The Molecular Revolution in Cutaneous Biology: Identification of Skin Disease Genes

Lynn M. Boyden[1] and Keith A. Choate[1,2]

The past 30 years have witnessed a marked expansion in the power to discover genes relevant to human disease. From the development of DNA markers and PCR in the 1980s, fluorescent DNA labeling and microarray genotyping in the 1990s, and publication of the complete human genome and the rise of high-throughput sequencing in the 2000s, each technological advance has accelerated our ability to clone and characterize coding sequences, map traits and disorders to specific loci, and discover pathogenic mutations.

In hypothesis-driven functional identification of disease genes, specific genes are examined as putative candidates on the basis of existing knowledge of the biology of the gene and disorder, sometimes aided by model organism studies. In contrast, positional cloning requires no prior knowledge about either disease pathobiology or the causative gene. The era of positional cloning of disease genes began in the 1980s with a groundbreaking article describing how variations in genomic DNA could be used as genetic markers that permit distinction of the two alleles at specific chromosomal loci (Botstein et al., 1980). This provides the means to track inheritance of alleles through pedigrees and thereby look for markers for which the inheritance pattern matches that of an inherited disease or trait. These markers can be polymorphisms without functional significance, including restriction fragment length polymorphisms, short tandem repeats (also known as microsatellites), or single nucleotide polymorphisms. Technologies to genotype each have provided incremental advances in mapping efficiency and granularity. Although the markers themselves do not usually represent pathogenic mutations, examination of their inheritance patterns via linkage analysis can provide definitive evidence of the chromosomal position of a causative gene, the significance of which may be represented as a logarithm-of-odds ratio score. Subsequent efforts to discover pathogenic mutations are thereby restricted to a specific locus, facilitating a tractable search. In the past, a successful mapping project often had to be followed by cloning, sequencing, and assembly of overlapping genomic DNA fragments to characterize the genes within the linkage interval, which could then be screened for mutations. The completion of the Human Genome Project obviated this laborious and time-consuming cloning and characterization and enabled rapid transition from linkage analysis to mutation discovery. Concurrently, as sequencing technologies advanced and costs declined, mutation detection moved from methods based on the migration properties of DNA (e.g., heteroduplex and single-strand conformation polymorphism analysis) to direct identification via Sanger or high-throughput DNA sequencing.

Over 6,000 Mendelian disorders have been described, and approximately 10% of these feature skin abnormalities. Genetic investigation of these genodermatoses has led to fundamental biological insights into epidermal homeostasis, differentiation, and renewal. Although many hundreds of genes have been associated with skin diseases, here we highlight some of the seminal reports, with a focus on three groups of cardinal disorders as archetypes of the progress and scope of genetic discovery.

SKIN FRAGILITY AND BLISTERING DISORDERS

The early 1990s saw an explosion in the genetic understanding of skin fragility and blistering disorders, with many of the major genetic causes discovered over the course of only a few years. This progress was enabled largely by "reverse genetics" mouse models and by linkage analysis performed solely with markers specific to particular candidate gene loci, strategies particularly amenable to this group of relatively well-understood diseases and genes.

In January 1991, a mouse model transgenic for a truncated human keratin 14 (*KRT14*) suggested that mutations in it and in the gene for its intermediate filament partner, keratin 5 (*KRT5*), were the likely causes of dominant epidermolysis bullosa simplex (Vassar et al., 1991). Linkage analysis performed with markers specific to the keratin gene clusters on chromosomes 12 (Ryynänen et al., 1991b) and 17 (Bonifas et al., 1991) validated this hypothesis, and soon thereafter mutations causing human epidermolysis bullosa simplex were reported in both *KRT14* (Bonifas et al., 1991) and *KRT5* (Lane et al., 1992).

Discovery that mutations in the genes encoding keratin 10 (*KRT10*) and its intermediate filament partner, keratin 1 (*KRT1*), cause the dominant disorder epidermolytic hyperkeratosis was even more rapid. Linkage to the *KRT1* locus (Compton et al., 1992), a transgenic mouse expressing a truncated form of human *KRT10* (Fuchs et al., 1992), and causative human mutations in *KRT1* (Chipev et al., 1992; Rothnagel et al., 1992) and *KRT10* (Cheng et al., 1992;

[1]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA; and [2]Departments of Dermatology and Pathology, Yale University School of Medicine, New Haven, Connecticut, USA

Correspondence: Keith A. Choate, Department of Dermatology, Yale University School of Medicine, 333 Cedar St., New Haven, Connecticut, 06520. E-mail: keith.choate@yale.edu

Abbreviations: ADULT, acro-dermato-ungual-lacrimal-tooth; ARCI, autosomal recessive congenital ichthyosis; DOK, disorders of keratinization; ED, ectodermal dysplasia

Received 30 September 2016; accepted 11 November 2016

Journal of Investigative Dermatology (2017) **137,** e61–e65; doi:10.1016/j.jid.2016.11.019

Rothnagel et al., 1992) were published over the course of a single summer.

