The Molecular Revolution in Cutaneous Biology: Keratin Genes and their Associated Disease: Diversity, Opportunities, and Challenges

Pierre A. Coulombe[1,2,3,4,5]

The abundance of keratin proteins and the filaments they form in surface epithelia has long been appreciated. This said, the remarkable diversity of keratin proteins and the notion that they are encoded by one of the largest gene families in the human genome has come to the fore relatively recently, coinciding with the sequencing of whole genomes. This complexity has generated some practical challenges, notably in terms of nomenclature and tractability. More importantly, however, studies of keratin have seeded the discovery of the genetic basis for a large number of genodermatoses and continue to provide a unique perspective on and insight into epithelial cells and tissues, whether normal or diseased.

Journal of Investigative Dermatology (2017) **137,** e67–e71; doi:10.1016/j.jid.2016.04.039

Keratins are now understood by all to be intermediate filament-forming proteins encoded by two subgroupings of conserved genes, type I and type II, that are regulated in a pairwise-, differentiation-, and context-dependent fashion in all types of epithelial tissues (Schweizer et al., 2006). When coupled with their differentiation-dependent regulation, the diversity and differential immunogenicity of keratin proteins provides an unrivaled array of biomarkers with which to typify epithelial cells in healthy and diseased tissues. For instance, most of us are familiar with expression of the keratin (K) 5-K14 pair reflecting a progenitor state in stratified and pseudostratified epithelia, of the K1-K10 pair reflecting an early stage of terminal differentiation in the interfollicular epidermis, and of the K6-K16 pair reflecting a state of keratinocyte activation and/or alternative differentiation whenever a complex epithelium is under "duress" (Fuchs, 1995; Mansbridge et al., 1987; Sun et al., 1983).

[1]*Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA;* [2]*Department of Biological Chemistry, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA;* [3]*Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA;* [4]*Department of Dermatology, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA; and* [5]*The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, Maryland, USA*

Correspondence: Pierre A. Coulombe, Department of Biochemistry and Molecular Biology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, Maryland, USA. E-mail: coulombe@jhu.edu

Abbreviations: IF, intermediate filaments; K, keratin

Received 3 December 2015; revised 24 March 2016; accepted 19 April 2016

The discovery of keratin mutations as the underlying genetic defect in epidermolysis bullosa simplex (Bonifas et al., 1991; Coulombe et al., 1991), now 25 years old, has seeded an explosion of knowledge regarding the genetic etiology and pathophysiology of a very large number of diseases affecting skin and/or other tissues (Coulombe et al., 2009; Omary et al., 2004; Szeverenyi et al., 2008) (see Table 1 for a partial account of diseases arising from mutations in keratin genes). This contribution summarizes how we arrived at a tally of 54 functional keratin genes in the human genome and comments on the origins of the official nomenclature in use to designate these remarkable genes and proteins.

Intermediate filaments (IFs) were discovered by Holtzer and colleagues, in developing skeletal muscle, in the late 1960s (Ishikawa et al., 1968). Although it took nearly 10 years for researchers to fully uncover the status of the then well-known keratin filaments as epithelial-specific IFs (Osborn et al., 1977; Sun and Green, 1978a), the late 1970s and early 1980s proved to be an incredibly fertile period of discovery in keratin-related research. In a span of 5–6 years, we learned that purified native epidermal keratins could self-assemble into 10-nm filaments as obligate heteropolymers in vitro (Aebi et al., 1983; Steinert et al., 1976), that keratin proteins were structurally diverse and very abundant elements in keratinocytes in primary culture as well as in epidermis in situ (Fuchs and Green, 1978; Sun and Green, 1978b), and that they are encoded by two distinct types of mRNAs (corresponding to "acidic" and "basic" keratins) regulated in a pairwise- and differentiation-related fashion in epidermis (Fuchs and Green, 1979; 1980; Fuchs et al., 1981; Woodcock-Mitchell et al., 1982). In the early 1980s, K14 was the first of many IFs to be completely sequenced at both the cDNA and genomic (gene) levels (Hanukoglu and Fuchs, 1982; Marchuk et al., 1985). Such gene and cDNA cloning efforts made it considerably easier to relate individual members to one another in the emerging family of IF sequences, including the affirmation of multiple subtypes of IF sequences, the recognition of hair keratins as bona fide IFs (Powell et al., 1983), the notion of a conserved substructure of subclasses of IF genes and of their corresponding protein products, and the beginnings of evolutionary insight (Fuchs et al., 1981). The remarkable arrangement of keratin genes into large and compact clusters was first uncovered in the chick genome (Molloy et al., 1982) and later confirmed in the human (Romano et al., 1988). All of these elements, however impressive, are a partial account of all the important realizations that

Table 1. Examples of the variability in clinical presentation associated with mutations in the *KRT5/KRT14, KRT1/KRT10,* and *KRT6/KRT16/KRT17* pairings[1]

