,

- 122:747-754.
51. Grundmann U, Romisch J, Siebold B, Bohn H, Amann E: **Cloning and expression of a cDNA encoding human placental protein 11, a putative serine protease with diagnostic significance as a tumor marker.** *DNA Cell Biol* 1990, **9**:243-250.
 52. Hagiwara K, Kikuchi T, Endo Y, Huqun , Usui K, Takahashi M, Shibata N, Kusakabe T, Xin H, Hoshi S, et al.: **Mouse SWAMI and SWAM2 are antibacterial proteins composed of a single whey acidic protein motif.** *J Immunol* 2003, **170**:1973-1979.
 53. Cousin X, Hotelier T, Giles K, Lievin P, Toutant JP, Chatonnet A: **The alpha/beta fold family of proteins database and the cholinesterase gene server ESTHER.** *Nucleic Acids Res* 1997, **25**:143-146.
 54. Du H, Witte DP, Grabowski GA: **Tissue and cellular specific expression of murine lysosomal acid lipase mRNA and protein.** *J Lipid Res* 1996, **37**:937-949.
 55. Anderson RA, Sando GN: **Cloning and expression of cDNA encoding human lysosomal acid lipase/cholesterol ester hydrolase. Similarities to gastric and lingual lipases.** *J Biol Chem* 1991, **266**:22479-22484.
 56. Bodmer MW, Angal S, Yarranton GT, Harris TJ, Lyons A, King DJ, Pieroni G, Riviere C, Verger R, Lowe PA: **Molecular cloning of a human gastric lipase and expression of the enzyme in yeast.** *Biochim Biophys Acta* 1987, **909**:237-244.
 57. Ohman H, Vahlquist A: **In vivo studies concerning a pH gradient in human stratum corneum and upper epidermis.** *Acta Derm Venereol* 1994, **74**:375-379.
 58. Hanson KM, Behne MJ, Barry NP, Mauro TM, Gratton E, Clegg RM: **Two-photon fluorescence lifetime imaging of the skin stratum corneum pH gradient.** *Biophys J* 2002, **83**:1682-1690.
 59. Fluhr JW, Mao-Qiang M, Brown BE, Wertz PW, Crumrine D, Sundberg JP, Feingold KR, Elias PM: **Glycerol regulates stratum corneum hydration in sebaceous gland deficient (asebia) mice.** *J Invest Dermatol* 2003, **120**:728-737.
 60. Norlen L, Nicander I, Lundh Rozell B, Ollmar S, Forslind B: **Inter- and intra-individual differences in human stratum corneum lipid content related to physical parameters of skin barrier function in vivo.** *J Invest Dermatol* 1999, **112**:72-77.
 61. Menon GK, Grayson S, Elias PM: **Cytochemical and biochemical localization of lipase and sphingomyelinase activity in mammalian epidermis.** *J Invest Dermatol* 1986, **86**:591-597.
 62. Lefevre C, Jobard F, Caux F, Bouadjar B, Karaduman A, Heilig R, Lakhdar H, Wollenberg A, Verret JL, Weissenbach J, et al.: **Mutations in CGI-58, the gene encoding a new protein of the esterase/lipase/thioesterase subfamily, in Chanarin-Dorfman syndrome.** *Am J Hum Genet* 2001, **69**:1002-1012.
 63. Kazantseva A, Goltsov A, Zinchenko R, Grigorenko AP, Abrukova AV, Moliaka YK, Kirillov AG, Guo Z, Lyle S, Ginter EK, et al.: **Human hair growth deficiency is linked to a genetic defect in the phospholipase gene LIPH.** *Science* 2006, **314**:982-985.
 64. Liu Z, Yang X, Tan F, Cullion K, Thiele CJ: **Molecular cloning and characterization of human Castor, a novel human gene upregulated during cell differentiation.** *Biochem Biophys Res Commun* 2006, **344**:834-844.
 65. Olson MV: **When less is more: gene loss as an engine of evolutionary change.** *Am J Hum Genet* 1999, **64**:18-23.
 66. Leerkes MR, Caballero OL, Mackay A, Torloni H, O'Hare MJ, Simpson AJ, de Souza SJ: **In silico comparison of the transcriptome-derived from purified normal breast cells and breast tumor cell lines reveals candidate upregulated genes in breast tumor cells.** *Genomics* 2002, **79**:257-265.
 67. **RepeatMasker Open-3.0** [<http://www.repeatmasker.org>]
 68. **PHRAP Software** [<http://www.phrap.org/>]
 69. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
 70. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132**:365-386.
 71. Corpet F: **Multiple sequence alignment with hierarchical clustering.** *Nucleic Acids Res* 1988, **16**:10881-10890.
 72. **TreeTop: Phylogenetic Tree Prediction** [http://www.genebee.msu.su/services/phtree_reduced.html]
 73. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al.: **Aligning multiple genomic sequences with the threaded blockset aligner.** *Genome Res* 2004, **14**:708-715.

**A large-scale, multi-technique approach
identifies 49 new players of human
epidermal differentiation**

(2009, soumis)

Nicolas R. Mattiuzzo, Eve Toulza, Nathalie Jonca, Guy Serre and Marina Guerrin.

Résumé

La formation de la couche cornée, élément indispensable à la barrière épidermique, requiert la mise en oeuvre d'un programme de différenciation extrêmement complexe et finement régulé, culminant au niveau de la couche granuleuse, dernière assise de cellules vivantes. L'étude du transcriptome du kératinocyte granuleux constitue donc une cible prioritaire pour une meilleure connaissance de ce tissu. Pourtant, la plupart des études publiées n'ont pas été menées *in vivo* ; de plus, seule une poignée de facteurs de transcription requis pour la fonction barrière a été identifiée.

Au moyen d'une méthode de dissociation des cellules à partir d'épiderme humain, nous avons étudié la régulation de l'expression génique au cours de la différenciation *in vivo*. Plusieurs approches à grande échelle ont été combinées pour identifier des candidats (puces à ADN pangénomiques, banque d'EST de kératinocytes granuleux, analyses bioinformatiques des promoteurs de 52 gènes exprimés au cours de la différenciation) et leur niveau d'expression a été quantifié dans les couches basales et granuleuses par PCR quantitative.

Parmi 3576 candidats potentiels exprimés par les kératinocytes granuleux, 298 ont été sélectionnés et testés, en mettant particulièrement l'accent sur l'identification de nouveaux facteurs de transcription, cruciaux pour la compréhension du programme de différenciation. L'expression de 157 gènes, dont 94 codant pour des facteurs de transcription, a pu être mise en évidence pour la première fois dans l'épiderme. Au total, 49 gènes sont surexprimés au cours de la différenciation et 20 sous-exprimés, avec notamment plus d'un tiers des facteurs de transcription régulé (20 positivement et 17 négativement). Parmi les gènes les plus notables figurent *FOXL1*, susceptible de réguler négativement l'expression des héparan sulfate protéoglycans, *GATA4*, dont le paralogue *GATA3* joue un rôle majeur dans la fonction barrière, et *HOPX*, exprimé au cours de la différenciation terminale des cellules du cristallin. L'étude a également permis de mettre en évidence l'expression d'autres gènes codant pour des protéines structurales, telle la nouvelle kératine acide *K23*, ou encore des protéines potentiellement impliqués dans diverses fonctions telles que la signalisation, le métabolisme lipidique, la polarité cellulaire et le trafic vésiculaire, *etc.*

L'identification d'un grand nombre de facteurs de transcription, régulés ou non, ouvre de nouvelles voies à explorer pour décrypter la régulation génique complexe mise en jeu dans la différenciation terminale de l'épiderme. Elle permettra également de mieux comprendre certaines pathologies complexes caractérisées par des anomalies du programme de différenciation, telles que le psoriasis, la dermatite atopique ou les ichtyoses.

A large-scale, multi-technique approach identifies 49 new players of human epidermal differentiation

Running title: human epidermal differentiation program

Keywords: cell differentiation, transcription factors, gene regulatory networks

Manuscript: 3069 words, 4 Tables, 2 Figures, Supporting information.

Authors: N. R. Mattiuzzo, E. Toulza, N. Jonca, G. Serre and M. Guerrin*

Institution: UMR 5165 "Epidermis Differentiation and Rheumatoid Autoimmunity", CNRS – University of Toulouse, IFR 30 (INSERM - CNRS - Toulouse III University - CHU), CHU Purpan, 31059 Toulouse, France

* Corresponding author. Tel: +33 561 158 433. Fax: +33 561 499 036. E-mail: mweber@udear.cnrs.fr

Funding sources: This work was supported in part by grants from the Société Française de Dermatologie, the Centre National de la Recherche Scientifique and the University of Toulouse (Toulouse III). N.R.M. was supported by a fellowship of the Ministère de l'Enseignement Supérieur et de la Recherche.

Conflict of interest: none declared

Abbreviations used: GK, granular keratinocyte; BK, basal keratinocyte; EST, expressed sequence tag; ORESTE, open reading frame EST; qRT-PCR, quantitative real-time PCR; IHC, immunohistochemistry; TF, transcription factor.

Summary

Background At the latest stage of terminal differentiation, granular keratinocytes (GKs) undergo cornification, a programmed cell death required for a functional skin barrier. A complex gene regulatory network orchestrates the underlying biochemical modifications, but very few transcription factors specific for this program have been identified.

Objectives We describe a large-scale, multi-technique approach performed on cells purified from normal human epidermis, and focusing primarily on identifying key regulators.

Methods We combined data from microarrays analysis on cell fractions enriched in GKs or basal keratinocytes, from an EST library from GKs, and from in silico promoter analysis for 52 differentiation-associated genes.

Results Among 3576 potential GK-expressed genes, 298 candidates were selected and half could be profiled for the first time in the epidermis by real-time PCR. Results highlight various GK functions including lipid metabolism and secretion, cell signaling and transcriptional regulation. In total, 49 new genes upregulated with differentiation were identified, and 37 transcription factors (out of 94 detected) were found to be positively or negatively regulated, suggesting a function during the final steps of GK differentiation.

Conclusions These new leads contribute to a better understanding of gene networks involved in barrier function, and may help decipher some differentiation disorders observed in many skin diseases, including ichthyosis.

Introduction

The epidermis is a highly specialized, pluristratified epithelium composed primarily of keratinocytes widely dedicated to the barrier function. Genes known to be involved in barrier formation are mainly expressed in the last living cells: granular keratinocytes (GKs) produce more than 30 different epidermal-specific proteins including well-known structural components that aggregate into characteristic cytosolic granules. GKs give rise to corneocytes that accumulate in the cornified layer, ensuring the barrier function. After cornification, transcription and translation are no longer detectable and organelles disappear.

Thus, components of the stratum corneum responsible for most of the protective functions of the epidermis are produced by GKs.

Transcription factors (TFs) regulating cornification remain largely unknown. Biochemical techniques have thus far made possible to identify only four TFs involved in late differentiation and barrier formation: *KLF4*, *GATA3*, *GRHL3* and *PRDM1* (Blimp-1).¹⁻⁴ Few large-scale studies were previously performed using diverse models: pioneering work on skin differentiation initiated transcriptome studies of in vitro differentiated keratinocytes and total human epidermis.⁵⁻⁷ Other studies were carried out using macroarrays or microarrays,⁸⁻¹¹ with different in vitro models (proliferating or differentiating keratinocytes and reconstructed epidermis) and total skin. Interpretation of these results is often made difficult by the presence of diverse cell types or differentiation states in the sample, or by the use of in vitro models which do not fully recapitulate the complexity of the cellular environment. Analyses of purified samples from ex vivo material are thus expected to provide more relevant and functional results. Radoja and colleagues used microarrays that targeted ~10,000 genes to compare the transcriptomes of total epidermis and of two cell fractions purified by integrin beta-4 expression levels (i.e., basal vs. suprabasal cells).¹² These authors reported a large number of regulated genes, but GKs were largely absent from the suprabasal fraction due to the purification process.

By cDNA sequencing our laboratory has previously identified 3,387 genes expressed in GKs purified from normal human skin by a trypsin-based approach.¹³ However, the EST database alone gave no information about gene expression levels, nor allowed a comparison with basal keratinocytes to identify genes potentially required for barrier function. For this reason, we combined EST database and results from microarrays experiments and in silico promoter analyses, to select a large number of candidate genes in order to study their regulation during in vivo GK differentiation by quantitative real-time PCR (qRT-PCR).

Materials and Methods

Ethics statement

Human skin samples were obtained from healthy donors undergoing previously scheduled abdominal plastic surgery (Professor J.P. Chavoin, "Service de Chirurgie Plastique, Reconstructrice et Esthétique", Centre Hospitalier Universitaire Rangueil, Toulouse). Before anonymous skin samples were provided to our lab, informed written consent was obtained from donors, as recommended by local ethics committee (CHU Toulouse, France) and in accordance with Helsinki principles.

Skin samples and RNA extraction

Skin samples obtained from six healthy subjects were processed as previously described.¹³ Briefly, after dermis separation, epidermal fragments were incubated in trypsin-EDTA three times at 4°C for 15 min. Successively dissociated cells were named T1 (mainly basal keratinocytes or BKs), T2 and T3, while residual fragments constituted the T4 fraction (predominantly composed of GKs and cornified layer). Total RNA from BKs (T1) and GKs (T4) were isolated and concentrated using RNeasy columns (Qiagen). RNA quality, purity and concentration were checked on agarose gels and on the RNA 6000 Nano LabChip with 2100 Bioanalyzer (Agilent). Samples used for analyses showed $OD_{260/280} > 1.65$, a 28S/18S ratio > 1.6 , and no visible degradation of rRNA or genomic DNA contamination.

Microarrays experiments

Total RNA isolated from BKs and GKs were reverse-transcribed, differentially labelled with Cy3 or Cy5 and cohybridized on pangenomic microarrays consisting of 25,248 different 50mers targeting 22,458 genes (platform GPL1946 in Gene Expression Omnibus (GEO) database).¹⁴ Experiments were performed at the Biochip platform of the Génopole (Génotoul) in Toulouse, France (<http://biopuce.insa-toulouse.fr/sommaire.php>) and arrays were used according to the manufacturer's instructions for non-amplified RNA (see Supporting information). Microarrays data are MIAME compliant and raw data are available from the GEO database under accession number GSE13263. Data were normalized using the LOWESS method with Biplot software (<http://biopuce.insa-toulouse.fr/ExperimentExplorer/doc/>). SAM software was used for false discovery rate (FDR) analysis.¹⁵

Immunohistochemistry

Normal human skin samples were fixed 24h in formalin, dehydrated 24h in 70% ethanol and embedded in paraffin. Sections (4 μ m) were processed for immunohistochemical analysis. ImmPRESS anti-rabbit Ig (peroxidase) kit (Vector Laboratories) was used with the polyclonal rabbit antibodies against MAPT (Abcam) or SH3KBP1 (Sigma-Aldrich) prepared at the manufacturers' recommended dilutions. Images were taken using a Nikon eclipse 22 80i microscope equipped with a Nikon DXM 1200C digital camera and NIS image analysis software.

Quantitative real-time PCR

All primer pairs (Additional Table 2) generated 100-250 bp amplicons spanning across exons, except for 8 intronless genes. Primer sequences were designed using the Primer3 software,¹⁶ and BLAST analysis ensured the lack of any similarity to other human sequences.¹⁷ Reverse transcription was performed by standard procedures using 1 μ g of total RNA and a mixture of oligo(dT) and random hexamers. Primer pairs were systematically tested by classical PCR ($T_m=60^\circ\text{C}$, 38 cycles) on total epidermis cDNA to verify amplification of a single product of expected size. Quantification assays were performed using the Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen) and the 7300 Real-Time PCR System with corresponding software (Applied Biosystems). Expression levels were calibrated using transcript levels of the beta-2-microglobulin (*B2M*) housekeeping gene. The $2^{-\Delta\Delta\text{Ct}}$ method calculated the relative level of gene expression between BK and GK fractions.¹⁸ Dissociation curves were verified, and amplicons were systematically sequenced to attest the specificity of quantifications.

Promoter analyses

Human promoter sequences (and vertebrate orthologues) were retrieved for 52 previously described differentiation-associated genes selected from the literature, and analyzed with Genomatix Suite tools (Genomatix Software GmbH) (see Supporting

information). When a given transcription factor binding site was found in a minimum of three different promoters, corresponding TF or family was selected for qRT-PCR analyses.

Results

Large-scale identification of differentiation-specific genes

To optimize the discovery of new proteins involved in the differentiation, data compiled from several sources and techniques were used to select a large number of gene candidates for subsequent qRT-PCR analyses (Fig. 1). First, pangenomic oligonucleotide microarrays were used for global transcriptional profiling. Complementary DNAs prepared from GKs and BKs isolated from six individuals were cohybridized onto microarrays (see Material and Methods and Supporting information). Statistical analysis, using a local False Discovery Rate (FDR) threshold of 0.05 and a two-fold ratio threshold, revealed 203 upregulated genes and 3 downregulated (Fig. S1 & Table S1). Among the upregulated genes, 23 were known players of epidermal differentiation, while 180 were not previously known to be expressed in the human epidermis. Therefore, the microarray analysis provided us with the first series of gene candidates.

The second source for potential differentiation-associated genes was our EST dataset obtained from the GK fraction, which matched 3,387 different genes.¹³

A third strategy to select additional candidates was to identify transcription factor binding sites within promoter sequences of 52 genes expressed in differentiated keratinocytes, focusing on conserved non-coding sequences within the 2.8 kb proximal region. Recurrent binding sites were identified, and several modules (i.e., evolutionarily conserved combinations of binding sites) could be found within the majority of promoters examined (see Fig. S2 for an example). Thus, 88 TFs belonging to gene families underscored by the comparative promoter analysis were selected for qRT-PCR experiments, together with potential cofactors identified from the literature.

Data filtering of all three sources (3576 genes) was performed using the UniGene database, so that all genes represented by less than 200 ESTs were selected for subsequent

qRT-PCR analyses. Indeed, the number of ESTs for a given gene represents how many times it was sequenced from all human libraries. Thus, a small number suggests that the gene has a tissue-specific expression pattern, indicative of a differentiation program. Among the remaining genes our filtering focused on the regulation of the terminal differentiation program (potential transcription factors, cofactors, proteins linked to signal transduction). We also selected a few genes involved in lipid metabolism and programmed cell death, potentially important for the permeability barrier and cornification. Regardless of their function, all genes strongly represented in UniGene (>1500 ESTs) were discarded.

Gene expression profiling by qRT-PCR

In total, 298 candidates were tested by qRT-PCR analysis to quantify expression during keratinocyte differentiation. Among these, 216 (72.5%) were detected by classical PCR, confirming their expression in the epidermis. Furthermore, 157 were reliably quantified by qRT-PCR and confirmed by sequencing (see Table S2 for the sequences of primers). Indeed, the use of primers located in different exons associated with sequencing constitutes a definite evidence of their expression in normal epidermis. Among these genes, 49 were overexpressed in the GK fraction ($2.1 < R(\text{ratio}) < 745$; Table 1) and 86 showed no differential expression (Table 2), whereas 22 were more highly expressed in the BK fraction ($0.14 < R < 0.49$, Table 3). Among the 49 upregulated genes, 9 were selected from the ORESTEs database, 19 as a result of microarray experiments, 6 using Genomatix tools, and one from literature. Only 8 genes were retained from both microarrays and ORESTEs database, and 6 from both ORESTEs and Genomatix tools. This low redundancy highlights the complementarity of the sources, and thus the relevance of a multi-technique approach.

As we focused on these functions, upregulated genes comprise several transcriptional regulators (20 genes) and signal proteins (8 genes). Other functional categories are represented by small gene numbers. Four genes (*SH3KBP1*, *LRBA*, *EHD4* and *D4S234E*) are probably involved in vesicle trafficking and cell polarity. Four transporters potentially important for GK differentiation were also identified: a calcium/sodium exchanger

(*SLC8A1*), a calcium-activated chloride channel (*CLCA4*), a zinc transporter (*SLC39A6*) and a proton pump subunit (*ATP6V0A1*). Three genes involved in lipid synthesis or transport (*APOF*, *MVD* and *BBOX1*) were upregulated in GKs. Finally, several structural proteins were previously undescribed in the epidermis: a new epidermal acidic keratin (*KRT23*), the microtubule-associated protein Tau (*MAPT*) and calicin (*CCIN*), a basic actin-binding protein of the sperm head cytoskeleton. Remarkably, *KRT23* displayed the highest *R*-value (745) among the newly identified differentiation-associated genes. Furthermore, transmembrane proteins (*CRELD1*, *TMEM45A*, *LOC57228*) and hyaluronoglucosaminidase *HYAL2* were also specifically expressed in GKs.

As the expression ratios for these new differentiation-associated genes exhibited a large variation ($2.1 < R < 745$), we wondered whether choosing $R=2$ as the lower limit was judicious. To address this question, GK/BK ratios of 20 previously identified differentiation-specific genes were measured by qRT-PCR (Table 4). With the exception of *KLF4* ($R=1.7$), all these genes were differentially detected with ratios ranging from 3.8 (*GJB5*) to 5,800 (*FLG2/IFPS*). This comparison reinforced our identification of 49 novel differentiation-associated genes, and furthermore suggested that only a few false negatives ($1.7 < R < 2$) may have been overlooked.

Furthermore, two of the 49 new genes upregulated with differentiation were tested by immunohistochemistry (IHC) (Fig. 2). The microtubule-associated protein Tau (*MAPT*), which presents a modest qRT-PCR ratio ($R=6.1$), displays a suprabasal gradient moderately reinforced in the granular layer at the protein level (Fig. 2a). SH3-domain kinase binding protein 1 (*SH3KBP1*), which was strongly overexpressed at mRNA level ($R=64$), appears to be restricted to the granular layer (Fig. 2b). These two validations by IHC on normal human skin strengthen our gene expression profiling by qRT-PCR, and suggest that genes displaying the highest qRT-PCR ratios are more likely to have a GK-restricted expression, while others may be expressed earlier (i.e. in all suprabasal layers or progressively upregulated through differentiation).

Discussion

An integrated approach to identify novel human differentiation-associated genes

Basal (B) and suprabasal (S) keratinocyte fractions separated by integrin beta-4 expression levels were analysed in a previous transcriptome study of human epidermis.¹² However, as most GKs remain attached to the stratum corneum after extensive trypsin treatment, GKs were largely absent from the S fraction. Accordingly, the authors noted that several gene transcripts were overexpressed in the S fraction relative to the B fraction, but were underrepresented when compared to total epidermis. Therefore, our study complements these previous findings and directly identifies new GK expressed genes. We discuss the potential roles of these new differentiation players based on function.

Potential roles of TF in the epidermis

Our approach allowed identification of differentiation-associated genes that belong to a variety of functional categories (Table 1). Transcription factors represented 170 of 298 candidate genes, with 94 detected in the epidermis for the first time and 20 overexpressed in GKs. Therefore, TFs constituted the major category of newly identified genes and displayed the highest ratios (except for *KRT23*). Eight of the ten TF genes with a ratio greater than 20 had less than 150 ESTs listed in the UniGene database, suggesting a restricted expression pattern.

We describe for the first time the expression of five forkhead family members in the epidermis: *FOXL1* and *FOXC1* had high ratios ($R=24$ and 22.7 , respectively); *FOXJ3* and *FOXP1* were equally detected in all epidermal layers ($R=0.7$); and *FOXK2* was preferentially expressed in the basal layer ($R=0.37$). It is to note that mutations in *FOXN1* cause severe T-cell immunodeficiency, accompanied by the nude phenotype in mice or alopecia and nail dystrophy in humans (OMIM 601705).

An epidermal-specific deletion of *GATA3* in mice leads to perinatal lethality caused by differentiation defects, especially in lipid synthesis.² Here, we show for the first time that *GATA4* is also expressed in GKs with an expression ratio of 43 when compared to basal

keratinocytes. Moreover, we evidenced the expression of *ZFPM2*, an inhibitor of GATA-4 function in cardiomyocytes,¹⁹ preferentially in basal keratinocytes, and three potential GATA4 target genes (*ANKRD1*, *NFE2* and *HOPX*) overexpressed during differentiation. *HOPX* was recently identified as a late marker of terminal differentiation of lens epithelial cells,²⁰ a process that ends with denucleatosis, a specialized cell death distinct from classical apoptosis. It is thus tempting to speculate that *HOPX* could be also important for cornification.

Signal transduction

We found that 3 regulators of intracellular calcium concentration previously unidentified in the epidermis were upregulated in GKs. *ITPKC* ($R=3.4$) promotes the transformation of IP_3 into IP_4 , both of which participate in calcium homeostasis. *SLC8A1* ($R=16.2$) encodes a sodium/calcium exchanger. Finally, *RCAN1* ($R=5.5$), a regulator of calcineurin-dependent pathway that regulates exocytosis in mouse chromaffin cells,²¹ could be involved in the extrusion of epidermal lamellar bodies.

Other overexpressed genes involved in signal transduction were linked to cell survival and death (*ARL11*, *PLEKHA2*, *MAP3K7IP1*), and possibly to adhesion and migration (*RASGRP1*, *CTNNBP1*, *GMIP*).

Cytoskeleton

Mechanical resistance represents another important aspect of proper barrier formation in the epidermis. We unexpectedly identified keratin 23 (*KRT23*) as a novel acidic epidermal cytokeratin strongly expressed in GKs, displaying the highest ratio ($R=745$) among the new genes. *KRT23* expression was described in sodium butyrate-treated human pancreatic cancer cells.²²

Lipid metabolism and barrier function

Two genes detected in the epidermis are part of the de novo cholesterol synthesis pathway: diphosphomevalonate decarboxylase (*MVD*, $R=4.8$) and emopamil-binding protein, a sterol isomerase (*EBP*, $R=1.5$). *EBP* is associated with X-linked chondrodysplasia punctata 2 (*CDPX2*, OMIM 302960), a human disease that includes striated hyperkeratosis and hair defects.²³ Surprisingly, to our best knowledge, the expression of this gene has never been reported in the epidermis.

We were particularly intrigued by the overexpression of *APOF* ($R=41$) observed in GKs. *APOF* encodes a minor plasma apolipoprotein which blocks cholesteryl ester and triglyceride transfer to lipoproteins. The biological significance of *APOF* overexpression in GKs remains presently unclear, but this unexpected finding underlines the role of lipid transfer during cornification.

Cell polarity, vesicle trafficking and lamellar bodies

Lamellar bodies are secreted from the apical pole of GKs, highlighting the importance of cell polarity and vesicle trafficking in these cells. Here, we show that several genes involved in these processes are overexpressed in GKs: *LRBA* ($R=3.1$) and *SH3KBP1* ($R=64$) for polarized vesicle trafficking,^{24,25} *EHD4* ($R=2.7$) and *D4S234E* ($R=2.4$) for membrane recycling/endocytosis.^{26,27} Interestingly, *MAPT*, a gene encoding the microtubule-associated protein Tau expressed primarily in neurons and oligodendrocytes, is overexpressed in GKs ($R=6.1$). By affecting microtubule stability and motor protein activity,²⁸ Tau may regulate vesicle trafficking in GKs.

Lamellar bodies are known to be acidic. Accordingly, we found that *ATP6V0A1* ($R=2.7$) was overexpressed in GKs. *ATP6V1A* (*ATP6A1*) was previously shown to be upregulated in the suprabasal layer, but not in GKs.¹² These genes encode components of the vacuolar-type H⁺ ATPase, a multimeric enzyme that mediates acidification of organelles, suggesting a function for this pump in the acidification of lamellar bodies.

In total, the present work constitutes the first comparison of expression level between granular and basal keratinocytes for 157 genes. Among them, 49 appeared overexpressed in GKs, including 20 TFs that call for further investigations to determine their contribution to the establishment of the barrier function. Furthermore, the potential role of a novel keratin, KRT23, and of several enzymes implicated in lipid metabolism has been highlighted. These advances made in the field of keratinocyte differentiation may contribute to the deciphering of complex diseases such as atopic dermatitis, psoriasis and ichthyoses, which are characterized by deregulated differentiation programs.

Acknowledgements

We give special thanks to M.-T. Ribouchon and R. Oger for excellent technical assistance. We are grateful to C. Offer for sequencing the qRT-PCR fragments at the sequencing platform of IFR30, Toulouse, and to V. Leberre, L. Trouilh and S. Sokol for their technical support to microarray experiments and statistical analysis at the Genotoul biochips platform, Toulouse. We thank I. El Gannab and M. Gutiérrez Martos for their contribution to qRT-PCR and sequencing, respectively.

References

- 1 Segre JA, Bauer C, Fuchs E. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet* 1999; **22**: 356-60.
- 2 de Guzman Strong C, Wertz PW, Wang C et al. Lipid defect underlies selective skin barrier impairment of an epidermal-specific deletion of Gata-3. *J Cell Biol* 2006; **175**: 661-70.
- 3 Yu Z, Lin KK, Bhandari A et al. The Grainyhead-like epithelial transactivator Get-1/Grhl3 regulates epidermal terminal differentiation and interacts functionally with LMO4. *Dev Biol* 2006; **299**: 122-36.
- 4 Magnusdottir E, Kalachikov S, Mizukoshi K et al. Epidermal terminal differentiation depends on B lymphocyte-induced maturation protein-1. *Proc Natl Acad Sci U S A* 2007; **104**: 14988-93.
- 5 Jansen BJ, van Ruissen F, de Jongh G et al. Serial analysis of gene expression in differentiated cultures of human epidermal keratinocytes. *J Invest Dermatol* 2001; **116**: 12-22.
- 6 van Ruissen F, Jansen BJ, de Jongh GJ et al. Differential gene expression in premalignant human epidermis revealed by cluster analysis of serial analysis of gene expression (SAGE) libraries. *Faseb J* 2002; **16**: 246-8.

- 7 van Ruissen F, Jansen BJ, de Jongh GJ et al. A partial transcriptome of human epidermis. *Genomics* 2002; **79**: 671-8.
- 8 Bernard FX, Pedretti N, Rosdy M et al. Comparison of gene expression profiles in human keratinocyte mono-layer cultures, reconstituted epidermis and normal human skin; transcriptional effects of retinoid treatments in reconstituted human epidermis. *Exp Dermatol* 2002; **11**: 59-74.
- 9 Gazel A, Ramphal P, Rosdy M et al. Transcriptional profiling of epidermal keratinocytes: comparison of genes expressed in skin, cultured keratinocytes, and reconstituted epidermis, using large DNA microarrays. *J Invest Dermatol* 2003; **121**: 1459-68.
- 10 Mehul B, Asselineau D, Bernard D et al. Gene expression profiles of three different models of reconstructed human epidermis and classical cultures of keratinocytes using cDNA arrays. *Arch Dermatol Res* 2004; **296**: 145-56.
- 11 Lemaitre G, Lamartine J, Pitaval A et al. Expression profiling of genes and proteins in HaCaT keratinocytes: proliferating versus differentiated state. *J Cell Biochem* 2004; **93**: 1048-62.
- 12 Radoja N, Gazel A, Banno T et al. Transcriptional profiling of epidermal differentiation. *Physiol Genomics* 2006; **27**: 65-78.
- 13 Toulza E, Mattiuzzo NR, Galliano MF et al. Large-scale identification of human genes implicated in epidermal barrier function. *Genome Biol* 2007; **8**: R107.
- 14 Le Brigand K, Russell R, Moreilhon C et al. An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes. *Nucleic Acids Res* 2006; **34**: e87.
- 15 Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001; **98**: 5116-21.
- 16 Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000; **132**: 365-86.
- 17 Altschul SF, Madden TL, Schaffer AA et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389-402.
- 18 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001; **25**: 402-8.
- 19 Svensson EC, Tufts RL, Polk CE et al. Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. *Proc Natl Acad Sci U S A* 1999; **96**: 956-61.
- 20 Vasiliev O, Rhodes SJ, Beebe DC. Identification and expression of Hop, an atypical homeobox gene expressed late in lens fiber cell terminal differentiation. *Mol Vis* 2007; **13**: 114-24.
- 21 Keating DJ, Dubach D, Zanin MP et al. DSCR1/RCAN1 regulates vesicle exocytosis and fusion pore kinetics: implications for Down syndrome and Alzheimer's disease. *Hum Mol Genet* 2008; **17**: 1020-30.
- 22 Zhang JS, Wang L, Huang H et al. Keratin 23 (K23), a novel acidic keratin, is highly induced by histone deacetylase inhibitors during differentiation of pancreatic cancer cells. *Genes Chromosomes Cancer* 2001; **30**: 123-35.
- 23 Derry JM, Gormally E, Means GD et al. Mutations in a delta 8-delta 7 sterol isomerase in the tattered mouse and X-linked dominant chondrodysplasia punctata. *Nat Genet* 1999; **22**: 286-90.
- 24 Wang JW, Howson J, Haller E et al. Identification of a novel lipopolysaccharide-inducible gene with key features of both A kinase anchor proteins and chs1/beige proteins. *J Immunol* 2001; **166**: 4586-95.

- 25 Havrylov S, Ichioka F, Powell K et al. Adaptor protein Ruk/CIN85 is associated with a subset of COPI-coated membranes of the Golgi complex. *Traffic* 2008; **9**: 798-812.
- 26 Sharma M, Naslavsky N, Caplan S. A role for EHD4 in the regulation of early endosomal transport. *Traffic* 2008; **9**: 995-1018.
- 27 Steiner P, Sarria JC, Glauser L et al. Modulation of receptor cycling by neuron-enriched endosomal protein of 21 kD. *J Cell Biol* 2002; **157**: 1197-209.
- 28 Dixit R, Ross JL, Goldman YE et al. Differential regulation of dynein and kinesin motor proteins by tau. *Science* 2008; **319**: 1086-9.

Table 1. Genes upregulated in the granular layer: GK/BK expression ratios determined by qRT-PCR.

Gene symbol	Full name (alias / previous name)	no. of UniGene ESTs	Technique(s) of identification	GK/BK ratio
<i>ABHD11</i>	abhydrolase domain containing 11 (WBSCR21)	150	ORESTES, microarrays	4.9
<i>ANKRD1</i>	ankyrin repeat domain 1 (cardiac muscle)	162	Microarrays*	41
<i>APOF</i>	apolipoprotein F (lipid transfer inhibitor protein, LTIP)	14	Microarrays*	41
<i>ARL11</i>	ADP-ribosylation factor-like protein 11 (ARLTS1)	20	Microarrays	9
<i>ATF6</i>	activating transcription factor 6	452	Literature	2.1
<i>ATP6V0A1</i>	ATPase, H ⁺ transporting, lysosomal V0 subunit a1	1352	ORESTES, microarrays	2.7
<i>BBOX1</i>	butyrobetaine (gamma), 2-oxoglutarate dioxygenase (gamma-butyrobetaine hydroxylase) 1	134	Genomatix (pt)	5.2
<i>BHLHB2</i>	basic helix-loop-helix domain containing, class B, 2	556	ORESTES	3.2
<i>CCIN</i>	calicin	189	Microarrays	5.6
<i>CLCA4</i>	chloride channel, calcium activated, family member 4	95	Microarrays	2.1
<i>CRELD1</i>	cysteine-rich with EGF-like domains 1 (AVSD2)	262	Microarrays*	6.9
<i>CTNBP1</i>	catenin, beta interacting protein 1 (ICAT)	321	ORESTES	9.1
<i>D4S234E</i>	DNA segment on chromosome 4 (unique) 234 expressed sequence (NEEP21)	997	ORESTES	2.4
<i>EHD4</i>	EH-domain containing 4 (PAST4)	186	ORESTES	2.7
<i>FBXO38</i>	F-box protein 38 (MoKA)	265	Microarrays*	2.8
<i>FGFR1OP2</i>	FGFR1 oncogene partner 2	290	Microarrays	2.2
<i>FOXC1</i>	forkhead box C1	115	ORESTES, Genomatix	22.7
<i>FOXL1</i>	forkhead box L1	17	Microarrays	24
<i>GATA4</i>	GATA binding protein 4	129	Microarrays*	43
<i>GMIP</i>	GEM interacting protein	160	Microarrays	2.1
<i>HOPX</i>	HOP homeobox	482	ORESTES, microarrays	23.1
<i>HSPC159</i>	hematopoietic stem cell precursor (galectin-related protein GRP)	190	ORESTES, microarrays	15.6
<i>HYAL2</i>	hyaluronoglucosaminidase 2	424	Microarrays*	2.6
<i>INSM1</i>	insulinoma-associated 1 (IA-1)	105	Genomatix	59.9
<i>ITPKC</i>	inositol 1,4,5-trisphosphate 3-kinase C	205	Microarrays	3.4
<i>KLF14</i>	Kruppel-like factor 14 (BTEB5)	5	Genomatix	41.3
<i>KRT23</i>	keratin 23 (histone deacetylase inducible)	320	ORESTES	745
<i>LOC57228</i>	small trans-membrane and glycosylated protein (SMAGP)	168	Microarrays*	10.4
<i>LRBA</i>	LPS-responsive vesicle trafficking, beach and anchor containing (CDC4L)	306	ORESTES, Genomatix (pt)	3.1
<i>MAP3K7IP1</i>	mitogen-activated protein kinase kinase 7-interacting protein 1 (TAB1)	421	Microarrays*	2.1
<i>MAPT</i>	microtubule-associated protein tau	475	Microarrays*	6.1
<i>MAX</i>	MYC-associated factor X (splicing variant 4)	434	ORESTES, Genomatix	5.4
<i>MBD1</i>	methyl-CpG binding domain protein 1	551	ORESTES	3.6
<i>MVD</i>	mevalonate (diphospho) decarboxylase	358	Microarrays	4.8
<i>NFE2</i>	nuclear factor (erythroid-derived 2), 45kDa	112	ORESTES, Genomatix	76.9
<i>NMI</i>	N-myc and STAT interactor	176	ORESTES	2.3
<i>NR2C1</i>	nuclear receptor subfamily 2, group C, member 1	303	Genomatix	7.7
<i>OTX1</i>	orthodenticle homeobox 1	45	ORESTES	4.3
<i>PHF2</i>	PHD finger protein 2	178	Microarrays*	7.5
<i>PLEKHA2</i>	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 2	257	ORESTES, microarrays	8.1
<i>RASGRP1</i>	RAS guanyl releasing protein 1 (calcium and DAG-regulated)	134	Microarrays*	31.8
<i>RCAN1</i>	regulator of calcineurin 1 (DSCR1)	595	ORESTES	5.5
<i>RFX5</i>	regulatory factor X subunit 5	485	ORESTES, Genomatix	4.5
<i>SH3KBP1</i>	SH3-domain kinase binding protein 1	397	ORESTES, microarrays	64
<i>SIM1</i>	single-minded homolog 1 (Drosophila)	18	Genomatix	46.8
<i>SLC39A6</i>	solute carrier family 39 (zinc transporter), member 6	392	ORESTES, microarrays	4
<i>SLC8A1</i>	solute carrier family 8 (sodium/calcium exchanger), member 1	179	ORESTES, Genomatix (pt)	16.2
<i>TMEM45A</i>	transmembrane protein 45A	257	ORESTES, microarrays	26.4
<i>ZNF239</i>	Zinc finger protein 239 (homologous to mouse MOK-2)	79	Genomatix	66.9

* based on preliminary microarrays analysis; FDR > 0.05. (pt) putative target gene. Names of genes encoding transcription factors are written in bold.

Table 2. Genes expressed equally in the basal and granular layers: GK/BK expression ratios determined by qRT-PCR

Gene symbol	Full name (alias / previous name)	no. of UniGene ESTs	Technique(s) of identification	GK/BK ratio
AATF	apoptosis antagonizing transcription factor	605	ORESTES	1
ACBD6	acyl-Coenzyme A binding domain containing 6	200	Microarrays	1
ACSL4	acyl-CoA synthetase long-chain family member 4 (FACL4, MRX63, MRX68)	392	Microarrays	0.9
AQR	aquarius homolog (mouse)	288	ORESTES	1.5
ARID2	AT rich interactive domain 2 (ARID, RFX-like)	206	ORESTES, Genomatix	1.1
ARID5B	AT rich interactive domain 5B (MRF1-like)	285	ORESTES, Genomatix	0.6
ARSD	arylsulfatase D	323	Microarrays*	1.4
ATF5	activating transcription factor 5	479	Genomatix	0.5
ATF6B	activating transcription factor 6 beta	217	ORESTES	1.1
AVEN	apoptosis, caspase activation inhibitor	77	ORESTES	1.3
BCL6	B-cell CLL/lymphoma 6 (zinc finger protein 51)	398	Genomatix	0.9
BPTF	bromodomain PHD finger transcription factor (FALZ)	455	ORESTES	1.3
BRF1	BRF1 homolog, subunit of RNA polymerase III TIFIIIB	322	Microarrays*	1
BTBD11	BTB (POZ) domain containing 11	95	Microarrays*	0.6
CBFA2T2	core-binding factor, runt domain, alpha subunit 2; translocated to, 2	439	ORESTES, Genomatix	0.6
CEBPZ	CCAAT/enhancer binding protein zeta	349	ORESTES, Genomatix	1.2
CHD2	chromodomain helicase DNA binding protein 2	923	ORESTES, microarrays	1.5
CNOT2	CCR4-NOT transcription complex, subunit 2	632	ORESTES	1.6
CST7	cystatin F (leukocystatin)	63	Microarrays	0.8
DDIT3	DNA damage-inducible transcript 3	312	Genomatix	0.6
DDX41	DEAD (Asp-Glu-Ala-Asp) box polypeptide 41	1416	Microarrays	1.6
EBP	emopamil binding protein (sterol isomerase) (CDPX2)	561	Microarrays	1.5
EGR1	early growth response 1	451	ORESTES, Genomatix	0.9
EXT2	exostosin 2 (N-acetylglucosaminyl-proteoglycan 4-beta-glucuronosyltransferase)	538	Microarrays*	0.8
FAM179B	family with sequence similarity 179, member B	140	Microarrays	0.7
FAR1	fatty acyl-CoA reductase 1 (MLSTD2)	363	Microarrays*	0.9
FLVCR1	feline leukemia virus subgroup C cellular receptor 1 (heme transporter)	73	Microarrays*	0.8
FOXJ3	forkhead box J3	314	ORESTES, Genomatix	0.7
FOXP1	forkhead box P1	458	ORESTES, Genomatix	0.7
HDAC8	histone deacetylase 8	223	Microarrays*	1.4
HLTF	helicase-like transcription factor	348	Genomatix	1
HOXA9	homeobox A9	62	Genomatix	0.6
ING4	inhibitor of growth family, member 4	280	ORESTES, Genomatix	1.6
IRF2	interferon regulatory factor 2	317	Genomatix	1
IRF7	interferon regulatory factor 7	304	Genomatix	1.7
ISL1	ISL LIM homeobox 1	79	Genomatix	0.7
KLF10	Kruppel-like factor 10 (TIEG1)	324	ORESTES	1.7
KLF13	Kruppel-like factor 13 (BTEB3)	253	ORESTES	1
LDB1	LIM domain binding 1	204	ORESTES	0.8
LLGL1	lethal giant larvae homolog 1 (DLG4, HUGL)	333	Microarrays*	0.9
PLA2G15	phospholipase A2, group XV (lysosomal phospholipase A2)	389	Microarrays*	1
MAFK	v-maf musculoaponeurotic fibrosarcoma oncogene homolog K	256	Literature	0.6
MAGED1	melanoma antigen family D. 1 (NRAGE)	1293	ORESTES	1
MAX	MYC-associated factor X (splicing variants 1-3 & 5)	434	ORESTES, Genomatix	1.7
MBD2	methyl-CpG binding domain protein 2	284	ORESTES	1.8
MBTPS1	membrane-bound transcription factor peptidase, site 1	804	ORESTES	0.9
MED4	mediator complex subunit 4 (VDRIP)	496	Microarrays*	1.9
NCOA6	nuclear receptor coactivator 6	362	Literature	1.6
NIPBL	Nipped-B homolog (Drosophila)	360	ORESTES	1.3
NR1D1	nuclear receptor subfamily 1, group D, member 1	196	ORESTES, Genomatix	1.2
NR4A1	nuclear receptor subfamily 4 group A member 1	677	ORESTES	1.9
NRBP1	nuclear receptor binding protein 1	931	ORESTES	1.6
ODZ1	odz, odd Oz/ten-m homolog 1(Drosophila)	37	Microarrays	0.6
OTUD7B	OTU domain-containing protein 7B (deubiquitinating cysteine protease, ZA20D1)	159	ORESTES	0.54
OXER1	oxoeicosanoid (OXE) receptor 1	12	Microarrays*	1.6
PAPOLG	poly(A) polymerase gamma	202	Microarrays*	1.8
PHF7	PHD finger protein 7	684	Genomatix	1.1

POU6F1	POU class 6 homeobox 1 (BRN5)	83	Genomatix	0.5
PRDM15	PR domain containing 15 (ZNF298)	170	ORESTES, microarrays	0.6
SCLY	selenocysteine lyase	470	Microarrays*	0.7
SCP2	sterol carrier protein 2 (SCP2 isoform)	1173	ORESTES	1.8
SCP2	sterol carrier protein 2 (SCPx isoform)	1173	ORESTES	1.2
SEZ6L2	seizure related 6 homolog (mouse)-like 2	380	Microarrays*	1.1
SKIV2L	superkiller viralicidic activity 2-like	286	Microarrays*	1.9
SLC30A6	solute carrier family 30 (zinc transporter), member 6	286	Microarrays	1.2
SLC35B2	solute carrier family 35, member B2 (3 -phosphoadenosine 5 -phosphosulfate transporter)	496	Microarrays	0.55
SLTM	SAFB-like transcription modulator	490	ORESTES	1.5
SSRP1	structure-specific recognition protein 1	696	ORESTES	0.8
TCFL5	transcription factor-like 5 (basic helix-loop-helix) [medial last exon]	121	ORESTES	1.5
TCFL5	transcription factor-like 5 (basic helix-loop-helix) [distal last exon]	121	ORESTES	1
TFAP2E	activating enhancer-binding protein 2 epsilon	3	Literature	1
TFCP2L1	transcription factor CP2-like protein 1	130	Genomatix	0.6
TLE1	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	546	Microarrays*	1.3
TRIM29	tripartite motif-containing 29 (ATDC)	772	ORESTES	0.8
TSPYL5	testis-specific Y-encoded-like protein 5	134	Microarrays*	1.2
UBP1	upstream binding protein 1 (LBP-1a)	298	ORESTES, Genomatix	1.3
VWA1	von Willebrand factor A domain containing 1 (WARP)	631	Microarrays*	1.6
ZBTB7B	zinc finger and BTB domain containing 7B (c-Krox)	105	ORESTES, Genomatix	0.6
ZEB2	zinc finger E-box binding homeobox 2 (ZFHX1B)	411	Microarrays*	1.1
ZFHX3	zinc finger homeobox 3 (ATBF1)	176	ORESTES, Genomatix	0.7
ZFPM1	zinc finger protein multitype 1 (FOG1)	40	Genomatix	0.9
ZHX2	zinc fingers and homeoboxes protein 2	151	ORESTES	1.3
ZNF143	zinc finger protein 143 (STAF)	204	Genomatix	1.2
ZNF230	zinc finger protein 230	76	Microarrays	1.3
ZNF335	zinc finger protein 335 (NIF-1)	187	ORESTES, microarrays	1.2
ZNF396	zinc finger protein 396 (ZSCAN14)	27	ORESTES	1.1

* based on preliminary microarrays analysis; FDR > 0.05. Names of genes encoding transcription factors are written in bold.

Table 3. Genes downregulated in the granular layer: GK/BK expression ratios determined by qRT-PCR.

Gene symbol	Full name (alias / previous name)	no. of UniGene ESTs	Technique(s) of identification	GK/BK ratio
BCL11A	B-cell CLL/lymphoma 11A (EVI9)	203	ORESTES, Genomatix	0.32
<i>DKFZP586H2123</i>	regeneration associated muscle protease (RAMP)	245	ORESTES	0.3
ETV5	ets variant gene 5 (ets-related molecule)	359	ORESTES	0.29
FOXK2	forkhead box protein K2 (ILF1)	508	Genomatix	0.37
<i>GRLF1</i>	glucocorticoid receptor DNA binding factor 1	490	ORESTES	0.47
HIVEP3	human immunodeficiency virus type I enhancer-binding protein 3	41	ORESTES	0.27
<i>IQGAP3</i>	IQ motif containing GTPase activating protein 3	77	ORESTES	0.18
MAML1	Mastermind-like protein 1	325	ORESTES	0.36
MAZ	MYC-associated zinc finger protein (purine-binding transcription factor)	746	Genomatix	0.36
MBD3	methyl-CpG binding domain protein 3	550	Microarrays	0.15
NCOR2	nuclear receptor corepressor 2 (SMRT)	544	ORESTES	0.38
<i>NDOR1</i>	NADPH dependent diflavin oxidoreductase 1	107	ORESTES, microarrays	0.42
NFE2L1	nuclear factor (erythroid-derived 2)-like 1	1022	ORESTES, Genomatix	0.14
PELP1	proline, glutamic acid and leucine rich protein 1	492	ORESTES	0.3
PKNOX1	PBX/knotted 1 homeobox 1	316	ORESTES, Genomatix	0.35
RREB1	ras responsive element binding protein 1	207	ORESTES, Genomatix	0.29
SOX12	SRY (sex determining region Y)-box 12	171	Microarrays	0.29
SUV39H1	suppressor of variegation 3-9 homolog 1 (histone H3-K9 methyltransferase 1)	241	Microarrays*	0.4
<i>TCIRG1</i>	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 subunit A3	268	Microarrays	0.49
TEF	thyrotrophic embryonic factor	348	ORESTES, Genomatix	0.17
ZFPM2	zinc finger protein multitype 2 (FOG2, modulator of GATA4)	55	Literature	0.33
ZNF76	zinc finger protein 76	232	Genomatix	0.24

* based on preliminary microarrays analysis; FDR > 0.05. Names of genes encoding transcription factors are written in bold.