The genes for both dystrophic and junctional epidermolysis bullosa were similarly hypothesized as likely candidates on the basis of previously described biology and were proven to be responsible in short order. The gene encoding type VII collagen (*COL7A1*) was proposed as an intuitive candidate for dystrophic epidermolysis bullosa, and *COL7A1*-specific markers were used to map the dominant form of the disease (Ryynänen et al., 1991a) before the gene's coding sequence was published (Parente et al., 1991). *COL7A1* markers were subsequently used to map the recessive form of dystrophic epidermolysis bullosa to the same locus (Hovnanian et al., 1992), and causative mutations in *COL7A1* for recessive (Christiano et al., 1993) and dominant (Christiano et al., 1994) dystrophic epidermolysis bullosa were published in the following 2 years. Only a few months later, and with the same candidate gene approach and rapid pace that surprisingly put pathogenesis reports ahead of those describing a gene's cloning and characterization, recessive junctional epidermolysis bullosa was found to be caused by mutations in the genes encoding the three subunit polypeptides of laminin-5 (*LAMA3* [Kivirikko et al., 1995], *LAMB3* [Pulkkinen et al., 1994b, 1995], and *LAMC2* [Aberdam et al., 1994; Pulkkinen et al., 1994a]) and the collagen gene *COL17A1* (McGrath et al., 1995). Mutations in *LAMB3* account for most junctional epidermolysis bullosa cases, because of the prevalence of two nonsense mutations that arise recurrently within mutational hotspots (Kivirikko et al., 1996).

ECTODERMAL DYSPLASIAS

Four important genes for ectodermal dysplasias (EDs) with an epidermal component were identified by different strategies over the course of a decade.

X-linked recessive ED was described by Charles Darwin, who noted that the disorder was present solely in males and was without male-to-male transmission (Darwin, 1875). A century later, at least 12 reports of translocations, rearrangements, deletions, and linkage analyses, published between 1986 and 1996, were used to fine-map the disease to an interval on the X chromosome containing the gene encoding ectodysplasin A (*EDA*), and additional subjects were subsequently shown to have point mutations in *EDA* (Kere et al., 1996). Consequently, the gene encoding the EDA receptor (*EDAR*) became a likely candidate for autosomal dominant and recessive forms of ED. A mouse model was created (Headon and Overbeek, 1999), and causative mutations for both forms were found in *EDAR* (Monreal et al., 1999).

The same year, limb-mammary syndrome, a dominant form of ED, was mapped to a locus on chromosome 3 (van Bokhoven et al., 1999). In contrast to the blistering disorders described, there had been no obvious candidate gene to pursue with a targeted query, so as with most successfully mapped autosomal Mendelian diseases, a genome-wide positional cloning strategy was pursued. However, the microsatellite marker genotyping, although state-of-the-art for its time, was laborious and slow by today's standards; consequently, a positive linkage result on chromosome 3 was obtained before markers to all of the chromosomes had been genotyped. The subsequent premature termination of a genome-wide scan would be unimaginable in the more modern era of array-based genotyping of hundreds of thousands of markers simultaneously. Regardless, the mapping of limb-mammary syndrome provided a plausible candidate region for a phenotypically similar form of ED, the dominant genodermatosis acro-dermato-ungual-lacrimal-tooth (ADULT) syndrome. Targeted genotyping of 19 markers in a large kindred mapped ADULT syndrome to the same locus as limb-mammary syndrome (Propping et al., 2000). Shortly thereafter, mutations in the gene for p63 (*TP63*) were reported to cause both ADULT syndrome (Amiel et al., 2001; Duijf et al., 2002) and limb-mammary syndrome (van Bokhoven et al., 2001).

Homozygosity mapping is a strategy that capitalizes on the increased linkage information provided by offspring of a consanguineous union, in whom the occurrence of a rare recessive disease strongly implies that a causative mutation is homozygous by descent from an ancestor shared by both of the proband's parents and therefore will be located within a region of consecutive homozygous markers (Lander and Botstein, 1987). This method was used to show that mutations in a Wnt gene (*WNT10A*) cause odonto-onycho-dermal dysplasia, a recessive form of ED (Adaimy et al., 2007).

NONEPIDERMOLYTIC DISORDERS OF KERATINIZATION (DOK)

The first Mendelian skin disease solved was a form of ichthyosis with X-linked recessive inheritance and a lack of steroid sulfatase activity (Webster et al., 1978). These observations facilitated the discovery that affected subjects were hemizygous for large deletions on the X chromosome that included the gene for steroid sulfatase isozyme S (*STS*) (Bonifas et al., 1987). Another major breakthrough occurred in the early 1990s, when the first gene for lamellar ichthyosis, an autosomal recessive congenital ichthyosis (ARCI), was discovered. Like most of the genes described above, the gene encoding transglutaminase 1 (*TGM1*) was considered to be a likely candidate for lamellar ichthyosis (Hohl et al., 1993), which was subsequently mapped to the *TGM1* locus (Russell et al., 1994); pathogenic mutations in *TGM1* were described shortly thereafter (Huber et al., 1995, Russell et al., 1995). This discovery was critical, not only for its biological insights into the most common DOK, but also for its future therapeutic potential. A few years ago topical enzyme replacement therapy was shown to restore TGM1 activity and considerably improve the ichthyosis phenotype in TGM1-deficient skin grafts (Aufenvenne et al., 2013).

Although mutations in *TGM1* account for approximately 20% of DOK, this group of disorders is otherwise notable for its significant genotypic and phenotypic heterogeneity. Lest the above accounts create the impression that most Mendelian gene identification has occurred in discrete and rapid bursts, facilitated by functional insights into the disease and strong candidate gene hypotheses, the discoveries of the vast majority of other DOK genes show otherwise. Most of these genes were pursued with trajectories more typical of most Mendelian diseases, for which putative candidate genes are usually not intuitive, and which consequently have been tackled with genome-wide positional cloning. Progress toward genetic understanding of the diverse spectrum of