Keratin Genes	Disease	Inheritance	OMIM #[2]	Comments
KRT5 or *KRT14*	EBS Dowling Meara (severe)	AD	131760	Classical EBS presentations dominated by trauma-induced bullous skin blisters, reflecting the fragility of basal layer keratinocytes. The most severe subtype (Dowling-Meara) is typified by clustered arrangement of lesions and presence of insoluble keratin aggregates in basal keratinocytes.
	EBS—generalized	AD	131900	
	EBS—localized (relatively mild)	AD	131800	
	ESB recessive	AR	601001	
KRT5	EBS—mottled pigmentation	AD	131960	Skin blistering is accompanied by mottled pigmentation of the trunk and limbs and is strongly associated with a specific mutation in *KRT5*, Pro25-Leu.
KRT5	Dowling-Degos disease	AD	179850	Typified by disfiguring reticulate hyperpigmentation and dark hyperkeratotic papules in skin flexural region, due to *KRT5* haploinsufficiency.
KRT5	EBS with migratory circinate erythema	AD	609352	Skin blistering is accompanied by an unusual migratory and circular erythema, spreading "away" from blisters. Associated with frameshift mutations altering the tail domain of K5.
KRT14	Dermatopathia pigmentosa reticularis	AD	125595	Similar syndromes showing reticular hyperpigmentation, hypohidrosis (reduced sweating), PPK-like lesions, and poor dermatoglyphics. Due to premature stop codon located very early in the *KRT14* coding sequence.
KRT14	Naegali-Franceschetti-Jadassohn	AD	161000	
KRT1 or *KRT10*	Epidermolytic hyperkeratosis	AD	113800	Related conditions typified by presence of erythroderma and flakiness, hyperkeratosis. Microscopy shows lysis (fragility) of keratinocytes in the suprabasal layers and hyperproliferation in the basal layer of epidermis.
	Cyclic ichthyosis with EHK	AD	607602	
	Ichthyosis hystrix (Curth Macklin)	AD	146590	
KRT1	Palmoplantar keratoderma (PPK)	AD	144200 600962 607654	Applies to three clinically distinct subtypes of PPK (in which, as implied by the name, lesions preferentially occur on palm and sole skin): Epidermolytic, nonepidermolytic, and striate.
KRT10	Congenital reticular ichthyosiform erythroderma (also known as ichthyosis with confetti)	AD	609165	Typified by slowly enlarging islands of normal skin surrounded by erythematous ichthyotic patches in a reticulated pattern. Accompanied by peculiar ultrastructural changes and caused by a frameshift mutation in *KRT10*.
KRT6A or *KRT16*	Pachyonychia congenita, type 1 (PC-1)	AD	167200	Presentation dominated by nail dystrophy, severe PPK, oral leukoplakia. Follicular keratosis, hoarseness.
KRT6B or *KRT17*	Pachyonychia congenita, type 2 (PC-2)	AD	167210	Similar to PC-1, but additionally typified by natal teeth, pili torti (twisted hair), and steatocystic skin lesions (steatocystoma multiplex).
KRT6C	See comment	AD	167200 600962	*KRT6* is a relatively poorly expressed *KRT6* paralog. Disease presentation related to that of PC-1 and/or NEPPK.
KRT16	Palmoplantar keratoderma, nonepidermolytic (NEPPK, focal)	AD	600962	See description for relevant *KRT1* entry.
KRT17	Steatocystoma multiplex	AD	184500	Related to PC-2, but without much nail involvement. Dominated by steatocystic lesions that originate from sebaceous glands all over the body.

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; EBS, epidermolysis bullosa simplex; K, keratin.

[1]The table does not include all of the diseases that have been attributed to mutations in these keratin genes to date. Figure 2 relates the chromosomal location of these keratin genes.

[2]OMIM # refers to the numbering in the Online Mendelian Inheritance in Man catalog of human gene and genetic disorders. The information related in this table originates in the OMIM catalog (http://omim.org; see McKusick, 1998) and in the Human Intermediate Filament Mutation Database maintained at the Institute of Medical Biology in Singapore (http://www.interfil.org/index.php; see Szeverenyi et al., 2008).

were made in a short period of progress spanning the late 1970s to the early 1980s. It is also worth mentioning that the study of hair keratins in sheep by Australian scientists, at the gene and proteins levels, has contributed greatly to progress in the field before and during that period (Dowling and Sparrow, 1991; Rogers and Powell, 1993).

In a landmark article, Moll et al. (1982) proposed a nomenclature that could accommodate and name the array of 19 so-called "cytokeratins" that had been inventoried, at the time, on the basis of a combination of immunological, molecular, and biochemical criteria; remarkably, this effort preceded the availability of keratin cDNA or gene sequences. These researchers, in particular, had accumulated a wealth of information from a systematic characterization of the keratin-rich insoluble protein fraction prepared from several human tissues and cell lines (normal and diseased) and from their analysis via two-dimensional PAGE, in which proteins are first resolved by charge in a first "dimension" and then by mass in a second "dimension." Moll et al. (1982) had the clever idea of representing all 19 cytokeratins in a virtual two-dimensional PAGE (Figure 1), enabling them to name individual cytokeratins according to their