Table 4. Previously described markers of suprabasal layers: GK/BK expression ratios determined by qRT-PCR

Gene symbol	Full name (alias / previous name)	no. of UniGene ESTs	GK/BK ratio (qRT-PCR)
<i>A2ML1</i>	alpha-2-macroglobulin-like 1	70	43
<i>CDSN</i>	corneodesmosin	134	3,225
<i>CNFN</i>	cornifelin	29	775
<i>CSTA</i>	cystatin A (stefin A)	179	16
<i>DLX3</i>	distal-less homeobox 3	93	3.7
<i>DMKN</i>	dermokine (<i>ZD52F10</i>)	446	390
<i>DSG1</i>	desmoglein 1	89	4.5
<i>FLG</i>	filaggrin	114	2,155
<i>FLG2</i>	filaggrin family member 2 (Ifapsoriasin, <i>IFPS</i>)	61	5,795
<i>GJB5</i>	gap junction protein, beta 5 (connexin 31.1, <i>Cx31.1</i>)	49	3.8
<i>IVL</i>	involucrin	73	204
<i>KLF4</i>	Kruppel-like factor 4 (Gut-enriched kruppel-like factor)	218	1.7 (<i>ns</i>)
<i>KLK5</i>	kallikrein-related peptidase 5 (Stratum corneum tryptic enzyme, SCTE)	53	23
<i>KLK7</i>	kallikrein-related peptidase 7 (Stratum corneum chymotryptic enzyme, SCCE)	91	80
<i>LOR</i>	loricrin	17	3,735
<i>RORA</i>	RAR-related orphan receptor A	62	7.2
<i>SCEL</i>	sciellin	134	34
<i>SERPINB5</i>	serpin peptidase inhibitor, clade B (ovalbumin), member 5 (maspin)	213	6.9
<i>SERPINB8</i>	serpin peptidase inhibitor, clade B (ovalbumin), member 8	105	4.9
<i>SPINK5</i>	serine peptidase inhibitor, Kazal type 5	138	44

ns: not significant (ratio < 2). Names of genes encoding transcription factors are written in bold.

FIGURE 1

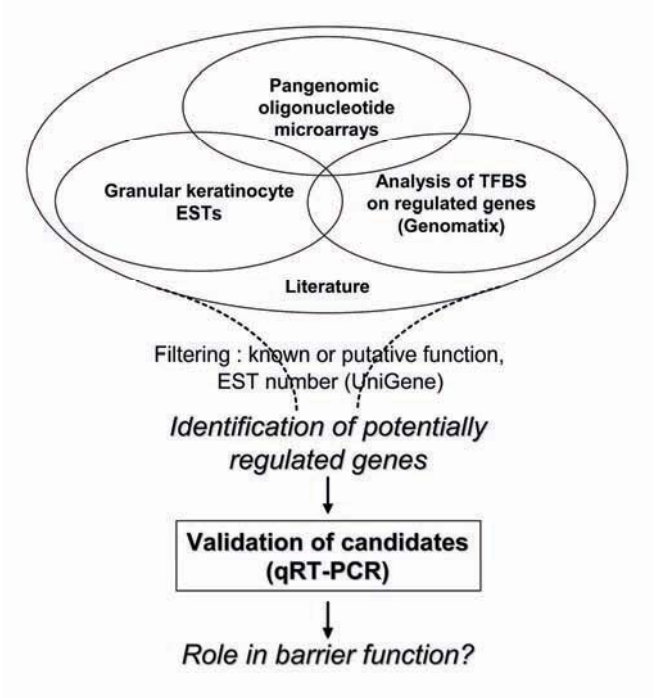


FIGURE 2

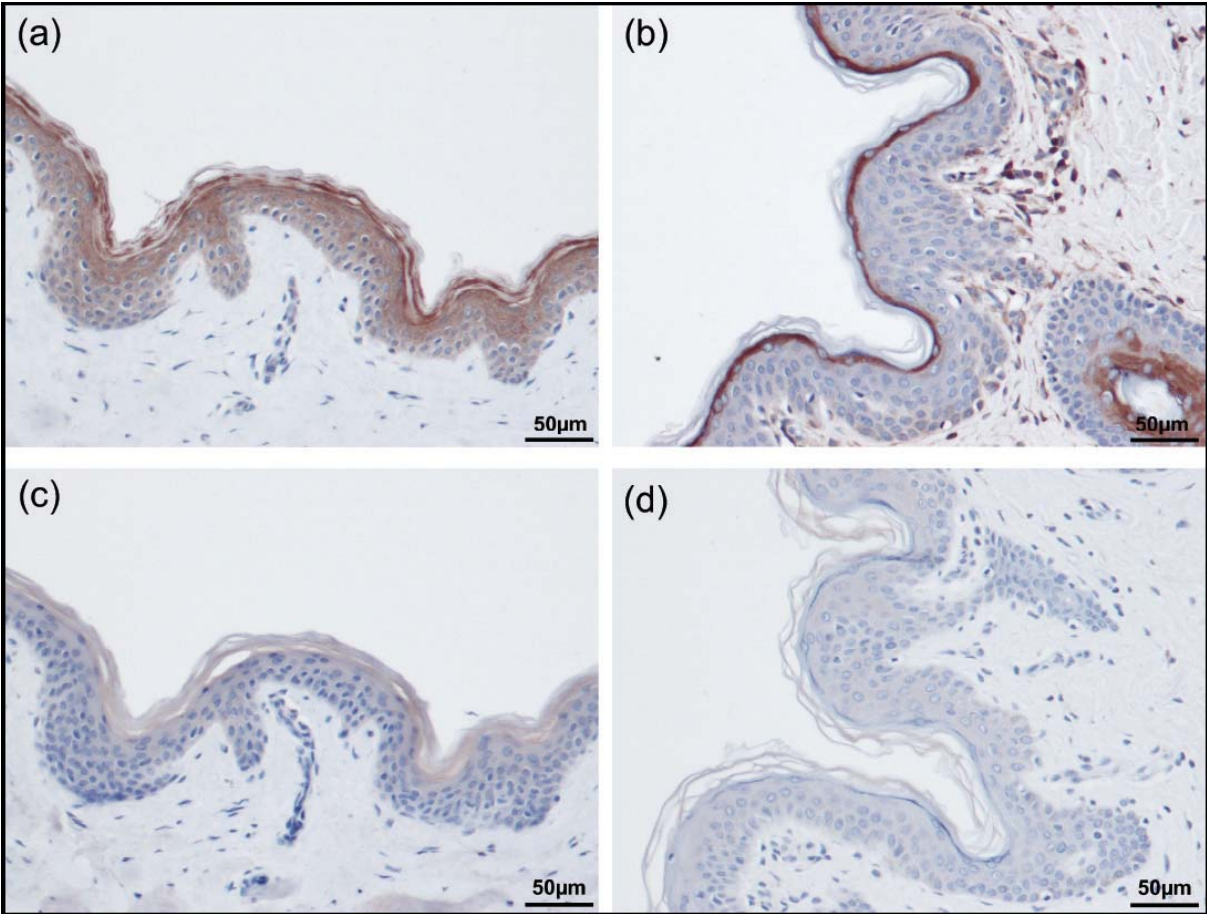


Figure Legends

Fig 1. Design of the large-scale multi-technique approach. Data from several sources were compiled and filtered to specifically target genes most relevant for barrier function. Expression levels during differentiation were determined by qRT-PCR.

Fig 2. Validation of new differentiation-associated genes by immunohistochemistry on normal human skin sections. MAPT protein (a) is mainly detected in suprabasal layers, and the staining is moderately reinforced in the granular layer. SH3KBP1 protein (b) is restricted to the granular layer. Negative controls (c-d) consisted in incubations in the absence of the primary antibody on corresponding serial sections. Bar = 50 μ m.

Supporting information

Supplementary Results and Material & Methods (microarrays).

Fig S1. Microarray results from six independent samples.

Table S1. Differentially expressed genes listed according to Significance Analysis of Microarrays (SAM).

Supplementary Material & Methods (in silico promoter analyses).

Fig S2. Example of promoter analysis for DMKN- β / γ orthologs using Genomatix tools.

Table S2. Sequences of primers used for qRT-PCR essays.

SUPPORTING INFORMATION

Supplementary Results and Material & Methods (microarrays).

Figure S1. Microarray results from six independent samples.

Table S1. Differentially expressed genes listed according to Significance Analysis of Microarrays (SAM). Known markers of basal or differentiated layers are written in green or red bold, respectively.

Supplementary Material & Methods (in silico promoter analyses).

Figure S2. Example of promoter analysis for DMKN- β / γ orthologs using Genomatix tools.

Table S2. Sequences of primers used for qRT-PCR essays.

Supplementary Results and Material & Methods (Microarrays).

Pangenomic oligonucleotide microarrays were used for global transcriptional profiling. Differentially labeled cDNAs prepared from GKs and BKs from six individuals were cohybridized onto microarrays. Data processing revealed that 11,520 spots, corresponding to 10,824 different genes, were significantly detected. For statistical analysis of differentially expressed genes, local False Discovery Rate (FDR) using the Significance Analysis of Microarrays (SAM) method was considered rather than p -values. With a delta value of 1.462 and a two-fold ratio threshold, 567 genes appeared to be up- or downregulated in the GK fraction (Figure S1). By further imposing a local FDR below 0.05, this number was restricted to 206 genes (Table S1), with 203 upregulated and 3 downregulated. Among the 203 upregulated genes, 23 were known players of epidermal differentiation, while the others were not previously known to be expressed in the human epidermis.

Bioanalyzer experiments, as well as cDNA labeling, microarray hybridizations and scanning, were performed at the Biochips platform of the Génopole Toulouse Midi-Pyrénées (Genotoul) in Toulouse, France (<http://biopuce.insa-toulouse.fr/sommaire.php>). All microarrays data are MIAME compliant and available from the Gene Expression Omnibus (GEO) database under accession number GSE13263. Six independent samples and nine arrays were used in this study: four BK samples were labeled with Cy3 and hybridized with the corresponding Cy5-labelled GK sample; reverse labeling was performed for three arrays (called swap arrays) in order to assess potential biases due to the dye; and the two remaining BK and GK samples were labeled with Cy5 and Cy3, respectively. Non-commercial pangenomic oligonucleotide microarrays (platform GPL1946 in GEO database) produced by the RNG/MRC for the Resogen Program¹ consisted of 25,248 different 50mers targeting 22,458 genes and used according to the manufacturer's instructions for non-amplified RNA. Briefly, 10-20µg of total RNA were reverse transcribed with Superscript II (18064-022, Invitrogen) in the presence of oligo-dT₁₂₋₁₈ primers (18418-012, Invitrogen) and aminoallyl-dUTP (A0410, Sigma-Aldrich). Single-stranded cDNA were purified using the cleanup module of the Labelstar Array Kit (28904, Qiagen) and coupled with monofunctional

NHS ester Cy3 or Cy5 dye (RPN5661, Amersham/GE Healthcare). After purification, Cy3- and Cy5-labelled cDNAs were pooled and vacuum-concentrated. Following dilution into 50µl of hybridization buffer (2x SSC, 0.1% SDS and 0.1% salmon sperm DNA) (15632-011, Invitrogen), cDNA were deposited onto microarrays under coverslips, and slides were placed in a hybridization chamber for 16 hours at 50°C. After four washes, slides were digitized with an Axon GenePix 4000A scanner. Image analysis was performed with the Axon GenePix Pro 3.0 software. Raw data were imported into Biplot software developed at the Genotoul biochips platform (<http://biopuce.insa-toulouse.fr/ExperimentExplorer/doc/>). Data from the three pairs of "swap arrays" were merged and considered as three simple arrays for further analyses. Only spots with signal intensities greater than the average signal for all buffer spots (i.e., spots without oligonucleotides) plus 2 SD and detected by more than half of the arrays (i.e., with at least four values) were analyzed. Local background-subtracted median intensities were log-transformed and normalized by the LOWESS method, and Student's *t*-tests were carried out on differentially expressed genes. SAM software² was used on normalized data for false discovery rate (FDR) analysis.

References

- 1 Le Brigand K, Russell R, Moreilhon C et al. An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes. *Nucleic Acids Res* 2006; **34**: e87.
- 2 Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001; **98**: 5116-21.

Figure S1. Microarray results from six independent samples. (a) MA plot. Mean GK/BK ratios and average intensities are represented for each of the 11,552 detected spots. A higher number of upregulated genes (589) exceeded the 2-fold threshold (red lines), when compared to downregulated genes (88). (b) SAM plotsheet. With a 2-fold ratio threshold and a delta value of 1.462, 567 genes were considered to be significantly up- or downregulated (red and green dots, respectively). Only genes (203 upregulated and 3 downregulated) with a local FDR lower than 0.05 were retained for further analysis.

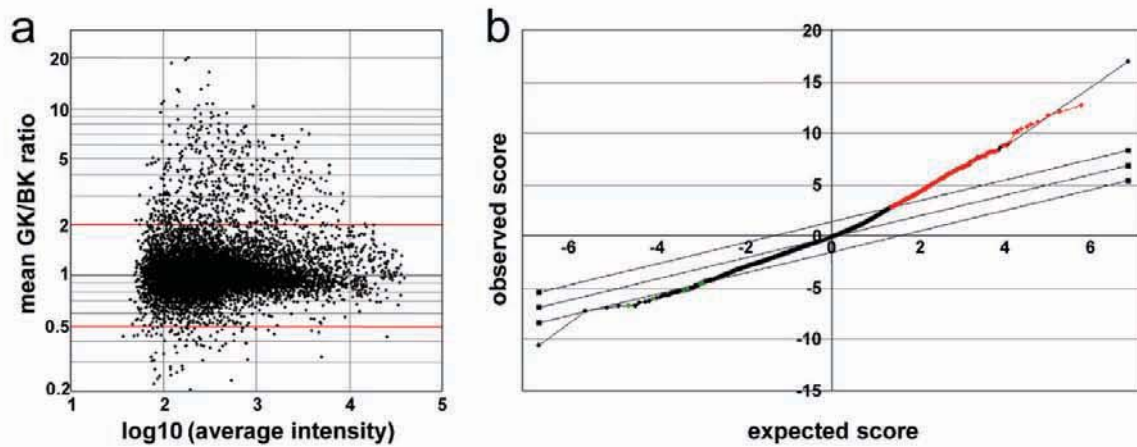


Table S1. Differentially expressed genes listed according to Significance Analysis of Microarrays (SAM).

Known markers of suprabasal and basal layers of epidermis are written in red and green, respectively.

Gene Name	Full name (alias/previous name)	Fold Change (GK vs BK)
CCIN	calicin	27.58
PCDH12	protocadherin 12	26.03
APOL5	apolipoprotein L, 5	19.71
CDSN	corneodesmosin	16.98
SPDYA	speedy homolog A (Drosophila)	16.41
BC039551	Homo sapiens, clone IMAGE:5762932, mRNA	16.37
LCE2B	late cornified envelope 2B	15.2
BMP1	bone morphogenetic protein 1	14.75
FGFR1OP2	FGFR1 oncogene partner 2	14.02
AK090694	Homo sapiens cDNA FLJ33375	13.94
LOC347509	similar to ribosomal protein L7-like 1	12.32
AK095117	Homo sapiens cDNA FLJ37798	12.06
KRTDAP	keratinocyte differentiation-associated protein	11.38
SCEL	sclellin	10.62
CALML5	calmodulin-like 5	10.39
SLC39A6	solute carrier family 39 (zinc transporter), member 6	10.24
ARL11	ADP-ribosylation factor-like 11	10.23
TMEM45A	transmembrane protein 45A	10.1
ABHD11	abhydrolase domain containing 11	10.01
TLN2	taln 2	9.95
FLG	filaggrin	9.59
SLC35B2	solute carrier family 35, member B2	9.47
AF462653	Homo sapiens isolate BRAS immunoglobulin light chain variable region mRNA	9.12
RESP18	regulated endocrine-specific protein 18	9.12
KIAA0423	uncharacterized protein KIAA0423	8.67
EGF	epidermal growth factor (beta-urogastrone)	8.35
MBD3	methyl-CpG binding domain protein 3	8.28
TOR1B	torsin family 1, member B	8.27
TEX13B	testis expressed 13B	8.23
C1orf68	Skin-specific protein 32 (xp32, LEP7)	7.96
AL117559	Homo sapiens mRNA AL117559	7.82
LCE3C	late cornified envelope 3C	7.79
TUBB2A	tubulin, beta 2A	7.75
KRT2	keratin 2 (epidermal ichthyosis bullosa of Siemens)	7.59
GPSM3	G-protein signalling modulator 3 (AGS3-like, C. elegans)	7.55
SBSN	suprabasin	7.47
LOR	loricrin	7.3
SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	7.27
SPINK5	serine protease inhibitor, Kazal type 5	7.2
CRCT1	cysteine-rich C-terminal 1 (C1orf42, NICE-1)	7.04
FLG	filaggrin (alternative probe)	7.03
DDX41	DEAD (Asp-Glu-Ala-Asp) box polypeptide 41	6.85
CHD2	chromodomain helicase DNA binding protein 2	6.83
HOXD13	homeobox D13	6.73
C12orf10	chromosome 12 open reading frame 10 (MYG1)	6.7
C10orf65	chromosome 10 open reading frame 65	6.69
FABP5	fatty acid binding protein 5 (psoriasis-associated)	6.56
BAIAP2	Brain-specific angiogenesis inhibitor 1-associated protein 2	6.52
HSD17B7P2	hydroxysteroid (17-beta) dehydrogenase 7 pseudogene 2	6.5
CLCA4	chloride channel, calcium activated, family member 4	6.43
C20orf181	chromosome 20 open reading frame 181	6.18
TIMM8A	translocase of inner mitochondrial membrane 8 homolog A (yeast)	6.17
TAF13	TAF13 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 18kDa	6.05
CMTM1	CKLF-like MARVEL transmembrane domain containing 1	5.76
DMKN	dermokine	5.67
AK095308	Homo sapiens cDNA FLJ37989	5.63
NUP205	nucleoporin 205kDa	5.4
CDC14B	cell division cycle 14 homolog B (S. cerevisiae)	5.4
AKAP11	A kinase (PRKA) anchor protein 11	5.37
TCIRG1	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 subunit A3	5.33
AK021962	Homo sapiens cDNA FLJ11900	5.22
FAM26E	family with sequence similarity 26, member E	5.21
SOX12	SRY (sex determining region Y)-box 12	5.21
KLK5	kallikrein-related peptidase 5	5.1
HSU79275	hypothetical protein LOC27105	5.05
KLK7	kallikrein-related peptidase 7	5.04
IQUB	IQ motif and ubiquitin domain containing	4.98
SPRR2C	small proline-rich protein 2C	4.77
BAGE	B melanoma antigen	4.73
LOC255783	hypothetical protein LOC255783	4.71

IGL@	immunoglobulin lambda locus	4.67
MPP6	membrane protein, palmitoylated 6 (MAGUK p55 subfamily member 6)	4.67
CSTA	cystatin A (stefin A)	4.58
LOC201181	zinc finger protein 385 C	4.58
HSPC159	Galectin-related protein	4.58
QRICH2	glutamine rich 2	4.56
NDOR1	NADPH dependent diflavin oxidoreductase 1	4.52
SFRS8	splicing factor, arginine/serine-rich 8 (suppressor-of-white-apricot homolog, Drosophila)	4.51
AK024255	Homo sapiens cDNA FLJ14193	4.49
CASQ1	calsequestrin 1 (fast-twitch, skeletal muscle)	4.39
SLC30A6	solute carrier family 30 (zinc transporter), member 6	4.38
KRTAP10-10	keratin associated protein 10-10	4.36
DUSP22	dual specificity phosphatase 22	4.32
TM4SF19	transmembrane 4 L six family member 19	4.31
NDRG4	NDRG family member 4	4.27
BC040863	Homo sapiens, clone IMAGE:5541055, mRNA	4.26
RNU2	U2 small nuclear RNA	4.25
FRMD7	FERM domain containing 7	4.23
SH3KBP1	SH3-domain kinase binding protein 1	4.23
PES1	pescafillo homolog 1, containing BRCT domain (zebrafish)	4.22
VWA3A	von Willebrand factor A domain containing 3A	4.18
NEUROD4	neurogenic differentiation 4	4.17
AK023412	Homo sapiens cDNA FLJ13350	4.1
LYPLA2	lysophospholipase II	4.09
EBF2	early B-cell factor 2	4.03
ATRN	attractin	4.02
AK098705	Homo sapiens cDNA FLJ25839	4.01
CRAT	carnitine acetyltransferase	4.01
EBP	emopamil binding protein (sterol isomerase)	3.95
SCGB1C1	secretoglobulin, family 1C, member 1	3.76
LOC201617	hypothetical protein LOC201617	3.71
ZNF230	zinc finger protein 230	3.66
AK094394	Homo sapiens cDNA FLJ37075	3.64
HOPX	HOP homeobox	3.61
ECM1	extracellular matrix protein 1	3.57
ACBD6	acyl-Coenzyme A binding domain containing 6	3.54
CNFN	cornifelin	3.5
TSSK1B	testis-specific serine kinase 1B	3.47
KRTAP19-5	keratin associated protein 19-5	3.4
NAT8	N-acetyltransferase 8	3.38
DCST2	DC-STAMP domain containing 2	3.35
HIATL1	hippocampus abundant transcript-like 1	3.35
ARHGAP23	Rho GTPase activating protein 23	3.27
CST7	cystatin F (leukocystatin)	3.26
OR6S1	olfactory receptor, family 6, subfamily S, member 1	3.25
HCG4P6	HLA complex group 4 pseudogene 6	3.22
CST2	cystatin SA	3.17
SLC25A11	solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier), member 11	3.17
TMEM150	transmembrane protein 150	3.16
BC043255	Homo sapiens cDNA FLJ32931	3.11
CA4	carbonic anhydrase IV	3.11
KRTAP19-3	keratin associated protein 19-3	3.07
C3orf1	chromosome 3 open reading frame 1	3.07
ZNF335	zinc finger protein 335	3.04
BGN	biglycan	3.03
CASP6	caspase 6	3.01
GSTO1	glutathione S-transferase omega 1	3.00
WDR69	WD repeat domain 69	3.00
MAP7D1	MAP7 domain containing 1	2.96
SLC39A10	solute carrier family 39 (zinc transporter), member 10	2.94
GJB3	gap junction protein, beta 3, 31kDa	2.94
LDLRAD2	low density lipoprotein receptor class A domain containing 2	2.93
PLEKHA2	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 2	2.92
FAM125A	family with sequence similarity 125, member A	2.9
LILRB4	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 4	2.89
ACSL4	acyl-CoA synthetase long-chain family member 4	2.89
RAB7B	RAB7B, member RAS oncogene family	2.77
KCNH6	potassium voltage-gated channel, subfamily H (eag-related), member 6	2.76
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	2.73
UBIAD1	UbiA prenyltransferase domain containing 1	2.71
RBKS	ribokinase	2.7
SPATA21	spermatogenesis associated 21	2.7
ITPKC	inositol 1,4,5-trisphosphate 3-kinase C	2.66

CG012	hypothetical gene CG012	2.66
C17orf37	chromosome 17 open reading frame 37	2.65
ANAPC2	anaphase promoting complex subunit 2	2.61
ODZ1	odz, odd Oz/ten-m homolog 1(Drosophila)	2.59
HERC1	hect (homologous to the E6-AP (UBE3A) carboxyl terminus) domain and RCC1 (CHC1)-like domain 1	2.58
ORC6L	origin recognition complex, subunit 6 like (yeast)	2.56
NLRX1	NLR family member X1	2.54
POMGNT1	protein O-linked mannose beta1,2-N-acetylglucosaminyltransferase	2.52
TMEM106B	transmembrane protein 106B	2.51
GMIP	GEM interacting protein	2.5
CRTC1	CREB regulated transcription coactivator 1	2.48
SAT1	spermidine/spermine N1-acetyltransferase 1	2.43
LYRM4	LYR motif containing 4	2.42
LOC96610	hypothetical gene LOC96610	2.41
FLJ31958	hypothetical protein FLJ31958	2.39
ISX	intestine-specific homeobox	2.38
TRD@	T cell receptor delta locus	2.37
LOC283483	hypothetical protein LOC283483	2.37
FOXL1	forkhead box L1	2.35
FHOD1	formin homology 2 domain containing 1	2.34
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	2.34
TRIM46	tripartite motif-containing 46	2.32
AANAT	arylalkylamine N-acetyltransferase	2.31
BCL2	B-cell CLL/lymphoma 2	2.29
PTS	6-pyruvoyltetrahydropterin synthase	2.29
FAM26F	family with sequence similarity 26, member F	2.28
PSMB8	proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional peptidase 7)	2.28
DSG1	desmoglein 1	2.27
SPTBN2	spectrin, beta, non-erythrocytic 2	2.27
ZNF585B	zinc finger protein 585B	2.26
SPRR2B	small proline-rich protein 2B	2.26
FLJ33590	hypothetical protein FLJ33590	2.25
PPFIA3	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein, alpha 3	2.24
HIST1H4B	histone cluster 1, H4b	2.23
C13orf33	chromosome 13 open reading frame 33	2.23
TMEM160	transmembrane protein 160	2.23
LRP6	low density lipoprotein receptor-related protein 6	2.22
DOCK2	dedicator of cytokinesis 2	2.22
PRDM15	PR domain containing 15	2.21
MVD	mevalonate (diphospho) decarboxylase	2.21
MGC39900	hypothetical protein MGC39900	2.21
C15orf21	chromosome 15 open reading frame 21	2.2
TFF2	trefoil factor 2 (spasmodic protein 1)	2.2
AK096857	hypothetical LOC646999	2.19
TMEM163	transmembrane protein 163	2.17
CDKL3	cyclin-dependent kinase-like 3	2.14
MYO15B	Myosin XVb pseudogene	2.14
LOC401152	HCV F-transactivated protein 1	2.12
BHLHB8	basic helix-loop-helix domain containing, class B, 8	2.11
CYLC2	cylicin, basic protein of sperm head cytoskeleton 2	2.1
AA776719	Soares_testis_NHT Homo sapiens cDNA clone 1292851	2.1
MAN1B1	mannosidase, alpha, class 1B, member 1	2.1
CDC42BPB	CDC42 binding protein kinase beta (DMPK-like)	2.08
LOC134111	FLJ25076 similar to CG4502-PA	2.06
TBC1D9B	TBC1 domain family, member 9B (with GRAM domain)	2.04
MRPL28	mitochondrial ribosomal protein L28	2.04
FLJ14100	hypothetical protein FLJ14100	2.04
BC036424	Homo sapiens cDNA clone IMAGE:4798730	2.03
ATP6V0A1	ATPase, H+ transporting, lysosomal V0 subunit a1	2.03
C20orf77	chromosome 20 open reading frame 77	2.02
EEF1E1	eukaryotic translation elongation factor 1 epsilon 1	2.00
COL17A1	collagen, type XVII, alpha 1	0.49
HS6ST3	heparan sulfate 6-O-sulfotransferase 3	0.37
KRT5	keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)	0.35

Supplementary Material & Methods (promoter analyses).

A set of 52 genes upregulated during keratinocyte differentiation was selected from the literature: *A2ML1*, *ABCA12*, *ASAH3*, *ASPRV1* (Saspase), *CASP14*, *CASR*, *CDSN*, *CLDN6*, *CRNN*, *CSTA*, *CTSC*, *CTSL2*, *DLX3*, *DMKN* (beta/gamma isoforms), *DSG1*, *EVPL*, *FABP5*, *FLG*, *FLG2*, *GRHL3*, *GSDM1*, *HRNR*, *IVL*, *KDAP*, *KLF4*, *KLK5*, *KLK7*, *KPRP*, *KRT1*, *KRT10*, *KRT2*, *LASS3*, *LCE1C*, *LIPK*, *LIPM*, *LIPN*, *LOR*, *POU2F3*, *RPTN*, *SBSN*, *SCEL*, *SERPINB5*, *SLPI*, *SPINK5*, *SPRR2G*, *SPRR4*, *TGM1*, *TGM3*, *TGM5*, *THHL1*, *TJP1*, *TLR9*. Promoter sequences of these genes for human and at least two other vertebrate species (mouse, rat, dog, cow, horse, occasionally armadillo or opossum, depending on sequence availability and conservation level) were retrieved 2300 bp upstream to 500 bp downstream from the transcription start site using the UCSC Genome Browser.¹ These sequences were analyzed with GEMS launcher tools from Genomatix Suite (Genomatix Software GmbH, Munich, Germany) as follows. For each gene, conserved transcription factor binding sites (TFBS) and modules (i.e., combinations of TFBS conserved in space and orientation) were searched using the Dialign/DialignTF and Frameworker tools, respectively.² When a given TFBS or module was found in a minimum of three different promoters, corresponding TFs were selected for qRT-PCR analyses.

1 Kent WJ, Sugnet CW, Furey TS et al. The human genome browser at UCSC. *Genome Res* 2002; **12**: 996-1006.

2 Cartharius K, Frech K, Grote K et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 2005; **21**: 2933-42.

Figure S2. Example of promoter analysis for DMKN- β - γ orthologs using Genomatix tools. (a) The UCSC Genome Browser shows a high level of conservation between the human promoter and those of other species, particularly for three conserved regions (CR, red lines). (b) Using the Frameworker tool, conserved combinations of transcription factor binding sites (colored thumbs) were identified within two of the most conserved regions. (c) Corresponding transcription factor families (from Genomatix Matrix Library) for CR2 and CR3 include known players of epidermal differentiation including ETS factors (V\$ETSF), AP1 and AP1-related proteins (V\$AP1F, V\$AP1R), as well as other TF families previously uncharacterized in the epidermis.

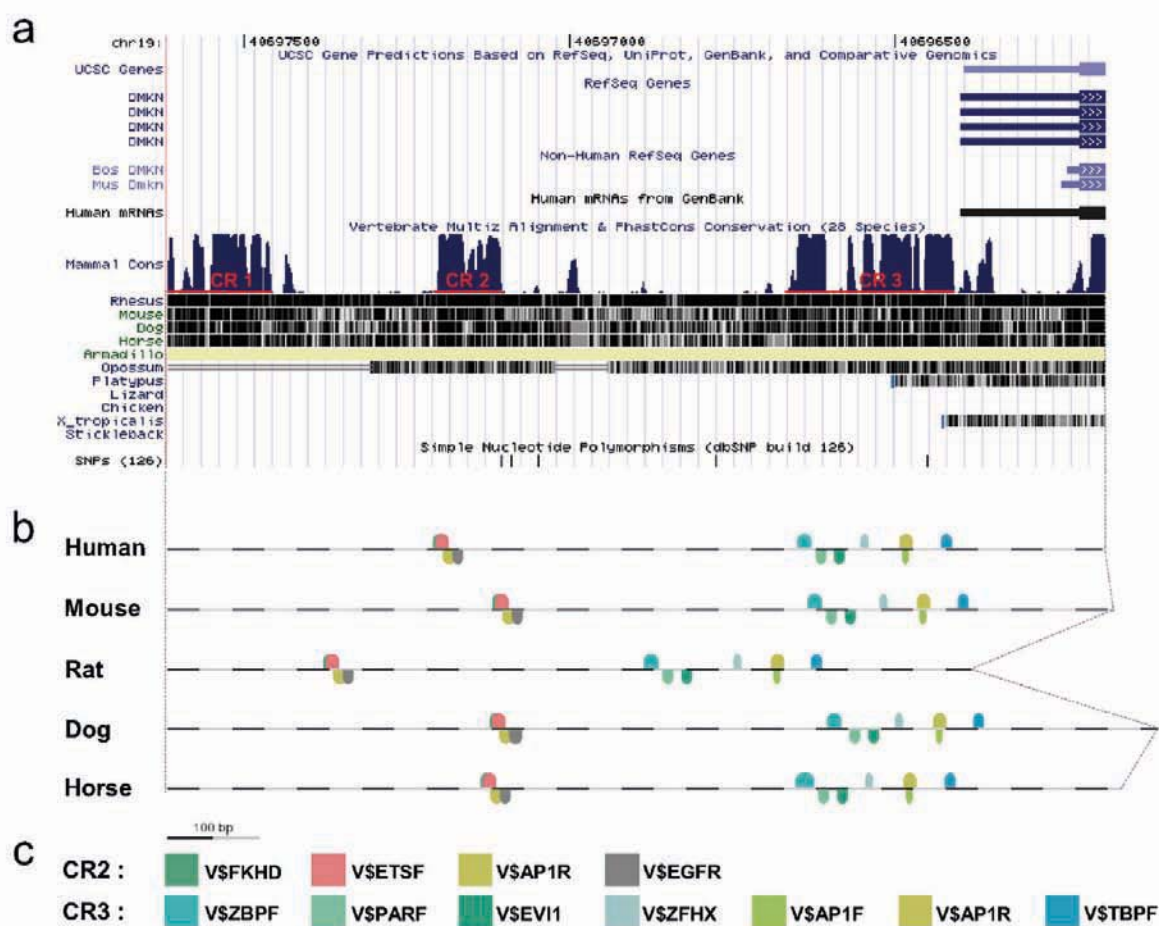


Table S2. Sequences of primers used for qRT-PCR essays.

Gene symbol	Forward primer	Reverse primer
<i>A2ML1</i>	CCCTGGTGAAGAAGGTTGAA	GTTGCCTGTTTCATCTGGTAGG
<i>AATF</i>	AGATAGGAAAGCCAGCAAAGG	GCCAAAAGAGAGCGGTACA
<i>ABHD11</i>	GCTCTTCGGCAGCAAACCTA	GTCCTGGCTCATGATCTCGT
<i>ACBD6</i>	CCATCAAATCGAAAAATGTGG	AGCTGTTTGGCCTTCATTGT
<i>ACSL4</i>	GGTTGACACTTTTGGCACAAC	GGAATTTCAAATCGCTCCAA
<i>ANKRD1</i>	AGAACTGTGCTGGGAAGACG	GCCTTCAAATGCCAGTGA
<i>APOF</i>	GGACTGTGTGGGTACTCTGCT	ACTTTGGGTGCAGAAATTGG
<i>AQR</i>	TGCCCCAGATGGCAA	CAGTCATGGCCTCTTCTTCTG
<i>ARID2</i>	TCCGACTAACAGCTGCCTTAATA	CATAAAGGCATTTGGCAAGG
<i>ARID5B</i>	CTCCAATCAAACCTCGGAAA	AGGCTCAGGGATGCTTTTCT
<i>ARL11</i>	CTAATAGTAACACAGAGGATTCAGCAGT	GTGGCCCTTCAGCTTGAAA
<i>ARSD</i>	ATGTGGAGGAGATGGACTGG	ACTGGTCTCTCCAGGGTTT
<i>ATF5</i>	CAGCCATGGAGTCTTCCACT	GGAGTGACATGGCTGTAGCA
<i>ATF6</i>	ACCACCCATAACAAGACCACA	CTGGTGCCATCACCTGACA
<i>ATF6B</i>	CCTCATTGAAACCGTCCAG	CTTCCAGGACCTCCTCTCCT
<i>ATP6V0A1</i>	CTGAGCGTGAAGAGCTTGG	CGGTCCCGCTGTAGAATTT
<i>AVEN</i>	CACAGAGGGGAACAGATTTCA	GGGCTCGAACCAATAACTCA
<i>B2M</i>	GAATTCACCCCCACTGAAAA	AGCAAGCAAGCAGAATTTGG
<i>BBOX1</i>	AAGATGAATCCAGGTGATGTGA	TTTCCATTCTCCACCCTCTG
<i>BCL11A</i>	GCCCCAAACAGGAACACATA	TAGCTGCTGGGCTCATCTTT
<i>BCL6</i>	CGGAGCCAGATTTGTACAGG	CTGGCTTTTGTGACGGAAAT
<i>BHLHB2</i>	GACCGGATTAACAGTGCAT	TGTCTGGAAACCTGAGCAGA
<i>BPTF</i>	TCCCCATTTTACCAGTGTGC	AGTTCAGACAACCAGATTCTTGC
<i>BRF1</i>	AGGAAGCTGTCACGAAGGAG	CTTGGGAGCAAAGCCTCTC
<i>BTBD11</i>	CATGGAGCTTCTGTCTGCTG	AGCTGCTTGAATGCTTCGTT
<i>CBFA2T2</i>	AGAGGATGCTTTCCTCGTCA	TGCAGGTTCTGACCACAGAG
<i>CCIN</i>	TCACCTTCCCCATTGAGTTC	ACTGAGGTGGTCTTGTGCT
<i>CDSN</i>	ACTGCTGCTGGCTGGTCT	AGAGCTTCTGGCAGTGGAAA
<i>CEBPZ</i>	GTGAGAGCGTTCCAGAACTTG	ATGGCCAAACTCTTCAGCAG
<i>CHD2</i>	TCACACAACCTTCACCCCTCA	TTGCCACCACCATAGTTGAA
<i>CLCA4</i>	GCCAATGTGACTGCTTTTATT	TCCATGAGCCCGAACTTTTA
<i>CLCA4 (alternative primer pair)</i>	GTAGTGAACGGGGAAATTGAAG	TATTGGTCAGGCAAGGGGAAG
<i>CNFN</i>	GCGATCCTGCAGCTATGTC	AGACAGGCATGTCGTTGCAG
<i>CNOT2</i>	TTTGGGGCAAGTTTATACGG	GTTGGTACCCTGTTGTTGG
<i>CRELD1</i>	GAGCTGAGTGAGGAGCTGGT	CTCTCTGTTCCCCAGGAC
<i>CST7</i>	TCTGCTGCCTGGTCTTGAG	GTATCTGGCTGCTTGGAGGA
<i>CSTA</i>	ACGGAAAATTGGAAGCTGTG	GCTAAAAGCCCGTCAGCTC
<i>CTNNBIP1</i>	ATGGGATCAAACCTGACAGC	GCTACTGCCTCCGGTCTTC
<i>D4S234E</i>	GATGGGTTTCGTCCTCAAGAA	ACTGACATCCAGGGCGATAC
<i>DDIT3</i>	AGCCAAAATCAGAGCTGGAA	CAGTGTCCCGAAGGAGAAAG
<i>DDX41</i>	TGTCCGTGAAAGAGCAGATG	TGAATCTTCTTCGGCATGGT
<i>DKFZP586H2123</i>	GGAGAGTGGTCAGGGAAACA	AAGGCCGCTGAGTATAGCTG
<i>DLX3</i>	CAGACGGTGAACCCCTACAC	TCACCGACACTGGGTCTCCT
<i>DMKN</i>	GACACGGAGCAGATGCTGT	ACGGAGTCCCCAGACCTC
<i>DSG1</i>	AAATGCAAGATCTGGGAGGA	TTAGATGGGCGTCTTTCATC
<i>EBP</i>	TTGCAGTGTGTGGGTTTATT	GATGGTTTCCATGCACACTG
<i>EGR1</i>	TGACCGCAGAGTCTTTTCTCCT	TGGGTTGGTCATGCTCACTA
<i>EHD4</i>	CAAGCTCAACGACCTCATCA	TTCAGCGAGTGGAAATTTGGT
<i>ETV5</i>	GGCATGGAGTTCAAGCTGAT	GCATCTGGGTACAGACAAAT
<i>EXT2</i>	CCTGAGTGACAGCCATAGA	TCACATTACAGCCTCCACCA
<i>FAM179B</i>	CCTTCAGCAAAAGGAAGACG	ACGAGGGACTGATTTTCTGGT

<i>FAR1</i>	TGAGCCTCTGCGTCTAGGAT	AGCTTTCTGCCTCACCAAAA
<i>FBXO38</i>	TGACCAATTGTGGAATCACAG	TCCATGTTTCAGTTTCTTGCAC
<i>FGFR1OP2</i>	TGGGAATCCAGCAAGAAAAC	TGGAGTGCTGCTCTTTAACTTC
<i>FLG</i>	TGAACAAGGTTACATTTATTGC	TGGATTCTTCAGGATTTGCC
<i>FLG2</i>	TCCAGTTCTGAAGAACCCAGAT	CTTTGACCCTGAAGCTTTGC
<i>FLVCR1</i>	ATTGGCTCAAGGAAAAGCTCA	CTTGGTCTGTGGGACTGTCA
<i>FOXC1</i>	AAGAAGGAAGGCGAGAGGAG	CAGGGCAGATCACCTAAGA
<i>FOXJ3</i>	CAGGTTTCAGTTTGCCGATCT	TGTTCTGAGTGCTGAGTG
<i>FOXK2</i>	TGGGTCTATCCTCCAGAACG	CTCACCCCTGAACGGGATTTA
<i>FOXL1</i>	TTCCACGACTGGGTTCTACC	TTTTTTGAGAGGTTAATGAAAAACA
<i>FOXP1</i>	GGCAGATCTCCTATGCAAGC	CCCCGATAGTCACTCCATGT
<i>GATA4</i>	GTGTCCCAGACGTTCTCAGTC	GGGAAGAGGGAAGATTACGC
<i>GJB5</i>	CAGGAGAGCCCTGAGGAGTA	TTGCAGTCGAAGTCCTTGTG
<i>GMIP</i>	ACCACAGTACCCTGGAGCAG	ACTGTGTCCCCAGGAGTGAG
<i>GRLF1</i>	CTGTGCTGACTCCAGAGAAGC	AGGTCCAGGTTGTGGTCTTG
<i>HDAC8</i>	CACCGAATCCAACAAATCCT	ACTTGATGCCCTTGGAAAT
<i>HIVEP3</i>	CACAGCAAAAAGTGCCAAGA	GTCTGAGTCCGAGTCTCCA
<i>HLTF</i>	CATGGCCCCAAAAGAAAAGAG	AGCACTGATCTTCAGCAGCA
<i>HOPX</i>	TCAACAAGGTCGACAAGCAC	TCTGTGACGGATCTGCACTC
<i>HOXA9</i>	CCCATCGATCCCAATAACC	ATTTTCATCCTGCGGTTCTG
<i>HSPC159</i>	ATTCCAGACCAGCCATTCAG	TGGCACGTGATCCTATTTGA
<i>HYAL2</i>	ACCACAAGCACGGAGACCT	CAGGAAGGTAAGTGGCACTGG
<i>ING4</i>	CGTGCGCACAAGTCCTG	TGGAACCACTCAATGGAACA
<i>INSM1</i>	GGAACCCCCACTTTTACGTT	CCAGACCGCATTTCTAGACC
<i>IQGAP3</i>	CCGCCTCTCACTTCAGAAAC	TGACAGCCACACCCTCATA
<i>IRF2</i>	AGTCCCATCTGGACAGCAAC	TGGGCACACTATCAGTCGTT
<i>IRF7</i>	GAGTCTTCTTCCAAGAGCTGGT	CCAGGGTCCAGCTTCAC
<i>ISL1</i>	CAGCAACTGGTCAATTTTTCA	TTCCCACTTTCTCCAACAGG
<i>ITPKC</i>	ACCTCTACCCTGGGCTTCC	CCGTCTTGCCGAAGTCTATC
<i>IVL</i>	TGCCTCAGCCTTACTGTGAGT	TGGGTATTGACTGGAGGAGG
<i>KLF10</i>	CATCTGTAGCCACCCAGGAT	ATCAGAACGGGCAAACCTC
<i>KLF13</i>	AGCACAAAGTGCCACTACGC	AACTTCTTCTCGCCCGTGT
<i>KLF14</i>	ACCAAGGCACAGACTGGTTC	ACGGGAGTTCATCCCTTCTT
<i>KLF4</i>	CCCACACAGGTGAGAAACCT	ATGTGTAAGGCGAGGTGGTC
<i>KLK5</i>	AGTCAGAAAAGGTGCGAGGA	TGAACTTGAGAGGTTCTGTG
<i>KLK7</i>	CAGGACTGCACGAAGGTTTACA	CCCCCTGAGTCAACCATTGC
<i>KRT23</i>	AGAAGCATCGAGACTTGGACA	CTTCTCCAGGTGGGTTTTGA
<i>LDB1</i>	AACTACCTCCGACTCTGTGTGATAC	TGGGCTCTTCTTCTGCTGT
<i>LLGL1</i>	GGGACGTCACAGTGGAAGAT	GATAGTTCTGGCCTTCCATTT
<i>LOC57228</i>	GTGACAATGACCAGCCTCCT	AGAGCAGGGTGAGGAAGACA
<i>LOR</i>	GTGCTTTGGGCTCTCCTTC	GAGTAGCCGCAGCCAGAAC
<i>LRBA</i>	GACAGAGGAGTGGTCTGTGGT	CCACCGGTTAAAGTCGTTGT
<i>MAFK</i>	AAGGCATTAAAGGTCAAGAAGGA	GTAGCCGCGGTTCTTGAGT
<i>MAGED1</i>	GACATCCCCCTCCTTGAGAT	TTTTCTGAACCTCTGCAATGAAT
<i>MAML1</i>	GGCAACAGCACCTTCTCG	GGATGGACCCAAGGTGTTT
<i>MAP3K7IP1</i>	CATGACCCTGCTAGTGAGGA	TGGACTGCAGGGTTAAGGTC
<i>MAPT</i>	AAGGTGACCTCCAAGTGTGG	AGGTCAGCTTGTGGGTTTCA
<i>MAX</i>	GAACGAAAACGTAGGGACCA	TGCTGGTGTGTGGTTTTT
<i>MAX</i>	GAACGAAAACGTAGGGACCA	AAGGTGTGGCATTCTGCAT
<i>MAZ</i>	CCTTCAAATGTGAGAAATGTGAG	CTGTGCACCTTCATGTGGTC
<i>MBD1</i>	TACCAGAGCCCCACAGGA	TTGAAGGCTTCTTTTCGCTTC
<i>MBD2</i>	ATGAATGAACAGCCACGTCA	GCACTGGCAACAGCAGATAA
<i>MBD3</i>	ACTAGCACCATGCCCATCAC	TTCTGCTTCTGATGTCCT
<i>MBTPS1</i>	AGAAGCGTGAGCTGGTGAAT	TCTATGTAGCTGGGGCTCAAA

<i>MED4</i>	TTCCACTAATGGCGTGAATG	CCAGGAGGTTCCAACAAAA
<i>MVD</i>	TGGCGGCAGTCACTTGTA	TGAAGTCCTTGCTGATGACG
<i>NCOA6</i>	GGCCTCTGAAAGTGTGGAAA	ATCAAGTCGCAGTCTGCTT
<i>NCOR2</i>	CAAGCTGACCGAGAGCAACT	ACCGCCTGGCTTCTATAGGT
<i>NDOR1</i>	GGGCCAGACTGGAACTTCT	CCGGTGCTGCACATATACTTT
<i>NFE2</i>	GAGCTGCAGGGTCTGAATG	GGAGCTCATAAGGTGGTGA
<i>NFE2L1</i>	CTTCCCTGCACAGTTTCCA	GGGCACTGTACAGGATTTCACT
<i>NIPBL</i>	GCACTTCAGGGTCAATTGAGAA	GTCCAGGAGCCACTGTAGGA
<i>NMI</i>	TCACGTTTGTGGAGATTGGA	CCTGTCAGAAGCACTGTCCTC
<i>NR1D1</i>	CACCTTTGAGGTGCTGATGG	GAAGCGAATTCTCCATGC
<i>NR2C1</i>	GAGTCTGCCTCCAGACTGCT	CATTCATCACTTGCCAGCAC
<i>NR4A1</i>	GCTTCTTCAAGCGCACAGT	CGGACAACCTTCTCCACCAT
<i>NRBP1</i>	GAGATTCAGGGCAATGGAGA	AATGCTGGGTGGAACAGAAG
<i>ODZ1</i>	TCGGGATTGGTGACGTTTAT	GCTGGCTGTGTAGGAGAAGC
<i>OTUD7B</i>	ATGGCAGAAGGAGTGGAAATG	TGAGCAAGGACAAAGACGTG
<i>OTX1</i>	CCAAGACTCGCTACCTGAC	GAGCTAGAGGACGAGGCAGA
<i>OXER1</i>	ATCATCTTTGGCATGGCTTC	GGAGGAAGTTGGGGCTAGAG
<i>PAPOLG</i>	AAGAGATCCCATTCCCCATC	TGACACGGATGTTGTTTGCT
<i>PELP1</i>	GTGAGCAGTCAGCATTGGAA	AAAAGCCAGCCCCTAAAGAG
<i>PHF2</i>	TGCCAGAGTCAAGAAGGAGAG	AATGGCTTCTGTGTGCTG
<i>PHF7</i>	GCTATGCCTTCGAGAACCCTG	ATCCATGGAAGCCTCTGTTG
<i>PKNOX1</i>	CATCCCTACCCAACAGAGGA	TTTGGGGTCTCTGAACAAC
<i>PLA2G15</i>	CTGGCTTCAGGAGACAACAA	CCTGGAAGAACTTGCGGTAG
<i>PLEKHA2</i>	CGTGAAGCAAGGGAATGTG	GCTGTGCATGTCTTCTGGACT
<i>POU6F1</i>	CCATCTGCCGGTTCGAG	CCCACAACTCCATCAGGTT
<i>PRDM15</i>	ACGTTCAAGCTCACACACGAC	TCCTCTTCTGGCTTCTGCTT
<i>RASGRP1</i>	AGCCCAGTCTTGGTCAGAAA	GAACAGTCACCCTGCTCCAT
<i>RCAN1</i>	ACTTCAGCAACCCCTTCTCC	GCTGCGTGCAATTCATACTTT
<i>RFX5</i>	GGCTIGCTGAAGAGGACGA	CGGGCTTCCAGATCCTTAG
<i>RORA</i>	ACACCAGCATCAGGCTTCTT	GGTCTGCCACGTTATCTGCT
<i>RREB1</i>	CTCCCACACAGGGGAAAG	GCTGTGGGTGGACTCATTCT
<i>SCEL</i>	CCCAAGGATGGATATCAGGA	GCAAGTAGAATGGCAGCAAA
<i>SCLY</i>	TGGGGCACAAAGTTTATGGT	GCAATCATTGGGGTGTCTC
<i>SCP2</i>	TTAGGCATTGGAGGAGCTGT	TGATTGTGCAGTCAGCCTTC
<i>SCP2</i>	TCAAAGGGACACCCACTAGG	ACAGCTCCTCCAATGCCTAA
<i>SERPINB5</i>	CTACCCAAGGATGTGGAGGA	CAAGCCTTGGGATCAATCAT
<i>SERPINB8</i>	ATGCGGATGAGGTACACACC	TCTGAATTTGTCCAGGCTTTG
<i>SEZ6L2</i>	CCTCTACGTGGAGCTGCTG	ACTCAGGGTCCGTGGTAGTG
<i>SH3KBP1</i>	GAAGGACCAGCAGAAACGAG	GGGGAAGATTCTCCTCTTGG
<i>SIM1</i>	GATCCATGGGCGAGGTC	AATGGAAGCCAGTTGGTCTG
<i>SKIV2L</i>	CCCTCAAGCAGGGAATAGAAC	CGGGCCCACTCATATAACAAC
<i>SLC30A6</i>	AGACAACACCACCCCATGTT	CCATTTGTTCAATTGGCATCTC
<i>SLC35B2</i>	TGGTACCTGGCTACCTCCTG	ACTCTTCTCCTGCAGCACACC
<i>SLC39A6</i>	GCTCTAGGAGGCCTGTATTTCA	TTCTCCTCATTGTTGAAAGTTGA
<i>SLC8A1</i>	CACCATTGGCCTGAAAGATT	TTCCCAGGAAGACATTCCACC
<i>SLTM</i>	GTGGATCACAGCACTATCCTGA	GCTTAAATCCGGAGCCACTT
<i>SOX12</i>	ATAGCCCCTTTTCCGGTATG	TCCCTTCTCTACTGCTGGA
<i>SPINK5</i>	AGGGAAGTGTGAGGAGAGCA	TCCATTCCCAAAGCTCTGT
<i>SSRP1</i>	AGGAGAATGCCAATGACAGC	CTTCTCATCCCGTCACTGT
<i>SUV39H1</i>	ATTCGCAAGAACAGCTTCGT	ACACGTCTCCACGTAGTCC
<i>TCFL5</i>	CATCCAGGAAAGACATGGAGA	CACAGCCATTTGTGATTTGC
<i>TCFL5</i>	CGCATTTGCTGTGATGAGTT	ACCGCAAATACGCTCTCAA
<i>TCIRG1</i>	CCGCTACCAGGAGGTCAAC	GCCCCTGAAGAAAGTCTGC
<i>TEF</i>	GAGTCTGCCAGCTCTCCAC	TCAGGTCCTCCTCAGCAAAC

<i>TFAP2E</i>	ACGTCTGTGAGACGGAGTTC	CTGAGCCATCAAGTCTGCAA
<i>TFCP2L1</i>	CAGCACATCCACCGAGTCTA	ACTGCTGCTCAGAGTCCACA
<i>TLE1</i>	CCTACCATTCCCTCAAACCTG	GGTTGAGGGTGTTGATCTGG
<i>TMEM45A</i>	TGCATGGGATCTGATGGAT	TCCAATTCTGAGGAGCAGAG
<i>TRIM29</i>	ACTACACCTCCCGGGTCTG	ACGTAGAATGGCCGGTAGTG
<i>TSPYL5</i>	CCCAATCCCTTGAGTTCTA	ATCAGTTGGATTGGCTCACC
<i>UBP1</i>	AGCAGTGCAAGCGAGAATG	TGTGGATGCCATCACTACTTTC
<i>VWA1</i>	TGACCTGCACATCATTGTCC	CAGCTCCAGCACATAGTAGCC
<i>ZBTB7B</i>	ATTCACCAGGAACGACAAGC	CAGTTCTGGCCTTTGAGGTG
<i>ZEB2</i>	TCCTGTCTGTCTCGAAAAA	GCCTTGAGTGCTCGATAAGG
<i>ZFHX3</i>	AGCAGCAAGGGATGTTTGAC	CAAGTTCGGCTTAGGAGACG
<i>ZFPM1</i>	GTGAAGAAGGAGCCAGCAGA	TCTTTGGGCTTCTCGTCTGT
<i>ZFPM2</i>	GCTTCTATTTTGCCACAGC	ACAGGCTGGAAATCTGATGG
<i>ZHX2</i>	GGGAGGAGGATGCGATCT	CGTTGATCAGCACAGCAGTT
<i>ZNF143</i>	CACAATGGTAACGCAGGATG	CCCATTTCTGATGAGGCAGT
<i>ZNF230</i>	CCTGCGGTTGCTACATAACC	ATGGAGCACAAAGGAAGGTG
<i>ZNF239</i>	TCCTCCTGGCTAGTTTTGGA	CACACTGTCCCTGTTTCTGG
<i>ZNF239 (alternative primer pair)</i>	AGCCCATTCAGGTCTTCACA	CCCATCTGAGAAACAAGAACAA
<i>ZNF335</i>	CCGAGCTGAGATGGAGAGTC	CAAAGCTGCACTGGCTACAC
<i>ZNF396</i>	AGCATCATGGGAGCTTCAGT	TCCAAGAAGGGTTTCCTTGTC
<i>ZNF76</i>	GCCAGAAGCCAACCTTCATGT	ATCCTCGAGCTGGATCACC

III. DISCUSSION

Technique des ORESTEs

La technique des ORESTEs repose sur l'utilisation d'amorces non spécifiques, arbitrairement choisies afin d'amplifier une sous-population d'ARN indépendamment de leur abondance relative. La stratégie employée vise à produire un ensemble de minibanques afin d'aboutir à une couverture maximale du transcriptome. Au cours de l'étude, plus de 150 minibanques ont été produites au total, correspondant à 22 585 clones, chaque amorce produisant entre 100 et 200 clones.

Dans un premier temps, 2200 clones ont été produits afin d'optimiser la procédure. Les premiers clonages produisant 10% de vecteurs vides, une stratégie de criblage LacZ/X-gal a été employée afin de réduire le nombre de clones vides à moins de 5%.

Une première série d'ADNc clonés dans des vecteurs pCRII-TOPO par clonage A/T a été séquencée par le service commun de l'IFR30, puis les 2200 séquences ont été annotées manuellement. Une des stratégies employées a été de cartographier les séquences en les implémentant sur le Genome Browser, un outil extrêmement utile pour visualiser l'organisation du génome et la structure des gènes, développé par l'Université de Santa Cruz, Californie (UCSC). L'intégration des séquences dans ce navigateur permet ainsi d'identifier aisément les gènes correspondants s'ils sont déjà décrits, les éventuels variants d'épissage tissu-spécifiques, les séquences générées dans d'autres tissus, etc. L'ajout de cette piste personnalisée (« custom track ») dans ce navigateur est rapidement devenu un moyen aisé de consulter les résultats de façon ciblée (famille de gènes, région génomique, séquence particulière) et a été mis à disposition de la communauté (<http://genome.ucsc.edu/cgi-bin/hgTracks?org=human&hgt.custom> Text=<http://udear.cnrs.free.fr/orestes.txt>).

Par la suite, 22 000 clones supplémentaires ont été produits et séquencés à grande échelle grâce à un financement en réponse à un appel d'offre du Génoscope d'Evry. Le volume des données imposant une analyse informatique, une collaboration avec le Centre de BioInformatique de Bordeaux (CBiB) a permis de réaliser l'assemblage des séquences (contigs), ainsi qu'une première annotation et une conversion des séquences en coordonnées génomiques pour l'intégration au Genome Browser.

Afin de réduire le nombre de clones correspondant à des ARN ribosomiques, deux cycles successifs de purification des ARN totaux sur billes magnétiques couplées à des amorces oligo(dT) ont été effectués. La proportion finale de séquences correspondant à des ARNr est de 8%, contre 7,8% pour les données produites par le consortium brésilien à l'origine de la méthode (Camargo et al. 2001). Les ARN d'origine mitochondriale représentaient quant à eux près de 10% du total.

Au total, les 24 200 séquences produites ont permis d'identifier, après élimination des séquences contaminantes (ADN génomique, ARN ribosomique, vecteur...), 3387 gènes différents, avec une moyenne de 4,6 séquences par gène, soit une redondance assez faible, montrant bien le pouvoir normalisateur intrinsèque de cette méthode. Cette redondance est encore plus faible (3 séquences par gène) pour les 2200 premières séquences produites à petite échelle. En effet, les clones étaient systématiquement criblés par PCR et électrophorèse, afin que les clonages générant un grand nombre de fragments de même taille, potentiellement redondants, soient écartés. De plus, le séquençage progressif des minibanques permettait de ne pas analyser en totalité les plus redondantes. Toutefois, la faible redondance indique probablement que cette description du transcriptome n'est pas exhaustive, suggérant un nombre total de gènes exprimés plus important (cf. conclusion de la discussion).

Malgré la normalisation intrinsèque de la technique des ORESTEs et l'élimination avant le clonage des produits d'amplification préférentielle (par l'excision des éventuelles bandes visibles après migration sur gel), quelques rares gènes sont représentés plus fortement, 9 d'entre eux dépassant les 100 séquences. C'est le cas notamment de trois gènes connus pour être fortement exprimés par le kératinocyte granuleux (la kératine 1, la filaggrine, et la protéine d'enveloppe LEP7), de deux gènes codant pour des protéines ubiquistes (la protéine ribosomique RPS8 et le facteur d'élongation EEF1A1), mais aussi de gènes peu caractérisés tels que l'ifapsoriasine (filaggrine 2), l'inhibiteur de protéases à sérine SERPINA12 ou la tétraspanine 5.

Parmi les gènes découverts et jamais caractérisés auparavant dans l'épiderme, deux ont fait l'objet d'une étude plus poussée. Ainsi, le gène *DMKN* codant pour la dermokine, représenté par 217 ORESTEs, a été identifié chez la souris en tant que membre d'un locus de trois gènes (*KDAP*, *DMKN* et *SBSN*) codant pour des protéines sécrétées dans les couches suprabasales de l'épiderme, le SSC (Stratified epithelium-

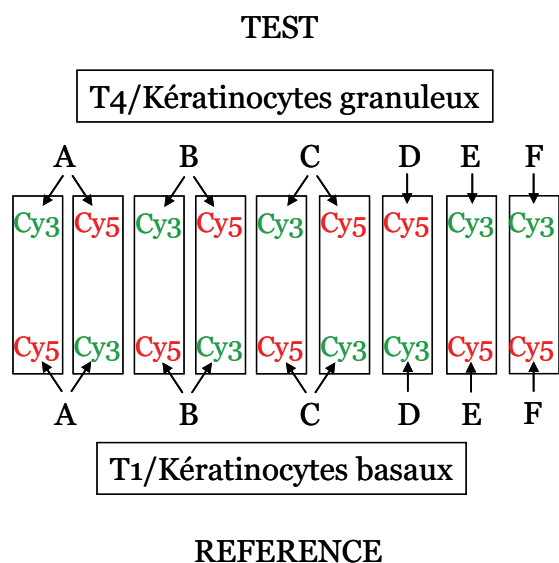
Secreted protein gene Complex) (Matsui et al. 2004). Le gène *DMKN* possède trois promoteurs distincts, deux exons terminaux différents, et les messagers subissent un épissage alternatif complexe (Toulza et al. 2006). Parmi les quatre classes de transcrits identifiées (*DMKN* α , β , γ et δ), les isoformes β et γ sont spécifiques du kératinocyte granuleux, alors que l'isoforme α est exprimée dans toutes les couches de l'épiderme et dans le placenta, tandis que l'isoforme δ est ubiquiste.

Un autre gène, *A2ML1*, a fait l'objet d'une caractérisation au sein de notre équipe. Identifié à partir de six séquences d'ORESTE, le gène code un inhibiteur de protéase de la famille des α 2-macroglobulines, dont l'expression dans l'épiderme est restreinte à la couche granuleuse (Galliano et al. 2006). Des transcrits ont été détectés par PCR sur des ADNc commerciaux préparés à partir de placenta, de thymus et de testicule, bien que la protéine soit en dessous du seuil de détection dans des extraits protéiques commerciaux de placenta et de testicule. D'autre part, parmi les 70 EST présents dans les banques de données, 5 proviennent de placenta, 7 de testicule, et les autres ont été clonés en grande majorité à partir de tissus exprimant un programme de différenciation proche de celui de l'épiderme comme la langue, l'oesophage, l'hypopharynx ou des carcinomes épidermoïdes.

Puces à ADN

Pour poursuivre le travail de description du transcriptome des kératinocytes granuleux et de sa régulation, une approche par puces à ADN a été choisie, principalement pour deux raisons. Premièrement, aucune méthode à grande échelle n'est réellement exhaustive, et la méthode des ORESTEs, bien que capable d'identifier des gènes très faiblement transcrits, n'a certainement pas couvert la totalité du transcriptome de la couche granuleuse de l'épiderme. De plus, cette approche n'est ni quantitative ni comparative, et ne discrimine donc pas les gènes spécifiques du kératinocyte granuleux, de ceux exprimés plus précocement, voire dans toutes les couches.

Il a donc été décidé d'employer des puces à ADN dites pangénomiques, afin de pousser plus loin la description du transcriptome des kératinocytes granuleux, et d'identifier de nouveaux gènes spécifiquement exprimés dans cette couche. La stratégie choisie a été de comparer uniquement deux des quatre fractions récupérées au moyen du procédé d'enrichissement décrit auparavant (Toulza et al. 2007). Les fractions cellulaires T1 (correspondant aux cellules de la couche basale) et T4 (constituée de la couche granuleuse et cornée) ont été choisies pour être comparées sur puces. Ainsi, la fraction T4 sert d'échantillon test, tandis que la fraction T1 du même individu sert d'échantillon de référence afin de discriminer les gènes présents dans toutes les couches de ceux surexprimés au cours de la différenciation, voire spécifiques de la couche granuleuse. Pour les besoins de l'étude, six échantillons indépendants ont été utilisés, soit six couples T1-T4. Le marquage en fluorescence des ADNc, même indirect (après la réaction de transcription inverse), est susceptible d'introduire des biais selon le fluorochrome employé (cyanines Cy3 ou Cy5) ; il est donc nécessaire, autant que possible, d'équilibrer les marquages, c'est à dire d'avoir un nombre équivalent de marquages « Cy3 vs Cy5 » et « Cy5 vs Cy3 ». Le plan expérimental choisi est le « dye-switch », qui consiste à inverser le « sens » du marquage pour chaque nouvel individu (soit pour notre étude, trois lames « T1-Cy3/T4-Cy5 » et trois « T1-Cy5/T4-Cy3 »). Toutefois, pour trois des six couples T1-T4, des lames « swap » ont été réalisées. Le dye-swap consiste à réaliser systématiquement, à partir des mêmes ADNc, les deux marquages possibles. Leur cohybridation en parallèle permet ainsi d'obtenir deux lames « miroir » ou lames « swappées », soit « T1-Cy3/T4-Cy5 » et « T1-Cy5/T4-Cy3 » pour le même couple T1-T4. Si le dye-swap permet en toute rigueur d'équilibrer les marquages pour chaque



échantillon, il requiert une quantité d'ARN deux fois plus importante ; de plus, les échantillons étant non-indépendants deux à deux, l'approche est moins performante que le dye-switch d'un point de vue statistique (à nombre de lames égal). Au final, notre étude a donc été réalisée à partir de six individus et neuf lames (dont trois couples de lames swappées – voir figure ci-contre). Pour cette raison, les marquages ne sont pas tout à fait équilibrés

(quatre T1-Cy3/T4-Cy5 contre cinq T1-Cy5/T4-Cy3) ; toutefois, une comparaison entre les résultats incluant ou excluant la lame non équilibrée montre qu'il n'y a pas de déséquilibre majeur lié à la prise en compte de cette lame. Les analyses finales ont donc été réalisées à partir de neuf lames. Le plan expérimental final étant donc un mélange de dye-switch pour trois lames, et de dye-swap pour les six autres. Les lames swappées ont été fusionnées deux à deux afin d'être analysées comme trois lames virtuelles indépendantes. L'analyse finale repose donc sur six mesures indépendantes correspondant aux six échantillons.

Le choix des puces pangénomiques employées s'est porté sur des lames à oligonucléotides produites pour le programme Resogen. Soutenu par un consortium franco-britannique RNG (Réseau National des Génopoles) – MRC (Medical Research Council), le programme a été lancé pour permettre le développement et la commercialisation à prix coûtant de puces pangénomiques (humain et souris), afin que la communauté scientifique puisse disposer d'un outil d'analyse à grande échelle à un prix bien plus abordable que les offres commerciales disponibles auparavant. Les lames humaines, produites en France par trois plateformes (Nice/Sophia Antipolis, Evry et Strasbourg) comprennent 25 248 oligonucléotides (50mères) différents (pour un total de 26 496 dépôts/lame), ciblant 22 458 gènes connus, prédits ou ESTs du génome humain (Le Brigand et al. 2006; Le Brigand and Barbry 2007). Les quantités d'ARN de qualité optimale se sont avérées suffisantes pour ne pas avoir recours à une étape supplémentaire d'amplification in vitro, alternative intéressante lorsque les ARN sont limitants mais susceptible d'introduire des biais

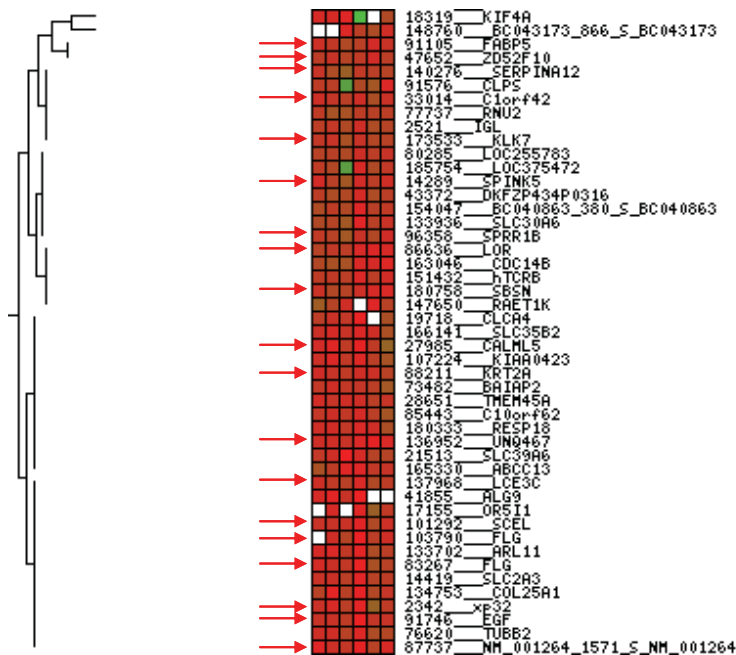
supplémentaires. Les lames ont été hybridées manuellement en chambre d'incubation et bain-marie à 50°C.

L'analyse des neuf lames réalisées a permis une détection significative (supérieure au niveau moyen du bruit de fond + 2 fois l'écart-type, sur au moins 4 des 6 échantillons) pour 11 520 sondes, soit 10 824 gènes différents. La première analyse statistique réalisée sur ces données (test *t* de Student) a révélé que 624 gènes étaient surexprimés dans la fraction T4 (ratio > 1,5) avec une p-value inférieure à 0,05, contre 141 sous-exprimés (ratio < 0,66 ; p < 0,05). Le biais clairement visible entre le nombre de gènes surexprimés et sous-exprimés peut s'expliquer par le fait qu'au cours du processus de dissociation et d'enrichissement en kératinocytes granuleux, une partie des kératinocytes basaux et épineux restent accrochés aux lambeaux épidermiques, ceux-ci se repliant parfois sur eux-mêmes et devenant inaccessibles à la trypsine. Pour cette raison, la fraction T4 (~KG) est partiellement contaminée par des kératinocytes basaux, alors que la fraction T1 (~KB) n'est pas contaminée par des kératinocytes granuleux, ce qui explique que les ratios des gènes surexprimés dans les KG sont généralement plus élevés que ceux des gènes sous-exprimés (*cf.* fig S1 *in* Mattiuzzo et al. 2009).

Les résultats de puces se sont avérés relativement cohérents en ce qui concerne des marqueurs déjà décrits de la différenciation, avec 31 gènes ayant un ratio supérieur à 1.5, dont 19 supérieurs à 4 (*CDSN*, *KDAP*, *LCE2B*, *SCEL*, *CALML5*, *LCE5A*, *LOR*, *FLG*, *SBSN*, *LCE3C*, *KRT2*, *FABP5*, *SPINK5*, *LCE4A*, *NICE-1*, *DMKN*, *SPRR1B*, *KLK7* et *KLK5*). Toutefois, de nombreux autres marqueurs sont non détectés par les sondes correspondantes (*A2ML1*, *GJB5*, *IVL*, *KLF4*, *KRT10*, *TGM1*, 3 & 5...), non différentiellement détectés (*KRT1*, *RORA*...) ou n'ont pas de sonde correspondante (ex : *FLG2/IFPS*). A l'exception de *KLF4*, ces marqueurs sont parfaitement quantifiables en PCR quantitative, avec des ratios parfois bien plus importants (ex : *LOR* affiche un ratio de 7,3 sur puces et de 3735 en PCRq) (Mattiuzzo et al. 2009). De plus, le test en PCR quantitative de 103 des 624 gènes potentiellement surexprimés n'a permis de confirmer les résultats que pour une faible proportion des candidats (17 sur 54 gènes quantifiables, soit 31%), le reste se répartissant entre gènes non-régulés (26 sur 54, 48%) voire sous-exprimés (11 sur 54, 20%). Dans le même temps, une approche purement bioinformatique (voir plus loin) a conduit au choix de 88 gènes codant pour des facteurs de transcription. Cette

approche a permis de quantifier 43 de ces gènes par PCRq, avec 9 surexprimés (21%) et 8 sous-exprimés (19%), soit des proportions assez proches de celles obtenues avec les 103 gènes précédents. Ce résultat est surprenant, les candidats sélectionnés à partir de données de puces étant supposés surexprimés dans leur grande majorité, alors que l'approche bioinformatique n'est que prédictive. Certaines contradictions peuvent provenir de biais dans la quantification par PCRq (ce point est discuté dans la section suivante). Toutefois, le degré de validation des données de PCRq (avec notamment 19 marqueurs de la différenciation confirmés par cette approche, sur 20 testés) va plutôt dans le sens de faux positifs dus à la technique de puces, plutôt que d'une quantification erronée par PCRq.

Plusieurs démarches ont été testées pour tenter de s'affranchir de ces faux positifs. Ainsi, le seuil de p-value a été abaissé de 0,05 à 0,01, améliorant la proportion de vrais positifs dans une faible mesure (passant de 31% de gènes confirmés à 47%), mais augmentant considérablement le nombre de faux négatifs (parmi les 624 gènes considérés comme surexprimés ($p < 0,05$) seuls 238 gènes ont une p-value inférieure à 0,01). Pour tenter d'améliorer la discrimination entre vrais et faux positifs, une approche alternative par clustering hiérarchique a été employée, le but étant ici d'identifier des gènes qui auraient un profil proche de celui des marqueurs connus de la différenciation. L'analyse par clustering hiérarchique ou par Self-Organizing Tree Algorithm a permis de montrer qu'un grand nombre de marqueurs connus étaient regroupés dans les mêmes branches du graphe (voir ci-



contre le détail d'un des deux principaux clusters, regroupant à lui seul 19 marqueurs de la différenciation – flèches rouges). Toutefois, là encore cette approche permet d'augmenter la proportion de vrais positifs (44%), mais augmente considérablement le nombre de faux négatifs. L'approche finalement choisie est celle du tri des gènes en

fonction du FDR (taux de fausse découverte), qui a pour avantage d'être plus stringent que le test de Student sans écarter de l'analyse un nombre trop important de gènes. Toutefois, le problème ne se situe pas dans la méthode d'analyse des données, mais dans les mesures elles-mêmes, et aucune des approches n'a permis de discriminer avec efficacité les vrais des faux positifs.

Si les données issues de PCR quantitative sont considérées comme des références, les ratios de puces non confirmés peuvent trouver plusieurs explications propres à la technique et au matériel biologique employé. La fiabilité des données de puces a pu être diminuée par la variabilité inter-échantillons liée à la fois aux différences inter-individuelles et aux fluctuations des niveaux d'enrichissement des fractions T1 et T4 en kératinocytes basaux et granuleux respectivement. Toutefois, cette variabilité concerne essentiellement des gènes qui ont une p-value très élevée au test *t* de Student ; or plusieurs ratios de gènes avec une p-value largement inférieure au seuil de 0,05, ou même 0,01, n'ont pu être validés par la PCRq. L'explication la plus plausible reste donc celle de quantifications biaisées (de façon reproductible) par des hybridations croisées. La spécificité *in silico* de certaines des sondes concernées semblant correcte, il est probable que le problème vienne en partie d'une stringence trop faible lors de l'hybridation. A ce jour, seuls quelques articles présentent des données issues des puces du RNG (Moreilhon et al. 2005; Chalabi et al. 2007; Lyons et al. 2007), ce qui est faible comparé aux 3500 puces et 100 projets concernés selon les seuls chiffres de 2006 (Le Brigand et al. 2006). Plusieurs publications ne concernent des versions « à façon » ne contenant que 1700 sondes (Hofman et al. 2007; Puissegur et al. 2007) ; de plus, les résultats ne sont parfois pas présentés (Colell et al. 2007). Une seule étude présente des résultats validés par une autre méthode (PCR quantitative) (Doghman et al. 2007). Ceci rend globalement difficile l'évaluation de la qualité de ces puces pangénomiques.

Au vu de ces résultats, l'emploi de puces pangénomiques s'est avéré certes décevant en tant que méthode présentée initialement comme semi-quantitative, mais malgré tout intéressante dans une approche de criblage en association avec d'autres méthodes, 27 des 49 nouveaux marqueurs de différenciation identifiés au cours de cette étude apparaissent surexprimés d'après les données de puces, dont 19 identifiés uniquement d'après cette méthode. Il n'est d'ailleurs pas rare depuis quelques années de voir la technique des puces à ADN présentée comme méthode de criblage pour rechercher des candidats plutôt que comme méthode de quantification de

l'expression génique, comme par exemple pour l'analyse du triple KO involucrine/envoplakine/périplakine (Sevilla et al. 2007).

Analyse in silico des promoteurs de gènes coexprimés

L'approche bioinformatique par recherche de TFBS a permis de sélectionner un grand nombre de candidats non identifiés par les autres méthodes, mais n'a pas rendu possible la reconstitution, même partielle, des réseaux de régulation mis en jeu. Les raisons potentielles sont multiples. Premièrement, la découverte d'un site de fixation pour un facteur de transcription dans une séquence régulatrice (ou supposées telle) ne constitue pas une preuve de l'implication du facteur en question, mais seulement une prédiction, un moyen d'augmenter la probabilité d'identifier un candidat potentiel pouvant expliquer la régulation du gène étudié. Inversement, l'absence de modèles (combinaisons de TFBS) nouveaux ou connus dans les promoteurs de gènes coexprimés n'implique pas nécessairement une absence de corégulation. Les gènes corégulés peuvent être notamment sous la dépendance d'une région régulant un locus ou LCR (Locus Controlling Region) pouvant agir sur de grandes étendues (ex : le locus des β -globines) ou simplement de régions régulatrices distantes (« enhancers » ou « silencers »). De plus, certaines combinaisons de TFBS peuvent être jugées non significatives par le logiciel si l'agencement des sites varie trop. L'outil FrameWorker de Genomatix, par exemple, ne considère que les modèles où l'espacement et surtout l'ordre et l'orientation des sites restent relativement conservés, alors qu'un agencement différent peut aboutir à une même régulation. Il est enfin possible que les gènes coexprimés ne soient pas réellement corégulés, mais soumis à des combinaisons différentes de facteurs de transcription présents dans le même type cellulaire. Au final, la prédiction de sites de fixation peut donc se heurter à la faible spécificité de certains facteurs de transcription, à la complexité de leurs interactions, et à la difficulté d'élaborer des logiciels tenant compte de l'ensemble des cas décrits en matière de régulation génique.

Pour pousser plus loin la caractérisation des réseaux de régulation impliqués dans le programme de différenciation kératinocytaire, il semble donc nécessaire de compléter cette première description par une approche expérimentale, par exemple en identifiant les cibles directes des FT par ChIP-on-chip ou ChIP-Seq. Aucune technique à grande échelle n'étant exhaustive ou exempte de faux positifs, de nombreux allers-retours entre approches à grande échelle et approches ciblées

restent nécessaires pour dresser un portrait aussi complet que possible. De plus, les approches par ChIP sont également susceptibles d'apporter des informations sur les liens complexes entre les facteurs de transcription, les réseaux d'une grande complexité impliquant généralement que des FT se régulent les uns les autres. Ainsi, des travaux menés sur *S. cerevisiae* (le modèle eucaryote le mieux décrit au niveau transcriptionnel) montrent que les gènes codant pour des FT sont régulés de manière plus fine que d'autres gènes au niveau transcriptionnel (Balaji et al. 2006). Ils présentent notamment un degré de corégulation plus important et un nombre de FT par gène régulé supérieur, ce qui n'est pas surprenant au vu de leur rôle crucial de « chefs d'orchestre » de la différenciation.

PCR quantitative

Plusieurs gènes mis en évidence par la technique des puces à ADN n'ont pu être confirmés en PCR quantitative. Il convient de distinguer ceux qui n'ont pu être quantifiés de façon fiable, et ceux qui affichent des ratios contradictoires. Pour les gènes non quantifiables, il peut s'agir d'un problème de sensibilité de la PCR quantitative et/ou d'amorces peu adaptées. Même si la qRT-PCR est théoriquement très sensible, certains gènes identifiés formellement par la méthode des ORESTES n'ont jamais pu être quantifiés, même en choisissant différents couples d'amorces. Il est probable que plusieurs candidats sélectionnés à partir des données de puces n'aient pu être quantifiés pour la même raison. D'une façon globale, sur 298 candidats sélectionnés (toutes sources confondues), 216 (72,5%) ont pu être amplifiés par PCR classique, et 157 par qRT-PCR (+ confirmation par séquençage), la spécificité et/ou la sensibilité étant insuffisante pour les autres gènes.

Il est également possible que certains gènes « détectés » sur puces ne soient pas réellement exprimés : afin de ne pas analyser les dépôts dont la fluorescence n'est pas significative, un seuil a été arbitrairement fixé pour chaque lame, à une valeur qui est égale au bruit de fond + 2 x écart-type. La distribution des intensités pour le bruit de fond obéissant à une loi normale, près de 98% des dépôts négatifs sont ainsi théoriquement éliminés de l'analyse, avec un risque limité d'écarter des dépôts réellement positifs. Il est cependant possible que les signaux les plus faibles, bien qu'étant significativement supérieurs au seuil fixé, correspondent à des hybridations croisées, et que les gènes ciblés ne soient donc pas exprimés.

Le cas des ratios contradictoires est plus complexe. Il est évidemment possible d'évoquer pour partie des erreurs de PCR quantitative. Il convient toutefois de noter plusieurs éléments.

Premièrement, pour la totalité des gènes quantifiés de façon fiable en PCRq (157 gènes « nouveaux » dans l'épiderme, plus 20 marqueurs connus), les amplicons issus de la PCRq ont été systématiquement validés par séquençage, afin de s'assurer de la spécificité de l'amplification. Pour tous, l'amplicon séquencé correspondait à l'amplicon prédit par PCR *in silico*.

Deux autres paramètres peuvent venir fausser les ratios de PCR quantitative dans des proportions variables : le choix du gène de ménage, et l'efficacité des amorces. Les valeurs de PCRq devant être normalisées par rapport aux concentrations des ADNc, il convient de choisir un (voire plusieurs) gène(s) « de

ménage » dont l'expression est la plus stable possible. Dans un premier temps l'ARN ribosomique 18S et l'ARNm de la galectine 7 (*LGALS7*) ont été utilisés comme gènes de normalisation. Toutefois, ils ont rapidement été abandonnés, le 18S étant exprimé trop fortement par rapport aux gènes à quantifier (d'où des problèmes de dilution), tandis que *LGALS7*, initialement décrit comme exprimé dans toutes les couches, s'est avéré légèrement surexprimé au cours de la différenciation. Sept autres gènes de ménage ont été testés par PCRq sur l'ensemble des fractions T1 à T4, et les résultats obtenus ont été analysés avec l'outil geNorm (Vandesompele et al. 2002) afin de déterminer quel gène possède l'expression la plus stable. Il ressort que le gène *B2M*, codant la β 2-microglobuline du CMH de classe I, est le gène de référence le plus adapté. Les biais de normalisation ne sont réellement gênants que pour les ratios très proches de 1, or la plupart des nouveaux marqueurs de la différenciation que nous avons identifiés ont des ratios supérieurs à 4, voire bien plus pour certains gènes comme *KRT23* (R=745). Ces éventuels biais sont donc négligeables et ne remettent pas en cause la surexpression des gènes identifiés.

L'efficacité des amorces est également importante pour toute quantification relative, la formule employée ($2^{\Delta\Delta Ct}$) n'étant qu'une version simplifiée de la formule exacte $(1+E)^{\Delta\Delta Ct}$, où E est l'efficacité de la PCRq. La formule simplifiée nécessite que l'efficacité des amorces du gène testé et du gène de référence soient toutes deux proches de 1 (100%). Dans le cas contraire, plus l'efficacité E est faible et l'écart de Ct élevé, plus le ratio calculé est surestimé, parfois de façon très importante. La table ci-dessous donne le pourcentage d'erreur suivant la valeur de E et du $\Delta\Delta Ct$, donné par la

$$\text{formule } 100 \times \frac{2^{\Delta\Delta Ct} - (1 + E)^{\Delta\Delta Ct}}{(1 + E)^{\Delta\Delta Ct}}.$$

		Efficacité						
		99%	95%	90%	85%	80%	75%	70%
$\Delta\Delta Ct$	1	1%	3%	5%	8%	11%	14%	18%
	2	1%	5%	11%	17%	23%	31%	38%
	3	2%	8%	17%	26%	37%	49%	63%
	3,32 (R = 10)	2%	9%	19%	30%	42%	56%	72%
	4	2%	11%	23%	37%	52%	71%	92%
	5	3%	13%	29%	48%	69%	95%	125%
	6	3%	16%	36%	60%	88%	123%	165%
	7	4%	19%	43%	73%	109%	155%	212%
	8	4%	22%	51%	87%	132%	191%	267%
	9	5%	26%	59%	102%	158%	233%	332%
10 (R ~1000)	5%	29%	67%	118%	187%	280%	408%	

Erreur $R_{\text{calculé}}$: <10% 10-25% 25-50% 50-100% >100%

Ainsi, dans le cas d'une efficacité faible (70%) et d'un écart de Ct élevé (10 cycles), le ratio calculé selon la formule simplifiée est de 1024, alors qu'en prenant en compte l'efficacité il n'est que de 201, soit une erreur de 408%. Au contraire, pour une efficacité correcte (95%) et un écart de Ct similaire, l'erreur reste inférieure à 30% (formule simplifiée : $R=1024$, f. exacte : $R=795$).

Des expériences de dilutions limites ont montré que l'efficacité moyenne de nos amorces pour les gènes *B2M*, *KRT14* et *KLK7* est de 96 %. Pour des raisons pratiques, l'efficacité des amorces des autres gènes n'a pas été évaluée, même si le parallélisme des courbes d'amplifications permet de supposer qu'elles ont une efficacité similaire (90-95%). Généralement, les amorces ayant une efficacité plus faible correspondent également à des amorces peu spécifiques ou dont les courbes d'amplification ne sont pas dans la gamme de cycle voulue (17-30 cycles), donc écartés de l'analyse. Au vu de l'efficacité moyenne de nos amorces, l'erreur liée à l'efficacité reste relativement faible pour les gènes que nous avons quantifiés, et ne remet pas en cause leur surexpression, même si le ratio peut être surévalué.

Au final, les quantifications en PCRq ont été réalisées en prenant en compte tous les biais possibles et paraissent d'un niveau de fiabilité satisfaisant. Toutefois, bien que la plupart des gènes impliqués dans la fonction barrière soient surexprimés à la fois au niveau de l'ARNm et de la protéine, une régulation post-transcriptionnelle est envisageable pour certains gènes. De plus, l'expression de ces nouveaux gènes peut débuter plus ou moins tardivement au cours de la différenciation. Ainsi, deux protéines testées en immunohistochimie ont montré des profils différents, avec MAPT (protéine Tau associée aux microtubules ; $R=6.1$) progressivement surexprimée au cours de la différenciation, et SH3KBP1 (une protéine adaptatrice potentiellement impliquée dans le trafic vésiculaire et/ou la signalisation ; $R=64$) restreinte à la couche granuleuse (Mattiuzzo et al. 2007). Chaque candidat identifié appelle donc à une validation du profil au niveau protéique, par exemple par immunohistochimie ou immunofluorescence.

Nouveaux acteurs de la différenciation

Les différentes approches discutées précédemment ont permis d'identifier un nombre important de gènes jamais décrits dans l'épiderme et surexprimés au cours de la différenciation. L'expression de chacun de ces gènes n'étant décrite qu'au niveau transcriptionnel, ces données appellent dans un premier temps une validation au niveau protéique, par immunohistochimie ou immunofluorescence, voire par microscopie électronique pour une localisation subcellulaire. Selon la fonction putative de ces gènes, plusieurs approches sont ensuite envisageables pour la caractérisation fonctionnelle de ces protéines. Des expériences de gain ou de perte de fonction, soit dans des modèles *in vitro* – kératinocytes en condition différenciantes ou épiderme reconstitué + siRNA ou transfections – soit *in vivo*, préférables pour étudier la fonction barrière – souris transgéniques, knock out... – sont possibles pour les gènes d'intérêt. Dans le cas de facteurs de transcription, les expériences de gain ou de perte de fonction ne permettent pas de distinguer les cibles directes des cibles indirectes lorsque l'on identifie des gènes différentiellement exprimés. Or l'identification des cibles directes, si possible *in vivo*, est d'un intérêt élevé pour comprendre leur fonction qui peut être différente d'un tissu à un autre. Pour cette raison, les expériences d'immunoprécipitation de chromatine suivie de puces à ADN ou de séquençage pour identifier les séquences génomiques fixées sont de plus en plus employées systématiquement, et constitue une approche de choix pour une partie des facteurs peu décrits que nous avons identifiés.

Nous avons choisi de discuter le rôle potentiel dans l'épiderme de certaines des protéines correspondant aux gènes identifiés dans nos deux études de transcriptome (Toulza et al. 2007 ; Mattiuzzo et al. 2009). Certains des gènes déjà présents dans la discussion du second article ne sont pas discutés ici. Dans tous les cas, le ratio R correspond au ratio KG/KB en qRT-PCR.

- Protéines de la matrice extracellulaire

Bien que la plupart des protéines de la matrice extracellulaire soient localisées au niveau de la membrane basale, la matrice extracellulaire est également impliquée dans la stratification, la différenciation kératinocytaire ainsi que la cicatrisation. Le protéoglycan membranaire Syndecan-1 et la glycoprotéine sécrétée ECM1b sont des exemples de protéines de la matrice surexprimées au cours de la différenciation (Sanderson et al. 1992; Smits et al. 2000). En plus du gène *EXT2* (non régulé) jouant

un rôle dans la synthèse des protéoglycans, nous avons également identifié *HYAL2* ($R=2.6$), un gène codant une enzyme capable d'hydrolyser l'acide hyaluronique. L'acide hyaluronique est un constituant majeur de la matrice extracellulaire, mais est absent de la couche granuleuse (Tammi et al. 1988). De plus, le gène *HAS1* (hyaluronan synthase 1) est surexprimé dans la couche basale de l'épiderme (Radoja et al. 2006). Ces données suggèrent un gradient opposé de synthèse et de dégradation de l'acide hyaluronique dans les couches de l'épiderme. La protéine *HYAL2* est généralement localisée au niveau des lysosomes et/ou de la membrane plasmique (selon le type cellulaire), et possède un pH optimal acide (pour revue : (Lepperdinger et al. 2001)), suggérant que l'enzyme pourrait être sécrétée par les kératinosomes (souvent comparés à des lysosomes sécrétoires) et dégrader l'acide hyaluronique présent dans la matrice de la couche granuleuse. Bien que le rôle de l'acide hyaluronique dans l'épiderme reste peu défini, sa capacité à retenir l'eau, médier des interactions intercellulaires ou encore lier des protéines extracellulaires tels que les facteurs de croissance, suggère un rôle dans la formation d'un environnement favorable pour la différenciation kératinocytaire (<http://www.glycoforum.gr.jp/science/hyaluronan/HA04/HA04E.html>). De plus, les produits de faible poids moléculaire issus de la dégradation de l'acide hyaluronique possèdent une large gamme d'effets, notamment l'induction de l'expression de hBD2 (β 2-defensine), un peptide antimicrobien naturellement présent dans la couche granuleuse (Ali et al. 2001; Gariboldi et al. 2008). Ces données suggèrent que de nouveaux composants ou modificateurs de la matrice extracellulaire restent à découvrir dans les couches suprabasales de l'épiderme. De part leurs multiples fonctions, notamment dans la séquestration et/ou la présentation de nombreux facteurs diffusibles, leur identification est primordiale pour comprendre le rôle joué non seulement dans la différenciation épidermique, mais également dans le cadre de processus tels que la cicatrisation et l'inflammation.

- Jonctions cellulaires et protéolyse

En raison de leur implication dans la cohésion tissulaire, les protéines de jonctions ainsi que les protéases et inhibiteurs de protéases, susceptibles de moduler leur durée de vie, jouent un rôle crucial dans la fonction barrière de l'épiderme.

Le séquençage des ORESTEs a permis d'identifier une nouvelle claudine, *CLDN23*, fortement surexprimée au cours de la différenciation ($R = 25$). Cette

claudine a également été identifiée comme spécifique des couches suprabasales de l'épiderme murin (April and Barsh 2006). Toutefois, son importance dans la formation des jonctions serrées épidermiques est inconnue.

La protéine placentaire P11, initialement décrite comme une protéase à sérine sécrétée, est fortement surexprimée dans les kératinocytes granuleux ($R = 35$) (Toulza et al. 2007). Si la protéine possède effectivement un peptide signal suggérant qu'elle est bien sécrétée, elle a été récemment décrite comme une endoribonucléase (Laneve et al. 2008), en accord avec la présence d'un domaine conservé pfam09412 (endoribonuclease XendoU). La protéine possède également un domaine somatomédine B semblable à celui de la vitronectine impliqué dans la liaison de l'inhibiteur de protéase PAI-1/SERPINE1. Une implication de la protéine P11 dans la desquamation ou dans un autre processus reste donc indéterminée. Une seconde protéase à sérine sécrétée, PAMR1, impliquée dans la régénération musculaire (Nakayama et al. 2004), s'est avérée sous-exprimée au cours de la différenciation ($R = 0,3$; Mattiuzzo et al. 2009), suggérant un rôle dans un processus autre que la desquamation.

Trois nouveaux inhibiteurs de protéases à sérine, SERPINA12 ($R = 50$), SERPINB7 ($R = 10$) et SERPINB12 ($R = 15$), ont été identifiés à partir des ORESTES, mais leur fonction dans la régulation de la desquamation est toujours inconnue. La serpine A12 est décrite comme une adipokine impliquée dans l'obésité et le diabète de type 2 (Kloting et al. 2006) ; la serpine B7 est fortement exprimée au niveau du glomérule par les cellules mésangiales et aurait pour cible potentielle la plasmine (Inagi et al. 2003) ; enfin, la fonction de la serpine B12 n'a jamais été explorée. Le gène *CST7*, codant pour la cystatine F/leucocystatine, un inhibiteur potentiel de la cathepsine C (Hamilton et al. 2008), est également exprimé mais ne semble pas régulé au cours de la différenciation ($R = 0,8$; Mattiuzzo et al. 2009).

La découverte de ces nouvelles protéines structurales, enzymes et régulateurs tend à prouver que les acteurs impliqués dans la régulation de la cohésion intercellulaire et intercornéocytaire, pourtant essentielle à la fonction du tissu, ne sont que partiellement décrits à ce jour.

- Protéines structurales

Parmi les protéines structurales nouvellement décrites figurent l'ifapsoriasine (IFPS). Situé sur le locus de l'EDC (1q21.3) entre les gènes *FLG* et *CRNN* (cornuline), le gène *IFPS* a été renommée *FLG2* en raison de la composition en acides aminés de la protéine prédite proche de celle de la filaggrine. Comme les autres membres de la famille *fused* à laquelle il appartient (composée des gènes *FLG*, *TCHH*, *RPTN*, *HRNR* et *CRNN*), le gène *FLG2* est composé de trois exons, comportant un cadre ouvert de lecture de grande taille (>7kb) qui code pour une protéine prédite de 2391 acides aminés (contre 4061 pour la filaggrine), riche en sérine (22%), glycine (20%), histidine (10%) et glutamine (10%). La protéine prédite comporte plusieurs domaines caractéristiques des protéines *fused* : un domaine N-terminal S100, suivi d'un domaine EF-hand de liaison au calcium, puis un large domaine constitué de deux séries de répétitions de 75 à 77 acides aminés (9 répétitions de type A et 14 de type B) (figure 64). Si la protéine prédite comporte dans son ensemble une faible homologie avec les autres protéines *fused*, les répétitions de type A sont toutefois

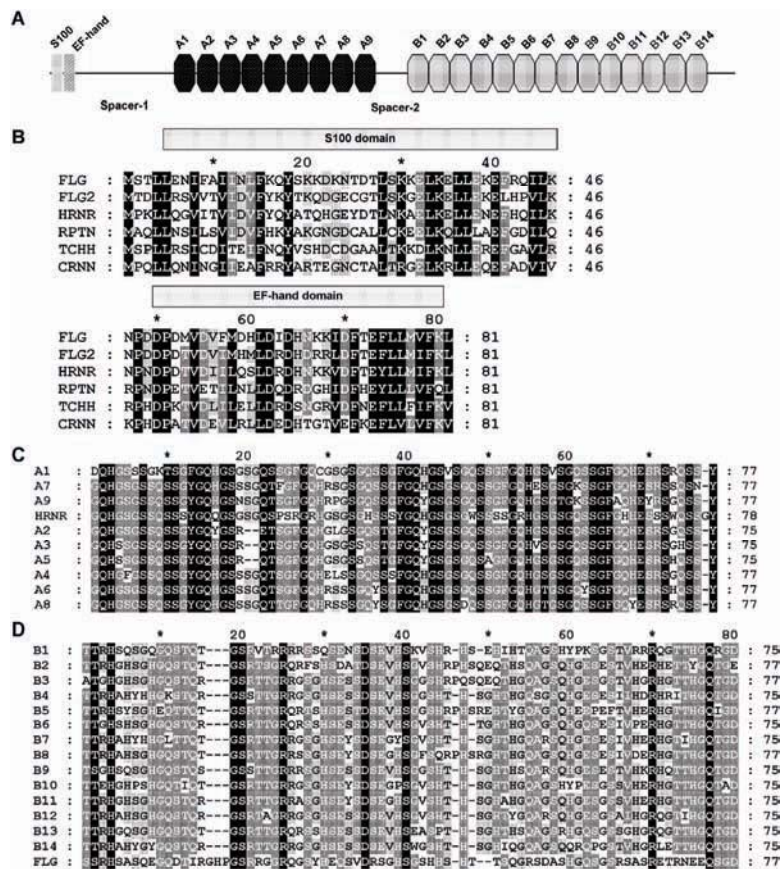


Figure 64. Structure du gène *FLG2/IFPS* (Wu et al. 2009).

fortement homologues de l'hornerine (50 à 77% d'identité), tandis que les répétitions de type B sont plutôt apparentées à la filaggrine (28 à 39% d'identité) (Wu et al. 2009). Enfin, l'expression et la localisation subcellulaire de l'ifapsoriasine rappelle fortement de celle de la filaggrine, puisqu'elle est détectée au niveau de la couche granuleuse dans les granules de kératohyaline, et de la couche cornée dans la matrice fibreuse.

La résistance mécanique est un aspect important de la fonction barrière. Nous avons identifié une nouvelle kératine acide épidermique, K23, fortement surexprimée dans les kératinocytes granuleux ($R=745$). Initialement décrite dans une lignée de cellules cancéreuses traitées au butyrate de sodium (Zhang et al. 2001), l'expression de *KRT23* est principalement détectable dans le placenta d'après les données de la banque UniGene (EST ProfileViewer Hs.9029). Des mutations des gènes *KRT1*, *KRT2* et *KRT10* étant responsables de diverses genodermatoses incluant des épidermolyses bulleuse et des kératodermies, ces données suggèrent une implication du gène *KRT23* dans des génodermatoses d'étiologie inconnue, même si une description plus poussée de l'expression de K23 dans l'épiderme reste nécessaire, par exemple par immunohistochimie.

En plus de l'ifapsoriasine et de la kératine 23, un autre gène codant une protéine structurale pourrait jouer un rôle au cours de la différenciation épidermique. Le gène CCIN ($R = 5,6$) code pour la calicine, composant principal du calice post-acrosomal de la thèque périnucléaire des spermatozoïdes (coque protéique entourant la région postérieure du noyau) (Lecuyer et al. 2000). En raison de sa capacité à se lier aux filaments d'actine et à participer à la rigidification de cette structure, il serait intéressant d'étudier le rôle potentiel de cette protéine au cours de la cornification. Il faut toutefois noter que sa composition en acides aminés (seulement 9,1% de glutamine ou lysine) et l'absence de séquence consensus n'en font pas un bon candidat pour l'action des transglutaminases.

- Lipides impliqués dans la fonction barrière

Les kératinocytes granuleux sont considérés comme des cellules sécrétoires en raison du grand nombre de corps lamellaires renfermant de nombreux lipides. Sécrété au pôle apical de la cellule en cours de cornification, le contenu de ces vésicules contribue à la formation de la barrière épidermique. Parmi les candidats que nous avons identifiés figurent des enzymes impliquées dans la synthèse et la

modification des lipides, particulièrement du cholestérol et des sphingolipides, ainsi que des protéines de transport.

- Synthèse et dégradation

Deux gènes impliqués dans la synthèse de novo du cholestérol (15-20% des lipides sécrétés) ont été détectés dans notre seconde étude (Mattiuzzo et al. 2009) : le gène codant la diphosphomevalonate decarboxylase (*MVD*, $R=4,8$) et celui codant l'emopamil-binding protein, une stérol isomérase (*EBP*, $R=1,5$). De nombreuses mutations du gène *EBP* sont associées à la chondrodysplasie ponctuée dominante liée à l'X de type 2 (CDPX2, OMIM 302960), une maladie humaine caractérisée notamment par une hyperkératose striée et des défauts capillaires (Derry et al. 1999). *MVD* code pour l'enzyme ultime de la voie du mévalonate (en amont d'*EBP*) dans la synthèse des stérols, et contrairement à *EBP* est surexprimé dans les kératinocytes granuleux. Bien que jamais décrit dans l'épiderme, *MVD* semble être un bon candidat pour des génodermatoses liées à un défaut de synthèse du cholestérol. Il est intéressant de noter que Radoja et ses collaborateurs ont décrit la surexpression dans les kératinocytes granuleux de plusieurs autres gènes liés à la synthèse des stérols, tels que *SC4MOL*, *DHCR7* et *OSBPL8*, tandis que d'autres gènes (*IDI1* et *HMGCR*) sont surexprimés dans la fraction suprabasale, mais pas spécialement dans les kératinocytes granuleux. Au final, ces données montrent que les gènes impliqués dans le métabolisme du cholestérol ne sont pas nécessairement tous surexprimés, ou peuvent l'être à différents stades de la différenciation.

En plus du cholestérol, les lipides sécrétés sont composées de 20 à 30% de sphingolipides, essentiellement des céramides qui sont soit néosynthétisés à partir d'acyl-CoA à longues chaînes et de sérine, soit issus de la dégradation d'autres sphingolipides. De façon intéressante, deux gènes impliqués en amont de la néosynthèse des céramides, *ACBD6* ($R=1$) qui code pour une protéine liant les acyl-CoA à longues chaînes, et *ACSL4* ($R=0.9$) qui est responsable de la production de ces acyl-CoA à longues chaînes, sont détectés indifféremment des couches de l'épiderme (Mattiuzzo et al. 2009). Toutefois, *LASS3*, une des enzymes impliquées dans la synthèse des céramides à partir de ces acylCoA et de sérine, est surexprimée avec un facteur 10 au cours de la différenciation (Toulza et al. 2007). Deux études postérieures ont confirmé ce résultat, indiquant que *LASS3* est bien la céramide synthase principale des kératinocytes différenciés (Mizutani et al. 2006; Mizutani et al. 2008). Les céramides peuvent également être produits par dégradation de la

sphingomyéline. Radoja et ses collaborateurs ont décrit la surexpression très faible dans les kératinocytes granuleux de *SMPD1* (ratio épiderme total / kératinocytes épineux = 1,8), qui code pour une sphingomyélinase, tandis que nous avons identifié *SMPD3* comme étant très fortement surexprimé au cours de la différenciation, avec un ratio KG/KB de 70 (Toulza et al. 2007), suggérant une implication de *SMPD3* (plutôt que de *SMPD1*) dans la production de céramides à partir de sphingomyéline. *SMPD3* est notamment connu pour être présent dans les kératinosomes (Rassner et al. 1999), et son expression est diminuée dans le cas de la dermatite atopique (Jensen et al. 2004). Au final, nos données suggèrent que la synthèse des céramides au niveau de la couche granuleuse implique plusieurs des voies métaboliques connues, mais aussi que certaines des enzymes impliquées, telle *ACSL4*, ne sont pas nécessairement régulées au cours de la différenciation. Enfin, il est intéressant de noter que *ASAH3*, un gène fortement surexprimé ($R = 300$; Toulza et al. 2007) codant pour la céramidase alcaline *hCER1* qui dégrade certaines espèces de céramides à longue chaîne, a récemment été impliqué directement dans la différenciation épidermique, probablement en raison de l'activation de voies de signalisation liées à la sphingosine et la sphingosine-1-phosphate (Sun et al. 2008).

De façon surprenante, nous avons également identifié deux gènes, *SLC22A5* (66 ORESTEs) et *BBOX1* ($R = 5,2$), respectivement impliqués dans le transport et la synthèse de carnitine. La carnitine est un composé impliqué dans le transport des acides gras jusqu'à la matrice mitochondriale, en vue d'une dégradation via la bêta-oxydation. La production d'énergie dans les kératinocytes reste peu décrite, et si des travaux anciens indiquent que la bêta-oxydation se produit bien au sein de l'épiderme (Anastasia and Conley 1977), il est peu probable que ce processus joue un rôle direct dans la fonction barrière.

- Transport

Bien que n'étant pas ou très peu surexprimés, deux gènes identifiés par notre étude, *SCP2* ($R=1,8$) et *SCPx* ($R=1,2$), sont susceptibles de jouer un rôle important dans le transfert intracellulaire de lipides, et plus particulièrement de cholestérol. *SCP2* and *SCPx* sont issues de l'usage de deux promoteurs alternatifs de la même unité de transcription. *SCP2* code pour un transporteur facilitant le transfert intermembranaire de stérols, augmentant notamment la sécrétion du cholestérol dans les sécrétions biliaires (Gallegos et al. 2001). Il est tentant de spéculer sur le rôle de *SCP2* dans le transfert du cholestérol vers les kératinosomes, même s'il semble que

le transporteur est également capable de lier les céramides (Atshaves et al. 2007). Le rôle de *SCPx*, qui code pour une thiolase de la matrice des peroxisomes, semble plus difficile à définir.

Plusieurs gènes semblent également impliqués dans le transport extracellulaire de lipides au sein de l'épiderme. La surexpression du gène *APOF* dans les kératinocytes granuleux ($R=41$) est particulièrement intéressante, d'autant que *APOD* et *APOE* ont également été décrits comme surexprimés dans les couches suprabasales (Radoja et al. 2006). *APOD* et *APOE* sont des apolipoprotéines respectivement impliquées dans les échanges de cholestérol via l'activation de l'enzyme responsable de la formation d'ester de cholestérol (plus hydrophobe), la lécitine cholestérol acyltransférase (*LCAT*), et dans le transport de lipides sous forme de chylomicrons (cholestérol et triglycérides). *APOF* est une apolipoprotéine mineure du plasma qui bloque le transfert des esters de cholestérol et des triglycérides vers les lipoprotéines, normalement médié par la Cholesteryl ester transfer protein (*CETP*; (Paromov and Morton 2003)). Le gène codant la *LCAT* est majoritairement transcrit dans la couche basale de l'épiderme de singe Rhesus (Smith et al. 1990), tandis que l'expression ou non du gène *CETP* reste inconnue. La fonction précise de chacune de ces protéines dans la biologie de l'épiderme reste à élucider. Toutefois, l'existence d'une machinerie probablement complexe régulant la disponibilité, la modification, le transport et peut être la clairance de plusieurs lipides sécrétés par les kératinocytes granuleux, particulièrement le cholestérol et ses dérivés, pourrait être un élément important de la régulation de la fonction barrière.

- Maturation et dégradation

Le cholestérol est partiellement sulfaté dans les couches suprabasales de l'épiderme, et le sulfate de cholestérol (2 à 6% des lipides intercornéocytaires) est ensuite converti en cholestérol par les arylsulfatases du stratum corneum. Nous avons confirmé l'expression du gène *ARSD*, codant pour une arylsulfatase, dans l'ensemble de l'épiderme ($R=1.4$). *ARSD* réside dans le même locus (Xp22.32) que deux autres gènes apparentés, *ARSE* et *STS* (*ARSC1*), impliqués respectivement dans la chondrodysplasie ponctuée dominante liée à l'X de type 1 (*CDPX1*, OMIM 302950) et l'ichtyose liée à l'X (OMIM 308100). En raison de sa proximité avec *ARSE*, il n'est pas exclu que *ARSD* soit également impliqué dans la *CDPX1* (Franco et al. 1995). *ARSD* étant exprimé dans la plupart des territoires anatomiques alors que *ARSE* n'y est pas détectable par RT-PCR (Dooley et al. 2000), une déficience en *ARSE* explique

difficilement les signes d'ichtyose des personnes atteintes de *CDPX1*, faisant de *ARSD* un meilleur gène candidat en cas de défaut épidermique. Il est intéressant de noter que *ARSF*, présent sur le même locus, est également surexprimé dans les couches suprabasales, y compris dans les kératinocytes granuleux (ratio épiderme/kér. épineux = 2,4) (Radoja et al. 2006).

Un certain nombre de gènes codant pour des lipases ont également été identifiées, notamment à partir des ORESTEs (Toulza et al. 2007). Les lipases sécrétées LIPK, LIPM et LIPN, toutes trois fortement surexprimées dans les kératinocytes granuleux ($R > 100$), ainsi que la lipase LIPJ (probablement non sécrétée), restent peu caractérisées. Même si le substrat et la fonction de ces lipases restent non décrits dans la littérature, leur homologie de séquence avec les lipases acides LIPA et LIPH, toutes deux codées par des gènes du même locus (10q23.31), suggère un rôle dans la maturation des triglycérides et/ou des esters de cholestérol, tous deux importants pour la fonction du stratum corneum. La fonction des phospholipases identifiées PNPLA1, PLA2G4E, PLA2G4F (Toulza et al. 2007) et PLA2G15 (Mattiuzzo et al. 2009) reste inconnue, mais le fait que seul *PNPLA1* soit surexprimé ($R = 20$) suggère que ces trois autres phospholipases ne jouent peut-être pas de rôle direct dans la fonction barrière.

- Signalisation cellulaire

Même si plusieurs protéines impliquées dans diverses voies de signalisation sont connues pour leur fonction dans la régulation de la prolifération et de la différenciation dans l'épiderme, les acteurs moléculaires des voies impliquées dans ces cascades complexes et souvent imbriquées restent peu décrites dans les kératinocytes différenciés. Même si leur fonction exacte est pour le moment inconnue, l'identification de nouveaux gènes impliqués dans de nombreuses voies reste primordiale pour espérer comprendre l'intégration des nombreux signaux intervenant dans la régulation fine du programme de différenciation.

Ainsi, plusieurs gènes impliqués dans les voies de signalisation des caténines, de Ras et de Src sont surexprimés dans les kératinocytes différenciés. *CTNNBIP1* ($R=9.1$) empêche l'interaction entre la caténine β -1 et les facteurs de transcription de la famille TCF, empêchant la transcription des gènes cibles (Tago et al. 2000). *RASGRP1* ($R=31,8$) code pour une GEF (GDP/GTP exchange factor) qui active spécifiquement Ras au niveau de l'appareil de Golgi et régule la voie ERK/MAP

kinases (Caloca et al. 2003). En aval de Ras, deux autres gènes (*GMIP* et *IQGAP3*) sont également régulés. *GMIP* code pour un répresseur de la voie CDC42/Rac/Rho (Hatzoglou et al. 2007), et est légèrement surexprimé dans les kératinocytes granuleux ($R=2,1$), tandis que *IQGAP3*, qui code pour un effecteur en aval de CDC42 et Rac1, est essentiellement exprimé dans la couche basale ($R=0,18$). Enfin, en aval de Src, *GRLF1* – qui code pour la RhoGAP (GTPase activating protein) p190 – est sous-exprimé dans les kératinocytes granuleux ($R=0,47$). Ces gènes sont susceptibles d’être impliqués dans la régulation de l’adhérence et de la migration, mais aussi dans la survie cellulaire, Ras et Src ayant potentiellement une fonction de protection contre l’anoïkis (apoptose déclenchée par la perte de contact avec la lame basale) (Frisch and Francis 1994).

Plusieurs régulateurs de la concentration intracellulaire de calcium ont également été identifiés. ITPKC ($R=3,4$) est responsable de la transformation de l’inositol 1,4,5-trisphosphate (IP_3) en inositol 1,3,4,5-tetrakisphosphate (IP_4), tous deux impliqués dans l’homéostasie calcique. Le canal sodique ENaC étant requis pour une fonction barrière efficace, le sodium est probablement impliqué lui aussi dans la différenciation kératinocytaire (Mauro et al. 2002). *SLC8A1* ($R=16,2$) codant pour un échangeur membranaire sodium/calcium, son action pourrait être double. Enfin, RCAN1 ($R=5,5$) inhibe la voie de la calcineurine (Rothermel et al. 2000) et régule l’exocytose dans les cellules chromaffines murines (Keating et al. 2008), suggérant une implication de RCAN1 dans la sécrétion des kératinosomes.

Le gène *GSDMA*, dont le ratio de surexpression est très élevé ($R = 800$; Toulza et al. 2007), code pour la gasdermine 1, une protéine ayant un rôle antitumoral (Saeki et al. 2000) et impliquée dans la signalisation du TGF β (Saeki2007), mais dont la fonction au cours de la différenciation épidermique reste inconnue. Un anticorps dirigé contre plusieurs gasdermines conduit à un marquage restreint à la couche granuleuse de l’épiderme humain, à la gaine épithéliale interne et la tige pileuse des follicules pileux et aux sébocytes matures (Lunny et al. 2005). Si la *GSDM3* joue un rôle dans la différenciation de ces annexes épidermiques, *GSDM1* semble être la principale gasdermine exprimée dans les couches différenciées de l’épithélium de l’oesophage murin – dont les cellules présentent une différenciation très proche de celle de l’épiderme (Saeki et al. 2009).

Les kératinocytes peuvent subir deux types de mort cellulaire programmée : la cornification, point culminant de la différenciation, mais également apoptose,

notamment dans le cas d'une exposition important aux ultraviolets. Le résultat de ces deux morts est radicalement différent, l'apoptose entraînant la dégradation du corps cellulaire tandis que la cornification permet la formation de la couche cornée indispensable à la fonction barrière. Il n'est pas donc surprenant que les kératinocytes expriment de nombreux régulateurs (positifs ou négatifs) de la (des) mort(s) cellulaire(s). Si le gène antiapoptotique *AVEN* est détecté dans toutes les couches ($R=1.3$), le suppresseur de tumeurs *ARL11* (ADP-ribosylation factor-like 11 or *ARLTS1*), qui agit sur plusieurs voies de signalisation pro-apoptotiques (Yendamuri et al. 2008) est surexprimé ($R=9$; Mattiuzzo et al. 2009). La fonction de *PLEKHA2* ($R=8,1$) reste obscure : le gène code une protéine à domaine PH (Pleckstrin Homology) recruté à la membrane plasmique par le PI(3,4)P2 au cours de la signalisation de la PI3 kinase (Dowler et al. 2000; Ma et al. 2008). La voie de la PI3K joue un rôle important dans la protection des kératinocytes contre l'anoïkis, probablement via l'activation de la kinase AKT1 (Calautti et al. 2005; Thrash et al. 2006). Cependant, le rôle potentiel de *PLEKHA2* dans la voie PI3K/AKT1 reste à déterminer. Le gène *MAP3K7IP1* ($R=2,1$) code pour la protéine TAB1 qui interagit avec la kinase TAK1/MAP3K7 requise pour l'homéostasie épidermique et la survie kératinocytaire (Sayama et al. 2006). Toutefois, le rôle de TAB1 dans la signalisation de TAK1 est peut-être accessoire (Shim et al. 2005). Le gène codant la protéine CARD14 (Caspase Recruitment Domain family member 14) est également surexprimé au cours de la différenciation kératinocytaire avec un facteur x8 (Toulza et al. 2007). Cette protéine est capable, en réponse à divers stimuli dont l'activation des PKC, de former un complexe multiprotéique avec BCL10, TRAF6, TAK1 et TAB1 ou TAB2. Ce complexe active ensuite l'IkappaB kinase et donc la signalisation de NF-kB (Bertin et al. 2001).

- Facteurs de transcription

Le travail initial de séquençage des ORESTEs a permis d'identifier un nombre important de facteurs de transcription, pour la plupart jamais décrits dans l'épiderme. Le but de cette première étude étant de dresser un inventaire des gènes exprimés dans les kératinocytes granuleux, un nombre limité de gènes a été testé par PCR quantitative pour évaluer le profil d'expression des gènes identifiés. La priorité a été donnée aux gènes les moins décrits dans les banques de séquences (donc les plus susceptibles d'être spécifiques de la différenciation épidermique), et surtout aux

gènes les plus facilement reliables à la fonction barrière : lipases et autres gènes du métabolisme des lipides, protéines structurales, protéases et inhibiteurs de protéases, etc. Un seul facteur de transcription, *CASZ1*, figure parmi les gènes surexprimés identifiés dans notre première étude du transcriptome kératinocytaire (Toulza et al. 2007). Sa fonction reste à ce jour totalement inconnue, de même que sa spécificité de séquence vis à vis de l'ADN. D'une façon générale, la recherche d'un rôle pour les facteurs de transcription identifiés se heurte à plusieurs obstacles. La majorité des facteurs identifiés sont très peu caractérisés, même dans d'autres types cellulaires. De plus, lorsqu'une fonction est connue dans un autre tissu, rien n'indique que les gènes cibles de ce facteur seront les mêmes dans l'épiderme : même si la spécificité de séquence reste la même, certains cofacteurs essentiels peuvent être absents, et la chromatine peut être condensée et rendre inaccessible certains sites de fixation décrits, ou inversement rendre accessibles de nouveaux sites.

Pour les facteurs les moins décrits dans la littérature, la méthode généralement privilégiée pour analyser leur fonction est la perte ou le gain de fonction (sur cellules ou animaux), et surtout l'identification des gènes cibles, généralement par immunoprécipitation de chromatine suivie de l'identification des séquences par microarray (ChIP-on-chip) ou par séquençage (ChIP-seq). Pour les facteurs mieux décrits dans d'autres systèmes, une stratégie susceptible de livrer quelques indices sur la fonction est la recherche de l'expression de cofacteurs ou de gènes cibles connus. Cette approche est forcément limitée, puisque basée sur la comparaison de types cellulaires parfois très distincts, mais a permis de montrer que certains des nouveaux FT surexprimés au cours de la différenciation kératinocytaire (et donc régulateurs probables de ce programme génétique) sont également exprimés au cours d'autres programmes, comme la différenciation des cardiomyocytes (*GATA4*, *ZFPM2*, *ANKRD1* et *HOPX*) ou l'hématopoïèse (*NFE2*, de même que *GATA3* déjà décrit).

Si l'on se base sur le nombre de gènes codant pour des protéines contenant un domaine de liaison à l'ADN, le nombre estimé de facteurs de transcription chez l'être humain est d'environ 2600 (Babu et al. 2004). Ce chiffre exclut les nombreuses protéines adaptatrices susceptibles d'influer grandement sur l'action de ces facteurs sans se lier directement à l'ADN. Alors que cette immense diversité de facteurs suggère une faible redondance (il existe déjà plus de 3 millions de combinaisons de deux facteurs de transcription, et près de 3 milliards de combinaisons de trois

facteurs), il semble qu'un même facteur de transcription non-ubiquiste soit fréquemment « réutilisé » dans des contextes et types cellulaires parfois très différents. Il serait par exemple intéressant de savoir si la différenciation des kératinocytes et des cardiomyocytes, deux types cellulaires soumis à des contraintes mécaniques importantes, partagent (en plus de l'expression des protéines desmosomales) un ou plusieurs modules de régulation transcriptionnelle, comme suggéré par nos résultats (*ANKRD1* et *HOPX* sont ainsi régulés par des membres des familles *GATA* et *ZFPM* dans les cardiomyocytes).

Seule une analyse à grande échelle (et sur de multiples tissus) des sites réellement fixés par chaque facteur de transcription est susceptible d'aboutir à une description suffisante pour pouvoir véritablement comparer entre eux plusieurs programmes de différenciation et évaluer leur degré de « recyclage ». Si un tel projet est actuellement en cours chez certains organismes modèles (*C. elegans* et *D. melanogaster*) dans le cadre du projet modENCODE (<http://www.modencode.org/>), la tâche sera extrêmement longue et coûteuse dans le cas d'un génome aussi complexe que celui des mammifères, puisqu'elle nécessiterait de développer et/ou de valider des anticorps dirigés contre chaque facteur de transcription, puis de réaliser du ChIP-on-chip ou du ChIP-seq pour chaque tissu, stade développemental, etc.

Conclusion

D'une façon générale, les techniques discutées précédemment n'ont pas les mêmes sensibilités et spécificités. Ceci explique que leur emploi sur un même type cellulaire, loin d'être redondant, conduit plutôt à des résultats complémentaires. Ainsi, si l'on compare la contribution de chacune des techniques dans l'identification des 49 nouveaux marqueurs de la différenciation présentés dans le second article, 9 proviennent des ORESTEs, 19 des puces pangénomiques, 6 de l'analyse bioinformatique et 1 de l'analyse bibliographique, tandis que seulement 8 sont issus à la fois des ORESTEs et des puces, et 6 des ORESTEs et de l'analyse bioinformatique. La redondance entre les différentes bases apparaît donc faible (28,6%), ce qui est également confirmé par la confrontation des données d'ORESTEs et de puces : sur les 3387 et 10824 gènes identifiés respectivement, seuls 1930 sont en commun (en plus des 150 ORESTEs sans sonde correspondante sur les puces), soit une redondance de 60% ou 18% selon la technique considérée.

Combiner plusieurs techniques complémentaires permet également de faire une estimation du nombre de gènes exprimés dans l'épiderme. Toutes couches confondues, les différentes techniques ont permis d'identifier environ 12300 gènes potentiellement exprimés (dont 3466 avec certitude – ORESTEs et/ou PCRq + séquençage des amplicons). De nombreuses études à grande échelle ne rendent disponible que le nombre de gènes considérés comme différentiellement exprimés, et non le total des gènes détectés. Seule l'étude réalisée au moyen de puces pangénomiques sur de l'épiderme murin (April and Barsh 2006) cite le chiffre de 16715 gènes détectés. On peut donc estimer que le nombre de gènes exprimés dans l'épiderme (humain ou murin) se situe aux alentours de 14500 gènes. Comparé aux 25000 gènes du génome humain ce chiffre peut sembler important, surtout pour un tissu majoritairement composé de cellules spécialisées et engagées dans un processus de mort cellulaire programmée. Il reste dans le même ordre de grandeur (>10000) que plusieurs autres études menées à l'aide de puces pangénomiques sur divers tissus ou cellules – pour exemple (Akavia et al. 2008; Kobayashi et al. 2008). Toutefois, il est également possible que la technique des puces à ADN conduise à une surévaluation du nombre de gènes exprimés.

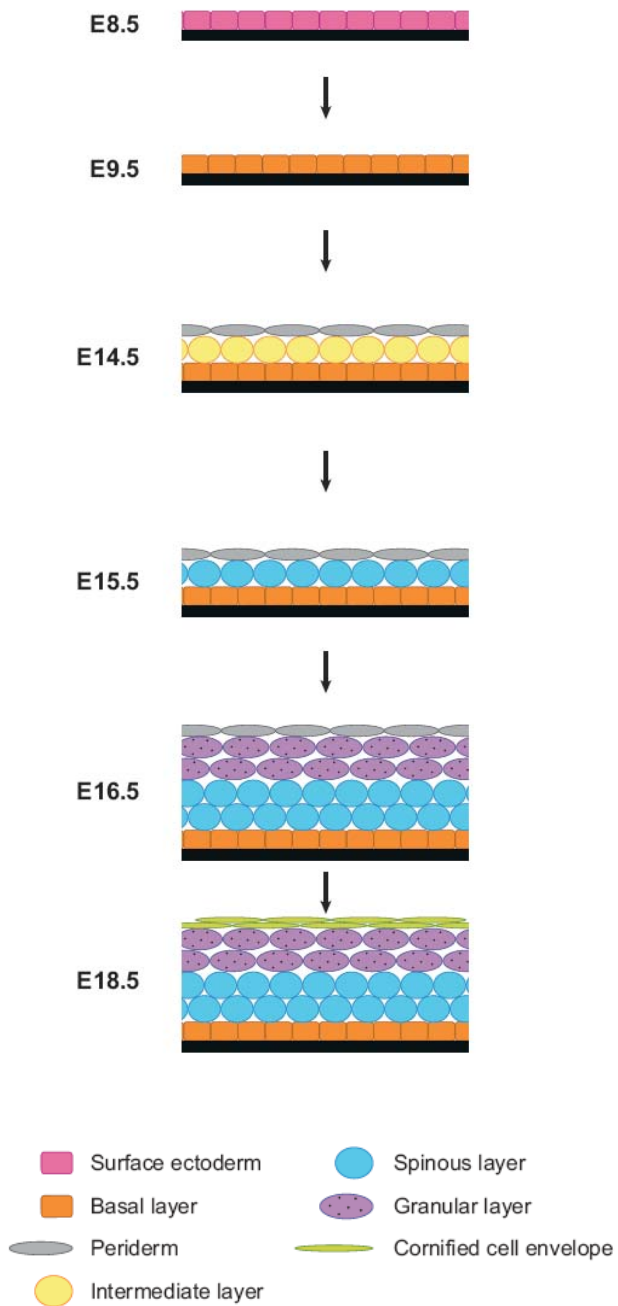
L'étude d'April & Barsh a également permis de distinguer des gènes spécifiques de la couche basale (2097) ou des couches suprabasales (883) sur les 16715 gènes détectés au total. Comme pour notre étude, la majorité des gènes ne

semblent donc pas régulés. Les différentes approches que nous avons employées ont permis d'identifier au total 82 nouveaux gènes associés à la différenciation (Toulza et al. 2007; Mattiuzzo et al. 2009). Si l'on se base sur la première analyse des puces pangénomiques (624 gènes avec $R > 1,5$ et $p < 0,05$) et que l'on extrapole aux 1450 ORESTEs non détectés sur puces, on atteint un chiffre de 695 gènes environ, assez proche des 883 gènes de l'étude d'April & Barsh. Toutefois, en tenant compte des résultats de PCR quantitative (27 gènes confirmés sur 64 choisis à partir de puces), cette estimation chute à 293 gènes environ. Toutefois, le nombre réel de gènes spécifiques du kératinocyte granuleux reste difficile à évaluer, les hybridations croisées biaisant le nombre de gènes exprimés tandis qu'au contraire certains transcrits sont probablement en-dessous du seuil de détection des puces à ADN. De plus, le nombre de faux-négatifs (gènes déclarés non-régulés alors qu'ils sont différentiellement exprimés) est quasi impossible à évaluer. Par ailleurs, l'étude d'April & Barsh met en évidence un plus grand nombre de gènes exprimés dans la couche basale que dans les couches suprabasales, ce qui est logique pour des cellules de type et d'état prolifératif variés, et qui confirme le biais de notre méthode d'enrichissement puisque nous avons plus de difficulté à identifier des gènes surexprimés dans la couche basale (*cf.* discussion sur les puces pangénomiques).

Au final, le choix d'une approche multi-technique peut sembler paradoxal dans le sens où chacune présente des difficultés techniques non négligeables et nécessite de prendre un maximum de précautions, tant au moment de la réalisation qu'à celui de l'interprétation des résultats. Cependant, la plupart des techniques actuelles (à l'exception peut-être du séquençage nouvelle génération) ne permettant pas réellement de décrire un système de façon exhaustive, cette stratégie possède l'avantage indéniable d'être un des meilleurs moyens de dresser un inventaire aussi complet que possible de l'état du transcriptome d'un type cellulaire.

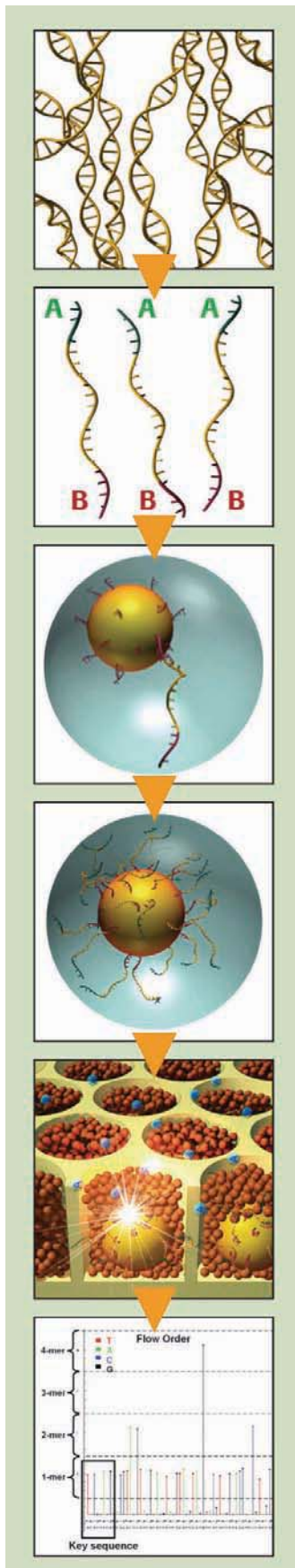
IV. ANNEXES

Annexe 1 : Stratification épidermique



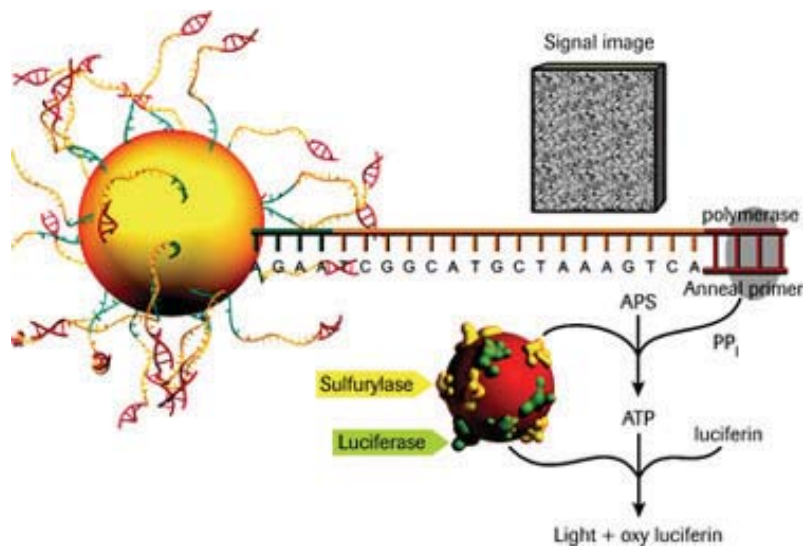
Représentation schématique des étapes de la stratification épidermique au cours du développement murin (Koster and Roop 2007).

Annexe 2 : le séquençage nouvelle génération.

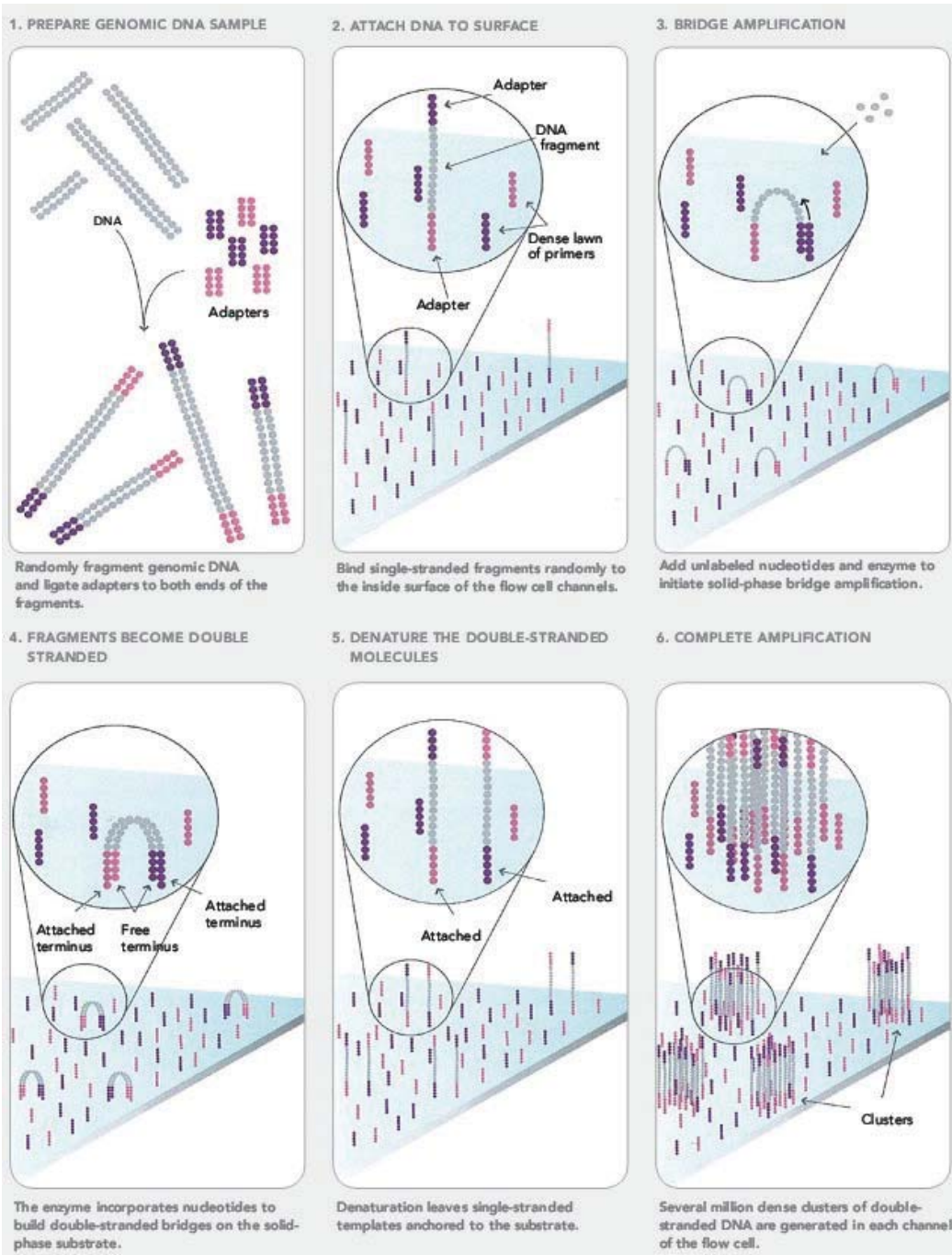


La technologie 454, commercialisée par 454 Life Sciences (Roche Applied Science), couple une méthode de PCR en émulsion et un pyroséquençage. Dans un premier temps, des fragments de 300 à 800 paires de bases sont générés à partir d'ADN génomique (ou autre), puis des adaptateurs différents sont ajoutés à chaque extrémité des fragments. Un des adaptateurs étant biotinylé, les fragments sont immobilisés en dilution limite sur des billes portant de la streptavidine, de sorte à ne pas avoir plus d'un fragment par bille. Puis, les billes sont placées dans une émulsion, où chaque bille est isolée dans une gouttelette qui va tenir lieu de « microréacteur » pour la PCR. Cette étape d'amplification *in vitro*, très performante, permet de générer jusqu'à 10^7 fragments par bille, à partir d'un seul fragment d'ADN initial. Pour l'étape de pyroséquençage, chaque bille est isolée dans un puit d'une plaque de très haute densité (les PicoTiter plates™ employées sont formées d'un très grand nombre de fibres optiques formant un large support dans lequel sont creusés plusieurs centaines de milliers de puits de $44\mu\text{m}$ de large), puis des microbilles portant deux enzymes, une ATP sulfurylase et une luciférase, sont ajoutées. En effet, le pyroséquençage est basé non pas sur l'arrêt de la synthèse d'un brin d'ADN par des nucléotides « terminateurs » comme pour la technique de Sanger, mais par la détection en temps réel de l'incorporation des nucléotides. Durant un cycle, chaque dNTP possible (dATP, dTTP, dGTP ou dCTP) est ajoutée non pas simultanément, mais séquentiellement ; s'il correspond à la base complémentaire du brin séquencé, l'ADN polymérase présente dans le puit allonge le brin néosynthétisé en incorporant la base, ce qui entraîne la libération d'un pyrophosphate inorganique (PPi). L'ATP

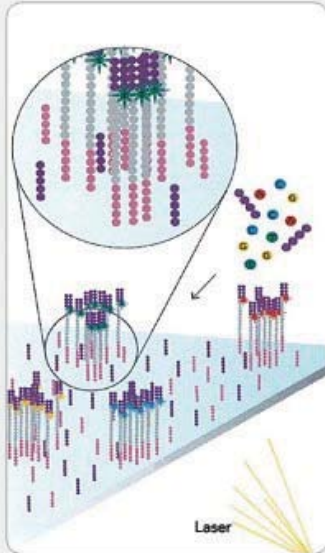
sulfurylase emploie alors ce PPi pour produire de l'ATP, qui est employé par la luciférase pour transformer la luciférine en oxyluciférine (une apyrase dégrade les nucléotides en surplus). Cette réaction s'accompagne d'une émission de lumière, qui est proportionnelle à la quantité de PPi libéré, donc au nombre de bases incorporées simultanément (si deux bases successives sont identiques, deux fois plus de lumière sera émise, *etc.*). La lumière est détectée en temps réel par un capteur CCD directement couplé aux fibres optiques composant le support ; ainsi, à chaque ajout d'un dNTP, le capteur CCD produit une image qui témoigne de la lumière émise dans chaque puit. Ces données brutes permettent, après normalisation des images, de reconstruire un « pyrogramme » correspondant à l'enchaînement des bases dans la séquence pour chacun des puits de la machine. Ainsi, la version la plus récente de la plateforme 454, le Q20, permet de générer plus d'un million de séquences de 400 bases de long par « run ».



La seconde approche à avoir été développée est la technologie Solexa (Illumina). L'étape d'amplification est ici réalisée en phase solide (bridge amplification) : l'ADN est fragmenté, couplé à des adaptateurs différents à chaque extrémité, puis déposé sous forme simple brin sur la Cluster Station, un support spécialement développé pour permettre l'amplification et le séquençage des brins. Chaque molécule se fixe aléatoirement sur le support ; or, celui-ci est recouvert d'oligonucléotides complémentaires des adaptateurs portés par les fragments. Ainsi, plusieurs cycles de dénaturation/renaturation/extension en présence d'ADN polymérase et de nucléotides (non marqués) permettent l'amplification locale (Cluster) d'une molécule d'ADN (jusqu'à 1000 copies par cluster, et 80-100 millions de clusters par support). L'étape de séquençage consiste à employer des amorces complémentaires d'un des adaptateurs, et ajouter simultanément quatre terminateurs réversibles marqués à des fluorochromes différents. Ainsi, après lavage, la base ajoutée est identifiée en fonction de sa longueur d'onde d'émission, puis le blocage est reversé et un nouveau cycle permet de déterminer la base suivante, *etc.* La réaction, bien qu'automatisée, s'avère toutefois lente : 2 jours pour déterminer des séquences de 35 pb (soit jusqu'à 3,5 Gigabases séquencées), les versions les plus récentes permettant de générer jusqu'à 120 millions de séquences d'une longueur de 75 pb, voire 2 x 75 pb grâce à la technique de Paired-End diTag (initialement développée pour l'immunoprécipitation de chromatine – CHIP-PET, cette technique consiste à ne garder que les extrémités des fragments d'ADN, à les rabouter par ligation et à les séquencer ensemble afin d'avoir les « bornes » des fragments), pour un temps de réaction pouvant atteindre 8 jours pour 15 Gigabases.



7. DETERMINE FIRST BASE



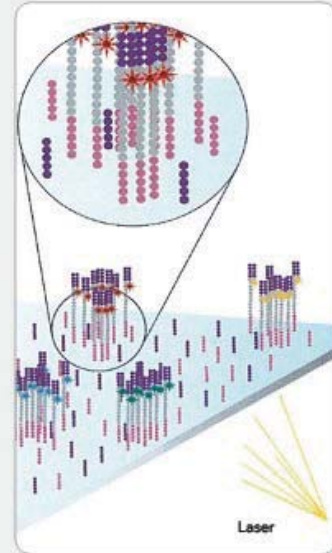
First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE



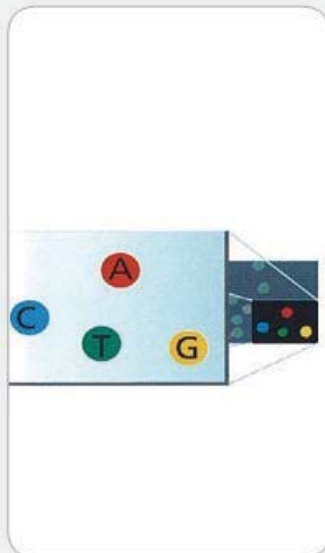
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE



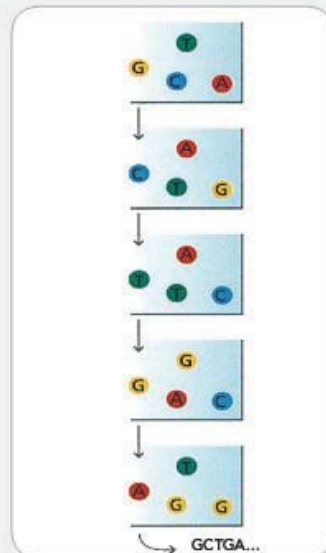
Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

10. IMAGE SECOND CHEMISTRY CYCLE



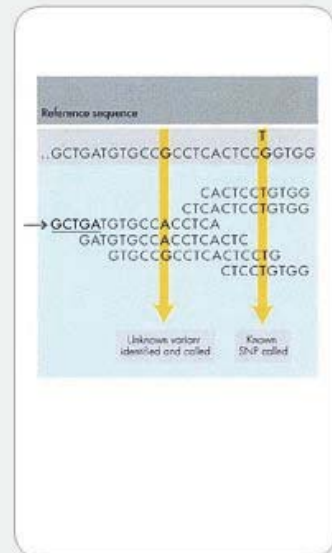
After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES



Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

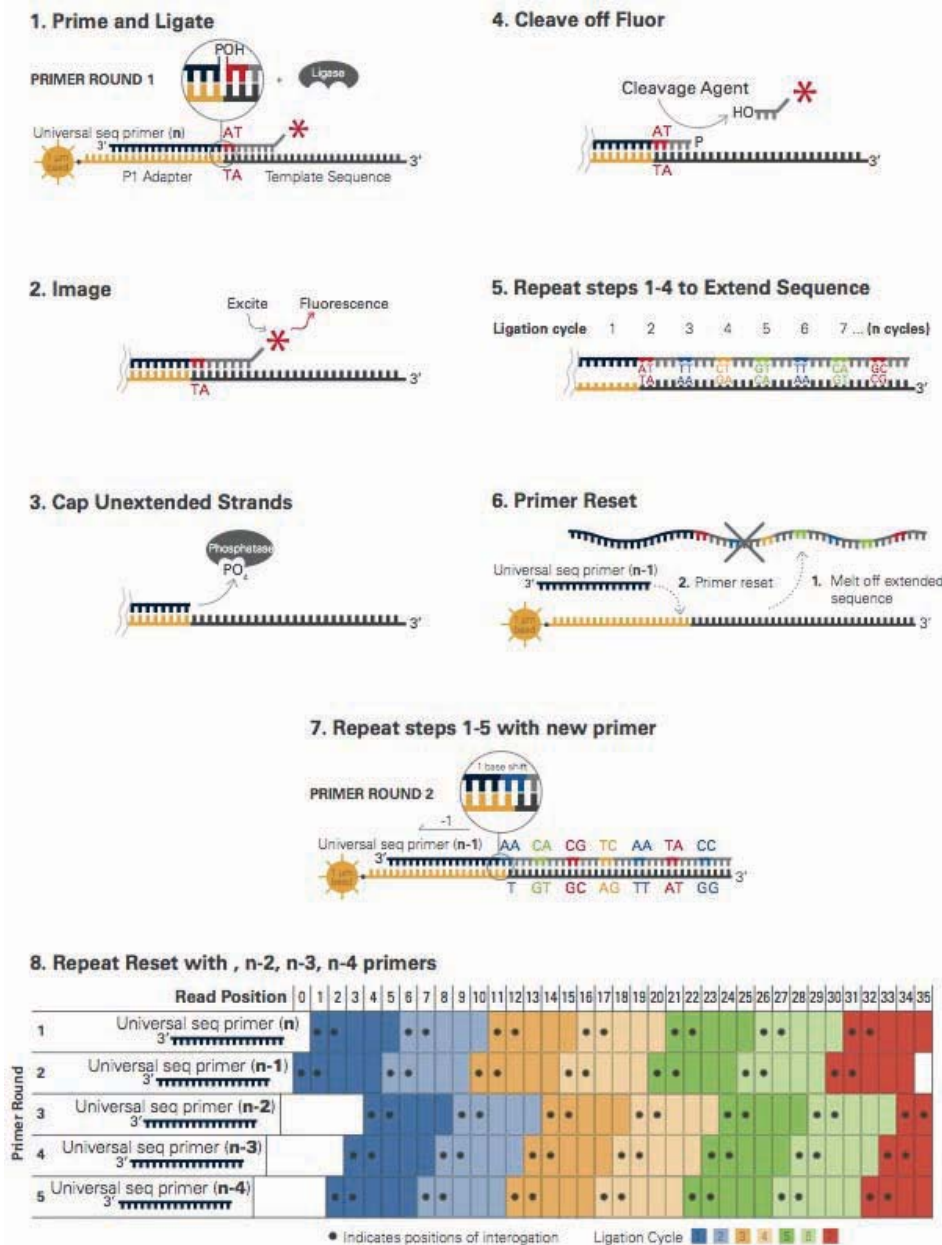
12. ALIGN DATA



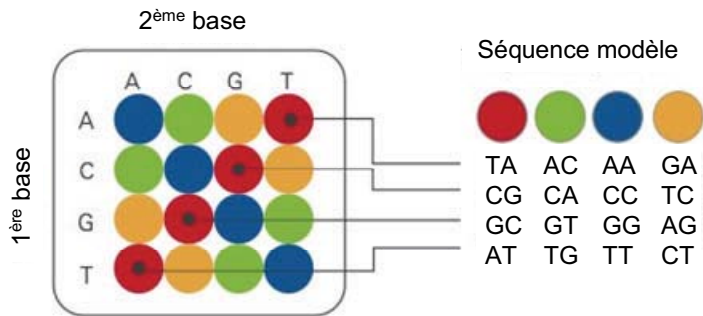
Align data, compare to a reference, and identify sequence differences.

La troisième approche, nommée SOLiD (Applied Biosystems), peut être rapprochée du 454 pour sa stratégie d'amplification par PCR en émulsion, mais l'étape de séquençage est radicalement différente, puisque effectuée en absence de polymérase. Après la fragmentation de l'ADN et l'ajout d'adaptateurs différents aux extrémités, les fragments sont capturés par des billes et amplifiés par PCR en émulsion. Puis, les extrémités 3' des fragments sont modifiées afin de permettre la fixation covalente avec le support de verre sur lequel les billes sont déposées (plusieurs dizaines de millions de billes par lame de verre). Une fois immobilisés sur le support, les fragments sont hybridés avec une amorce universelle de longueur n , complémentaire d'un des adaptateurs. Puis, au lieu d'employer une ADN polymérase comme précédemment décrit, la technique SOLiD fait appel à un mélange d'oligonucléotides marqués et utilise une ligase pour déterminer la séquence du brin d'ADN. Il est crucial de décrire la conception des oligonucléotides pour bien appréhender la technique SOLiD. Chaque oligo, de formule 3' **XX**-**NNN**-**YYY** 5', possède un des 16 dinucléotides spécifiques sur les bases 1 et 2 (**XX**), est dégénéré sur les bases 3-5 (**NNN**), ainsi que sur les bases 6-8 (**YYY**) couplées à un des quatre fluorochromes employés pour marquer les oligos. Seules les bases **XX** donnent une information sur la séquence d'ADN, mais seulement après plusieurs étapes (puisque'il y a seulement 4 fluorochromes pour 16 dinucléotides). Dans un premier temps, après l'ajout de l'amorce universelle de longueur n , le mélange d'oligos fluorescents (8-mers) est hybridé, et celui dont les deux premières bases (**XX**) sont complémentaires du brin d'ADN immédiatement à la suite de l'amorce universelle se trouve accroché à sa suite par une ligase ; les bases 1-2 de la séquence sont donc interrogées. Un scanner détecte alors les niveaux de fluorescence pour chacun des quatre fluorochromes, sur l'ensemble du support. Puis, un clivage chimique libère le fluorochrome ainsi que les bases 6-8, laissant un phosphate libre en 5' ; les étapes précédentes (oligos, ligation, scan, phosphatase, clivage) sont ensuite répétées pendant 5 à 7 cycles, permettant d'interroger les bases 6-7, 11-12, 16-17, etc., de la séquence d'ADN. Puis, l'ensemble est « réinitialisé » en déshybridant l'amorce universelle de longueur n ainsi que les extensions, et une nouvelle amorce universelle de longueur $n-1$ est hybridée, permettant alors d'interroger les bases 1, 5-6, 10-11, 15-16, etc. L'opération est ensuite répétée avec une amorce de longueur $n-2$, puis $n-3$ et $n-4$, la succession de cinq cycles complets d'hybridation et d'extension par ligation permettant de générer un lot de données qui se chevauchent et couvrent environ 35

bases continues. Ainsi, chaque base de la séquence est interrogée par deux oligonucléotides différents, ce qui permet d'une part de connaître son identité (puisque seuls 4 fluorochromes sont employés pour 16 dinucléotides, une seule lecture ne peut identifier le dinucléotide en question), et d'autre part de distinguer les erreurs de lectures (reconnues dans une seule des deux phases) des polymorphismes sur un nucléotide (détectés dans les deux phases).

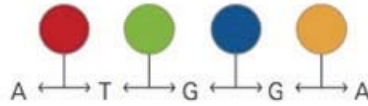


Dinucléotides encodés par chaque couleur



Double interrogation

Avec un encodage à deux bases, chacune est définie deux fois



SNP site indicated by 2 adjacent color changes



Single color change is typically a measurement error



1 Base Deletion



Insertion



Annexe 3 : Vers l'intégration des données d'expression : de la connaissance à la modélisation

Un des intérêts des approches à grande échelle est la possibilité (théorique) des combiner les résultats de plusieurs études ou de les comparer. En réalité, de telles « méta-analyses » sont généralement compliquées par l'emploi de techniques différentes et par l'absence d'uniformisation dans l'analyse et l'annotation des résultats. Ainsi, même lorsqu'il s'agit de comparer des données de diverses puces à ADN par exemple, l'obstacle que représente la diversité des techniques (puces à oligonucléotides courts, longs, ADNc... ; hybridation simple ou double...) et des annotations employées par les fabricants (nom de gène, symbole officiel ou alias, référence de la sonde propre au fabricant...) rend la comparaison complexe et augmente le risque d'erreur, d'où l'intérêt d'employer une ontologie précise. Bien que se situant à un degré supérieur, l'intégration des données, c'est-à-dire l'assimilation et l'assemblage de nombreuses informations de *-omique* de nature souvent différente (transcriptomique, protéomique, métabolomique...) en un portrait cohérent des événements réels, se heurte également au même problème de la diversité de nature et d'annotation des données.

L'intégration représente pourtant un des défis les plus cruciaux pour la bioinformatique et plus largement la biologie cellulaire et moléculaire, le but ultime étant de parvenir à une description aussi complète que possible des événements ayant lieu à tous les niveaux – molécules, cellules, tissus, organes, organismes entiers et même populations – ceci afin de pouvoir modéliser les systèmes biologiques dans leur ensemble en tenant compte de leur exceptionnelle complexité. Si les outils sont théoriquement disponibles ou à portée de développement, l'absence quasi complète de nomenclature commune à l'ensemble des techniques de génomique fonctionnelle à haut débit, ainsi que le volume et la diversité de nature des informations générées par ces techniques, rendent cette intégration impossible pour des équipes ou des laboratoires isolés. Pour faciliter cette étape d'intégration, un premier pas a été franchi vers un formatage universel des données grâce aux standards définis par la MGED Society, non seulement pour les données de puces à ADN (standard MIAME) mais plus largement pour toutes les données d'expression génique avec MAGE (MicroArray and Gene Expression), de séquençage avec MINSEQE (Minimal Information about a high-throughput SEQuencing Experiment), de protéomique,

génomique, *etc*, l'ensemble des standards étant regroupé sous le projet MIBBI (Minimum Information for Biological and Biomedical Investigations, <http://www.mibbi.org/>). L'emploi – aujourd'hui très marginal – de LIMS (Laboratory Information Management System), sorte de cahier de manipulation numérique permettant un suivi parfait de toutes les étapes depuis la collecte des échantillons jusqu'aux méthodes d'analyse employées, facilitera à plus long terme l'élimination des données peu fiables déposées dans les banques et la comparaison entre études.

A ce jour, plusieurs projets publics de grande ampleur, l'un français (SIDR, Standard-based Infrastructure with Distributed Resources, lancé par l'INIST – l'Institut de l'Information Scientifique et Technique du CNRS ; <http://www.sidr-isb.eu/>), l'autre européen (ELIXIR, European Life Sciences Infrastructure for Biological Information lancé par l'EMBL-EBI ; <http://www.elixir-europe.org/>), visent à collecter, structurer et pérenniser les données biologiques de toute nature, et fournir des outils pour comparer et intégrer ces données afin de mieux comprendre des processus complexes qui ne sont que très partiellement décrits par des études isolées.

Annexe4 : Etude du rôle de la cornéodesmosine in vivo (modèles murin)

Dans le cadre de l'étude du rôle de la cornéodesmosine par inactivation du gène *cdsn* chez la souris, j'ai réalisé des comparaisons des niveaux d'expression de plusieurs gènes entre des peaux de souris sauvages ou KO, par PCR quantitative. Ce travail a fait l'objet de la publication qui suit, parue dans the Journal of Cell Science en 2009.

JOURNAL OF CELL SCIENCE
122, 2699-2709 (2009)

**Corneodesmosin gene ablation induces
lethal skin barrier disruption and hair
follicle degeneration related to desmosome
dysfunction**

Emilie Leclerc, Anne Huchenoq, **Nicolas R. Mattiuzzo**, Daniel Metzger, Pierre Chambon, Norbert B. Ghyselinck, Guy Serre, Nathalie Jonca and Marina Guerrin.

Résumé

Afin d'analyser le rôle de la CDSN *in vivo*, notre équipe a développé un modèle d'inactivation somatique basé sur le système CreLox et ciblant le gène *Cdsn* chez la souris, soit très précocement pendant le développement, soit spécifiquement dans l'épiderme chez l'adulte.

Les souris hétérozygotes pour la mutation ne présentent aucun phénotype apparent. Au contraire, les souris homozygotes *Cdsn*^{-/-} souffrent d'une extrême fragilité épidermique et meurent quelques heures après la naissance. Des tests fonctionnels montrent un trouble majeur de la perméabilité épidermique et une perte de la résistance mécanique de la couche cornée. Les analyses histologiques et ultrastructurales montrent que la cohésion épidermique est rompue à l'interface couche granuleuse / couche cornée. Pour évaluer à plus long terme l'évolution de l'épiderme des souris *Cdsn*^{-/-}, des greffes de peau sur souris Nude ont été réalisées. Au niveau des greffons, l'épiderme présente d'abord une hyperprolifération compensatoire et une hyperkératose parakératosique associées à un trouble persistant de la barrière, avant que le greffon ne finisse par dégénérer. De plus, les follicules pileux, où la *Cdsn* est également exprimée (dans la gaine épithéliale interne), subissent une dégénérescence progressive qui aboutit à leur disparition après huit semaines.

Chez la souris adulte, l'inactivation au niveau de l'épiderme (par traitement topique) du gène *Cdsn* conduit à des anomalies histologiques similaires et à un défaut finalement létal de la barrière épidermique.

L'ensemble de ces résultats montre que la *Cdsn* n'est pas nécessaire pour le maintien de la cohésion épidermique *in utero*, mais qu'elle est par contre indispensable tout au long de la vie pour maintenir l'intégrité des desmosomes et donc la cohésion intercellulaire dans l'épiderme. Le renforcement des propriétés adhésives que cette protéine confère aux cornéodesmosomes semble également nécessaire au maintien de l'architecture des follicules pileux.

Corneodesmosin gene ablation induces lethal skin-barrier disruption and hair-follicle degeneration related to desmosome dysfunction

Emilie A. Leclerc¹, Anne Huchencq¹, Nicolas R. Mattiuzzo¹, Daniel Metzger², Pierre Chambon², Norbert B. Ghyselinck², Guy Serre^{1,*}, Nathalie Jonca^{1,‡} and Marina Guerrin^{1,‡}

¹UMR 5165 'Différenciation Epidermique et Autoimmunité Rhumatoïde' (UDEAR), CNRS – Université Toulouse III, IFR150, INSERM, CHU PURPAN, Place du Dr Baylac, TSA 40031, F-31059 Toulouse, Cedex 9, France

²IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), Inserm U596 – CNRS UMR 7104 – Université Louis Pasteur – Collège de France, Illkirch, F-67400 France

*Author for correspondence (e-mail: Guy.Serre@udear.cnrs.fr)

‡These authors contributed equally to this work

Accepted 1 May 2009

Journal of Cell Science 122, 2699–2709 Published by The Company of Biologists 2009
doi:10.1242/jcs.050302

Summary

Corneodesmosin (CDSN) is specific to desmosomes of epithelia undergoing cornification, mainly the epidermis and the inner root sheath of the hair follicles. *CDSN* nonsense mutations are associated with hypotrichosis simplex of the scalp, a rare disease that leads to complete baldness in young adults. CDSN displays adhesive properties, mostly attributable to its N-terminal glycine-rich domain, and is sequentially proteolyzed as corneocytes migrate towards the skin surface. *K14*-promoter driven Cre-mediated deletion of *Cdsn* in mice resulted in neonatal death as a result of epidermal tearing upon minor mechanical stress. Ultrastructural analyses revealed a desmosomal break at the interface between the living and cornified layers. After grafting onto nude mice, knockout skin showed a chronic defect in the epidermal permeability barrier. The epidermis was first hyperproliferative with a thick cornified

layer, then, both the epidermis and the hair follicles degenerated. In adults, *Cdsn* deletion resulted in similar histological abnormalities and in a lethal barrier defect. We demonstrate that *Cdsn* is not essential for skin-barrier formation in utero, but is vital throughout life to preserve this barrier by maintaining desmosome integrity. The strong adhesive function that the protein confers on corneodesmosomes also seems necessary for maintaining the architecture of the hair follicle.

Supplementary material available online at
<http://jcs.biologists.org/cgi/content/full/122/15/2699/DC1>

Key words: Desmosome, Epidermis, Hypotrichosis simplex of the scalp, Knockout mice, Psoriasis

Introduction

Lessons from human diseases and engineered mouse models have established the importance of desmosomes in tissues subjected to high mechanical stress. Indeed, these intercellular junctions provide strong cohesion between adjacent cells. Mutations in genes encoding desmosomal components are invariably responsible for skin diseases, often associated with heart defects (for reviews, see Kottke et al., 2006; Garrod and Chidgey, 2008). In the epidermis, desmosomes undergo profound morphological and biochemical changes throughout keratinocyte differentiation. In particular, during cornification, a specific cell death process that transforms keratinocytes into anucleated, flattened corneocytes, considerable modifications of the desmosomes occur, accounting for their designation as corneodesmosomes (Serre et al., 1991).

Corneodesmosin (CDSN) is a protein specific to desmosomes that will undergo transformation into corneodesmosomes, which, in humans, comprises desmosomes of the epidermis, the three layers of the inner root sheath (IRS) of the hair follicles and the hard palate epithelium. Indeed, the protein is not expressed in noncornified squamous epithelia (Serre et al., 1991). After its secretion by granular keratinocytes via the lamellar granules, CDSN is incorporated into the desmoglea of the desmosomes shortly before

their transformation into corneodesmosomes, most probably reinforcing the cell-cell adhesion mediated by the desmosomal cadherins already present in the structure. In the course of stratum corneum maturation, CDSN is progressively degraded, its degradation being a prerequisite for desquamation (Simon et al., 2001).

Cloning of its cDNA has revealed that *CDSN* is located on chromosome 6, in the major psoriasis susceptibility locus PSORS1 (Guerrin et al., 1998). Some *CDSN* single nucleotide polymorphisms (SNPs) have been associated with psoriasis in many genetic studies (reviewed by Capon et al., 2002), and a recent study revealed that only two genes from PSORS1, *HLA*C and *CDSN*, yield protein alleles that are unique to risk haplotypes (Nair et al., 2006). However, according to the latter study and others, the exact identity of the PSORS1 gene remains controversial (Veal et al., 2002; Capon et al., 2003; Helms et al., 2005; Orrù et al., 2005; Nair et al., 2006). Intriguingly, the only monogenic disease identified so far to be associated with mutations in *CDSN* revealed a hair phenotype and did not affect the epidermis (Levy-Nissenbaum et al., 2003); hypotrichosis simplex of the scalp (HSS; OMIM 146520) is a rare autosomal-dominant disease characterized by progressive loss of scalp hair resulting in almost complete baldness by the third decade.

To date, three different nonsense mutations have been reported in four families from Israel, Denmark and Mexico (Lai-Cheong et al., 2007). These mutations lead to the production of a truncated form of CDSN that has been detected as deposits beneath the epidermis basement membrane and deeper in the dermis, at the periphery of the hair follicles (Levy-Nissenbaum et al., 2003). Although the pathophysiological mechanism of this disorder is not understood, CDSN obviously has an important role in hair physiology.

A striking feature of CDSN is its very high serine and glycine content (27.5% and 16%, respectively), particularly at the N-terminus of the protein (residues 60-171). It has been suggested that similar serine- and glycine-rich domains described at both termini of keratins form structural motifs, the so-called 'glycine loops', that mediate intermolecular adhesion by acting like Velcro (Steinert et al., 1991). Furthermore, it has been demonstrated, in vivo and in vitro, that CDSN displays homophilic adhesive properties (Jonca et al., 2002; Caubet et al., 2004). Moreover, recombinant CDSN forms highly stable homo-oligomers that dissociate only partially in 8 M urea, possibly reflecting *cis* and/or *trans* interactions in vivo. The N-terminal glycine loop domain was shown to be necessary and sufficient, and thus responsible, for both the oligomerization of CDSN and its homophilic adhesive properties in vitro (Caubet et al., 2004).

In vivo, CDSN could thus mediate cell-cell adhesion by reinforcing the resistance of the desmosomes. It has also been proposed that CDSN contributes to the structural modifications that characterize the transformation of desmosomes into corneodesmosomes (Guerrin et al., 1998) or prevents premature desquamation by protecting desmosomal proteins from proteolysis (Lundström et al., 1994). Inactivation of mouse *Cdsn* by mutational insertion was very recently reported to induce neonatal death (Matsumoto et al., 2008), and defective corneodesmosome formation was proposed to be responsible for stratum corneum detachment.

We developed two epidermal-selective *Cdsn*-deficient mouse models. Somatic *Cdsn* inactivation in mouse epidermis was achieved using human keratin 14 (*K14*)-promoter driven Cre-mediated deletion (Li et al., 2001). Temporally controlled *Cdsn* deletion in epidermal keratinocytes was achieved using a tamoxifen-inducible Cre recombinase under the control of the same promoter (Li et al., 2000). Our results confirmed the vital role of CDSN in newborns. We also show that CDSN function is essential in adult skin and that the functional defect lies in an impaired resistance of corneodesmosomes to mechanical stress. This supports the idea that CDSN mainly acts as an adhesive protein.

Results

Generation of *Cdsn*-deficient mice

To investigate the role of CDSN in the barrier function and physiology of the hair follicle, we developed *Cdsn*-deficient mice by using *K14*-promoter driven Cre-mediated *loxP* recombination (Fig. 1A). The human *KRT14* promoter is active during mouse epidermal fetal development as early as embryonic day (E)9.5, and is strongly activated by E14.5 onwards (Vassar et al., 1989). Similarly to many other markers of epidermis differentiation, *Cdsn* mRNA was first detected at E16.5 and becomes prominent at E18.5 (H.G., unpublished results). Therefore, the chosen strategy would induce the deletion of the second exon of *Cdsn* before the gene is expressed in the developing epidermis and in hair follicles.

We first generated mice bearing floxed *Cdsn* alleles (*Cdsn*^{fl/fl}), which were bred to homozygosity. These animals appeared normal,

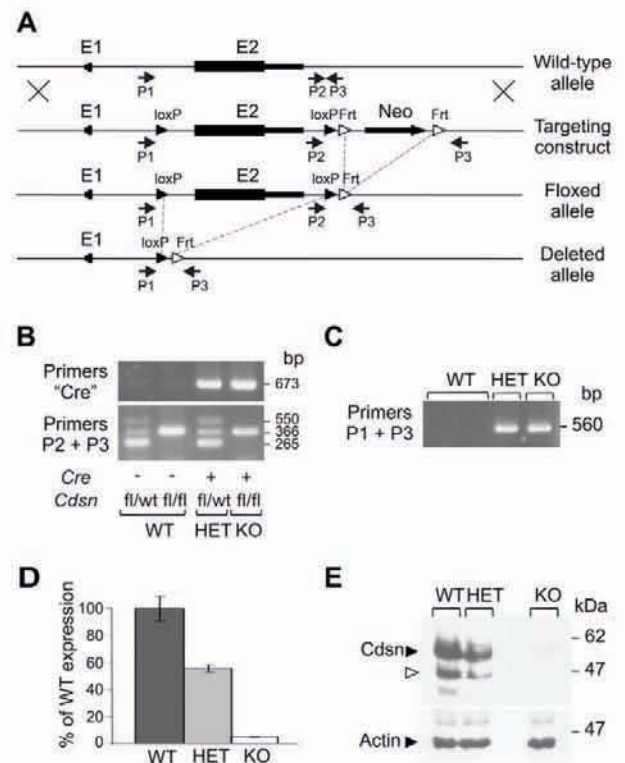


Fig. 1. Gene targeting of *Cdsn*. (A) Schematic drawing of the *Cdsn* wild-type locus, the targeting construct (containing the neomycin selector gene), the floxed allele, and the deleted allele obtained after Cre-mediated excision. The floxed part of *Cdsn* (3 kb) encompasses the second exon, so only the first exon, which encodes 28 amino-acids mostly corresponding to the signal peptide, is preserved after Cre-mediated recombination. Black boxes represent exons, *loxP* and *Frt* sites are depicted as black and open triangles, respectively. Primers for PCR genotyping (P1-P3) are indicated with arrows. (B) PCR analysis of tail DNA from offspring of *Cdsn*^{fl/fl} mice crossed with *K14-Cre*^{tg0}:*Cdsn*^{fl/wt} (HET) mice. Specific primers were used to detect the *K14-Cre* transgene (primers 'Cre'). Primers P2 and P3 were used to detect wild-type (265 bp) or floxed (366 bp) *Cdsn* alleles. A 550 bp band was detected only in heterozygous pups, which corresponds to the amplification of partial hybridization between 265 bp and 366 bp amplicons, as verified by sequencing. (C) Primers P1 and P3 allow the amplification of a 560 bp fragment when the exon 2 of *Cdsn* is excised. The experimental conditions used for this PCR do not allow amplification of the non-excised 3.7 kb fragment still containing the second exon. (D) Quantitative RT-PCR analysis of *Cdsn* expression on dorsal skin samples from WT, HET and KO neonates. Values represent relative RNA level from the WT, after normalization with *Dsg3* transcript level. Values are mean \pm s.e.m. (E) Equal amounts of protein extracts from WT, HET and KO neonate dorsal skin were separated by SDS-PAGE and subjected to western blot analysis using the anti-CDSN F28-27 mAb or an anti-actin mAb. In protein extracts from WT and HET animals, F28-27 reacts with two major polypeptides of about 54 kDa and 45 kDa, which correspond to the full-length (black triangle) and a proteolytically processed form of CDSN (open triangle), respectively.

indicating that the genetic manipulation did not affect the function of the floxed gene. Generation of *Cdsn*-deficient mice was obtained by crossing *Cdsn*^{fl/fl} mice with *K14-Cre* transgenic mice (Li et al., 2001). The resulting animals bearing the *K14-Cre* and heterozygous for the floxed *Cdsn* allele were selected and mated again with *Cdsn*^{fl/fl} mice. Four different genotypes were thus obtained (Fig. 1B): *K14-Cre*^{0/0}:*Cdsn*^{wt/fl} and *K14-Cre*^{0/0}:*Cdsn*^{fl/fl}, both denoted hereafter as wild-type (WT), *K14-Cre*^{tg0}:*Cdsn*^{wt/fl}, denoted hereafter

as heterozygous mice (HET), and *K14-Cre^{tg0};Cdsn^{fl/fl}*, denoted as knockout (KO) mice. The genotyping of 232 newborns from 12 intercrosses revealed that WT, HET and KO pups were produced in the expected mendelian ratios, indicating no embryonic lethality for the KO mice. Excision of the floxed *Cdsn* allele in the skin of newborn pups was controlled by PCR analysis (Fig. 1C). As expected, in skin of KO neonates, *Cdsn* mRNA level measured by quantitative real-time PCR was very low, confirming efficient ablation of *Cdsn* (Fig. 1D). In HET pup skin, *Cdsn* expression was roughly 50% of the WT level, demonstrating the absence of dose compensation. Western blot analysis (Fig. 1E) using the F28-27 anti-CDSN monoclonal antibody (mAb) confirmed the absence of CDSN in the skin of KO animals, whereas the full-length 54 kDa protein was detected in skin extracts from WT and HET animals. The smaller amount of protein in HET mice is consistent with results of RT-PCR analysis, which indicate inactivation of one *Cdsn* allele. A lower molecular mass peptide of ~45 kDa was also detected in both WT and HET pup extracts, which corresponds to a first step of the CDSN proteolytic cleavage in the course of stratum corneum maturation (Simon et al., 1997).

Inactivation of CDSN in mouse skin results in early postnatal death

KO pups suffered from severe skin detachment (Fig. 2A) and died within 1 hour of birth or were immediately eliminated by their mother. To delineate the causes of neonatal lethality, we systematically performed caesarian deliveries at E18.5. Under these conditions, KO embryos were indistinguishable from their littermates when handled carefully (Fig. 2B). However, during grooming by surrogate mothers we observed the rapid appearance of skin detachment that was similar to that observed after natural birth. Skin rupture started in the ventral area, paws and snout and extended to the flanks (Fig. 2C). The peeled skin was red, shiny, and became somewhat sticky to the touch. On the dorsal part of the pups, the skin retained a normal macroscopic aspect until death, which occurred within 2 hours. A standard procedure for analysis of the phenotype was thus established: after caesarian delivery, all neonates were given to a surrogate mother and euthanized after 1 hour.

Inactivation of CDSN in mouse skin leads to a drastic barrier defect

To measure the effects of the mutation on skin permeability, we performed a Toluidine blue penetration assay on E18.5 embryos (Fig. 3A). When the assessment was carried out immediately after caesarian delivery, before any skin phenotype onset, complete dye exclusion was observed in KO pups, as in their control littermates, reflecting an intact barrier at that stage (Fig. 3A, left panels). However, when mice were given to a surrogate mother for 1 hour, large areas showed toluidine blue penetration, particularly in the ventral part of the embryos (Fig. 3A, right panels). Thus, as soon as a skin phenotype became macroscopically visible in KO animals, the dye extensively penetrated the skin. This drastic perturbation of the barrier could be perfectly superimposed with the zones of detachment of the stratum corneum in KO animals (see Fig. 2C). Similarly, transepidermal water loss (TEWL) measured on whole pups was similar at first in WT, HET and KO pups immediately after caesarian delivery but dramatically increased by a factor greater than ten in KO mice after phenotype onset (Fig. 3B). To maintain conditions of minimal mechanical stress, KO neonates were isolated at 37°C for 4 hours, immediately after caesarian section. Under these

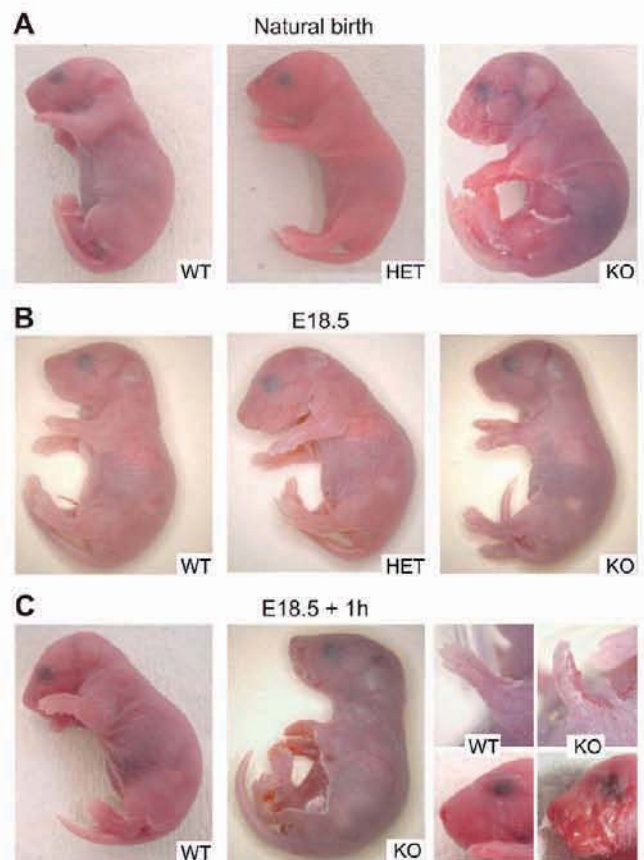


Fig. 2. Mice lacking CDSN show a drastic skin tear leading to postnatal death. (A) Macroscopic appearance of WT, HET and KO pups born at full term. KO pups die within 1 hour of birth and show considerable skin detachment, particularly in the ventral area, paws and snout, whereas HET newborns show no visible difference from WT. (B) Macroscopic appearance of WT, HET and KO mice delivered by caesarian section at E18.5. Immediately after birth, KO neonates are indistinguishable from the WT and HET littermates. (C) When neonates were left in surrogate mother's care for 1 hour, the skin phenotype of KO pups appears and death occurs rapidly. Higher magnifications of paw and snout of WT and KO neonates are shown on the right.

conditions, only one out of nine KO pups died. The remaining pups did not develop any noteworthy skin detachment, and five of them maintained a TEWL value that was roughly similar to that of the WT and HET newborns, between 0 and 4 g m⁻² per hour. These experiments demonstrate that the skin phenotype is closely linked to mechanical stress from the environment. To assess the mechanical resistance of the skin, the amount of protein detached from the skin by tape stripping was measured. Three times more protein was extracted from KO skin than from WT or HET skin (Fig. 3C). Hence, even macroscopically intact KO skin is extremely fragile, and tears under mechanical stress.

Histological analysis of KO mouse skin reveals a detachment of the stratum corneum from the living layers of epidermis
In zones of skin rupture, such as the ventral region, a complete absence of the stratum corneum was observed (Fig. 4A). In some areas, the skin abrasion was more pronounced and also concerned part of the living layers (data not shown). On the dorsal skin, the

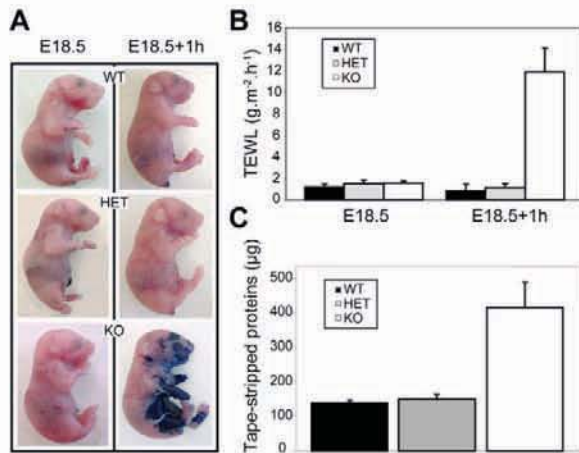


Fig. 3. Mice lacking CDSN exhibit a drastic barrier defect and decreased mechanical resistance of the skin. (A) Barrier-dependent dye exclusion assay (Toluidine blue) performed on WT, HET and KO neonates. Pups obtained by caesarian delivery (E18.5) were placed with surrogate mother for 1 hour until the skin phenotype appears (E18.5+1h). (B) TEWL assay performed on whole WT, HET or KO neonates immediately after caesarian delivery (E18.5) and after phenotype onset (E18.5+1h). (C) Mechanical resistance assay of the stratum corneum from neonates. Tape stripping was performed on the dorsal skin of WT, HET or KO neonates before phenotype onset, and the proteins present on the tape were quantified by colorimetric assay. Values are mean \pm s.e.m.

stratum corneum was still present, and looked similar to that of the WT. However, even if the macroscopic aspect was normal, some blisters were observed between the stratum granulosum and stratum corneum (data not shown). Although we could not discriminate between blisters induced by natural mechanical stress or during sample preparation, our results suggest an extreme fragility of the intercellular links. No acanthosis or hyperkeratosis was observed in either dorsal or ventral KO epidermis. Immunohistochemical labeling of skin sections with F28-27 mAb confirmed the complete absence of CDSN in the stratum granulosum and IRS of hair follicles in KO mice (Fig. 4B). Transcriptional activity of the human *KRT14* promoter in transgenic mice is largely restricted to external epithelia, with a reduced expression in more internal stratified tissues (Wang et al., 1997). In accordance with these findings, CDSN was still detected in the tongue, the hard palate and the esophagus of KO newborns (Fig. 4C; and data not shown).

Corneodesmosomes deficient for CDSN display normal morphology but impaired functionality

Transmission electron microscopy performed on dorsal skin of WT and KO neonates revealed a similar organization of the granular layer. The thickness of stratum corneum from KO mice was also unchanged, with 10 to 13 layers (Fig. 5A,B). Some blisters, already detected by optical microscopy, were present between the stratum granulosum and stratum corneum (Fig. 5J; and data not shown). The ultrastructure of desmosomes and corneodesmosomes was analyzed at higher magnification. In granular keratinocytes from both WT and KO animals, the same characteristic symmetrical trilamellar structure of regular desmosomes was observed (Fig. 5C,F). Morphological changes accompanying cornification occurred similarly in both WT and KO mice. Typical transitional junctions were found, with classical dense fibrils of keratin filaments attached

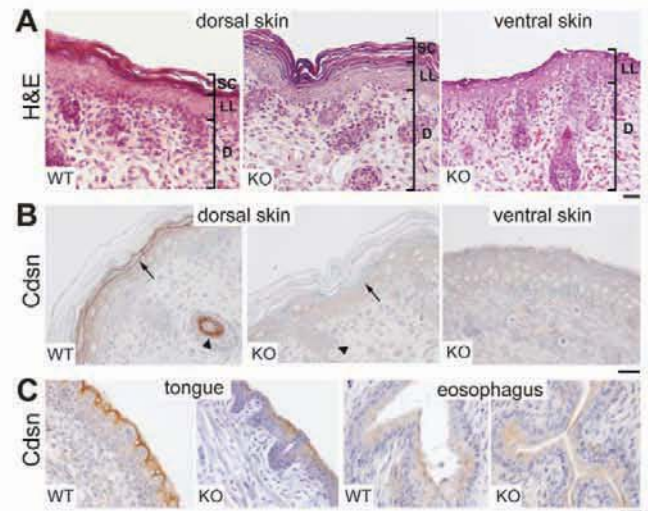


Fig. 4. Histological analysis of cornified epithelia in *Cdsn* KO mice. (A) Representative sections stained with hematoxylin and eosin (H&E), from dorsal skin of WT and KO neonates, or from ventral skin of KO neonates. The staining shows complete detachment of the SC from the ventral epidermis of the KO mouse. D, dermis; LL, living layers of epidermis; SC, cornified layer. (B) Immunohistochemical staining with F28-27 mAb (CDSN) of sections from dorsal skin of WT and KO neonates, or from ventral skin of KO neonates, confirmed the absence of CDSN in the stratum granulosum of the epidermis (arrows) and the IRS of the hair follicles (arrowheads) of KO pups. (C) Immunohistochemical staining with F28-27 mAb (CDSN) of sections from tongue and esophagus of WT and KO neonates showing that *Cdsn* excision is not efficient in internal epithelia. Scale bars: 20 μ m.

to the electron-dense intracellular desmosomal plaque in the granular keratinocyte, densification of the intercellular portion, and an electron-dense line at the other side of the junction corresponding to the cornified envelope of the upper transitional keratinocyte (Fig. 5D,G). The number of these transitional desmosomes at the interface between the stratum granulosum and stratum corneum was not significantly different in WT and KO samples. We made counts on 30 fields at magnification $\times 20,000$, corresponding to the length of 3-4 corneocytes, and found 59 transitional desmosomes in the KO skin section versus 61 in the WT. Hence, in newborn KO mice, *Cdsn* deficiency seems to have no consequences on the transformation of desmosomes into corneodesmosomes. In the cornified layers, the morphology of the corneodesmosomes from KO mouse also appeared classical (Fig. 5E,H). By contrast, numerous split desmosomes could be observed at the transition from stratum granulosum to stratum corneum, either isolated (Fig. 5I) or adjacent in zones of detachment (Fig. 5J). The main structure of the split transitional desmosomes remained attached to the granular keratinocyte, and no cell lysis was observed. Desmosome breaking seemed to occur specifically at the stratum granulosum to stratum corneum transition, because no split corneodesmosomes were observed higher in the stratum corneum. Thus, abnormalities due to the absence of CDSN seem to occur only when cornification takes place, and lie in adhesive defects.

Absence of CDSN does not affect expression of differentiation markers and desmosomal proteins in neonatal epidermis
Histological analysis revealed no thickening of the living layers in the epidermis from KO neonate dorsal skin. To further investigate

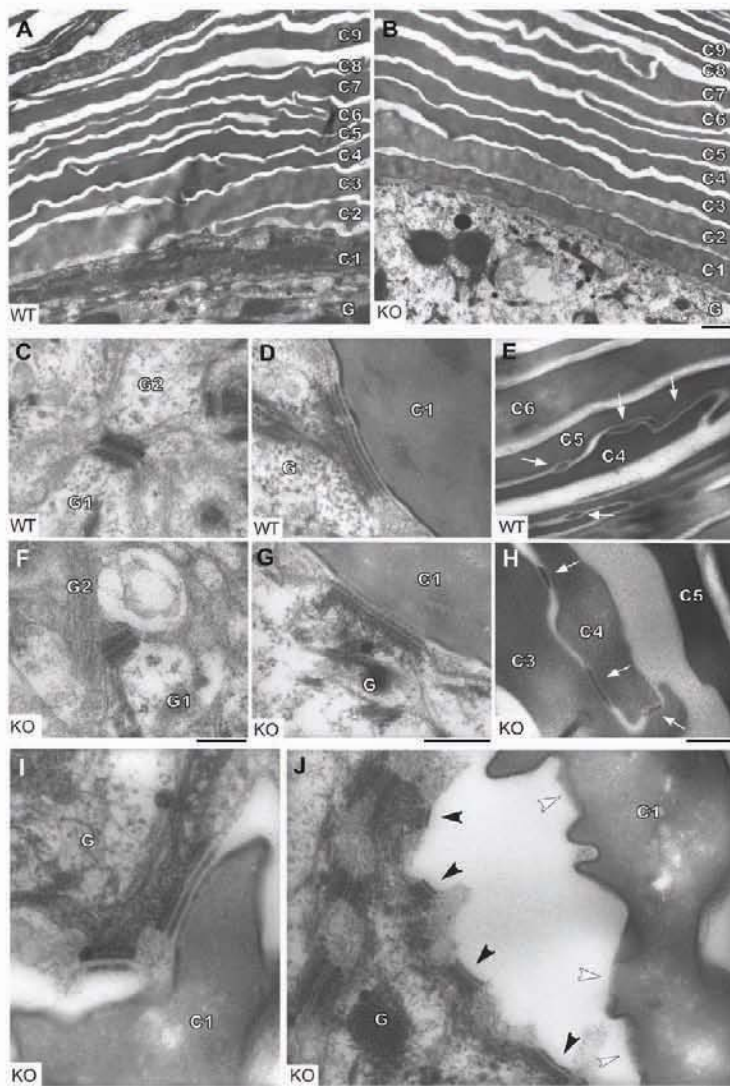


Fig. 5. Ultrastructural abnormalities in KO neonate epidermis. Dorsal skin of WT and KO mice was processed for transmission EM. (A,B) At low magnification, the cells at the junction between the granular and cornified layers have no obvious abnormality. G, G1, G2, granular keratinocytes; C1 to C9, corneocytes, with C1 being adjacent to the granular layer and C9 being nearest the skin surface. (C-H) The ultrastructure of desmosomes and corneodesmosomes from WT (C-E) and KO (F-H) epidermis analyzed at higher magnification. Regular desmosomes in the stratum granulosum (C,F), transitional desmosomes at the stratum granulosum to stratum corneum interface (D,G), and corneodesmosomes (white arrows) in the stratum corneum (E,H), show the same classical morphology in WT and KO epidermis. (I,J) Detachment between granular keratinocyte and corneocyte is frequently observed in KO epidermis. Split desmosomes isolated (I) or adjacent (J) can be seen. The mid-dense intercellular plugs of the detached transitional desmosomes remain associated with the granular keratinocyte (black arrowhead), whereas an interruption in the electron-dense line corresponding to the cornified envelope is clearly visible at the surface of the corneocyte, at the site where the desmosome was attached (white arrowhead). Scale bars: 2 μ m (A,B), 0.2 μ m (C-I).

the effect of the null mutation on epidermal differentiation, we examined the expression of differentiation markers and desmosomal proteins by western blot, quantitative real-time PCR and immunohistochemistry (Fig. 6). No significant differences in the amount of early (keratin K10, involucrin, desmoglein 1) and late (loricrin, filaggrin) differentiation markers were observed in skin from WT, HET or KO neonates by western blot analysis (Fig. 6A). Expression level analysis by qRT-PCR of *Dsc1*, *Dsg1a* and *Dsg1b*, coding for desmosomal proteins present in the suprabasal layers of epidermis, did not reveal any differences between the dorsal skin of WT and KO neonates (Fig. 6B). Immunohistochemical analyses (Fig. 6C) were consistent with the western blot and qRT-PCR data. Again, no differences in the distribution or the expression level of differentiation markers were observed between WT and KO neonate dorsal skin. As expected, we observed a dramatic reduction in the expression level of all the differentiation markers in the ventral area of KO animals (data not shown), which is consistent with the sudden loss of the most suprabasal keratinocytes once the stratum corneum had been shed (see Fig. 4A). In conclusion, the absence of CDSN does not affect the expression of differentiation markers and

desmosomal proteins as long as the integrity of the epidermis is preserved.

Long-term phenotype of mouse skin lacking CDSN after grafting onto nude mice

The drastic barrier defect that occurs in KO mice does not allow the long-term consequences of *Cdsn* excision to be analyzed in the epidermis and hair follicle. To address this issue, full-thickness dorsal skin from WT and KO neonates was grafted onto the back of nude mice (Fig. 7). Absence of CDSN in the epidermis and IRS of hair follicles from KO grafts was controlled by immunohistochemistry (data not shown). Macroscopically, newborn KO skin developed some sparse hair that progressively disappeared and became totally absent 9 weeks after grafting. In parallel, the graft became hard, devoid of softness, and formed a scab that was continually renewed (Fig. 7A). In contrast to the WT, KO epidermis first showed a prominent hyperplasia with papillomatosis (Fig. 7B). Hyperkeratosis and parakeratosis were obvious and infiltration of inflammatory cells was observed in the dermis, together with dermal cysts, which characteristically result from degenerative hair follicles

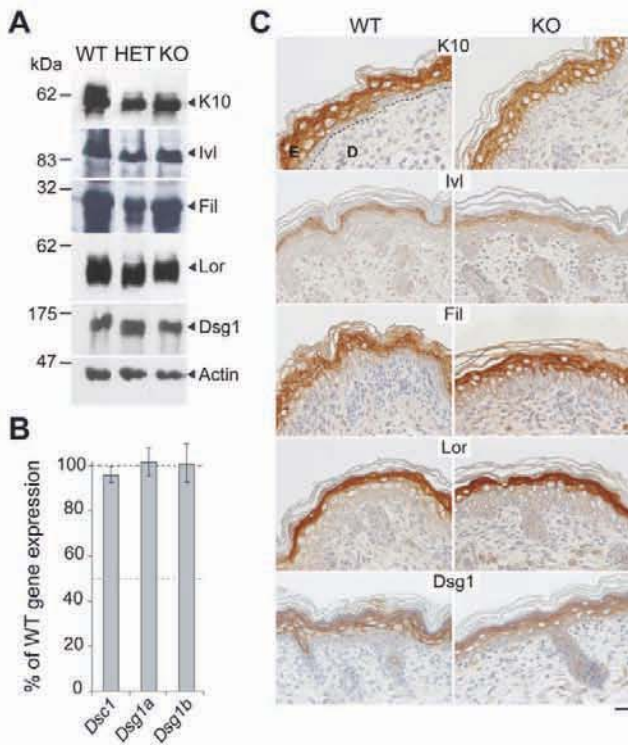


Fig. 6. Expression analysis of epidermal differentiation markers and desmosomal proteins in neonatal KO mice. (A) Equal amounts of protein extract from WT, HET and KO neonate dorsal skin separated by SDS-PAGE and analyzed by western blot with antibodies directed against keratin 10 (K10), involucrin (Ivl), filaggrin (Fil), loricrin (Lor), desmoglein 1 (Dsg1) or actin. (B) Quantitative RT-PCR analysis of genes encoding desmosomal proteins. *Dsc1*, *Dsg1a* and *Dsg1b* expression analyzed in dorsal skin biopsies from KO neonates. Values are shown relative to RNA level in the WT, after normalization with *Dsg3* transcripts, to take the epidermal area of each sample into account. Values are mean \pm s.e.m. (C) Immunohistochemical labeling of dorsal skin sections from WT and KO neonates using antibodies specific to keratin 10 (K10), involucrin (Ivl), filaggrin (Fil), loricrin (Lor) and desmoglein 1 (Dsg1) shows a similar expression level and localization of the differentiation markers in WT and KO pups. D, dermis; E, epidermis; dotted line, dermoepidermal junction. Scale bar: 20 μ m.

(Fig. 7D). Concomitantly, immunohistochemistry revealed high expression levels of keratin K6 and increased expression of several epidermal differentiation markers, such as involucrin, filaggrin, loricrin and K10 (Fig. 7C,E, left panel; and data not shown). From 8 weeks after grafting, despite its hyperproliferative state, the epidermis started to disappear in some central areas, replaced by a scab directly covering the dermis (Fig. 7B). By 9 weeks post grafting, this phenomenon had amplified, ending with the total absence of the grafted epidermis. In parallel, hair follicles became scarce and then totally disappeared. Concomitantly, the epidermis of the nude mice located at the edge of the initial graft became thicker, producing keratinocytes that seemed to migrate to the grafted zone. Interestingly, these keratinocytes strongly expressed the hyperproliferative keratin K6 (Fig. 7C,E, right panel). They also highly expressed CDSN, confirming that they originated from nude epidermis (data not shown). Hence, *Cdsn* deficiency resulted in chronic ulceration of the grafted skin. To study potential barrier defects at the level of the graft, TEWL measurements were performed. Values measured from 4 to 9 weeks post grafting ranged

from 1.5 to 5 g m^{-2} per hour for the WT grafts versus 11.5 to 22.5 g m^{-2} per hour for the KO grafts, suggesting a strong permeability defect. Histological and immunohistological data suggested that the KO graft first behaved like a healing epidermis. However, whereas wound healing leads to barrier restoration and a return of keratinocytes to a normal differentiation program, the KO graft was unable to produce an efficient barrier. However, the epidermis of nude mice at the border of the scab showed acanthosis and perfectly mimicked a healing epidermis.

CDSN is also required in adult mice for epidermis integrity and barrier function

To investigate the consequences of CDSN loss in adult mice, we developed an additional model using a *K14*-promoter driven CreER^{T2} recombinase, the activity of which is efficiently induced by 4-hydroxy-tamoxifen (OHT), but not natural estrogen receptor ligands (Li et al., 2000). Mice homozygous for the floxed *Cdsn* allele and transgenic for *K14CreER^{T2}tg/0* were generated following the same breeding protocol described previously. Resulting animals were wild-type mice (*K14CreER^{T2}tg/0*; *Cdsn^{w/w}*) or *K14CreER^{T2}tg/0*; *Cdsn^{f/f}*, heterozygous pre-mutant mice (*K14CreER^{T2}tg/0*; *Cdsn^{w/f}*) and homozygous pre-mutant mice (*K14CreER^{T2}tg/0*; *Cdsn^{f/f}*, denoted hereafter HOMO). In the absence of treatment, HOMO mice did not develop any skin or hair phenotype for at least 6 months, denoting no leakiness of uninduced CreER^{T2}. Generalized or localized temporally controlled *Cdsn* excision was achieved in the skin of adult HOMO mice by intraperitoneal injection of tamoxifen (Tam) or topical application of OHT, respectively. In both cases, immunodetection of CDSN revealed a loss of *Cdsn* expression in epidermal keratinocytes and in the IRS of hair follicles (Fig. 8C; Fig. 9C). HOMO mice injected with Tam or locally treated with OHT developed a skin phenotype with variable kinetics according to the individuals but, once the phenotype had initiated, it progressed within 5 days to a critical situation. Wild-type mice treated in the same conditions never developed any phenotype, denoting the absence of toxicity of Tam or OHT in our experimental conditions. We attributed the variation in the kinetic of the HOMO mice phenotype onset to the genetic background heterogeneity of the mice. For this reason, we used TEWL values to evaluate the onset and progress of the phenotype.

Typically, TEWL values reached 4–5 g m^{-2} per hour when the first scales became visible, and rose to 33 g m^{-2} per hour as the scales became more numerous and were shed. After generalized induction of *Cdsn* deletion, the scales were first visible on the ventral side and snout, and then extended to the back and the whole body, consistent with the level of mechanical stress applied to these different body areas. Within 5 days of the abnormal skin phenotype onset, mice became prostrate, lost weight and were euthanized. A representative aspect of a ventral skin region with a TEWL value of 13 g m^{-2} per hour is presented in Fig. 8. At the histological level, acanthosis, hyperkeratosis and parakeratosis of the epidermis accompanied scaling (Fig. 8A,B,D, left panel). Topical applications of OHT were made, with the aim of analyzing the phenotype for a longer period of time (Fig. 9). Skin phenotype progression is presented at early (Fig. 9, middle panels) or late (Fig. 9, right panels) stages in comparison with the control (Fig. 9, left panels). The early stage showed a moderate barrier defect (TEWL of 11 g m^{-2} per hour) associated with the development of acanthosis, hyperkeratosis and parakeratosis, together with a strong expression of the hyperproliferative keratin K6 (Fig. 9, middle panel). At the late

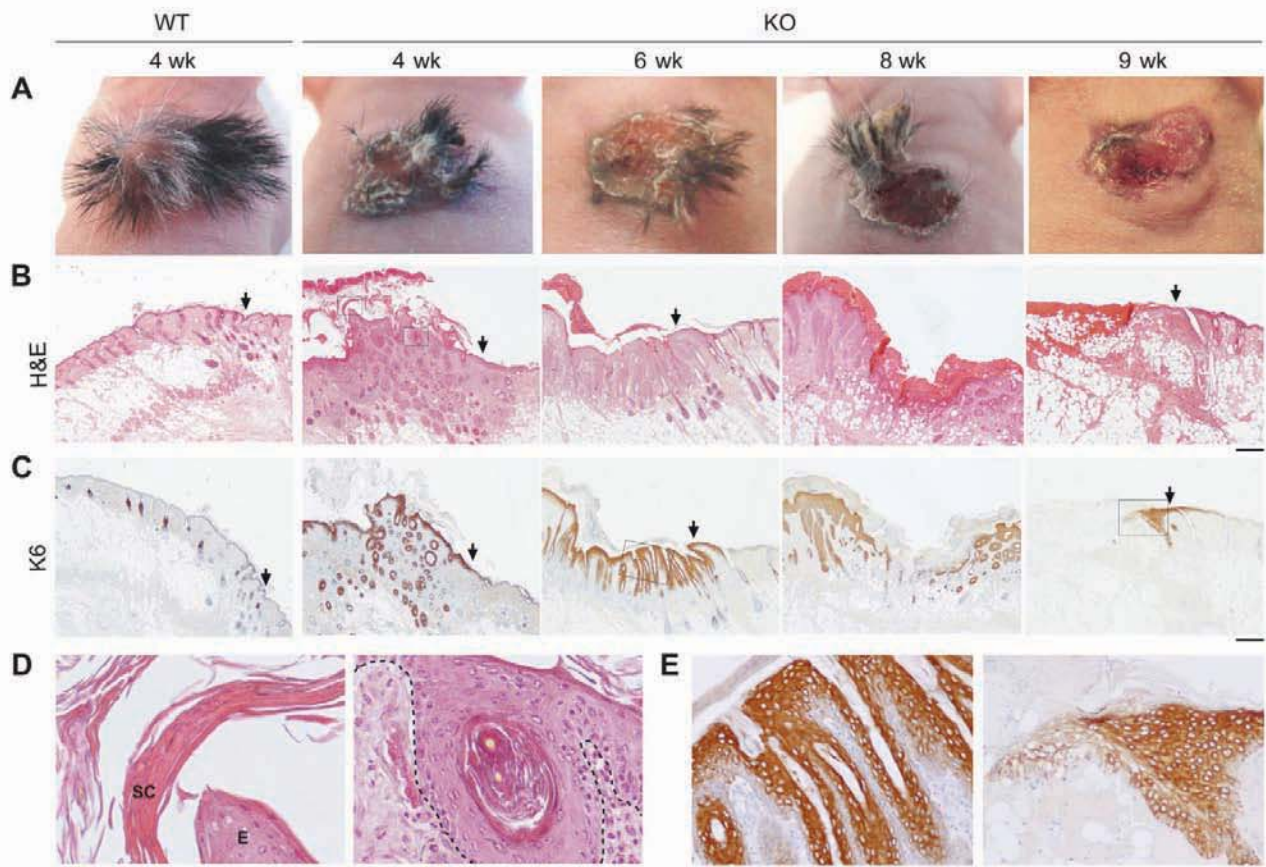


Fig. 7. Transplantation of *Cdsn*-deficient skin onto nude mice. Dorsal skin from WT or KO neonates was grafted onto the back of nude mice. Four weeks after grafting is shown as a representative stage for the WT graft, whereas 4, 6, 8 and 9 weeks after grafting are shown for the KO graft. (A) Macroscopic appearance of the transplanted skin. Grafts from KO mice develop few hairs, which become rarer and finally disappear. The skin is red, forms a scab and looks like a chronic healing area. Sections of WT or KO skin grafted onto nude mice were stained with hematoxylin and eosin (H&E) (B) or processed for immunohistochemistry with an antibody specific to K6 (C). Morphology of the grafted skin from WT mice is similar to that of adult mice skin. From 4 to 8 weeks, the KO grafted epidermis becomes acanthotic and develops prominent papillomatosis. The increasing K6 expression in epidermal keratinocytes confirms its hyperproliferative state. The KO grafted epidermis also develops prominent hyperkeratosis and parakeratosis. In the KO grafted skin, hair follicles become cystic then progressively disappear. They are totally absent by 9 weeks after grafting. Inflammatory infiltrate in the dermis from KO skin is obvious by 4 weeks after grafting, and diminishes in the course of the following weeks. Nine weeks after grafting, only a scab persists at the place of the KO grafted skin. K6 is only detected at the border of the graft, where the epidermis of the nude mice is thicker, denoting its hyperproliferative state. In B and C, the graft border is indicated with a vertical arrow. (D) Histological higher magnifications corresponding to the areas boxed in B show hyperkeratosis and parakeratosis (left) and dermal cyst formation (right). (E) Higher magnifications of skin sections after immunostaining with K6, corresponding to areas boxed in C, show hyperproliferative region (left) and edge of the KO graft (right). SC, stratum corneum; E, epidermis. Scale bars: 100 μ m (B,C), 50 μ m (D,E).

stage, corresponding to a TEWL value of 33 g m⁻² per hour, ulcerations appeared and were associated with the shedding of large scales; histological analysis revealed almost complete disappearance of the epidermis, replaced by a scab (Fig. 9, right panel). At the early stages, moderate inflammation was observed, which we further characterized by immunohistochemistry for CD3 and F4/80 (supplementary material Fig. S1). It mainly consisted of CD3⁺ T cells infiltrating the epidermis (supplementary material Fig. S1A) and macrophages and Langerhans cells in the dermis, at proximity of the epidermis (supplementary material Fig. S1B). The increasing barrier defect led to a typical inflammatory response, with ulceration, dilation of capillaries and subcomeal abscesses with neutrophils (see Fig. 9B, right panel).

Localized temporally controlled *Cdsn* excision in adult mouse skin also induced a notable morphological modification of the hair follicles. At early stages, they appeared to be distended and

associated with hypertrophic sebaceous glands (Fig. 9B, middle panel). This was also the case when generalized *Cdsn* excision was induced (Fig. 8D). At later stages, hair follicles disappeared, as evidenced by the absence of K6 labeling (Fig. 9B,D, right panel). These observations are reminiscent of the neonatal-skin-grafted model, which is characterized by cyst formation and hair-follicle degeneration.

Discussion

In this study, we performed conditional ablation of *Cdsn* in mouse skin using *K14*-promoter driven Cre-mediated *loxP* recombination. Our results demonstrated that CDSN is necessary for epidermis integrity and barrier function in adult and neonatal mouse skin. *Cdsn* is also excised in the IRS of the hair follicles. At birth, hair follicles from KO mice appeared normal. However, they degenerated in KO

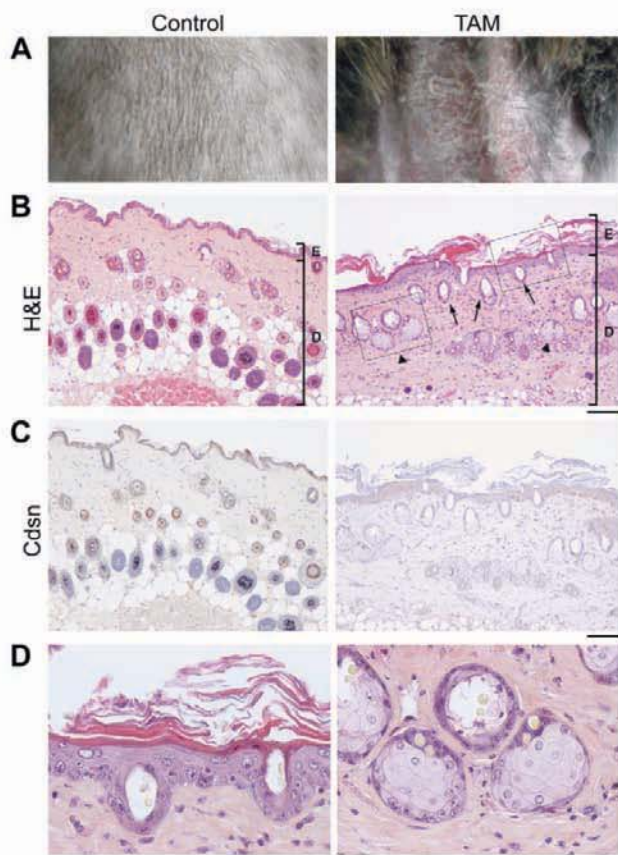


Fig. 8. Phenotype of generalized temporally controlled *Cdsn* KO in adult mouse skin. Shaved adult HOMO mice were injected with vehicle only (control) or with Tam (TAM) for 5 consecutive days, and checked daily for phenotype onset by clinical observation and TEWL measurement. (A) Macroscopic appearance of the mice. The ventral skin of the control mouse has no phenotype, whereas scales are visible on the ventral area of the TAM mouse. (B) Representative skin sections from ventral areas of control and TAM mice were stained with hematoxylin and eosin (H&E), showing acanthosis, hyperkeratosis and parakeratosis of the TAM mouse epidermis. In the dermis of TAM mouse, the hair follicles display an altered morphology; their upper part from the infundibulum to the skin surface is distended (arrows), and they are associated with hypertrophic sebaceous glands (arrowheads). D, dermis; E, epidermis. (C) Immunohistochemical staining with F28-27 mAb (CDSN) of ventral skin sections from control and TAM mice shows an effective loss of CDSN expression in the stratum granulosum of the epidermis and the IRS of the hair follicles in TAM mouse. (D) Higher magnification corresponding to the areas boxed in B shows in detail hyperproliferation, hyperkeratosis and abnormally distended hair follicles (left panel), hypertrophic sebaceous glands (right panel). Scale bars: 100 μ m (B,C), 20 μ m (D).

grafted skin as well as in the adult KO models, revealing that CDSN is also required for hair-follicle integrity.

Cdsn inactivation is useful to examine the function of CDSN within the corneodesmosome. As previously shown (Matsumoto et al., 2008), the major consequence observed here was an epidermal tear at sites of trauma and friction. Our detailed analysis of the neonatal phenotype showed that, although the greatly reduced mechanical resistance of the stratum corneum is an intrinsic feature caused by *Cdsn* deficiency, epidermal tear, leading to lethal barrier defect, occurred only under mechanical stress encountered after birth. At the histological level, blisters were found as soon as

cornification occurred, that is, at the stratum granulosum to stratum corneum transition. Ultrastructural analysis revealed numerous split junctions with the main structure remaining attached to the granular keratinocytes, suggesting that the cohesive defect lies in the upper side of the junction. Thus, although CDSN is already present in the extracellular core of desmosomes from the stratum granulosum, it seems to have a fundamental role only when cornification is complete. Intriguingly, we did not observe thinning of single or bundled corneocytes throughout the stratum corneum, but detachment of the whole stratum corneum from the subjacent stratum granulosum. The stratum granulosum to stratum corneum interface might actually be the most fragile zone because it links two epidermal layers with different junctional organizations and mechanical characteristics: desmosomes and keratin intermediate filaments that are organized in taut cables in the stratum granulosum and rigid cornified envelopes linked by corneodesmosomes in the stratum corneum.

The underlying molecular mechanism by which CDSN assumes its adhesive function within the corneodesmosome has not been fully elucidated. In vivo, CDSN is known to be covalently linked to the cornified envelope (Serre et al., 1991). On the one hand, the previously demonstrated homophilic adhesive properties of CDSN and the strong resistance of aggregates formed by bacterially recombinant CDSN to highly denaturing conditions (Jonca et al., 2002; Caubet et al., 2004) suggest that, in vivo, CDSN reinforces cohesion by its own adhesive properties. On the other hand, adhesion provided by glycine-loop domains has been suggested to mediate reversible and constantly adjustable intermolecular links similar to Velcro (Steinert et al., 1991). Consistent with this, CDSN could give the junction the elasticity essential to prevent breaking as soon as the cell envelope rigidifies. Unlike the recent *Cdsn* inactivation study (Matsumoto et al., 2008), our ultrastructural analysis did not show any significant differences in the number of transitional desmosomes between WT and KO neonates. Moreover, the electron density of the corneodesmosomes appeared unchanged (Fig. 5). Thus, CDSN does not seem indispensable for the morphogenesis of corneodesmosomes, but appears essential to their function as adhesive structures.

Our *Cdsn* KO mice developed a phenotype that was very close to that displayed by *Spink5*^{-/-} mice deficient for the serine protease inhibitor Lekt1. These mice exhibited fragile stratum corneum and perinatal death due to dehydration, with detachment of the stratum corneum from the subjacent epidermis layers and split desmosomes (Yang et al., 2004; Descargues et al., 2005). It was proposed that, in the absence of Lekt1, the premature proteolysis of CDSN or Dsg1 contributed significantly to desmosomal fragility, epidermal detachment and skin-barrier defects. It is interesting to note that, in both models, the breaking points mainly lie in the junction between living and cornified layers. This emphasizes the importance of corneodesmosomes as cornification takes place. However, the timing of the barrier defect differs between the two models: E18.5 for *Spink5*^{-/-} mice but only after birth for our somatic *Cdsn*^{-/-} mouse model. It is tempting to suggest that in the *Spink5*^{-/-} model, the biochemical defect (absence of protease inhibitor) induces a dysfunction of the corneodesmosomes in utero, as soon as cornification occurs. By contrast, the absence of CDSN leads to a mechanical deficiency and thus the barrier impairment only takes place in the environment experienced after birth.

Grafting experiments demonstrated that the barrier defect extended postnatally. Skin from *Cdsn*-deficient grafts first developed acanthosis and hyperkeratosis. Increased expression of various

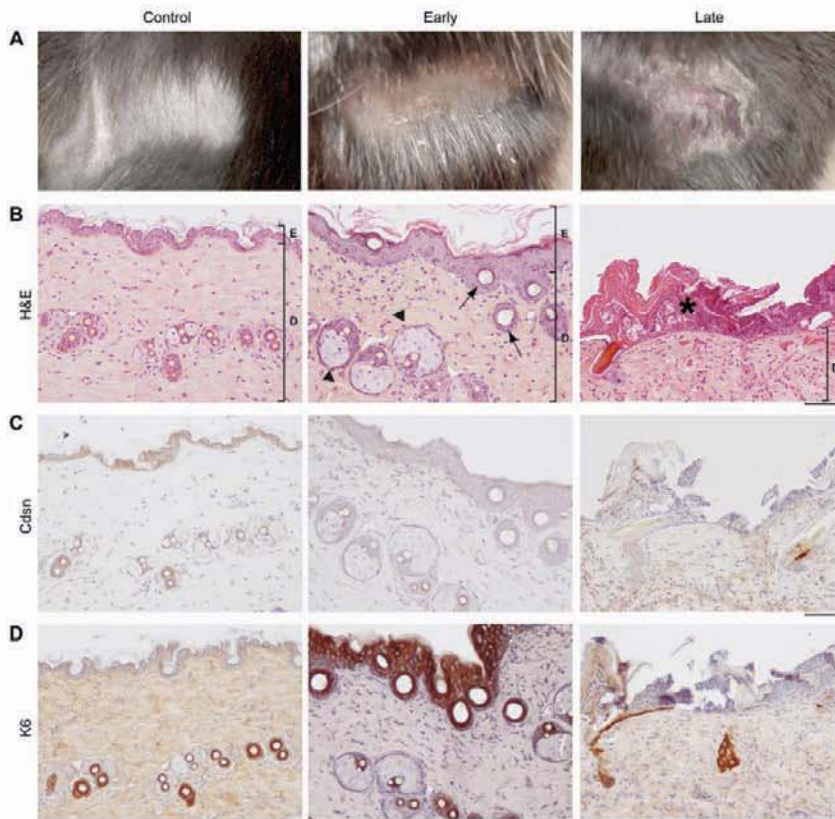


Fig. 9. Phenotype of localized temporally controlled *Cdsn* KO in adult mouse skin. Adult HOMO mice were shaved on both flanks and OHT was applied on the right flank for 5 consecutive days. The treated areas were checked daily for phenotype onset by clinical observation and TEWL measurement. (A) Macroscopic appearance of untreated left flank (control) or OHT-treated right flank from representative animals. Right flank of mice euthanized when the phenotype was moderate (early) or more severe (late), are presented. The scaling becomes more marked and ulceration of the skin appears as the phenotype progresses. (B) Representative skin sections from control, early or late flanks stained with hematoxylin and eosin (H&E). At early stage (middle panel), note the development of acanthosis, hyperkeratosis and parakeratosis, as well as the altered morphology of the hair follicles, which are distended from the infundibulum to the skin surface (arrows) and associated with hypertrophic sebaceous glands (arrowheads). At the late stage (right panel), hair follicles are no longer visible and a scab replaces the epidermis (asterisk). D, dermis; E, epidermis. (C,D) Skin sections from control, early or late flanks processed for immunohistochemical staining with F28-27 mAb (CDSN) (C) or with an antibody specific to K6 (D). Absence of labeling in early and late flanks confirms the efficient *Cdsn* excision in HOMO mouse skin treated with OHT, and the strong K6 expression confirms the hyperproliferative state of the HOMO mice epidermis. Scale bars: 50 μ m.

differentiation markers (involucrin, K10), and induced expression of K6, confirmed the altered differentiation and the hyperproliferative state of the grafted epidermis. Hyperproliferation and acanthosis are thought to be compensatory responses to impaired epidermal barrier (Proksch et al., 1991). In the case of our *Cdsn*-deficient graft, these compensatory mechanisms appeared to be ineffective, and were followed by a complete disappearance of the epidermis. This demonstrates that CDSN is necessary for maintaining the integrity and barrier function of the postnatal epidermis. In our inducible models, the epidermal phenotype was very similar to that of our *Cdsn* KO skin graft model, although it progressed faster (within days). This might be due to the different location of the zone of interest, the interscapular area in the case of the grafts, and the flank in the inducible model, which is subjected to higher mechanical stress. Finally, in our inducible *Cdsn* KO models, barrier restoration remained unsuccessful, compromising the vital prognosis of the mice when the whole skin area was affected. Thus, CDSN seems to have a vital role in adult epidermis too.

Acanthosis, hyperkeratosis and parakeratosis are histological features more generally encountered in inflammatory epidermis, such as during wound healing or in psoriatic lesions. Although wound healing is transitory and rapidly leads to orthokeratotic epidermis and barrier repair, psoriatic epidermis maintains acanthosis, hyperkeratosis and parakeratosis with barrier defect. The situation has been described as a chronically persistent epidermal-healing state (Nickoloff et al., 2006). Our model of induction of *Cdsn* KO in adult mouse showed a persistent barrier defect similar to that observed in psoriasis. However, although we observed

acanthosis, hyperkeratosis and parakeratosis at early stages, there was no formation of prominent and elongated rete ridges, only a weak inflammatory cell infiltrate, and the phenotype evolved towards ulceration and disappearance of the epidermis. In psoriasis, the barrier defect is thought to be associated with intrinsic immunological defects. The lack of any distinctive feature at the immunological level in our inducible KO model might contribute to the differences in the intensity of the inflammatory response. Finally, unlike our *Cdsn*-deficient mouse models, psoriasis is characterized by a significant increase in CDSN expression which is observed in multiple living layers and stratum corneum from lesional psoriatic epidermis (Hafték et al., 1997; Allen et al., 2001). However, the psoriasis-associated SNPs of *CDSN* could affect the function of the corresponding proteins, thus participating in the development of hyperproliferation as a compensatory mechanism to the barrier impairment. Assessment of the functional consequence of the amino acid substitution in alleles unique to risk haplotypes is needed to further investigate the putative involvement of CDSN in psoriasis pathophysiology.

Spatially controlled somatic *Cdsn* inactivation in mouse epidermis, using *K14*-promoter driven Cre-mediated deletion, also allowed inactivation of *Cdsn* in the developing hair follicle. Mice heterozygous for the excised *Cdsn* allele were indistinguishable from their wild-type littermates. In particular, they did not develop any hair phenotype up to 8 months. This suggests that the rare autosomal dominant human disease HSS is not caused by *Cdsn* haploinsufficiency. Moreover, recent work demonstrated that truncated CDSN, which accumulates abnormally as high-order

assemblies in the dermis of patients, organizes into proto-fibrillar structures in vitro and is toxic to keratinocytes (Cécile Caubet, Luc Bousset and G.S. et al., unpublished). Common features between CDSN mutant aggregates and amyloid deposits support HSS as a new potential amyloid disease and allow a better understanding of the HSS pathophysiology. Analysis of skin sections from newborn *Cdsn* KO mice showed a similar morphology and number of hair follicle to those in skin from WT littermates. The KO neonates also displayed normal whiskers. The only difference was that, in KO pup skin, no hair follicles expressed CDSN in the IRS. Therefore, CDSN seems to be dispensable for hair-follicle morphogenesis, as was the case for barrier establishment in the absence of external stress. The progressive degeneration of hair follicles and the formation of cysts observed from 4 weeks after grafting, together with the altered morphology and disappearance of hair follicles in adult mouse skin induced for *Cdsn* excision, suggest that CDSN is necessary for normal hair-follicle integrity. Mutation of another gene encoding a structural protein of the IRS, *DSG4*, is responsible for localized autosomal recessive hypotrichosis (LAH; OMIM 607903) (Kljuic et al., 2003). LAH is allelic with the lanceolate hair (lah) mouse, which fails to grow any normal hair and completely lacks vibrissae. Finally, the inactivation of genes encoding other desmosomal cadherins also leads to hair-follicle abnormalities. Mice deficient for *Dsg3* or *Dsc3*, normally detected in the outer root sheath of the hair follicle, suffer from acantholysis between the two cell layers surrounding the telogen hair club, leading to telogen hair loss (Koch et al., 1998; Chen et al., 2008). Absence of *Dsc1* in the IRS of hair follicles from *Dsc1*^{-/-} mice induces localized hair loss associated with formation of utricular and dermal cysts denoting hair-follicle degeneration (Chidgey et al., 2001). Altogether, these data emphasize the importance of desmosomes and corneodesmosomes in hair-follicle integrity. In our different models, the hyperproliferative and inflammatory environment could aggravate the hair-follicle phenotype. A mouse model with specific inactivation of *Cdsn* in the IRS would be useful for an in-depth study of the consequences of *Cdsn* deletion on hair-follicle integrity.

Materials and Methods

Mice, tamoxifen treatment and skin grafting

Animals were handled according to the institutional guidelines and policies. The *Cdsn* mutant mouse line was established at the MCI/ICS (Mouse Clinical Institute/Institut Clinique de la Souris, Illkirch, France; <http://www-mci.u-strasbg.fr>). The targeting vector was constructed as follows. A 3 kb fragment encompassing most of the coding region (exon 2) and corresponding to the floxed region was amplified by PCR (from 129S2/SvPas genomic DNA) and subcloned in an MCI proprietary vector, resulting in step 1 plasmid. The 5' (4.7 kb) and the 3' (3.5 kb) homologous arms were subcloned successively into the step 1 plasmid to generate the final targeting vector. All the coding regions were sequenced. The MCI vector has a flipped neomycin-resistance cassette. The linearized construct was electroporated in 129S2/SvPas mouse embryonic stem (ES) cells. After selection, targeted clones were identified by PCR using external primers and further confirmed by Southern blot with both Neo and 5' external probes. One positive ES clone was injected into C57BL/6J blastocysts, and derived male chimeras gave germline transmission. Chimeric offspring were mated to FLPer (flipper) mice that express the Flp recombinase to excise the neomycin selection marker that was flanked by Frt sites. *K14-Cre* and *K14-CreER²* were described (Indra et al., 1999; Li et al., 2001). All experiments were performed on isolated mice. Tamoxifen (Tam) and 4-hydroxy-tamoxifen (OHT) (Sigma) were prepared as previously described (Indra et al., 1999). For *K14-CreER²* transgene activation, 8- to 10-week-old transgenic mice were either intraperitoneally injected with Tam (0.1 mg in 100 μ l sunflower oil), or subjected to topical application of OHT (0.02 mg in 50 μ l ethanol) on the shaved flank, for five consecutive days. For skin grafting, the dorsal skin harvested from KO ($n=12$) or WT ($n=9$) pups was grafted onto the back (interscapular region) of 7- to 8-week-old Rj:NMRI nude mice (Janvier, France) using the skin-flap technique (Barrandon et al., 1988). We removed the flap 2 weeks after grafting to bring the graft into an aerial environment, and started analysis 2 weeks later.

Primary antibodies

Rabbit polyclonal antibodies against K10, K6, filaggrin, involucrin, loricrin (all from Covance Research Products), Dsg1 (Santa Cruz Biotechnology), mouse mAb MAB1501 against actin (Chemicon International), rat mAb against CD3 (clone CD3-12, Serotec) and rat mAb against F4/80 (clone BM8, Serotec), were diluted following the manufacturers' recommendations. Mouse mAb F28-27, raised against human CDSN, crossreacting with mouse CDSN, was used as previously described (Serre et al., 1991).

Epidermal protein extraction and western blot analysis

Frozen tissues were homogenized in extraction buffer (40 mM Tris-HCl, 10 mM EDTA, 8 M urea, 0.25 mM PMSF, pH 7.4) with 0.1% v/v protease inhibitor cocktail (Sigma) by the FastPrep system (MP Biomedicals). Equal quantities of protein were separated by 12.5% SDS-PAGE and transferred to nitrocellulose membrane. The blots were probed with primary antibodies and HRP-conjugated secondary antibodies (Invitrogen). Detection was performed with ECL reagent (Amersham Pharmacia Biotech).

Histological and immunohistological analysis

Tissue samples were fixed for 24 hours in Bouin's fixative, dehydrated for 24 hours in 70% ethanol and embedded in paraffin. Sections (4 μ m) were stained with hematoxylin and eosin or processed for immunohistochemical analysis. Rabbit polyclonal antibodies were detected with the appropriate ImmPRESS anti-rabbit Ig (peroxidase) kit whereas the Vector MOM peroxidase kit was used to detect CDSN (Vector Laboratories). Images were taken using a Nikon eclipse 80i microscope equipped with a Nikon DXM 1200C digital camera and NIS image analysis software.

Quantitative real-time PCR experiments

Skin samples were frozen immediately after dissection and total RNAs were isolated according to RNeasy mini columns protocol (Qiagen). Reverse transcription was performed using a combination of oligo(dT) and random hexamers. Primer sequences (supplementary material Table S1) were designed using Primer3 software to generate amplicons of 100-250 bp encompassing different exons. BLAST analysis (Altschul et al., 1997) ensured the absence of similarity to any other mouse sequence. Amplification assays were performed with the 7300 Real Time PCR System (Applied Biosystems) using the Sybr qPCR SuperMix with ROX (Invitrogen). Fluorescence was quantified as C_t (threshold cycle) values. Samples were analyzed in triplicate, with differences between the three C_t values lower than 0.3. Relative levels of gene expression between samples were determined using $2^{-(\Delta\Delta C_t)}$ (Livak and Schmittgen, 2001) with desmoglein 3 gene (*Dsg3*) expression for normalization. Control wells (without template cDNA) emitted no significant fluorescence. Specificity was assessed by sequencing the qRT-PCR amplicons.

Skin-barrier function assays

TEWL was measured using an EPI Evaporimeter (Servo Med, AB Stockholm, Sweden). The dye exclusion assay was performed as previously described (Hardman et al., 1998; Marshall et al., 2000). The mice were then photographed using a Sony DSC-W50 camera. For quantification of stratum corneum removal by tape stripping, a quarter of a D-Squame disk (22 mm diameter, Monaderm) was placed on the flank of neonates and stripped off, and the process was repeated once on the same area. Tape disc fragments were incubated with shaking in 250 μ l extraction buffer (40 mM Tris-HCl, 10 mM EDTA, 8 M urea, 50 mM dithiothreitol, pH 7.4) for 1 hour at 70°C, then centrifuged for 5 minutes at 12,000 r.p.m. and the protein content was measured in the supernatant using a colorimetric protein assay (Bio-Rad).

Transmission electron microscopy

Specimens were fixed for at least 4 hours in 2% glutaraldehyde in Sorensen's buffer (pH 7.4), washed with Sorensen's buffer, cut into pieces of ~ 1 mm³, postfixed for 1 hour at room temperature in 1% osmium, 250 mM saccharose in Sorensen's buffer (pH 7.4), dehydrated through graded ethanol solutions, transferred into propylene oxide and embedded in araldite resin. Ultrathin 90 nm sections cut on a Reichert Ultratuc ultramicrotome were stained with uranyl acetate and lead citrate and examined with an electron microscope (Hitachi HU12A).

The floxed mouse mutant line was established at the Mouse Clinical Institute (Institut Clinique de la Souris, MCI/ICS) in the Targeted Mutagenesis and Transgenesis Department with funds from the GIS-Institut des Maladies Rares. The authors would like to thank H. Gallinaro for participating in the targeting strategy, and M. Ribouchon for excellent technical assistance. We thank staff of INSERM-IFR150, and more particularly M. Calise, S. Appolinaire, A. Tridon, P. Aregui, G. Marsal from the animal facilities, F. Capilla and T. Al Saati from the technical platform of 'histopathologie expérimentale', C. Offer and H. Brun from the 'plateau de séquençage'. We also thank I. Fourquaux from the CMEAB for technical assistance with ultrastructural analysis. This study was supported by grants from the CNRS, Toulouse III University, the 'Société de Recherche Dermatologique' (SRD) and the 'Association Alopecia Areata' (AAA).

References

- Allen, M., Ishida-Yamamoto, A., McGrath, J., Davison, S., Iizuka, H., Simon, M., Guerrin, M., Hayday, A., Vaughan, R., Serre, G. et al. (2001). Corneodesmosin expression in psoriasis vulgaris differs from normal skin and other inflammatory skin disorders. *Lab. Invest.* **81**, 969-976.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
- Barrandon, Y., Li, V. and Green, H. (1988). New techniques for the grafting of cultured human epidermal cells onto athymic animals. *J. Invest. Dermatol.* **91**, 315-318.
- Capon, F., Munro, M., Barker, J. and Trembath, R. (2002). Searching for the major histocompatibility complex psoriasis susceptibility gene. *J. Invest. Dermatol.* **118**, 745-751.
- Capon, F., Toal, I. K., Evans, J. C., Allen, M. H., Patel, S., Tillman, D., Burden, D., Barker, J. N. and Trembath, R. C. (2003). Haplotype analysis of distantly related populations implicates corneodesmosin in psoriasis susceptibility. *J. Med. Genet.* **40**, 447-452.
- Caubet, C., Jonca, N., Lopez, F., Estève, J. P., Simon, M. and Serre, G. (2004). Homologization of human corneodesmosin is mediated by its N-terminal glycine loop domain. *J. Invest. Dermatol.* **122**, 747-754.
- Chen, J., Den, Z. and Koch, P. J. (2008). Loss of desmocollin 3 in mice leads to epidermal blistering. *J. Cell Sci.* **121**, 2844-2849.
- Chidgey, M., Brakebusch, C., Gustafsson, E., Cruchley, A., Hail, C., Kirk, S., Merritt, A., North, A., Tselipis, C., Hewitt, J. et al. (2001). Mice lacking desmocollin 1 show epidermal fragility accompanied by barrier defects and abnormal differentiation. *J. Cell Biol.* **155**, 821-832.
- Descargues, P., Deraison, C., Bonnart, C., Kreft, M., Kishibe, M., Ishida-Yamamoto, A., Elias, P., Barrandon, Y., Zambruno, G., Sonnenberg, A. et al. (2005). Spink5-deficient mice mimic Netherton syndrome through degradation of desmoglein 1 by epidermal protease hyperactivity. *Nat. Genet.* **37**, 56-65.
- Garrod, D. and Chidgey, M. (2008). Desmosome structure, composition and function. *Biochim. Biophys. Acta* **1778**, 572-587.
- Guerrin, M., Simon, M., Montézin, M., Haftek, M., Vincent, C. and Serre, G. (1998). Expression cloning of human corneodesmosin proves its identity with the product of the S gene and allows improved characterization of its processing during keratinocyte differentiation. *J. Biol. Chem.* **273**, 22640-22647.
- Haftek, M., Simon, M., Kanitakis, J., Maréchal, S., Claudy, A., Serre, G. and Schmitt, D. (1997). Expression of corneodesmosin in the granular layer and stratum corneum of normal and diseased epidermis. *Br. J. Dermatol.* **137**, 864-873.
- Hardman, M. J., Sisi, P., Banbury, D. N. and Byrne, C. (1998). Patterned acquisition of skin barrier function during development. *Development* **125**, 1541-1552.
- Helms, C., Saccone, N. L., Cao, L., Daw, J. A., Cao, K., Hsu, T. M., Taillon-Miller, P., Duan, S., Gordon, D., Pierce, B. et al. (2005). Localization of PSORS1 to a haplotype block harboring HLA-C and distinct from corneodesmosin and HCR. *Hum. Genet.* **118**, 466-476.
- Indra, A. K., Warot, X., Brocard, J., Bornert, J. M., Xiao, J. H., Chambon, P. and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. *Nucleic Acids Res.* **27**, 4324-4327.
- Jonca, N., Guerrin, M., Hadjilova, K., Caubet, C., Gallinaro, H., Simon, M. and Serre, G. (2002). Corneodesmosin, a component of epidermal corneocyte desmosomes, displays homophilic adhesive properties. *J. Biol. Chem.* **277**, 5024-5029.
- Kljuic, A., Bazzi, H., Sundberg, J. P., Martínez-Mir, A., O'Shaughnessy, R., Mahoney, M. G., Levy, M., Montagutelli, X., Ahmad, W., Aita, V. M. et al. (2003). Desmoglein 4 in hair follicle differentiation and epidermal adhesion: evidence from inherited hypotrichosis and acquired pemphigus vulgaris. *Cell* **113**, 249-260.
- Koch, P. J., Mahoney, M. G., Cotsarelis, G., Rothenberger, K., Lavker, R. M. and Stanley, J. R. (1998). Desmoglein 3 anchors telogen hair in the follicle. *J. Cell Sci.* **111**, 2529-2537.
- Kottke, M. D., Delva, E. and Kowalezyk, A. P. (2006). The desmosome: cell science lessons from human diseases. *J. Cell Sci.* **119**, 797-806.
- Lai-Cheong, J. E., Arita, K. and McGrath, J. A. (2007). Genetic diseases of junctions. *J. Invest. Dermatol.* **127**, 2713-2725.
- Levy-Nissenbaum, E., Betz, R. C., Frydman, M., Simon, M., Lahat, H., Bakhan, T., Goldman, B., Bygum, A., Pierick, M., Hillmer, A. M. et al. (2003). Hypotrichosis simplex of the scalp is associated with nonsense mutations in CDSN encoding corneodesmosin. *Nat. Genet.* **34**, 151-153.
- Li, M., Indra, A. K., Warot, X., Brocard, J., Messaddeq, N., Kato, S., Metzger, D. and Chambon, P. (2000). Skin abnormalities generated by temporally controlled RXRalpha mutations in mouse epidermis. *Nature* **407**, 633-636.
- Li, M., Chiba, H., Warot, X., Messaddeq, N., Gérard, C., Chambon, P. and Metzger, D. (2001). RXR-alpha ablation in skin keratinocytes results in alopecia and epidermal alterations. *Development* **128**, 675-688.
- Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408.
- Lundström, A., Serre, G., Haftek, M. and Egelrud, T. (1994). Evidence for a role of corneodesmosin, a protein which may serve to modify desmosomes during cornification, in stratum corneum cell cohesion and desquamation. *Arch. Dermatol. Res.* **286**, 369-375.
- Marshall, D., Hardman, M. J. and Byrne, C. (2000). SPRR1 gene induction and barrier formation occur as coordinated moving fronts in terminally differentiating epithelia. *J. Invest. Dermatol.* **114**, 967-975.
- Matsumoto, M., Zhou, Y., Matsuo, S., Nakanishi, H., Hirose, K., Oura, H., Arase, S., Ishida-Yamamoto, A., Bando, Y., Izumi, K. et al. (2008). Targeted deletion of the murine corneodesmosin gene delineates its essential role in skin and hair physiology. *Proc. Natl. Acad. Sci. USA* **105**, 6720-6724.
- Nair, R. P., Stuart, P. E., Nistor, I., Hiremagalore, R., Chia, N. V., Jenisch, S., Weichenthal, M., Abecasis, G. R., Lim, H. W., Christophers, E. et al. (2006). Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am. J. Hum. Genet.* **78**, 827-851.
- Nickoloff, B. J., Bonish, B. K., Marble, D. J., Schriedel, K. A., DiPietro, L. A., Gordon, K. B. and Lingen, M. W. (2006). Lessons learned from psoriatic plaques concerning mechanisms of tissue repair, remodeling, and inflammation. *J. Invest. Dermatol. Symp. Proc.* **11**, 16-29.
- Orrù, S., Giurelli, E., Carcassi, C., Casula, M. and Contu, L. (2005). Mapping of the major psoriasis-susceptibility locus (PSORS1) in a 70-Kb interval around the corneodesmosin gene (CDSN). *Am. J. Hum. Genet.* **76**, 164-171.
- Proksch, E., Feingold, K. R., Man, M. Q. and Elias, P. M. (1991). Barrier function regulates epidermal DNA synthesis. *J. Clin. Invest.* **87**, 1668-1673.
- Serre, G., Mils, V., Haftek, M., Vincent, C., Croute, F., Reano, A., Ouhayoun, J. P., Bettinger, S. and Soleilhavoup, J. P. (1991). Identification of late differentiation antigens of human cornified epithelia, expressed in re-organized desmosomes and bound to cross-linked envelope. *J. Invest. Dermatol.* **97**, 1061-1072.
- Simon, M., Montézin, M., Guerrin, M., Durieux, J. J. and Serre, G. (1997). Characterization and purification of human corneodesmosin, an epidermal basic glycoprotein associated with corneocyte-specific modified desmosomes. *J. Biol. Chem.* **272**, 31770-31776.
- Simon, M., Jonca, N., Guerrin, M., Haftek, M., Bernard, D., Caubet, C., Egelrud, T., Schmidt, R. and Serre, G. (2001). Refined characterization of corneodesmosin proteolysis during terminal differentiation of human epidermis and its relationship to desquamation. *J. Biol. Chem.* **276**, 20292-20299.
- Steinert, P. M., Mack, J. W., Korge, B. P., Gan, S. Q., Haynes, S. R. and Steven, A. C. (1991). Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. *Int. J. Biol. Macromol.* **13**, 130-139.
- Vassar, R., Rosenberg, M., Ross, S., Tyner, A. and Fuchs, E. (1989). Tissue-specific and differentiation-specific expression of a human K14 keratin gene in transgenic mice. *Proc. Natl. Acad. Sci. USA* **86**, 1563-1567.
- Veal, C. D., Capon, F., Allen, M. H., Heath, E. K., Evans, J. C., Jones, A., Patel, S., Burden, D., Tillman, D., Barker, J. N. et al. (2002). Family-based analysis using a dense single-nucleotide polymorphism-based map defines genetic variation at PSORS1, the major psoriasis-susceptibility locus. *Am. J. Hum. Genet.* **71**, 554-564.
- Wang, X., Zinkel, S., Polonsky, K. and Fuchs, E. (1997). Transgenic studies with a keratin promoter-driven growth hormone transgene: prospects for gene therapy. *Proc. Natl. Acad. Sci. USA* **94**, 219-226.
- Yang, T., Liang, D., Koch, P. J., Hohl, D., Kheradmand, F. and Overbeek, P. A. (2004). Epidermal detachment, desmosomal dissociation, and destabilization of corneodesmosin in Spink5^{-/-} mice. *Genes Dev.* **18**, 2354-2358.

V. BIBLIOGRAPHIE

2003. The International HapMap Project. *Nature* **426**: 789-796.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merrill, C.R., Wu, A., Olde, B., Moreno, R.F., and et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**: 1651-1656.
- Adams, M.J., Reichel, M.B., King, I.A., Marsden, M.D., Greenwood, M.D., Thirlwell, H., Arnemann, J., Buxton, R.S., and Ali, R.R. 1998. Characterization of the regulatory regions in the human desmoglein genes encoding the pemphigus foliaceus and pemphigus vulgaris antigens. *Biochem J* **329** (Pt 1): 165-174.
- Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A., and Rubin, E.M. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol* **5**: e234.
- Aho, S., Li, K., Ryoo, Y., McGee, C., Ishida-Yamamoto, A., Uitto, J., and Klement, J.F. 2004. Periplakin gene targeting reveals a constituent of the cornified cell envelope dispensable for normal mouse development. *Mol Cell Biol* **24**: 6410-6418.
- Akavia, U.D., Veinblat, O., and Benayahu, D. 2008. Comparing the transcriptional profile of mesenchymal cells to cardiac and skeletal muscle cells. *J Cell Physiol* **216**: 663-672.
- Akiyama, M., Sugiyama-Nakagiri, Y., Sakai, K., McMillan, J.R., Goto, M., Arita, K., Tsuji-Abe, Y., Tabata, N., Matsuoka, K., Sasaki, R., Sawamura, D., and Shimizu, H. 2005. Mutations in lipid transporter ABCA12 in harlequin ichthyosis and functional recovery by corrective gene transfer. *J Clin Invest* **115**: 1777-1784.
- Al-Amoudi, A., Diez, D.C., Betts, M.J., and Frangakis, A.S. 2007. The molecular architecture of cadherins in native epidermal desmosomes. *Nature* **450**: 832-837.
- Alef, T., Torres, S., Hausser, I., Metze, D., Tursen, U., Lestringant, G.G., and Hennies, H.C. 2009. Ichthyosis, follicular atrophoderma, and hypotrichosis caused by mutations in ST14 is associated with impaired profilaggrin processing. *J Invest Dermatol* **129**: 862-869.
- Ali, R.S., Falconer, A., Ikram, M., Bissett, C.E., Cerio, R., and Quinn, A.G. 2001. Expression of the peptide antibiotics human beta defensin-1 and human beta defensin-2 in normal human skin. *J Invest Dermatol* **117**: 106-111.
- Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., and Staudt, L.M. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**: 503-511.
- Allen, C.T., Ricker, J.L., Chen, Z., and Van Waes, C. 2007. Role of activated nuclear factor-kappaB in the pathogenesis and therapy of squamous cell carcinoma of the head and neck. *Head Neck* **29**: 959-971.
- Altieri, D.C. 2008. Survivin, cancer networks and pathway-directed drug discovery. *Nat Rev Cancer* **8**: 61-70.
- Alwine, J.C., Kemp, D.J., and Stark, G.R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* **74**: 5350-5354.
- Anastasia, J.V. and Conley, J.P. 1977. The role of fatty acid oxidation in the epidermis. *J Invest Dermatol* **69**: 430-434.
- April, C.S. and Barsh, G.S. 2006. Skin layer-specific transcriptional profiles in normal and recessive yellow (Mc1re/Mc1re) mice. *Pigment Cell Res* **19**: 194-205.
- Arredondo, J., Chernyavsky, A.I., Webber, R.J., and Grando, S.A. 2005. Biological effects of SLURP-1 on human keratinocytes. *J Invest Dermatol* **125**: 1236-1241.

- Arredondo, J., Nguyen, V.T., Chernyavsky, A.I., Bercovich, D., Orr-Urtreger, A., Kummer, W., Lips, K., Vetter, D.E., and Grando, S.A. 2002. Central role of alpha7 nicotinic receptor in differentiation of the stratified squamous epithelium. *J Cell Biol* **159**: 325-336.
- Ashcroft, G.S., Lei, K., Jin, W., Longenecker, G., Kulkarni, A.B., Greenwell-Wild, T., Hale-Donze, H., McGrady, G., Song, X.Y., and Wahl, S.M. 2000. Secretory leukocyte protease inhibitor mediates non-redundant functions necessary for normal wound healing. *Nat Med* **6**: 1147-1153.
- Atshaves, B.P., Jefferson, J.R., McIntosh, A.L., Gallegos, A., McCann, B.M., Landrock, K.K., Kier, A.B., and Schroeder, F. 2007. Effect of sterol carrier protein-2 expression on sphingolipid distribution in plasma membrane lipid rafts/caveolae. *Lipids* **42**: 871-884.
- Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. 2004. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14**: 283-291.
- Balaji, S., Babu, M.M., Iyer, L.M., Luscombe, N.M., and Aravind, L. 2006. Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol* **360**: 213-227.
- Barrandon, Y. and Green, H. 1987. Three clonal types of keratinocyte with different capacities for multiplication. *Proc Natl Acad Sci U S A* **84**: 2302-2306.
- Bazzi, H., Getz, A., Mahoney, M.G., Ishida-Yamamoto, A., Langbein, L., Wahl, J.K., 3rd, and Christiano, A.M. 2006. Desmoglein 4 is expressed in highly differentiated keratinocytes and trichocytes in human epidermis and hair follicle. *Differentiation* **74**: 129-140.
- Bazzi, H., Martinez-Mir, A., Kljuic, A., and Christiano, A.M. 2005. Desmoglein 4 mutations underlie localized autosomal recessive hypotrichosis in humans, mice, and rats. *J Invest Dermatol Symp Proc* **10**: 222-224.
- Ben Selma, Z., Yilmaz, S., Schischmanoff, P.O., Blom, A., Ozogul, C., Laroche, L., and Caux, F. 2007. A novel S115G mutation of CGI-58 in a Turkish patient with Dorfman-Chanarin syndrome. *J Invest Dermatol* **127**: 2273-2276.
- Bernard, B.A. 2002. Hair biology: an update. *Int J Cosmet Sci* **24**: 13-16.
- Bernard, F.X., Pedretti, N., Rosdy, M., and Deguercy, A. 2002. Comparison of gene expression profiles in human keratinocyte mono-layer cultures, reconstituted epidermis and normal human skin; transcriptional effects of retinoid treatments in reconstituted human epidermis. *Exp Dermatol* **11**: 59-74.
- Berthou, C., Michel, L., Soulie, A., Jean-Louis, F., Flageul, B., Dubertret, L., Sigaux, F., Zhang, Y., and Sasportes, M. 1997. Acquisition of granzyme B and Fas ligand proteins by human keratinocytes contributes to epidermal cell defense. *J Immunol* **159**: 5293-5300.
- Bertin, J., Wang, L., Guo, Y., Jacobson, M.D., Poyet, J.L., Srinivasula, S.M., Merriam, S., DiStefano, P.S., and Alnemri, E.S. 2001. CARD11 and CARD14 are novel caspase recruitment domain (CARD)/membrane-associated guanylate kinase (MAGUK) family members that interact with BCL10 and activate NF-kappa B. *J Biol Chem* **276**: 11877-11882.
- Bickenbach, J.R. 1981. Identification and behavior of label-retaining cells in oral mucosa and skin. *J Dent Res* **60 Spec No C**: 1611-1620.
- Bikle, D.D., Ng, D., Oda, Y., Hanley, K., Feingold, K., and Xie, Z. 2002. The vitamin D response element of the involucrin gene mediates its regulation by 1,25-dihydroxyvitamin D3. *J Invest Dermatol* **119**: 1109-1113.

- Bikle, D.D., Oda, Y., and Xie, Z. 2004. Calcium and 1,25(OH)₂D: interacting drivers of epidermal differentiation. *J Steroid Biochem Mol Biol* **89-90**: 355-360.
- Bikle, D.D., Oda, Y., and Xie, Z. 2005. Vitamin D and skin cancer: a problem in gene regulation. *J Steroid Biochem Mol Biol* **97**: 83-91.
- Bikle, D.D. and Pillai, S. 1993. Vitamin D, calcium, and epidermal differentiation. *Endocr Rev* **14**: 3-19.
- Bishop, J.O., Morton, J.G., Rosbash, M., and Richardson, M. 1974. Three abundance classes in HeLa cell messenger RNA. *Nature* **250**: 199-204.
- Blanpain, C. and Fuchs, E. 2009. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nat Rev Mol Cell Biol* **10**: 207-217.
- Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., and Fuchs, E. 2004. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**: 635-648.
- Blenkiron, C., Goldstein, L.D., Thorne, N.P., Spiteri, I., Chin, S.F., Dunning, M.J., Barbosa-Morais, N.L., Teschendorff, A.E., Green, A.R., Ellis, I.O., Tavare, S., Caldas, C., and Miska, E.A. 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* **8**: R214.
- Boguski, M.S., Tolstoshev, C.M., and Bassett, D.E., Jr. 1994. Gene discovery in dbEST. *Science* **265**: 1993-1994.
- Borradori, L. and Sonnenberg, A. 1999. Structure and function of hemidesmosomes: more than simple adhesion complexes. *J Invest Dermatol* **112**: 411-418.
- Bowcock, A.M., Shannon, W., Du, F., Duncan, J., Cao, K., Aftergut, K., Catier, J., Fernandez-Vina, M.A., and Menter, A. 2001. Insights into psoriasis and other inflammatory diseases from large-scale gene expression studies. *Hum Mol Genet* **10**: 1793-1805.
- Brandner, J.M. 2008. Tight junctions and tight junction proteins in mammalian epidermis. *Eur J Pharm Biopharm.*
- Brattsand, M., Stefansson, K., Hubiche, T., Nilsson, S.K., and Egelrud, T. 2009. SPINK9: A Selective, Skin-Specific Kazal-Type Serine Protease Inhibitor. *J Invest Dermatol.*
- Braun, K.M., Niemann, C., Jensen, U.B., Sundberg, J.P., Silva-Vargas, V., and Watt, F.M. 2003. Manipulation of stem cell proliferation and lineage commitment: visualisation of label-retaining cells in wholemounts of mouse epidermis. *Development* **130**: 5241-5255.
- Brown, A.C., Kai, K., May, M.E., Brown, D.C., and Roopenian, D.C. 2004. ExQuest, a novel method for displaying quantitative gene expression from ESTs. *Genomics* **83**: 528-539.
- Cabral, A., Sayin, A., de Winter, S., Fischer, D.F., Pavel, S., and Backendorf, C. 2001a. SPRR4, a novel cornified envelope precursor: UV-dependent epidermal expression and selective incorporation into fragile envelopes. *J Cell Sci* **114**: 3837-3843.
- Cabral, A., Voskamp, P., Cleton-Jansen, A.M., South, A., Nizetic, D., and Backendorf, C. 2001b. Structural organization and regulation of the small proline-rich family of cornified envelope precursors suggest a role in adaptive barrier function. *J Biol Chem* **276**: 19231-19237.
- Calautti, E., Li, J., Saoncella, S., Brissette, J.L., and Goetinck, P.F. 2005. Phosphoinositide 3-kinase signaling to Akt promotes keratinocyte differentiation versus death. *J Biol Chem* **280**: 32856-32865.
- Caloca, M.J., Zugaza, J.L., and Bustelo, X.R. 2003. Exchange factors of the RasGRP family mediate Ras activation in the Golgi. *J Biol Chem* **278**: 33465-33473.
- Camargo, A.A., Samaia, H.P., Dias-Neto, E., Simao, D.F., Migotto, I.A., Briones, M.R., Costa, F.F., Nagai, M.A., Verjovski-Almeida, S., Zago, M.A., Andrade, L.E., Carrer,

- H., El-Dorry, H.F., Espreafico, E.M., Habr-Gama, A., Giannella-Neto, D., Goldman, G.H., Gruber, A., Hackel, C., Kimura, E.T., Maciel, R.M., Marie, S.K., Martins, E.A., Nobrega, M.P., Paco-Larson, M.L., Pardini, M.I., Pereira, G.G., Pesquero, J.B., Rodrigues, V., Rogatto, S.R., da Silva, I.D., Sogayar, M.C., Sonati, M.F., Tajara, E.H., Valentini, S.R., Alberto, F.L., Amaral, M.E., Aneas, I., Arnaldi, L.A., de Assis, A.M., Bengtson, M.H., Bergamo, N.A., Bombonato, V., de Camargo, M.E., Canevari, R.A., Carraro, D.M., Cerutti, J.M., Correa, M.L., Correa, R.F., Costa, M.C., Curcio, C., Hokama, P.O., Ferreira, A.J., Furuzawa, G.K., Gushiken, T., Ho, P.L., Kimura, E., Krieger, J.E., Leite, L.C., Majumder, P., Marins, M., Marques, E.R., Melo, A.S., Melo, M.B., Mestriner, C.A., Miracca, E.C., Miranda, D.C., Nascimento, A.L., Nobrega, F.G., Ojopi, E.P., Pandolfi, J.R., Pessoa, L.G., Prevedel, A.C., Rahal, P., Rainho, C.A., Reis, E.M., Ribeiro, M.L., da Ros, N., de Sa, R.G., Sales, M.M., Sant'anna, S.C., dos Santos, M.L., da Silva, A.M., da Silva, N.P., Silva, W.A., Jr., da Silveira, R.A., Sousa, J.F., Stecconi, D., Tsukumo, F., Valente, V., Soares, F., Moreira, E.S., Nunes, D.N., Correa, R.G., Zalcborg, H., Carvalho, A.F., Reis, L.F., Brentani, R.R., Simpson, A.J., and de Souza, S.J. 2001. The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc Natl Acad Sci U S A* **98**: 12103-12108.
- Candi, E., Schmidt, R., and Melino, G. 2005. The cornified envelope: a model of cell death in the skin. *Nat Rev Mol Cell Biol* **6**: 328-340.
- Capon, F., Allen, M.H., Ameen, M., Burden, A.D., Tillman, D., Barker, J.N., and Trembath, R.C. 2004. A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. *Hum Mol Genet* **13**: 2361-2368.
- Capon, F., Semprini, S., Chimenti, S., Fabrizi, G., Zambruno, G., Murgia, S., Carcassi, C., Fazio, M., Mingarelli, R., Dallapiccola, B., and Novelli, G. 2001. Fine mapping of the PSORS4 psoriasis susceptibility region on chromosome 1q21. *J Invest Dermatol* **116**: 728-730.
- Capon, F., Toal, I.K., Evans, J.C., Allen, M.H., Patel, S., Tillman, D., Burden, D., Barker, J.N., and Trembath, R.C. 2003. Haplotype analysis of distantly related populations implicates corneodesmosin in psoriasis susceptibility. *J Med Genet* **40**: 447-452.
- Carroll, J.M., Albers, K.M., Garlick, J.A., Harrington, R., and Taichman, L.B. 1993. Tissue- and stratum-specific expression of the human involucrin promoter in transgenic mice. *Proc Natl Acad Sci U S A* **90**: 10270-10274.
- Carroll, J.M., Romero, M.R., and Watt, F.M. 1995. Suprabasal integrin expression in the epidermis of transgenic mice results in developmental defects and a phenotype resembling psoriasis. *Cell* **83**: 957-968.
- Carter, N.P. 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* **39**: S16-21.
- Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., and Werner, T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**: 2933-2942.
- Cassidy, A.J., van Steensel, M.A., Steijlen, P.M., van Geel, M., van der Velden, J., Morley, S.M., Terrinoni, A., Melino, G., Candi, E., and McLean, W.H. 2005. A homozygous missense mutation in TGM5 abolishes epidermal transglutaminase 5 activity and causes acral peeling skin syndrome. *Am J Hum Genet* **77**: 909-917.
- Caubet, C., Jonca, N., Brattsand, M., Guerrin, M., Bernard, D., Schmidt, R., Egelrud, T., Simon, M., and Serre, G. 2004. Degradation of corneodesmosome proteins by two serine proteases of the kallikrein family, SCTE/KLK5/hK5 and SCCE/KLK7/hK7. *J Invest Dermatol* **122**: 1235-1244.

- Caux, F., Selma, Z.B., Laroche, L., Prud'homme, J.F., and Fischer, J. 2004. CGI-58/ABHD5 gene is mutated in Dorfman-Chanarin syndrome. *Am J Med Genet A* **129A**: 214.
- Chalabi, N., Satih, S., Delort, L., Bignon, Y.J., and Bernard-Gallon, D.J. 2007. Expression profiling by whole-genome microarray hybridization reveals differential gene expression in breast cancer cell lines after lycopene exposure. *Biochim Biophys Acta* **1769**: 124-130.
- Chalmel, F., Rolland, A.D., Niederhauser-Wiederkehr, C., Chung, S.S., Demougin, P., Gattiker, A., Moore, J., Patard, J.J., Wolgemuth, D.J., Jegou, B., and Prinig, M. 2007. The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A* **104**: 8346-8351.
- Chan, T.A., Hermeking, H., Lengauer, C., Kinzler, K.W., and Vogelstein, B. 1999. 14-3-3Sigma is required to prevent mitotic catastrophe after DNA damage. *Nature* **401**: 616-620.
- Chapman, S.J. and Walsh, A. 1990. Desmosomes, corneosomes and desquamation. An ultrastructural study of adult pig epidermis. *Arch Dermatol Res* **282**: 304-310.
- Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M., Irvine, A.D., Bonafe, J.L., Wilkinson, J., Taieb, A., Barrandon, Y., Harper, J.I., de Prost, Y., and Hovnanian, A. 2000. Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. *Nat Genet* **25**: 141-142.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., Loh, Y.H., Yeo, H.C., Yeo, Z.X., Narang, V., Govindarajan, K.R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W.K., Clarke, N.D., Wei, C.L., and Ng, H.H. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106-1117.
- Cheng, J., Syder, A.J., Yu, Q.C., Letai, A., Paller, A.S., and Fuchs, E. 1992. The genetic basis of epidermolytic hyperkeratosis: a disorder of differentiation-specific epidermal keratin genes. *Cell* **70**: 811-819.
- Chidgey, M., Brakebusch, C., Gustafsson, E., Cruchley, A., Hail, C., Kirk, S., Merritt, A., North, A., Tselepis, C., Hewitt, J., Byrne, C., Fassler, R., and Garrod, D. 2001. Mice lacking desmocollin 1 show epidermal fragility accompanied by barrier defects and abnormal differentiation. *J Cell Biol* **155**: 821-832.
- Chimienti, F., Hogg, R.C., Plantard, L., Lehmann, C., Brakch, N., Fischer, J., Huber, M., Bertrand, D., and Hohl, D. 2003. Identification of SLURP-1 as an epidermal neuromodulator explains the clinical phenotype of Mal de Meleda. *Hum Mol Genet* **12**: 3017-3024.
- Choi, M.J. and Maibach, H.I. 2005. Role of ceramides in barrier function of healthy and diseased skin. *Am J Clin Dermatol* **6**: 215-223.
- Chou, J., Lin, Y.C., Kim, J., You, L., Xu, Z., He, B., and Jablons, D.M. 2008. Nasopharyngeal carcinoma--review of the molecular mechanisms of tumorigenesis. *Head Neck* **30**: 946-963.
- Clayton, E., Doupe, D.P., Klein, A.M., Winton, D.J., Simons, B.D., and Jones, P.H. 2007. A single type of progenitor cell maintains normal epidermis. *Nature* **446**: 185-189.
- Cohen, C.D., Klingenhoff, A., Boucherot, A., Nitsche, A., Henger, A., Brunner, B., Schmid, H., Merkle, M., Saleem, M.A., Koller, K.P., Werner, T., Grone, H.J., Nelson, P.J., and Kretzler, M. 2006. Comparative promoter analysis allows de novo identification of specialized cell junction-associated proteins. *Proc Natl Acad Sci U S A* **103**: 5682-5687.
- Cole, J., Tsou, R., Wallace, K., Gibran, N., and Isik, F. 2001. Early gene expression profile of human skin to injury using high-density cDNA microarrays. *Wound Repair Regen* **9**: 360-370.

- Colell, A., Ricci, J.E., Tait, S., Milasta, S., Maurer, U., Bouchier-Hayes, L., Fitzgerald, P., Guio-Carrion, A., Waterhouse, N.J., Li, C.W., Mari, B., Barbry, P., Newmeyer, D.D., Beere, H.M., and Green, D.R. 2007. GAPDH and autophagy preserve survival after apoptotic cytochrome c release in the absence of caspase activation. *Cell* **129**: 983-997.
- Collin, C., Moll, R., Kubicka, S., Ouhayoun, J.P., and Franke, W.W. 1992. Characterization of human cytokeratin 2, an epidermal cytoskeletal protein synthesized late during differentiation. *Exp Cell Res* **202**: 132-141.
- Cooper, L., Johnson, C., Burslem, F., and Martin, P. 2005. Wound healing and inflammation genes revealed by array analysis of 'macrophageless' PU.1 null mice. *Genome Biol* **6**: R5.
- Cotsarelis, G., Sun, T.T., and Lavker, R.M. 1990. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**: 1329-1337.
- Crish, J.F., Gopalakrishnan, R., Bone, F., Gilliam, A.C., and Eckert, R.L. 2006. The distal and proximal regulatory regions of the involucrin gene promoter have distinct functions and are required for in vivo involucrin expression. *J Invest Dermatol* **126**: 305-314.
- Crish, J.F., Howard, J.M., Zaim, T.M., Murthy, S., and Eckert, R.L. 1993. Tissue-specific and differentiation-appropriate expression of the human involucrin gene in transgenic mice: an abnormal epidermal phenotype. *Differentiation* **53**: 191-200.
- Cui, Y., Wang, Q., Stormo, G.D., and Calvo, J.M. 1995. A consensus sequence for binding of Lrp to DNA. *J Bacteriol* **177**: 4872-4880.
- Dale, B.A., Holbrook, K.A., Fleckman, P., Kimball, J.R., Brumbaugh, S., and Sybert, V.P. 1990. Heterogeneity in harlequin ichthyosis, an inborn error of epidermal keratinization: variable morphology and structural protein expression and a defect in lamellar granules. *J Invest Dermatol* **94**: 6-18.
- de Cid, R., Riveira-Munoz, E., Zeeuwen, P.L., Robarge, J., Liao, W., Dannhauser, E.N., Giardina, E., Stuart, P.E., Nair, R., Helms, C., Escaramis, G., Ballana, E., Martin-Ezquerria, G., den Heijer, M., Kamsteeg, M., Joosten, I., Eichler, E.E., Lazaro, C., Pujol, R.M., Armengol, L., Abecasis, G., Elder, J.T., Novelli, G., Armour, J.A., Kwok, P.Y., Bowcock, A., Schalkwijk, J., and Estivill, X. 2009. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet* **41**: 211-215.
- de Guzman Strong, C., Wertz, P.W., Wang, C., Yang, F., Meltzer, P.S., Andl, T., Millar, S.E., Ho, I.C., Pai, S.Y., and Segre, J.A. 2006. Lipid defect underlies selective skin barrier impairment of an epidermal-specific deletion of Gata-3. *J Cell Biol* **175**: 661-670.
- de Juanes, S., Epp, N., Latzko, S., Neumann, M., Furstenberger, G., Hausser, I., Stark, H.J., and Krieg, P. 2009. Development of an ichthyosiform phenotype in Alox12b-deficient mouse skin transplants. *J Invest Dermatol* **129**: 1429-1436.
- De Laurenzi, V., Rogers, G.R., Hamrock, D.J., Marekov, L.N., Steinert, P.M., Compton, J.G., Markova, N., and Rizzo, W.B. 1996. Sjogren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. *Nat Genet* **12**: 52-57.
- Denecker, G., Hoste, E., Gilbert, B., Hochepped, T., Ovaere, P., Lippens, S., Van den Broecke, C., Van Damme, P., D'Herde, K., Hachem, J.P., Borgonie, G., Presland, R.B., Schoonjans, L., Libert, C., Vandekerckhove, J., Gevaert, K., Vandenaabeele, P., and Declercq, W. 2007. Caspase-14 protects against epidermal UVB photodamage and water loss. *Nat Cell Biol* **9**: 666-674.
- Deraison, C., Bonnart, C., Lopez, F., Besson, C., Robinson, R., Jayakumar, A., Wagberg, F., Brattsand, M., Hachem, J.P., Leonardsson, G., and Hovnanian, A. 2007. LEKTI

- fragments specifically inhibit KLK5, KLK7, and KLK14 and control desquamation through a pH-dependent interaction. *Mol Biol Cell* **18**: 3607-3619.
- Derry, J.M., Gormally, E., Means, G.D., Zhao, W., Meindl, A., Kelley, R.I., Boyd, Y., and Herman, G.E. 1999. Mutations in a delta 8-delta 7 sterol isomerase in the tattered mouse and X-linked dominant chondrodysplasia punctata. *Nat Genet* **22**: 286-290.
- Descargues, P., Deraison, C., Bonnart, C., Kreft, M., Kishibe, M., Ishida-Yamamoto, A., Elias, P., Barrandon, Y., Zambruno, G., Sonnenberg, A., and Hovnanian, A. 2005. Spink5-deficient mice mimic Netherton syndrome through degradation of desmoglein 1 by epidermal protease hyperactivity. *Nat Genet* **37**: 56-65.
- Dias Neto, E., Correa, R.G., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W., Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J. 2000. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc Natl Acad Sci U S A* **97**: 3491-3496.
- Dias Neto, E., Harrop, R., Correa-Oliveira, R., Wilson, R.A., Pena, S.D., and Simpson, A.J. 1997. Minilibraries constructed from cDNA generated by arbitrarily primed RT-PCR: an alternative to normalized libraries for the generation of ESTs from nanogram quantities of mRNA. *Gene* **186**: 135-142.
- Djalilian, A.R., McGaughey, D., Patel, S., Seo, E.Y., Yang, C., Cheng, J., Tomic, M., Sinha, S., Ishida-Yamamoto, A., and Segre, J.A. 2006. Connexin 26 regulates epidermal barrier and wound remodeling and promotes psoriasiform response. *J Clin Invest* **116**: 1243-1253.
- Djian, P., Easley, K., and Green, H. 2000. Targeted ablation of the murine involucrin gene. *J Cell Biol* **151**: 381-388.
- Doghman, M., Karpova, T., Rodrigues, G.A., Arhatte, M., De Moura, J., Cavalli, L.R., Virolle, V., Barbry, P., Zambetti, G.P., Figueiredo, B.C., Heckert, L.L., and Lalli, E. 2007. Increased steroidogenic factor-1 dosage triggers adrenocortical cell proliferation and cancer. *Mol Endocrinol* **21**: 2968-2987.
- Dooley, T.P., Haldeman-Cahill, R., Joiner, J., and Wilborn, T.W. 2000. Expression profiling of human sulfotransferase and sulfatase gene superfamilies in epithelial tissues and cultured cells. *Biochem Biophys Res Commun* **277**: 236-245.
- Dowler, S., Currie, R.A., Campbell, D.G., Deak, M., Kular, G., Downes, C.P., and Alessi, D.R. 2000. Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. *Biochem J* **351**: 19-31.
- Dudoit, S., Gentleman, R.C., and Quackenbush, J. 2003. Open source software for the analysis of microarray data. *Biotechniques* **Suppl**: 45-51.
- Duffin, K.C. and Krueger, G.G. 2009. Genetic variations in cytokines and cytokine receptors associated with psoriasis found by genome-wide association. *J Invest Dermatol* **129**: 827-833.
- Eckert, R.L., Crish, J.F., Efimova, T., Dashti, S.R., Deucher, A., Bone, F., Adhikary, G., Huang, G., Gopalakrishnan, R., and Balasubramanian, S. 2004. Regulation of involucrin gene expression. *J Invest Dermatol* **123**: 13-22.
- Eckert, R.L. and Green, H. 1986. Structure and evolution of the human involucrin gene. *Cell* **46**: 583-589.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**: 14863-14868.
- Ekholm, E. and Egelrud, T. 2000. Expression of stratum corneum chymotryptic enzyme in relation to other markers of epidermal differentiation in a skin explant model. *Exp Dermatol* **9**: 65-70.

- Elias, P., Ahn, S., Brown, B., Crumrine, D., and Feingold, K.R. 2002a. Origin of the epidermal calcium gradient: regulation by barrier status and role of active vs passive mechanisms. *J Invest Dermatol* **119**: 1269-1274.
- Elias, P.M., Ahn, S.K., Denda, M., Brown, B.E., Crumrine, D., Kimutai, L.K., Komuves, L., Lee, S.H., and Feingold, K.R. 2002b. Modulations in epidermal calcium regulate the expression of differentiation-specific markers. *J Invest Dermatol* **119**: 1128-1136.
- Elias, P.M., Crumrine, D., Rassner, U., Hachem, J.P., Menon, G.K., Man, W., Choy, M.H., Leyboldt, L., Feingold, K.R., and Williams, M.L. 2004. Basis for abnormal desquamation and permeability barrier dysfunction in RXLI. *J Invest Dermatol* **122**: 314-319.
- Fartasch, M., Bassukas, I.D., and Diepgen, T.L. 1993. Structural relationship between epidermal lipid lamellae, lamellar bodies and desmosomes in human epidermis: an ultrastructural study. *Br J Dermatol* **128**: 1-9.
- Fischer, J. 2009. Autosomal recessive congenital ichthyosis. *J Invest Dermatol* **129**: 1319-1321.
- Fischer, J., Bouadjar, B., Heilig, R., Huber, M., Lefevre, C., Jobard, F., Macari, F., Bakija-Konsuo, A., Ait-Belkacem, F., Weissenbach, J., Lathrop, M., Hohl, D., and Prud'homme, J.F. 2001. Mutations in the gene encoding SLURP-1 in Mal de Meleda. *Hum Mol Genet* **10**: 875-880.
- Fluhr, J.W., Man, M.Q., Hachem, J.P., Crumrine, D., Mauro, T.M., Elias, P.M., and Feingold, K.R. 2009. Topical peroxisome proliferator activated receptor activators accelerate postnatal stratum corneum acidification. *J Invest Dermatol* **129**: 365-374.
- Fodor, S.P., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P., and Adams, C.L. 1993. Multiplexed biochemical assays with biological chips. *Nature* **364**: 555-556.
- Forslind, B. 1986. Particle probe analysis in the study of skin physiology. *Scan Electron Microsc*: 1007-1014.
- Franco, B., Meroni, G., Parenti, G., Levilliers, J., Bernard, L., Gebbia, M., Cox, L., Maroteaux, P., Sheffield, L., Rappold, G.A., and et al. 1995. A cluster of sulfatase genes on Xp22.3: mutations in chondrodysplasia punctata (CDPX) and implications for warfarin embryopathy. *Cell* **81**: 15-25.
- Frisch, S.M. and Francis, H. 1994. Disruption of epithelial cell-matrix interactions induces apoptosis. *J Cell Biol* **124**: 619-626.
- Fuchs, E. and Weber, K. 1994. Intermediate filaments: structure, dynamics, function, and disease. *Annu Rev Biochem* **63**: 345-382.
- Furuse, M., Hata, M., Furuse, K., Yoshida, Y., Haratake, A., Sugitani, Y., Noda, T., Kubo, A., and Tsukita, S. 2002. Claudin-based tight junctions are crucial for the mammalian epidermal barrier: a lesson from claudin-1-deficient mice. *J Cell Biol* **156**: 1099-1111.
- Gallegos, A.M., Atshaves, B.P., Storey, S.M., Starodub, O., Petrescu, A.D., Huang, H., McIntosh, A.L., Martin, G.G., Chao, H., Kier, A.B., and Schroeder, F. 2001. Gene structure, intracellular localization, and functional roles of sterol carrier protein-2. *Prog Lipid Res* **40**: 498-563.
- Galliano, M.F., Toulza, E., Gallinaro, H., Jonca, N., Ishida-Yamamoto, A., Serre, G., and Guerrin, M. 2006. A novel protease inhibitor of the alpha2-macroglobulin family expressed in the human epidermis. *J Biol Chem* **281**: 5780-5789.
- Gan, S.Q., McBride, O.W., Idler, W.W., Markova, N., and Steinert, P.M. 1990. Organization, structure, and polymorphisms of the human profilaggrin gene. *Biochemistry* **29**: 9432-9440.
- Garber, K. 2008. Fixing the front end. *Nat Biotechnol* **26**: 1101-1104.
- Gariboldi, S., Palazzo, M., Zanobbio, L., Selleri, S., Sommariva, M., Sfondrini, L., Cavicchini, S., Balsari, A., and Rumio, C. 2008. Low molecular weight hyaluronic

- acid increases the self-defense of skin epithelium by induction of beta-defensin 2 via TLR2 and TLR4. *J Immunol* **181**: 2103-2110.
- Garrod, D.R., Berika, M.Y., Bardsley, W.F., Holmes, D., and Taberner, L. 2005. Hyper-adhesion in desmosomes: its regulation in wound healing and possible relationship to cadherin crystal structure. *J Cell Sci* **118**: 5743-5754.
- Gazel, A., Ramphal, P., Rosdy, M., De Wever, B., Tornier, C., Hosein, N., Lee, B., Tomic-Canic, M., and Blumenberg, M. 2003. Transcriptional profiling of epidermal keratinocytes: comparison of genes expressed in skin, cultured keratinocytes, and reconstituted epidermis, using large DNA microarrays. *J Invest Dermatol* **121**: 1459-1468.
- Ghazizadeh, S. and Taichman, L.B. 2005. Organization of stem cells and their progeny in human epidermis. *J Invest Dermatol* **124**: 367-372.
- Gibbs, S., Fijneman, R., Wiegant, J., van Kessel, A.G., van De Putte, P., and Backendorf, C. 1993. Molecular characterization and evolution of the SPRR family of keratinocyte differentiation markers encoding small proline-rich proteins. *Genomics* **16**: 630-637.
- Gibson, D.F., Ratnam, A.V., and Bikle, D.D. 1996. Evidence for separate control mechanisms at the message, protein, and enzyme activation levels for transglutaminase during calcium-induced differentiation of normal and transformed human keratinocytes. *J Invest Dermatol* **106**: 154-161.
- Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., and Joint, I. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE* **3**: e3042.
- Glas, A.M., Floore, A., Delahaye, L.J., Witteveen, A.T., Pover, R.C., Bakx, N., Lahti-Domenici, J.S., Bruinsma, T.J., Warmoes, M.O., Bernards, R., Wessels, L.F., and Van't Veer, L.J. 2006. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* **7**: 278.
- Groot, K.R., Sevilla, L.M., Nishi, K., DiColandrea, T., and Watt, F.M. 2004. Kazrin, a novel periplakin-interacting protein associated with desmosomes and the keratinocyte plasma membrane. *J Cell Biol* **166**: 653-659.
- Gudjonsson, J.E., Johnston, A., Dyson, M., Valdimarsson, H., and Elder, J.T. 2007. Mouse models of psoriasis. *J Invest Dermatol* **127**: 1292-1308.
- Gudjonsson, J.E., Karason, A., Runarsdottir, E.H., Antonsdottir, A.A., Hauksson, V.B., Jonsson, H.H., Gulcher, J., Stefansson, K., and Valdimarsson, H. 2006. Distinct clinical differences between HLA-Cw*0602 positive and negative psoriasis patients--an analysis of 1019 HLA-C- and HLA-B-typed patients. *J Invest Dermatol* **126**: 740-745.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**: 77-88.
- Guerrin, M., Simon, M., Montezin, M., Haftek, M., Vincent, C., and Serre, G. 1998. Expression cloning of human corneodesmosin proves its identity with the product of the S gene and allows improved characterization of its processing during keratinocyte differentiation. *J Biol Chem* **273**: 22640-22647.
- Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G., and Chee, M.S. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37**: 549-554.
- Guttormsen, J., Koster, M.I., Stevens, J.R., Roop, D.R., Williams, T., and Winger, Q.A. 2008. Disruption of epidermal specific gene expression and delayed skin development in AP-2 gamma mutant mice. *Dev Biol* **317**: 187-195.

- Haas, U., Podda, M., Behne, M., Gurrieri, S., Alonso, A., Furstenberger, G., Pfeilschifter, J., Lambeau, G., Gelb, M.H., and Kaszkin, M. 2005. Characterization and differentiation-dependent regulation of secreted phospholipases A in human keratinocytes and in healthy and psoriatic human skin. *J Invest Dermatol* **124**: 204-211.
- Hachem, J.P., Man, M.Q., Crumrine, D., Uchida, Y., Brown, B.E., Rogiers, V., Roseeuw, D., Feingold, K.R., and Elias, P.M. 2005. Sustained serine proteases activity by prolonged increase in pH leads to degradation of lipid processing enzymes and profound alterations of barrier function and stratum corneum integrity. *J Invest Dermatol* **125**: 510-520.
- Hachem, J.P., Wagberg, F., Schmuth, M., Crumrine, D., Lissens, W., Jayakumar, A., Houben, E., Mauro, T.M., Leonardsson, G., Brattsand, M., Egelrud, T., Roseeuw, D., Clayman, G.L., Feingold, K.R., Williams, M.L., and Elias, P.M. 2006. Serine protease activity and residual LEKTI expression determine phenotype in Netherton syndrome. *J Invest Dermatol* **126**: 1609-1621.
- Hadj-Rabia, S., Baala, L., Vabres, P., Hamel-Teillac, D., Jacquemin, E., Fabre, M., Lyonnet, S., De Prost, Y., Munnich, A., Hadchouel, M., and Smahi, A. 2004. Claudin-1 gene mutations in neonatal sclerosing cholangitis associated with ichthyosis: a tight junction disease. *Gastroenterology* **127**: 1386-1390.
- Haider, A.S., Peters, S.B., Kaporis, H., Cardinale, I., Fei, J., Ott, J., Blumenberg, M., Bowcock, A.M., Krueger, J.G., and Carucci, J.A. 2006. Genomic analysis defines a cancer-specific gene expression signature for human squamous cell carcinoma and distinguishes malignant hyperproliferation from benign hyperplasia. *J Invest Dermatol* **126**: 869-881.
- Halsall, J.A., Osborne, J.E., Pringle, J.H., and Hutchinson, P.E. 2005. Vitamin D receptor gene polymorphisms, particularly the novel A-1012G promoter polymorphism, are associated with vitamin D3 responsiveness and non-familial susceptibility in psoriasis. *Pharmacogenet Genomics* **15**: 349-355.
- Hamilton, G., Colbert, J.D., Schuettelkopf, A.W., and Watts, C. 2008. Cystatin F is a cathepsin C-directed protease inhibitor regulated by proteolysis. *Embo J* **27**: 499-508.
- Hanley, K., Wood, L., Ng, D.C., He, S.S., Lau, P., Moser, A., Elias, P.M., Bikle, D.D., Williams, M.L., and Feingold, K.R. 2001. Cholesterol sulfate stimulates involucrin transcription in keratinocytes by increasing Fra-1, Fra-2, and Jun D. *J Lipid Res* **42**: 390-398.
- Hansson, L., Stromqvist, M., Backman, A., Wallbrandt, P., Carlstein, A., and Egelrud, T. 1994. Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase. *J Biol Chem* **269**: 19420-19426.
- Harbers, M. and Carninci, P. 2005. Tag-based approaches for transcriptome research and genome annotation. *Nat Methods* **2**: 495-502.
- Hatzoglou, A., Ader, I., Spingard, A., Flanders, J., Saade, E., Leroy, I., Traver, S., Aresta, S., and de Gunzburg, J. 2007. Gem associates with Ezrin and acts via the Rho-GAP protein Gmip to down-regulate the Rho pathway. *Mol Biol Cell* **18**: 1242-1252.
- Heinzel, R., Appelhans, H., Gassen, G., Seemuller, U., Machleidt, W., Fritz, H., and Steffens, G. 1986. Molecular cloning and expression of cDNA for human antileukoprotease from cervix uterus. *Eur J Biochem* **160**: 61-67.
- Helms, C., Saccone, N.L., Cao, L., Daw, J.A., Cao, K., Hsu, T.M., Taillon-Miller, P., Duan, S., Gordon, D., Pierce, B., Ott, J., Rice, J., Fernandez-Vina, M.A., Kwok, P.Y., Menter, A., and Bowcock, A.M. 2005. Localization of PSORS1 to a haplotype block harboring HLA-C and distinct from corneodesmosin and HCR. *Hum Genet* **118**: 466-476.

- Henseler, T. and Christophers, E. 1985. Psoriasis of early and late onset: characterization of two types of psoriasis vulgaris. *J Am Acad Dermatol* **13**: 450-456.
- Herrero, J., Al-Shahrou, F., Diaz-Uriarte, R., Mateos, A., Vaquerizas, J.M., Santoyo, J., and Dopazo, J. 2003. GEPAS: A web-based resource for microarray gene expression data analysis. *Nucleic Acids Res* **31**: 3461-3467.
- Hofman, V.J., Moreilhon, C., Brest, P.D., Lassalle, S., Le Brigand, K., Sicard, D., Raymond, J., Lamarque, D., Hebuterne, X.A., Mari, B., Barbry, P.J., and Hofman, P.M. 2007. Gene expression profiling in human gastric mucosa infected with *Helicobacter pylori*. *Mod Pathol* **20**: 974-989.
- Hohl, D., Mehrel, T., Lichti, U., Turner, M.L., Roop, D.R., and Steinert, P.M. 1991. Characterization of human loricrin. Structure and function of a new class of epidermal cell envelope proteins. *J Biol Chem* **266**: 6626-6636.
- Horikoshi, T., Igarashi, S., Uchiwa, H., Brysk, H., and Brysk, M.M. 1999. Role of endogenous cathepsin D-like and chymotrypsin-like proteolysis in human epidermal desquamation. *Br J Dermatol* **141**: 453-459.
- Horsley, V., O'Carroll, D., Tooze, R., Ohinata, Y., Saitou, M., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., and Fuchs, E. 2006. Blimp1 defines a progenitor population that governs cellular input to the sebaceous gland. *Cell* **126**: 597-609.
- Huang, J., Sheng, H.H., Shen, T., Hu, Y.J., Xiao, H.S., Zhang, Q., Zhang, Q.H., and Han, Z.G. 2006. Correlation between genomic DNA copy number alterations and transcriptional expression in hepatitis B virus-associated hepatocellular carcinoma. *FEBS Lett* **580**: 3571-3581.
- Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephanants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., and Linsley, P.S. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**: 342-347.
- Hwang, J., Mehrani, T., Millar, S.E., and Morasso, M.I. 2008. Dlx3 is a crucial regulator of hair follicle differentiation and cycling. *Development* **135**: 3149-3159.
- Hwang, S.M., Ahn, S.K., Menon, G.K., Choi, E.H., and Lee, S.H. 2001. Basis of occlusive therapy in psoriasis: correcting defects in permeability barrier and calcium gradient. *Int J Dermatol* **40**: 223-231.
- Igarashi, S., Takizawa, T., Takizawa, T., Yasuda, Y., Uchiwa, H., Hayashi, S., Brysk, H., Robinson, J.M., Yamamoto, K., Brysk, M.M., and Horikoshi, T. 2004. Cathepsin D, but not cathepsin E, degrades desmosomes during epidermal desquamation. *Br J Dermatol* **151**: 355-361.
- Iizuka, H., Takahashi, H., Honma, M., and Ishida-Yamamoto, A. 2004. Unique keratinization process in psoriasis: late differentiation markers are abolished because of the premature cell death. *J Dermatol* **31**: 271-276.
- Inagi, R., Nangaku, M., Miyata, T., and Kurokawa, K. 2003. Mesangial cell-predominant functional gene, megsin. *Clin Exp Nephrol* **7**: 87-92.
- Ishida-Yamamoto, A. 2003. Loricrin keratoderma: a novel disease entity characterized by nuclear accumulation of mutant loricrin. *J Dermatol Sci* **31**: 3-8.
- Ishida-Yamamoto, A., McGrath, J.A., Lam, H., Iizuka, H., Friedman, R.A., and Christiano, A.M. 1997. The molecular pathology of progressive symmetric erythrokeratoderma: a frameshift mutation in the loricrin gene and perturbations in the cornified cell envelope. *Am J Hum Genet* **61**: 581-589.

- Ishida-Yamamoto, A., Simon, M., Kishibe, M., Miyauchi, Y., Takahashi, H., Yoshida, S., O'Brien, T.J., Serre, G., and Iizuka, H. 2004. Epidermal lamellar granules transport different cargoes as distinct aggregates. *J Invest Dermatol* **122**: 1137-1144.
- Ito, M., Liu, Y., Yang, Z., Nguyen, J., Liang, F., Morris, R.J., and Cotsarelis, G. 2005. Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat Med* **11**: 1351-1354.
- Jackson, B., Tilli, C.M., Hardman, M.J., Avilion, A.A., MacLeod, M.C., Ashcroft, G.S., and Byrne, C. 2005. Late cornified envelope family in differentiating epithelia--response to calcium and ultraviolet irradiation. *J Invest Dermatol* **124**: 1062-1070.
- Jang, S.I., Karaman-Jurukovska, N., Morasso, M.I., Steinert, P.M., and Markova, N.G. 2000. Complex interactions between epidermal POU domain and activator protein 1 transcription factors regulate the expression of the profilaggrin gene in normal human epidermal keratinocytes. *J Biol Chem* **275**: 15295-15304.
- Jang, S.I. and Steinert, P.M. 2002. Loricrin expression in cultured human keratinocytes is controlled by a complex interplay between transcription factors of the Sp1, CREB, AP1, and AP2 families. *J Biol Chem* **277**: 42268-42279.
- Jansen, B.J., van Ruissen, F., de Jongh, G., Zeeuwen, P.L., and Schalkwijk, J. 2001. Serial analysis of gene expression in differentiated cultures of human epidermal keratinocytes. *J Invest Dermatol* **116**: 12-22.
- Jaubert, J., Cheng, J., and Segre, J.A. 2003. Ectopic expression of kruppel like factor 4 (Klf4) accelerates formation of the epidermal permeability barrier. *Development* **130**: 2767-2777.
- Jensen, J.M., Folster-Holst, R., Baranowsky, A., Schunck, M., Winoto-Morbach, S., Neumann, C., Schutze, S., and Proksch, E. 2004. Impaired sphingomyelinase activity and epidermal differentiation in atopic dermatitis. *J Invest Dermatol* **122**: 1423-1431.
- Ji, Y., Lalli, M.J., Babu, G.J., Xu, Y., Kirkpatrick, D.L., Liu, L.H., Chiamvimonvat, N., Walsh, R.A., Shull, G.E., and Periasamy, M. 2000. Disruption of a single copy of the SERCA2 gene results in altered Ca²⁺ homeostasis and cardiomyocyte function. *J Biol Chem* **275**: 38073-38080.
- Jiang, Y.J., Kim, P., Elias, P.M., and Feingold, K.R. 2005. LXR and PPAR activators stimulate cholesterol sulfotransferase type 2 isoform 1b in human keratinocytes. *J Lipid Res* **46**: 2657-2666.
- Jiang, Y.J., Lu, B., Kim, P., Paragh, G., Schmitz, G., Elias, P.M., and Feingold, K.R. 2008. PPAR and LXR activators regulate ABCA12 expression in human keratinocytes. *J Invest Dermatol* **128**: 104-109.
- Jobard, F., Lefevre, C., Karaduman, A., Blanchet-Bardon, C., Emre, S., Weissenbach, J., Ozguc, M., Lathrop, M., Prud'homme, J.F., and Fischer, J. 2002. Lipoxygenase-3 (ALOXE3) and 12(R)-lipoxygenase (ALOX12B) are mutated in non-bullous congenital ichthyosiform erythroderma (NCIE) linked to chromosome 17p13.1. *Hum Mol Genet* **11**: 107-113.
- Jonca, N., Guerrin, M., Hadjiolova, K., Caubet, C., Gallinaro, H., Simon, M., and Serre, G. 2002. Corneodesmosin, a component of epidermal corneocyte desmosomes, displays homophilic adhesive properties. *J Biol Chem* **277**: 5024-5029.
- Jones, P. and Simons, B.D. 2008. Epidermal homeostasis: do committed progenitors work while stem cells sleep? *Nat Rev Mol Cell Biol* **9**: 82-88.
- Jullien, D. and Barker, J.N. 2006. Genetics of psoriasis. *JEADV* **20**: 42-51.
- Kafatos, F.C., Jones, C.W., and Efstratiadis, A. 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res* **7**: 1541-1552.

- Kainu, K., Kivinen, K., Zucchelli, M., Suomela, S., Kere, J., Inerot, A., Baker, B.S., Powles, A.V., Fry, L., Samuelsson, L., and Saarialho-Kere, U. 2009. Association of psoriasis to PGLYRP and SPRR genes at PSORS4 locus on 1q shows heterogeneity between Finnish, Swedish and Irish families. *Exp Dermatol* **18**: 109-115.
- Kamata, Y., Taniguchi, A., Yamamoto, M., Nomura, J., Ishihara, K., Takahara, H., Hibino, T., and Takeda, A. 2009. Neutral cysteine protease bleomycin hydrolase is essential for the breakdown of deiminated filaggrin into amino acids. *J Biol Chem*.
- Kaminski, W.E., Piehler, A., and Wenzel, J.J. 2006. ABC A-subfamily transporters: structure, function and disease. *Biochim Biophys Acta* **1762**: 510-524.
- Kazerounian, S. and Aho, S. 2003. Characterization of periphilin, a widespread, highly insoluble nuclear protein and potential constituent of the keratinocyte cornified envelope. *J Biol Chem* **278**: 36707-36717.
- Keating, D.J., Dubach, D., Zanin, M.P., Yu, Y., Martin, K., Zhao, Y.F., Chen, C., Porta, S., Arbones, M.L., Mittaz, L., and Pritchard, M.A. 2008. DSCR1/RCAN1 regulates vesicle exocytosis and fusion pore kinetics: implications for Down syndrome and Alzheimer's disease. *Hum Mol Genet* **17**: 1020-1030.
- Kelsell, D.P., Norgett, E.E., Unsworth, H., Teh, M.T., Cullup, T., Mein, C.A., Dopping-Hepenstal, P.J., Dale, B.A., Tadini, G., Fleckman, P., Stephens, K.G., Sybert, V.P., Mallory, S.B., North, B.V., Witt, D.R., Sprecher, E., Taylor, A.E., Ilchyshyn, A., Kennedy, C.T., Goodyear, H., Moss, C., Paige, D., Harper, J.I., Young, B.D., Leigh, I.M., Eady, R.A., and O'Toole, E.A. 2005. Mutations in ABCA12 underlie the severe congenital skin disease harlequin ichthyosis. *Am J Hum Genet* **76**: 794-803.
- Kharfi, M., El Fekih, N., Ammar, D., Jaafoura, H., Schwonbeck, S., van Steensel, M.A., Fazaa, B., Kamoun, M.R., and Fischer, J. 2009. A Missense Mutation in TGM5 Causes Acral Peeling Skin Syndrome in a Tunisian Family. *J Invest Dermatol*.
- Kimonis, V., DiGiovanna, J.J., Yang, J.M., Doyle, S.Z., Bale, S.J., and Compton, J.G. 1994. A mutation in the V1 end domain of keratin 1 in non-epidermolytic palmar-plantar keratoderma. *J Invest Dermatol* **103**: 764-769.
- Klar, J., Schweiger, M., Zimmerman, R., Zechner, R., Li, H., Torma, H., Vahlquist, A., Bouadjar, B., Dahl, N., and Fischer, J. 2009. Mutations in the fatty acid transport protein 4 gene cause the ichthyosis prematurity syndrome. *Am J Hum Genet* **85**: 248-253.
- Kljuic, A., Bazzi, H., Sundberg, J.P., Martinez-Mir, A., O'Shaughnessy, R., Mahoney, M.G., Levy, M., Montagutelli, X., Ahmad, W., Aita, V.M., Gordon, D., Uitto, J., Whiting, D., Ott, J., Fischer, S., Gilliam, T.C., Jahoda, C.A., Morris, R.J., Panteleyev, A.A., Nguyen, V.T., and Christiano, A.M. 2003. Desmoglein 4 in hair follicle differentiation and epidermal adhesion: evidence from inherited hypotrichosis and acquired pemphigus vulgaris. *Cell* **113**: 249-260.
- Kloting, N., Berndt, J., Kralisch, S., Kovacs, P., Fasshauer, M., Schon, M.R., Stumvoll, M., and Bluher, M. 2006. Vaspin gene expression in human adipose tissue: association with obesity and type 2 diabetes. *Biochem Biophys Res Commun* **339**: 430-436.
- Kobayashi, M.S., Asai, S., Ishikawa, K., Nishida, Y., Nagata, T., and Takahashi, Y. 2008. Global profiling of influence of intra-ischemic brain temperature on gene expression in rat brain. *Brain Res Rev* **58**: 171-191.
- Koch, P.J., de Viragh, P.A., Scharer, E., Bundman, D., Longley, M.A., Bickenbach, J., Kawachi, Y., Suga, Y., Zhou, Z., Huber, M., Hohl, D., Kartasova, T., Jarnik, M., Steven, A.C., and Roop, D.R. 2000. Lessons from loricrin-deficient mice: compensatory mechanisms maintaining skin barrier function in the absence of a major cornified envelope protein. *J Cell Biol* **151**: 389-400.

- Komuves, L., Oda, Y., Tu, C.L., Chang, W.H., Ho-Pao, C.L., Mauro, T., and Bikle, D.D. 2002. Epidermal expression of the full-length extracellular calcium-sensing receptor is required for normal keratinocyte differentiation. *J Cell Physiol* **192**: 45-54.
- Koria, P. and Andreadis, S.T. 2006. Epidermal morphogenesis: the transcriptional program of human keratinocytes during stratification. *J Invest Dermatol* **126**: 1834-1841.
- Koster, M.I. and Roop, D.R. 2007. Mechanisms regulating epithelial stratification. *Annu Rev Cell Dev Biol* **23**: 93-113.
- Kottke, M.D., Delva, E., and Kowalczyk, A.P. 2006. The desmosome: cell science lessons from human diseases. *J Cell Sci* **119**: 797-806.
- Kouzu, Y., Uzawa, K., Koike, H., Saito, K., Nakashima, D., Higo, M., Endo, Y., Kasamatsu, A., Shiiba, M., Bukawa, H., Yokoe, H., and Tanzawa, H. 2006. Overexpression of stathmin in oral squamous-cell carcinoma: correlation with tumour progression and poor prognosis. *Br J Cancer* **94**: 717-723.
- Kudryavtseva, E.I., Sugihara, T.M., Wang, N., Lasso, R.J., Gudnason, J.F., Lipkin, S.M., and Andersen, B. 2003. Identification and characterization of Grainyhead-like epithelial transactivator (GET-1), a novel mammalian Grainyhead-like factor. *Dev Dyn* **226**: 604-617.
- Kuramoto, N., Takizawa, T., Takizawa, T., Matsuki, M., Morioka, H., Robinson, J.M., and Yamanishi, K. 2002. Development of ichthyosiform skin compensates for defective permeability barrier function in mice lacking transglutaminase 1. *J Clin Invest* **109**: 243-250.
- Kurzen, H. and Schallreuter, K.U. 2004. Novel aspects in cutaneous biology of acetylcholine synthesis and acetylcholine receptors. *Exp Dermatol* **13 Suppl 4**: 27-30.
- Lai-Cheong, J.E., Arita, K., and McGrath, J.A. 2007. Genetic diseases of junctions. *J Invest Dermatol* **127**: 2713-2725.
- Lampe, M.A., Burlingame, A.L., Whitney, J., Williams, M.L., Brown, B.E., Roitman, E., and Elias, P.M. 1983. Human stratum corneum lipids: characterization and regional variations. *J Lipid Res* **24**: 120-130.
- Laneve, P., Gioia, U., Ragno, R., Altieri, F., Di Franco, C., Santini, T., Arceci, M., Bozzoni, I., and Caffarelli, E. 2008. The tumor marker human placental protein 11 is an endoribonuclease. *J Biol Chem* **283**: 34712-34719.
- Langbein, L., Heid, H.W., Moll, I., and Franke, W.W. 1993. Molecular characterization of the body site-specific human epidermal cytokeratin 9: cDNA cloning, amino acid sequence, and tissue specificity of gene expression. *Differentiation* **55**: 57-71.
- LaPres, J.J. and Hudson, L.G. 1996. Identification of a functional determinant of differentiation-dependent expression in the involucrin gene. *J Biol Chem* **271**: 23154-23160.
- Larderet, G., Fortunel, N.O., Vaigot, P., Cegalerba, M., Maltere, P., Zobiri, O., Gidrol, X., Waksman, G., and Martin, M.T. 2006. Human side population keratinocytes exhibit long-term proliferative potential and a specific gene expression profile and can form a pluristratified epidermis. *Stem Cells* **24**: 965-974.
- Le Behec, A., Zindy, P., Sierocinski, T., Petritis, D., Bihouee, A., Le Meur, N., Leger, J., and Theret, N. 2008. M@IA: a modular open-source application for microarray workflow and integrative datamining. *In Silico Biol* **8**: 63-69.
- Le Brigand, K. and Barbry, P. 2007. Mediante: a web-based microarray data manager. *Bioinformatics* **23**: 1304-1306.
- Le Brigand, K., Russell, R., Moreilhon, C., Rouillard, J.M., Jost, B., Amiot, F., Magnone, V., Bole-Feysot, C., Rostagno, P., Virolle, V., Defamie, V., Dessen, P., Williams, G., Lyons, P., Rios, G., Mari, B., Gulari, E., Kastner, P., Gidrol, X., Freeman, T.C., and

- Barbry, P. 2006. An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes. *Nucleic Acids Res* **34**: e87.
- Leclerc, E.A., Huchenq, A., Mattiuzzo, N.R., Metzger, D., Chambon, P., Ghyselinck, N.B., Serre, G., Jonca, N., and Guerrin, M. 2009. Corneodesmosin gene ablation induces lethal skin-barrier disruption and hair-follicle degeneration related to desmosome dysfunction. *J Cell Sci*.
- Lecuyer, C., Dacheux, J.L., Hermand, E., Mazeman, E., Rousseaux, J., and Rousseaux-Prevost, R. 2000. Actin-binding properties and colocalization with actin during spermiogenesis of mammalian sperm calicin. *Biol Reprod* **63**: 1801-1810.
- Lefevre, C., Audebert, S., Jobard, F., Bouadjar, B., Lakhdar, H., Boughdene-Stambouli, O., Blanchet-Bardon, C., Heilig, R., Foglio, M., Weissenbach, J., Lathrop, M., Prud'homme, J.F., and Fischer, J. 2003. Mutations in the transporter ABCA12 are associated with lamellar ichthyosis type 2. *Hum Mol Genet* **12**: 2369-2378.
- Lefevre, C., Bouadjar, B., Ferrand, V., Tadini, G., Megarbane, A., Lathrop, M., Prud'homme, J.F., and Fischer, J. 2006. Mutations in a new cytochrome P450 gene in lamellar ichthyosis type 3. *Hum Mol Genet* **15**: 767-776.
- Lefevre, C., Bouadjar, B., Karaduman, A., Jobard, F., Saker, S., Ozguc, M., Lathrop, M., Prud'homme, J.F., and Fischer, J. 2004. Mutations in ichthyin a new gene on chromosome 5q33 in a new form of autosomal recessive congenital ichthyosis. *Hum Mol Genet* **13**: 2473-2482.
- Lefevre, C., Jobard, F., Caux, F., Bouadjar, B., Karaduman, A., Heilig, R., Lakhdar, H., Wollenberg, A., Verret, J.L., Weissenbach, J., Ozguc, M., Lathrop, M., Prud'homme, J.F., and Fischer, J. 2001. Mutations in CGI-58, the gene encoding a new protein of the esterase/lipase/thioesterase subfamily, in Chanarin-Dorfman syndrome. *Am J Hum Genet* **69**: 1002-1012.
- Lemaitre, G., Lamartine, J., Pitaval, A., Vaigot, P., Garin, J., Bouet, S., Petat, C., Soularue, P., Gidrol, X., Martin, M.T., and Waksman, G. 2004. Expression profiling of genes and proteins in HaCaT keratinocytes: proliferating versus differentiated state. *J Cell Biochem* **93**: 1048-1062.
- Lepperdinger, G., Mullegger, J., and Kreil, G. 2001. Hyal2--less active, but more versatile? *Matrix Biol* **20**: 509-514.
- Lessin, S.R., Huebner, K., Isobe, M., Croce, C.M., and Steinert, P.M. 1988. Chromosomal mapping of human keratin genes: evidence of non-linkage. *J Invest Dermatol* **91**: 572-578.
- Lesueur, F., Oudot, T., Heath, S., Foglio, M., Lathrop, M., Prud'homme, J.F., and Fischer, J. 2007. ADAM33, a new candidate for psoriasis susceptibility. *PLoS One* **2**: e906.
- Levy-Nissenbaum, E., Betz, R.C., Frydman, M., Simon, M., Lahat, H., Bakhan, T., Goldman, B., Bygum, A., Pierick, M., Hillmer, A.M., Jonca, N., Toribio, J., Kruse, R., Dewald, G., Cichon, S., Kubisch, C., Guerrin, M., Serre, G., Nothen, M.M., and Pras, E. 2003. Hypotrichosis simplex of the scalp is associated with nonsense mutations in CDSN encoding corneodesmosin. *Nat Genet* **34**: 151-153.
- Levy, V., Lindon, C., Harfe, B.D., and Morgan, B.A. 2005. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. *Dev Cell* **9**: 855-861.
- Li, A., Simmons, P.J., and Kaur, P. 1998. Identification and isolation of candidate human keratinocyte stem cells based on cell surface phenotype. *Proc Natl Acad Sci U S A* **95**: 3902-3907.
- Li, Q., Lu, Q., Estepa, G., and Verma, I.M. 2005. Identification of 14-3-3sigma mutation causing cutaneous abnormality in repeated-epilation mutant mouse. *Proc Natl Acad Sci U S A* **102**: 15977-15982.

- Lin, J.C., Wang, W.Y., Chen, K.Y., Wei, Y.H., Liang, W.M., Jan, J.S., and Jiang, R.S. 2004. Quantification of plasma Epstein-Barr virus DNA in patients with advanced nasopharyngeal carcinoma. *N Engl J Med* **350**: 2461-2470.
- List, K., Szabo, R., Wertz, P.W., Segre, J., Haudenschild, C.C., Kim, S.Y., and Bugge, T.H. 2003. Loss of proteolytically processed filaggrin caused by epidermal deletion of Matriptase/MT-SP1. *J Cell Biol* **163**: 901-910.
- Litjens, S.H., de Pereda, J.M., and Sonnenberg, A. 2006. Current insights into the formation and breakdown of hemidesmosomes. *Trends Cell Biol* **16**: 376-383.
- Liu, Y., Helms, C., Liao, W., Zaba, L.C., Duan, S., Gardner, J., Wise, C., Miner, A., Malloy, M.J., Pullinger, C.R., Kane, J.P., Saccone, S., Worthington, J., Bruce, I., Kwok, P.Y., Menter, A., Krueger, J., Barton, A., Saccone, N.L., and Bowcock, A.M. 2008. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet* **4**: e1000041.
- Livak, K.J. and Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.
- Lowes, M.A., Bowcock, A.M., and Krueger, J.G. 2007. Pathogenesis and therapy of psoriasis. *Nature* **445**: 866-873.
- Lu, J., Goldstein, K.M., Chen, P., Huang, S., Gelbert, L.M., and Nagpal, S. 2005. Transcriptional profiling of keratinocytes reveals a vitamin D-regulated epidermal differentiation network. *J Invest Dermatol* **124**: 778-785.
- Lu, S., Saydak, M., Gentile, V., Stein, J.P., and Davies, P.J. 1995. Isolation and characterization of the human tissue transglutaminase gene promoter. *J Biol Chem* **270**: 9748-9756.
- Lundstrom, A. and Egelrud, T. 1990. Evidence that cell shedding from plantar stratum corneum in vitro involves endogenous proteolysis of the desmosomal protein desmoglein I. *J Invest Dermatol* **94**: 216-220.
- Lundstrom, A., Serre, G., Haftek, M., and Egelrud, T. 1994. Evidence for a role of corneodesmosin, a protein which may serve to modify desmosomes during cornification, in stratum corneum cell cohesion and desquamation. *Arch Dermatol Res* **286**: 369-375.
- Lunny, D.P., Weed, E., Nolan, P.M., Marquardt, A., Augustin, M., and Porter, R.M. 2005. Mutations in gasdermin 3 cause aberrant differentiation of the hair follicle and sebaceous gland. *J Invest Dermatol* **124**: 615-621.
- Lyle, S., Christofidou-Solomidou, M., Liu, Y., Elder, D.E., Albelda, S., and Cotsarelis, G. 1998. The C8/144B monoclonal antibody recognizes cytokeratin 15 and defines the location of human hair follicle stem cells. *J Cell Sci* **111 (Pt 21)**: 3179-3188.
- Lyons, P.A., Koukoulaki, M., Hatton, A., Doggett, K., Woffendin, H.B., Chaudhry, A.N., and Smith, K.G. 2007. Microarray analysis of human leucocyte subsets: the advantages of positive selection and rapid purification. *BMC Genomics* **8**: 64.
- Ma, K., Cheung, S.M., Marshall, A.J., and Duronio, V. 2008. PI(3,4,5)P3 and PI(3,4)P2 levels correlate with PKB/akt phosphorylation at Thr308 and Ser473, respectively; PI(3,4)P2 levels determine PKB activity. *Cell Signal* **20**: 684-694.
- Maatta, A., DiColandrea, T., Groot, K., and Watt, F.M. 2001. Gene targeting of envoplakin, a cytoskeletal linker protein and precursor of the epidermal cornified envelope. *Mol Cell Biol* **21**: 7047-7053.
- Madison, K.C., Sando, G.N., Howard, E.J., True, C.A., Gilbert, D., Swartzendruber, D.C., and Wertz, P.W. 1998. Lamellar granule biogenesis: a role for ceramide glucosyltransferase, lysosomal enzyme transport, and the Golgi. *J Invest Dermatol Symp Proc* **3**: 80-86.

- Maestrini, E., Monaco, A.P., McGrath, J.A., Ishida-Yamamoto, A., Camisa, C., Hovnanian, A., Weeks, D.E., Lathrop, M., Uitto, J., and Christiano, A.M. 1996. A molecular defect in loricerin, the major component of the cornified cell envelope, underlies Vohwinkel's syndrome. *Nat Genet* **13**: 70-77.
- Magert, H.J., Standker, L., Kreutzmann, P., Zucht, H.D., Reinecke, M., Sommerhoff, C.P., Fritz, H., and Forssmann, W.G. 1999. LEKTI, a novel 15-domain type of human serine proteinase inhibitor. *J Biol Chem* **274**: 21499-21502.
- Magnusdottir, E., Kalachikov, S., Mizukoshi, K., Savitsky, D., Ishida-Yamamoto, A., Panteleyev, A.A., and Calame, K. 2007. Epidermal terminal differentiation depends on B lymphocyte-induced maturation protein-1. *Proc Natl Acad Sci U S A* **104**: 14988-14993.
- Mao-Qiang, M., Feingold, K.R., Jain, M., and Elias, P.M. 1995. Extracellular processing of phospholipids is required for permeability barrier homeostasis. *J Lipid Res* **36**: 1925-1935.
- Marshall, H., Studer, M., Popperl, H., Aparicio, S., Kuroiwa, A., Brenner, S., and Krumlauf, R. 1994. A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1. *Nature* **370**: 567-571.
- Matsui, T., Hayashi-Kisumi, F., Kinoshita, Y., Katahira, S., Morita, K., Miyachi, Y., Ono, Y., Imai, T., Tanigawa, Y., Komiya, T., and Tsukita, S. 2004. Identification of novel keratinocyte-secreted peptides dermokine-alpha/-beta and a new stratified epithelium-secreted protein gene complex on human chromosome 19q13.1. *Genomics* **84**: 384-397.
- Matsuki, M., Yamashita, F., Ishida-Yamamoto, A., Yamada, K., Kinoshita, C., Fushiki, S., Ueda, E., Morishima, Y., Tabata, K., Yasuno, H., Hashida, M., Iizuka, H., Ikawa, M., Okabe, M., Kondoh, G., Kinoshita, T., Takeda, J., and Yamanishi, K. 1998. Defective stratum corneum and early neonatal death in mice lacking the gene for transglutaminase 1 (keratinocyte transglutaminase). *Proc Natl Acad Sci U S A* **95**: 1044-1049.
- Matsumoto, M., Zhou, Y., Matsuo, S., Nakanishi, H., Hirose, K., Oura, H., Arase, S., Ishida-Yamamoto, A., Bando, Y., Izumi, K., Kiyonari, H., Oshima, N., Nakayama, R., Matsushima, A., Hirota, F., Mouri, Y., Kuroda, N., Sano, S., and Chaplin, D.D. 2008. Targeted deletion of the murine corneodesmosin gene delineates its essential role in skin and hair physiology. *Proc Natl Acad Sci U S A* **105**: 6720-6724.
- Mattie, M.D., Benz, C.C., Bowers, J., Sensinger, K., Wong, L., Scott, G.K., Fedele, V., Ginzinger, D., Getts, R., and Haqq, C. 2006. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Mol Cancer* **5**: 24.
- Mauro, T., Guitard, M., Behne, M., Oda, Y., Crumrine, D., Komuves, L., Rassner, U., Elias, P.M., and Hummler, E. 2002. The ENaC channel is required for normal epidermal differentiation. *J Invest Dermatol* **118**: 589-594.
- Maytin, E.V., Lin, J.C., Krishnamurthy, R., Batchvarova, N., Ron, D., Mitchell, P.J., and Habener, J.F. 1999. Keratin 10 gene expression during differentiation of mouse epidermis requires transcription factors C/EBP and AP-2. *Dev Biol* **216**: 164-181.
- McGrath, J.A. 2008. Filaggrin and the great epidermal barrier grief. *Australas J Dermatol* **49**: 67-73; quiz 73-64.
- McLean, W.H., Eady, R.A., Dopping-Hepenstal, P.J., McMillan, J.R., Leigh, I.M., Navsaria, H.A., Higgins, C., Harper, J.I., Paige, D.G., Morley, S.M., and et al. 1994a. Mutations in the rod 1A domain of keratins 1 and 10 in bullous congenital ichthyosiform erythroderma (BCIE). *J Invest Dermatol* **102**: 24-30.

- McLean, W.H., Morley, S.M., Lane, E.B., Eady, R.A., Griffiths, W.A., Paige, D.G., Harper, J.I., Higgins, C., and Leigh, I.M. 1994b. Ichthyosis bullosa of Siemens--a disease involving keratin 2e. *J Invest Dermatol* **103**: 277-281.
- Meade, J.L., de Wynter, E.A., Brett, P., Sharif, S.M., Woods, C.G., Markham, A.F., and Cook, G.P. 2006. A family with Papillon-Lefevre syndrome reveals a requirement for cathepsin C in granzyme B activation and NK cell cytolytic activity. *Blood* **107**: 3665-3668.
- Medini, D., Serruto, D., Parkhill, J., Relman, D.A., Donati, C., Moxon, R., Falkow, S., and Rappuoli, R. 2008. Microbiology in the post-genomic era. *Nat Rev Microbiol* **6**: 419-430.
- Mehic, D., Bakiri, L., Ghannadan, M., Wagner, E.F., and Tschachler, E. 2005. Fos and jun proteins are specifically expressed during differentiation of human keratinocytes. *J Invest Dermatol* **124**: 212-220.
- Mehrel, T., Hohl, D., Rothnagel, J.A., Longley, M.A., Bundman, D., Cheng, C., Lichti, U., Bisher, M.E., Steven, A.C., Steinert, P.M., and et al. 1990. Identification of a major keratinocyte cell envelope protein, loricrin. *Cell* **61**: 1103-1112.
- Mehul, B., Asselineau, D., Bernard, D., Leclaire, J., Regnier, M., Schmidt, R., and Bernerd, F. 2004. Gene expression profiles of three different models of reconstructed human epidermis and classical cultures of keratinocytes using cDNA arrays. *Arch Dermatol Res* **296**: 145-156.
- Menon, G.K., Elias, P.M., Lee, S.H., and Feingold, K.R. 1992. Localization of calcium in murine epidermis following disruption and repair of the permeability barrier. *Cell Tissue Res* **270**: 503-512.
- Merritt, A.J., Berika, M.Y., Zhai, W., Kirk, S.E., Ji, B., Hardman, M.J., and Garrod, D.R. 2002. Suprabasal desmoglein 3 expression in the epidermis of transgenic mice results in hyperproliferation and abnormal differentiation. *Mol Cell Biol* **22**: 5846-5858.
- Mizutani, Y., Kihara, A., Chiba, H., Tojo, H., and Igarashi, Y. 2008. 2-Hydroxy-ceramide synthesis by ceramide synthase family: enzymatic basis for the preference of FA chain length. *J Lipid Res* **49**: 2356-2364.
- Mizutani, Y., Kihara, A., and Igarashi, Y. 2006. LASS3 (longevity assurance homologue 3) is a mainly testis-specific (dihydro)ceramide synthase with relatively broad substrate specificity. *Biochem J* **398**: 531-538.
- Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R. 2005. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1-15.
- Moran, J.L., Qiu, H., Turbe-Doan, A., Yun, Y., Boeglin, W.E., Brash, A.R., and Beier, D.R. 2007. A mouse mutation in the 12R-lipoxygenase, Alox12b, disrupts formation of the epidermal permeability barrier. *J Invest Dermatol* **127**: 1893-1897.
- Morasso, M.I., Grinberg, A., Robinson, G., Sargent, T.D., and Mahon, K.A. 1999. Placental failure in mice lacking the homeobox gene *Dlx3*. *Proc Natl Acad Sci U S A* **96**: 162-167.
- Morasso, M.I., Markova, N.G., and Sargent, T.D. 1996. Regulation of epidermal differentiation by a Distal-less homeodomain gene. *J Cell Biol* **135**: 1879-1887.
- Moreillon, C., Gras, D., Hologne, C., Bajolet, O., Cottrez, F., Magnone, V., Merten, M., Groux, H., Puchelle, E., and Barbry, P. 2005. Live *Staphylococcus aureus* and bacterial soluble factors induce different transcriptional responses in human airway cells. *Physiol Genomics* **20**: 244-255.
- Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., McDonald, H., Varhol, R., Jones, S., and Marra, M. 2008a. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45**: 81-94.

- Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C.J., and Marra, M.A. 2008b. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res* **18**: 610-621.
- Morris, R.J., Liu, Y., Marles, L., Yang, Z., Trempus, C., Li, S., Lin, J.S., Sawicki, J.A., and Cotsarelis, G. 2004. Capturing and profiling adult hair follicle stem cells. *Nat Biotechnol* **22**: 411-417.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621-628.
- Nachat, R., Mechin, M.C., Takahara, H., Chavanas, S., Charveron, M., Serre, G., and Simon, M. 2005. Peptidylarginine deiminase isoforms 1-3 are expressed in the epidermis and involved in the deimination of K1 and filaggrin. *J Invest Dermatol* **124**: 384-393.
- Nair, R.P., Duffin, K.C., Helms, C., Ding, J., Stuart, P.E., Goldgar, D., Gudjonsson, J.E., Li, Y., Tejasvi, T., Feng, B.J., Ruether, A., Schreiber, S., Weichenthal, M., Gladman, D., Rahman, P., Schrodi, S.J., Prahalad, S., Guthery, S.L., Fischer, J., Liao, W., Kwok, P.Y., Menter, A., Lathrop, G.M., Wise, C.A., Begovich, A.B., Voorhees, J.J., Elder, J.T., Krueger, G.G., Bowcock, A.M., and Abecasis, G.R. 2009. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet* **41**: 199-204.
- Nair, R.P., Stuart, P.E., Nistor, I., Hiremagalore, R., Chia, N.V., Jenisch, S., Weichenthal, M., Abecasis, G.R., Lim, H.W., Christophers, E., Voorhees, J.J., and Elder, J.T. 2006. Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am J Hum Genet* **78**: 827-851.
- Nakamura, S., Yang, C.S., Sakon, N., Ueda, M., Tougan, T., Yamashita, A., Goto, N., Takahashi, K., Yasunaga, T., Ikuta, K., Mizutani, T., Okamoto, Y., Tagami, M., Morita, R., Maeda, N., Kawai, J., Hayashizaki, Y., Nagai, Y., Horii, T., Iida, T., and Nakaya, T. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS ONE* **4**: e4219.
- Nakayama, Y., Nara, N., Kawakita, Y., Takeshima, Y., Arakawa, M., Katoh, M., Morita, S., Iwatsuki, K., Tanaka, K., Okamoto, S., Kitamura, T., Seki, N., Matsuda, R., Matsuo, M., Saito, K., and Hara, T. 2004. Cloning of cDNA encoding a regeneration-associated muscle protease whose expression is attenuated in cell lines derived from Duchenne muscular dystrophy patients. *Am J Pathol* **164**: 1773-1782.
- Neiva, K.G., Zhang, Z., Miyazawa, M., Warner, K.A., Karl, E., and Nor, J.E. 2009. Cross talk initiated by endothelial cells enhances migration and inhibits anoikis of squamous cell carcinoma cells through STAT3/Akt/ERK signaling. *Neoplasia* **11**: 583-593.
- Nemes, Z., Marekov, L.N., Fesus, L., and Steinert, P.M. 1999. A novel function for transglutaminase 1: attachment of long-chain omega-hydroxyceramides to involucrin by ester bond formation. *Proc Natl Acad Sci U S A* **96**: 8402-8407.
- Ng, D.C., Shafae, S., Lee, D., and Bikle, D.D. 2000. Requirement of an AP-1 site in the calcium response region of the involucrin promoter. *J Biol Chem* **275**: 24080-24088.
- Norlen, L. 2001. Skin barrier formation: the membrane folding model. *J Invest Dermatol* **117**: 823-829.
- Norlen, L., Al-Amoudi, A., and Dubochet, J. 2003. A cryotransmission electron microscopy study of skin barrier formation. *J Invest Dermatol* **120**: 555-560.
- Nukui, T., Ehama, R., Sakaguchi, M., Sonogawa, H., Katagiri, C., Hibino, T., and Huh, N.H. 2008. S100A8/A9, a key mediator for positive feedback growth stimulation of normal human keratinocytes. *J Cell Biochem* **104**: 453-464.

- Okunade, G.W., Miller, M.L., Azhar, M., Andringa, A., Sanford, L.P., Doetschman, T., Prasad, V., and Shull, G.E. 2007. Loss of the Atp2c1 secretory pathway Ca(2+)-ATPase (SPCA1) in mice causes Golgi stress, apoptosis, and midgestational death in homozygous embryos and squamous cell tumors in adult heterozygotes. *J Biol Chem* **282**: 26517-26527.
- Palmer, C.N., Irvine, A.D., Terron-Kwiatkowski, A., Zhao, Y., Liao, H., Lee, S.P., Goudie, D.R., Sandilands, A., Campbell, L.E., Smith, F.J., O'Regan, G.M., Watson, R.M., Cecil, J.E., Bale, S.J., Compton, J.G., DiGiovanna, J.J., Fleckman, P., Lewis-Jones, S., Arseculeratne, G., Sergeant, A., Munro, C.S., El Houate, B., McElreavey, K., Halkjaer, L.B., Bisgaard, H., Mukhopadhyay, S., and McLean, W.H. 2006. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* **38**: 441-446.
- Palmer, H.G., Anjos-Afonso, F., Carmeliet, G., Takeda, H., and Watt, F.M. 2008. The vitamin D receptor is a Wnt effector that controls hair follicle differentiation and specifies tumor type in adult epidermis. *PLoS ONE* **3**: e1483.
- Paromov, V.M. and Morton, R.E. 2003. Lipid transfer inhibitor protein defines the participation of high density lipoprotein subfractions in lipid transfer reactions mediated by cholesterol ester transfer protein (CETP). *J Biol Chem* **278**: 40859-40866.
- Patel, S., Xi, Z.F., Seo, E.Y., McGaughey, D., and Segre, J.A. 2006. Klf4 and corticosteroids activate an overlapping set of transcriptional targets to accelerate in utero epidermal barrier acquisition. *Proc Natl Acad Sci U S A* **103**: 18668-18673.
- Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., and Fodor, S.P. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* **91**: 5022-5026.
- Pellegrini, G., Dellambra, E., Golisano, O., Martinelli, E., Fantozzi, I., Bondanza, S., Ponzin, D., McKeon, F., and De Luca, M. 2001. p63 identifies keratinocyte stem cells. *Proc Natl Acad Sci U S A* **98**: 3156-3161.
- Pillai, S., Bikle, D.D., Mancianti, M.L., Cline, P., and Hincenbergs, M. 1990. Calcium regulation of growth and differentiation of normal human keratinocytes: modulation of differentiation competence by stages of growth and extracellular calcium. *J Cell Physiol* **143**: 294-302.
- Pillai, S., Bikle, D.D., Su, M.J., Ratnam, A., and Abe, J. 1995. 1,25-Dihydroxyvitamin D3 upregulates the phosphatidylinositol signaling pathway in human keratinocytes by increasing phospholipase C levels. *J Clin Invest* **96**: 602-609.
- Potten, C.S. 1974. The epidermal proliferative unit: the possible role of the central basal cell. *Cell Tissue Kinet* **7**: 77-88.
- Presland, R.B., Haydock, P.V., Fleckman, P., Nirunskisiri, W., and Dale, B.A. 1992. Characterization of the human epidermal profilaggrin gene. Genomic organization and identification of an S-100-like calcium binding domain at the amino terminus. *J Biol Chem* **267**: 23772-23781.
- Prowse, D.M., Lee, D., Weiner, L., Jiang, N., Magro, C.M., Baden, H.P., and Brissette, J.L. 1999. Ectopic expression of the nude gene induces hyperproliferation and defects in differentiation: implications for the self-renewal of cutaneous epithelia. *Dev Biol* **212**: 54-67.
- Puissegur, M.P., Lay, G., Gilleron, M., Botella, L., Nigou, J., Marrakchi, H., Mari, B., Duteyrat, J.L., Guerardel, Y., Kremer, L., Barbry, P., Puzo, G., and Altare, F. 2007. Mycobacterial lipomannan induces granuloma macrophage fusion via a TLR2-dependent, ADAM9- and beta1 integrin-mediated pathway. *J Immunol* **178**: 3161-3169.

- Pusztai, L., Mazouni, C., Anderson, K., Wu, Y., and Symmans, W.F. 2006. Molecular classification of breast cancer: limitations and potential. *Oncologist* **11**: 868-877.
- Radoja, N., Gazel, A., Banno, T., Yano, S., and Blumenberg, M. 2006. Transcriptional profiling of epidermal differentiation. *Physiol Genomics* **27**: 65-78.
- Rambaldi, A., Carlotti, E., Oldani, E., Della Starza, I., Baccarani, M., Cortelazzo, S., Lauria, F., Arcaini, L., Morra, E., Pulsoni, A., Rigacci, L., Rupolo, M., Zaja, F., Zinzani, P.L., Barbui, T., and Foa, R. 2005. Quantitative PCR of bone marrow BCL2/IgH+ cells at diagnosis predicts treatment response and long-term outcome in follicular non-Hodgkin lymphoma. *Blood* **105**: 3428-3433.
- Rassner, U., Feingold, K.R., Crumrine, D.A., and Elias, P.M. 1999. Coordinate assembly of lipids and enzyme proteins into epidermal lamellar bodies. *Tissue Cell* **31**: 489-498.
- Ratnam, A.V., Bikle, D.D., and Cho, J.K. 1999. 1,25 dihydroxyvitamin D3 enhances the calcium response of keratinocytes. *J Cell Physiol* **178**: 188-196.
- Reiner, A., Yekutieli, D., and Benjamini, Y. 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**: 368-375.
- Rice, R.H. and Green, H. 1977. The cornified envelope of terminally differentiated human epidermal keratinocytes consists of cross-linked protein. *Cell* **11**: 417-422.
- Roberts, D. and Marks, R. 1980. The determination of regional and age variations in the rate of desquamation: a comparison of four techniques. *J Invest Dermatol* **74**: 13-16.
- Rochat, A., Kobayashi, K., and Barrandon, Y. 1994. Location of stem cells of human hair follicles by clonal analysis. *Cell* **76**: 1063-1073.
- Roepman, P., Wessels, L.F., Kettelarij, N., Kemmeren, P., Miles, A.J., Lijnzaad, P., Tilanus, M.G., Koole, R., Hordijk, G.J., van der Vliet, P.C., Reinders, M.J., Slootweg, P.J., and Holstege, F.C. 2005. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet* **37**: 182-186.
- Rothermel, B., Vega, R.B., Yang, J., Wu, H., Bassel-Duby, R., and Williams, R.S. 2000. A protein encoded within the Down syndrome critical region is enriched in striated muscles and inhibits calcineurin signaling. *J Biol Chem* **275**: 8719-8725.
- Saeki, N., Kuwahara, Y., Sasaki, H., Satoh, H., and Shiroishi, T. 2000. Gasdermin (Gsdm) localizing to mouse Chromosome 11 is predominantly expressed in upper gastrointestinal tract but significantly suppressed in human gastric cancer cells. *Mamm Genome* **11**: 718-724.
- Saeki, N., Usui, T., Aoyagi, K., Kim, D.H., Sato, M., Mabuchi, T., Yanagihara, K., Ogawa, K., Sakamoto, H., Yoshida, T., and Sasaki, H. 2009. Distinctive expression and function of four GSDM family genes (GSDMA-D) in normal and malignant upper gastrointestinal epithelium. *Genes Chromosomes Cancer* **48**: 261-271.
- Sakabe, N.J., de Souza, J.E., Galante, P.A., de Oliveira, P.S., Passetti, F., Brentani, H., Osorio, E.C., Zaiats, A.C., Leerkes, M.R., Kitajima, J.P., Brentani, R.R., Strausberg, R.L., Simpson, A.J., and de Souza, S.J. 2003. ORESTES are enriched in rare exon usage variants affecting the encoded proteins. *C R Biol* **326**: 979-985.
- Sanderson, R.D., Hinkes, M.T., and Bernfield, M. 1992. Syndecan-1, a cell-surface proteoglycan, changes in size and abundance when keratinocytes stratify. *J Invest Dermatol* **99**: 390-396.
- Sato, J., Denda, M., Nakanishi, J., Nomura, J., and Koyama, J. 1998. Cholesterol sulfate inhibits proteases that are involved in desquamation of stratum corneum. *J Invest Dermatol* **111**: 189-193.
- Sayama, K., Hanakawa, Y., Nagai, H., Shirakata, Y., Dai, X., Hirakawa, S., Tokumaru, S., Tohyama, M., Yang, L., Sato, S., Shizuo, A., and Hashimoto, K. 2006. Transforming growth factor-beta-activated kinase 1 is essential for differentiation and the prevention of apoptosis in epidermis. *J Biol Chem* **281**: 22013-22020.

- Schalkwijk, J., Chang, A., Janssen, P., De Jongh, G.J., and Mier, P.D. 1990. Skin-derived antileucoproteases (SKALPs): characterization of two new elastase inhibitors from psoriatic epidermis. *Br J Dermatol* **122**: 631-641.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470.
- Schluter, H., Moll, I., Wolburg, H., and Franke, W.W. 2007. The different structures containing tight junction proteins in epidermal and other stratified epithelial cells, including squamous cell metaplasia. *Eur J Cell Biol* **86**: 645-655.
- Schmuth, M., Jiang, Y.J., Dubrac, S., Elias, P.M., and Feingold, K.R. 2008. Thematic review series: skin lipids. Peroxisome proliferator-activated receptors and liver X receptors in epidermal biology. *J Lipid Res* **49**: 499-509.
- Schweizer, J., Bowden, P.E., Coulombe, P.A., Langbein, L., Lane, E.B., Magin, T.M., Maltais, L., Omary, M.B., Parry, D.A., Rogers, M.A., and Wright, M.W. 2006. New consensus nomenclature for mammalian keratins. *J Cell Biol* **174**: 169-174.
- Scott, I.R., Harding, C.R., and Barrett, J.G. 1982. Histidine-rich protein of the keratohyalin granules. Source of the free amino acids, urocanic acid and pyrrolidone carboxylic acid in the stratum corneum. *Biochim Biophys Acta* **719**: 110-117.
- Segre, J.A., Bauer, C., and Fuchs, E. 1999. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet* **22**: 356-360.
- Serre, G., Mils, V., Haftek, M., Vincent, C., Croute, F., Reano, A., Ouhayoun, J.P., Bettinger, S., and Soleilhavoup, J.P. 1991. Identification of late differentiation antigens of human cornified epithelia, expressed in re-organized desmosomes and bound to cross-linked envelope. *J Invest Dermatol* **97**: 1061-1072.
- Serup, J. and Blichmann, C. 1987. Epidermal hydration of psoriasis plaques and the relation to scaling. Measurement of electrical conductance and transepidermal water loss. *Acta Derm Venereol* **67**: 357-359.
- Sevilla, L.M., Nachat, R., Groot, K.R., Klement, J.F., Uitto, J., Djian, P., Maatta, A., and Watt, F.M. 2007. Mice deficient in involucrin, envoplakin, and periplakin have a defective epidermal barrier. *J Cell Biol* **179**: 1599-1612.
- Sevilla, L.M., Nachat, R., Groot, K.R., and Watt, F.M. 2008. Kazrin regulates keratinocyte cytoskeletal networks, intercellular junctions and differentiation. *J Cell Sci* **121**: 3561-3569.
- Shields, J.M., Christy, R.J., and Yang, V.W. 1996. Identification and characterization of a gene encoding a gut-enriched Kruppel-like factor expressed during growth arrest. *J Biol Chem* **271**: 20009-20017.
- Shim, J.H., Xiao, C., Paschal, A.E., Bailey, S.T., Rao, P., Hayden, M.S., Lee, K.Y., Bussey, C., Steckel, M., Tanaka, N., Yamada, G., Akira, S., Matsumoto, K., and Ghosh, S. 2005. TAK1, but not TAB1 or TAB2, plays an essential role in multiple signaling pathways in vivo. *Genes Dev* **19**: 2668-2681.
- Shulenin, S., Noguee, L.M., Annilo, T., Wert, S.E., Whitsett, J.A., and Dean, M. 2004. ABCA3 gene mutations in newborns with fatal surfactant deficiency. *N Engl J Med* **350**: 1296-1303.
- Simon, M., Haftek, M., Sebbag, M., Montezin, M., Girbal-Neuhauser, E., Schmitt, D., and Serre, G. 1996. Evidence that filaggrin is a component of cornified cell envelopes in human plantar epidermis. *Biochem J* **317** (Pt 1): 173-177.
- Simon, M., Jonca, N., Guerrin, M., Haftek, M., Bernard, D., Caubet, C., Egelrud, T., Schmidt, R., and Serre, G. 2001. Refined characterization of corneodesmosin proteolysis during terminal differentiation of human epidermis and its relationship to desquamation. *J Biol Chem* **276**: 20292-20299.

- Singh, S.K., Nizard, C., Kurfurst, R., Bonte, F., Schnebert, S., and Tobin, D.J. 2008. The silver locus product (Silv/gp100/Pmel17) as a new tool for the analysis of melanosome transfer in human melanocyte-keratinocyte co-culture. *Exp Dermatol* **17**: 418-426.
- Skerrow, C.J., Clelland, D.G., and Skerrow, D. 1989. Changes to desmosomal antigens and lectin-binding sites during differentiation in normal human epidermis: a quantitative ultrastructural study. *J Cell Sci* **92 (Pt 4)**: 667-677.
- Smale, S.T. and Kadonaga, J.T. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449-479.
- Smith, F.J., Irvine, A.D., Terron-Kwiatkowski, A., Sandilands, A., Campbell, L.E., Zhao, Y., Liao, H., Evans, A.T., Goudie, D.R., Lewis-Jones, S., Arseculeratne, G., Munro, C.S., Sergeant, A., O'Regan, G., Bale, S.J., Compton, J.G., DiGiovanna, J.J., Presland, R.B., Fleckman, P., and McLean, W.H. 2006. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet* **38**: 337-342.
- Smith, K.M., Lawn, R.M., and Wilcox, J.N. 1990. Cellular localization of apolipoprotein D and lecithin:cholesterol acyltransferase mRNA in rhesus monkey tissues by in situ hybridization. *J Lipid Res* **31**: 995-1004.
- Smits, P., Poumay, Y., Karperien, M., Tylzanowski, P., Wauters, J., Huylebroeck, D., Ponc, M., and Merregaert, J. 2000. Differentiation-dependent alternative splicing and expression of the extracellular matrix protein 1 gene in human keratinocytes. *J Invest Dermatol* **114**: 718-724.
- Southern, E.M. 1974. An improved method for transferring nucleotides from electrophoresis strips to thin layers of ion-exchange cellulose. *Anal Biochem* **62**: 317-318.
- Spazierer, D., Fuchs, P., Reipert, S., Fischer, I., Schmuth, M., Lassmann, H., and Wiche, G. 2006. Epiplakin is dispensable for skin barrier function and for integrity of keratin network cytoarchitecture in simple and stratified epithelia. *Mol Cell Biol* **26**: 559-568.
- Sprecher, E., Ishida-Yamamoto, A., Becker, O.M., Marekov, L., Miller, C.J., Steinert, P.M., Neldner, K., and Richard, G. 2001. Evidence for novel functions of the keratin tail emerging from a mutation causing ichthyosis hystrix. *J Invest Dermatol* **116**: 511-519.
- Sriuranpong, V., Park, J.I., Amornphimoltham, P., Patel, V., Nelkin, B.D., and Gutkind, J.S. 2003. Epidermal growth factor receptor-independent constitutive activation of STAT3 in head and neck squamous cell carcinoma is mediated by the autocrine/paracrine stimulation of the interleukin 6/gp130 cytokine system. *Cancer Res* **63**: 2948-2956.
- Steinert, P.M., Cantieri, J.S., Teller, D.C., Lonsdale-Eccles, J.D., and Dale, B.A. 1981. Characterization of a class of cationic proteins that specifically interact with intermediate filaments. *Proc Natl Acad Sci U S A* **78**: 4097-4101.
- Steinert, P.M., Mack, J.W., Korge, B.P., Gan, S.Q., Haynes, S.R., and Steven, A.C. 1991. Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. *Int J Biol Macromol* **13**: 130-139.
- Steinert, P.M. and Marekov, L.N. 1995. The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodi-peptide cross-linked components of the human epidermal cornified cell envelope. *J Biol Chem* **270**: 17702-17711.
- Steinert, P.M. and Marekov, L.N. 1997. Direct evidence that involucrin is a major early isopeptide cross-linked component of the keratinocyte cornified cell envelope. *J Biol Chem* **272**: 2021-2030.
- Stern, R., Asari, A.A., and Sugahara, K.N. 2006. Hyaluronan fragments: an information-rich system. *Eur J Cell Biol* **85**: 699-715.
- Steven, A.C., Bisher, M.E., Roop, D.R., and Steinert, P.M. 1990. Biosynthetic pathways of filaggrin and loricrin--two major proteins expressed by terminally differentiated epidermal keratinocytes. *J Struct Biol* **104**: 150-162.

- Strimmer, K. 2008. A unified approach to false discovery rate estimation. *BMC Bioinformatics* **9**: 303.
- Su, M.J., Bikle, D.D., Mancianti, M.L., and Pillai, S. 1994. 1,25-Dihydroxyvitamin D3 potentiates the keratinocyte response to calcium. *J Biol Chem* **269**: 14723-14729.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keefe, S., Haas, S., Vingron, M., Lehrach, H., and Yaspo, M.L. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**: 956-960.
- Sun, W., Xu, R., Hu, W., Jin, J., Crellin, H.A., Bielawski, J., Szulc, Z.M., Thiers, B.H., Obeid, L.M., and Mao, C. 2008. Upregulation of the human alkaline ceramidase 1 and acid ceramidase mediates calcium-induced differentiation of epidermal keratinocytes. *J Invest Dermatol* **128**: 389-397.
- Swartzendruber, D.C., Wertz, P.W., Kitko, D.J., Madison, K.C., and Downing, D.T. 1989. Molecular models of the intercellular lipid lamellae in mammalian stratum corneum. *J Invest Dermatol* **92**: 251-257.
- Sybert, V.P., Dale, B.A., and Holbrook, K.A. 1985. Ichthyosis vulgaris: identification of a defect in synthesis of filaggrin correlated with an absence of keratohyaline granules. *J Invest Dermatol* **84**: 191-194.
- Syed, S.E., Trinnaman, B., Martin, S., Major, S., Hutchinson, J., and Magee, A.I. 2002. Molecular interactions between desmosomal cadherins. *Biochem J* **362**: 317-327.
- Tago, K., Nakamura, T., Nishita, M., Hyodo, J., Nagai, S., Murata, Y., Adachi, S., Ohwada, S., Morishita, Y., Shibuya, H., and Akiyama, T. 2000. Inhibition of Wnt signaling by ICAT, a novel beta-catenin-interacting protein. *Genes Dev* **14**: 1741-1749.
- Tammi, R., Ripellino, J.A., Margolis, R.U., and Tammi, M. 1988. Localization of epidermal hyaluronic acid using the hyaluronate binding region of cartilage proteoglycan as a specific probe. *J Invest Dermatol* **90**: 412-414.
- Tang, V.W. 2006. Proteomic and bioinformatic analysis of epithelial tight junction reveals an unexpected cluster of synaptic molecules. *Biol Direct* **1**: 37.
- Thrash, B.R., Menges, C.W., Pierce, R.H., and McCance, D.J. 2006. AKT1 provides an essential survival signal required for differentiation and stratification of primary human keratinocytes. *J Biol Chem* **281**: 12155-12162.
- Tiede, S., Kloepper, J.E., Bodo, E., Tiwari, S., Kruse, C., and Paus, R. 2007. Hair follicle stem cells: walking the maze. *Eur J Cell Biol* **86**: 355-376.
- Ting, S.B., Caddy, J., Hislop, N., Wilanowski, T., Auden, A., Zhao, L.L., Ellis, S., Kaur, P., Uchida, Y., Holleran, W.M., Elias, P.M., Cunningham, J.M., and Jane, S.M. 2005. A homolog of *Drosophila* grainy head is essential for epidermal integrity in mice. *Science* **308**: 411-413.
- Ting, S.B., Wilanowski, T., Auden, A., Hall, M., Voss, A.K., Thomas, T., Parekh, V., Cunningham, J.M., and Jane, S.M. 2003. Inositol- and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3. *Nat Med* **9**: 1513-1519.
- Toulza, E., Galliano, M.F., Jonca, N., Gallinaro, H., Mechin, M.C., Ishida-Yamamoto, A., Serre, G., and Guerrin, M. 2006. The human dermokine gene: description of novel isoforms with different tissue-specific expression and subcellular location. *J Invest Dermatol* **126**: 503-506.
- Toulza, E., Mattiuzzo, N.R., Galliano, M.F., Jonca, N., Dossat, C., Jacob, D., de Daruvar, A., Wincker, P., Serre, G., and Guerrin, M. 2007. Large-scale identification of human genes implicated in epidermal barrier function. *Genome Biol* **8**: R107.
- Trempus, C.S., Dang, H., Humble, M.M., Wei, S.J., Gerdes, M.J., Morris, R.J., Bortner, C.D., Cotsarelis, G., and Tennant, R.W. 2007. Comprehensive microarray transcriptome

- profiling of CD34-enriched mouse keratinocyte stem cells. *J Invest Dermatol* **127**: 2904-2907.
- Tschachler, E. 2007. Psoriasis: the epidermal component. *Clin Dermatol* **25**: 589-595.
- Tu, C.L., Chang, W., and Bikle, D.D. 2001. The extracellular calcium-sensing receptor is required for calcium-induced differentiation in human keratinocytes. *J Biol Chem* **276**: 41079-41085.
- Tu, C.L., Chang, W., and Bikle, D.D. 2007. The role of the calcium sensing receptor in regulating intracellular calcium handling in human epidermal keratinocytes. *J Invest Dermatol* **127**: 1074-1083.
- Tuerk, C. and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**: 505-510.
- Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. 2004. Defining the epithelial stem cell niche in skin. *Science* **303**: 359-363.
- Tummala, R., Romano, R.A., Fuchs, E., and Sinha, S. 2003. Molecular cloning and characterization of AP-2 epsilon, a fifth member of the AP-2 family. *Gene* **321**: 93-102.
- Turksen, K. and Troy, T.C. 2002. Permeability barrier dysfunction in transgenic mice overexpressing claudin 6. *Development* **129**: 1775-1784.
- van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., and Friend, S.H. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530-536.
- van Ruissen, F., Jansen, B.J., de Jongh, G.J., van Vlijmen-Willems, I.M., and Schalkwijk, J. 2002a. Differential gene expression in premalignant human epidermis revealed by cluster analysis of serial analysis of gene expression (SAGE) libraries. *Faseb J* **16**: 246-248.
- van Ruissen, F., Jansen, B.J., de Jongh, G.J., Zeeuwen, P.L., and Schalkwijk, J. 2002b. A partial transcriptome of human epidermis. *Genomics* **79**: 671-678.
- Van Tassel, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., and Sonstegard, T.S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* **5**: 247-252.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**: RESEARCH0034.
- Vendrell, J.A., Robertson, K.E., Ravel, P., Bray, S.E., Bajard, A., Purdie, C.A., Nguyen, C., Hadad, S.M., Bieche, I., Chabaud, S., Bachelot, T., Thompson, A.M., and Cohen, P.A. 2008. A candidate molecular signature associated with tamoxifen failure in primary breast cancer. *Breast Cancer Res* **10**: R88.
- Vigano, M.A., Lamartine, J., Testoni, B., Merico, D., Alotto, D., Castagnoli, C., Robert, A., Candi, E., Melino, G., Gidrol, X., and Mantovani, R. 2006. New p63 targets in keratinocytes identified by a genome-wide approach. *Embo J* **25**: 5105-5116.
- Wang, X., Bolotin, D., Chu, D.H., Polak, L., Williams, T., and Fuchs, E. 2006. AP-2alpha: a regulator of EGF receptor signaling and proliferation in skin epidermis. *J Cell Biol* **172**: 409-421.
- Wang, X., Pasolli, H.A., Williams, T., and Fuchs, E. 2008. AP-2 factors act in concert with Notch to orchestrate terminal differentiation in skin epidermis. *J Cell Biol* **183**: 37-48.
- Wang, Z., Gerstein, M., and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57-63.

- Welter, J.F., Crish, J.F., Agarwal, C., and Eckert, R.L. 1995. Fos-related antigen (Fra-1), junB, and junD activate human involucrin promoter transcription by binding to proximal and distal AP1 sites to mediate phorbol ester effects on promoter activity. *J Biol Chem* **270**: 12614-12622.
- Wiedow, O., Schroder, J.M., Gregory, H., Young, J.A., and Christophers, E. 1990. Elafin: an elastase-specific inhibitor of human skin. Purification, characterization, and complete amino acid sequence. *J Biol Chem* **265**: 14791-14795.
- Wilanowski, T., Caddy, J., Ting, S.B., Hislop, N.R., Cerruti, L., Auden, A., Zhao, L.L., Asquith, S., Ellis, S., Sinclair, R., Cunningham, J.M., and Jane, S.M. 2008. Perturbed desmosomal cadherin expression in grainy head-like 1-null mice. *Embo J* **27**: 886-897.
- Wrone-Smith, T. and Nickoloff, B.J. 1996. Dermal injection of immunocytes induces psoriasis. *J Clin Invest* **98**: 1878-1887.
- Wu, Z., Hansmann, B., Meyer-Hoffert, U., Glaser, R., and Schroder, J.M. 2009. Molecular identification and expression analysis of filaggrin-2, a member of the S100 fused-type protein family. *PLoS One* **4**: e5227.
- Xiao, H.S., Huang, Q.H., Zhang, F.X., Bao, L., Lu, Y.J., Guo, C., Yang, L., Huang, W.J., Fu, G., Xu, S.H., Cheng, X.P., Yan, Q., Zhu, Z.D., Zhang, X., Chen, Z., Han, Z.G., and Zhang, X. 2002. Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain. *Proc Natl Acad Sci U S A* **99**: 8360-8365.
- Xie, Z., Komuves, L., Yu, Q.C., Elalieh, H., Ng, D.C., Leary, C., Chang, S., Crumrine, D., Yoshizawa, T., Kato, S., and Bikle, D.D. 2002. Lack of the vitamin D receptor is associated with reduced epidermal differentiation and hair follicle growth. *J Invest Dermatol* **118**: 11-16.
- Xu, X.R., Huang, J., Xu, Z.G., Qian, B.Z., Zhu, Z.D., Yan, Q., Cai, T., Zhang, X., Xiao, H.S., Qu, J., Liu, F., Huang, Q.H., Cheng, Z.H., Li, N.G., Du, J.J., Hu, W., Shen, K.T., Lu, G., Fu, G., Zhong, M., Xu, S.H., Gu, W.Y., Huang, W., Zhao, X.T., Hu, G.X., Gu, J.R., Chen, Z., and Han, Z.G. 2001. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc Natl Acad Sci U S A* **98**: 15089-15094.
- Yamaguchi, T. and Osumi, T. 2009. Chanarin-Dorfman syndrome: deficiency in CGI-58, a lipid droplet-bound coactivator of lipase. *Biochim Biophys Acta* **1791**: 519-523.
- Yan, Q., Liu, J.P., and Li, D.W. 2006. Apoptosis in lens development and pathology. *Differentiation* **74**: 195-211.
- Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T., Calin, G.A., Liu, C.G., Croce, C.M., and Harris, C.C. 2006. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**: 189-198.
- Yang, T., Liang, D., Koch, P.J., Hohl, D., Kheradmand, F., and Overbeek, P.A. 2004. Epidermal detachment, desmosomal dissociation, and destabilization of corneodesmosin in Spink5^{-/-} mice. *Genes Dev* **18**: 2354-2358.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**: e15.
- Yano, S., Banno, T., Walsh, R., and Blumenberg, M. 2008. Transcriptional responses of human epidermal keratinocytes to cytokine interleukin-1. *J Cell Physiol* **214**: 1-13.
- Yazaki, J., Gregory, B.D., and Ecker, J.R. 2007. Mapping the genome landscape using tiling array technology. *Curr Opin Plant Biol* **10**: 534-542.

- Yendamuri, S., Trapasso, F., and Calin, G.A. 2008. ARLTS1 - a novel tumor suppressor gene. *Cancer Lett* **264**: 11-20.
- Yin, J.Q., Zhao, R.C., and Morris, K.V. 2008. Profiling microRNA expression with microarrays. *Trends Biotechnol* **26**: 70-76.
- Yu, Z., Lin, K.K., Bhandari, A., Spencer, J.A., Xu, X., Wang, N., Lu, Z., Gill, G.N., Roop, D.R., Wertz, P., and Andersen, B. 2006. The Grainyhead-like epithelial transactivator Get-1/Grhl3 regulates epidermal terminal differentiation and interacts functionally with LMO4. *Dev Biol* **299**: 122-136.
- Zeeuwen, P.L., Ishida-Yamamoto, A., van Vlijmen-Willems, I.M., Cheng, T., Bergers, M., Iizuka, H., and Schalkwijk, J. 2007. Colocalization of cystatin M/E and cathepsin V in lamellar granules and corneodesmosomes suggests a functional role in epidermal differentiation. *J Invest Dermatol* **127**: 120-128.
- Zeeuwen, P.L., van Vlijmen-Willems, I.M., Hendriks, W., Merckx, G.F., and Schalkwijk, J. 2002. A null mutation in the cystatin M/E gene of ichq mice causes juvenile lethality and defects in epidermal cornification. *Hum Mol Genet* **11**: 2867-2875.
- Zenz, R., Scheuch, H., Martin, P., Frank, C., Eferl, R., Kenner, L., Sibilina, M., and Wagner, E.F. 2003. c-Jun regulates eyelid closure and skin tumor development through EGFR signaling. *Dev Cell* **4**: 879-889.
- Zhang, J.S., Wang, L., Huang, H., Nelson, M., and Smith, D.I. 2001. Keratin 23 (K23), a novel acidic keratin, is highly induced by histone deacetylase inhibitors during differentiation of pancreatic cancer cells. *Genes Chromosomes Cancer* **30**: 123-135.
- Zhang, W., Morris, Q.D., Chang, R., Shai, O., Bakowski, M.A., Mitsakakis, N., Mohammad, N., Robinson, M.D., Zirngibl, R., Somogyi, E., Laurin, N., Eftekharpour, E., Sat, E., Grigull, J., Pan, Q., Peng, W.T., Krogan, N., Greenblatt, J., Fehlings, M., van der Kooy, D., Aubin, J., Bruneau, B.G., Rossant, J., Blencowe, B.J., Frey, B.J., and Hughes, T.R. 2004. The functional landscape of mouse gene expression. *J Biol* **3**: 21.
- Zhang, X.J., Huang, W., Yang, S., Sun, L.D., Zhang, F.Y., Zhu, Q.X., Zhang, F.R., Zhang, C., Du, W.H., Pu, X.M., Li, H., Xiao, F.L., Wang, Z.X., Cui, Y., Hao, F., Zheng, J., Yang, X.Q., Cheng, H., He, C.D., Liu, X.M., Xu, L.M., Zheng, H.F., Zhang, S.M., Zhang, J.Z., Wang, H.Y., Cheng, Y.L., Ji, B.H., Fang, Q.Y., Li, Y.Z., Zhou, F.S., Han, J.W., Quan, C., Chen, B., Liu, J.L., Lin, D., Fan, L., Zhang, A.P., Liu, S.X., Yang, C.J., Wang, P.G., Zhou, W.M., Lin, G.S., Wu, W.D., Fan, X., Gao, M., Yang, B.Q., Lu, W.S., Zhang, Z., Zhu, K.J., Shen, S.K., Li, M., Zhang, X.Y., Cao, T.T., Ren, W., Zhang, X., He, J., Tang, X.F., Lu, S., Yang, J.Q., Zhang, L., Wang, D.N., Yuan, F., Yin, X.Y., Huang, H.J., Wang, H.F., Lin, X.Y., and Liu, J.J. 2009. Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. *Nat Genet* **41**: 205-210.
- Zhao, X.S., Shin, D.M., Liu, L.H., Shull, G.E., and Muallem, S. 2001. Plasticity and adaptation of Ca²⁺ signaling and Ca²⁺-dependent exocytosis in SERCA2(+/-) mice. *Embo J* **20**: 2680-2689.

AUTEUR : Nicolas MATTIUZZO

TITRE : Étude du transcriptome kératinocytaire au cours du programme de différenciation de l'épiderme

DIRECTEUR DE THÈSE : Marina GUERRIN WEBER

LIEU ET DATE DE SOUTENANCE : Toulouse, le 23 octobre 2009

RÉSUMÉ

Au sein de l'épiderme, la fonction de barrière est assurée par la partie la plus superficielle, la couche cornée, constituée de cornéocytes, des cellules mortes très cohésives entourées d'une matrice lipidique. Un processus complexe de mort cellulaire programmée, la cornification, aboutit à la transformation des dernières cellules vivantes, les kératinocytes granuleux, en cornéocytes dépourvus d'activité transcriptionnelle et traductionnelle. Les kératinocytes granuleux constituent donc à la fois le stade ultime de la différenciation épidermique, mais aussi celui où sont produits les différents acteurs de la cornification et de la desquamation, donc de la fonction barrière.

Très peu d'études moléculaires systématiques ont été consacrées à l'épiderme humain, et seulement quatre facteurs de transcription spécifiques requis pour la fonction barrière ont été identifiés. Le travail présenté ici combine plusieurs techniques à grande échelle afin d'identifier de nouveaux gènes impliqués dans les étapes tardives de ce programme. Un procédé de purification de cellules à partir d'épiderme humain normal a été développé afin d'obtenir des fractions enrichies en kératinocytes de différents stades de différenciation. Ce procédé, couplé à la technique des ORESTEs, a permis dans un premier temps de produire 22 000 séquences et d'identifier 3387 gènes exprimés dans la couche granuleuse, parmi lesquels de nouveaux marqueurs de la différenciation impliqués dans diverses fonctions (protéine structurale, métabolisme et transport des lipides, protéase et inhibiteurs, etc.) Les profils d'expression génique des kératinocytes granuleux ont également été comparés avec ceux des kératinocytes basaux au moyen de puces à ADN pangénomiques, mettant en évidence 200 candidats. Afin d'identifier des facteurs de transcription supplémentaires, une recherche bioinformatique de sites de fixation conservés évolutivement a été réalisée sur les promoteurs de 52 marqueurs de la différenciation.

Les résultats issus de ces trois techniques ont été combinés et analysés afin de réduire à 300 gènes le nombre de candidats à valider. Pour 155 d'entre eux, leur profil d'expression au sein de l'épiderme a pu être analysé par PCR quantitative, permettant d'identifier 49 nouveaux marqueurs de la différenciation. De plus, sur 94 gènes quantifiés codant pour des facteurs de transcription, 37 apparaissent positivement ou négativement régulés, suggérant un rôle dans le contrôle des étapes tardives du programme de différenciation.

Ces résultats ouvrent de nombreuses voies et constituent donc une avancée importante pour la compréhension des mécanismes complexes mis en jeu dans ce réseau de régulation génique, tant dans le cadre physiologique que pour certaines pathologies caractérisées par une altération de la différenciation, comme dans le cas du psoriasis ou des ichthyoses.

MOTS CLÉS : Épiderme, différenciation, kératinocyte, transcriptome, bioanalyse, facteurs de transcription, expression génique.

DISCIPLINE ADMINISTRATIVE : Gènes, Cellules et Développement

LABORATOIRE : Unité Différenciation Épidermique et Autoimmunité Rhumatoïde
UMR 5165 CNRS – Université de Toulouse (Paul Sabatier)
CHU Purpan, place du Docteur Baylac
31059 Toulouse
