A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis

Cristina de Guzman Strong1, Sean Conlan1, Clayton B. Deming1, Jun Cheng1, Karen E. Sears2 and Julia A. Segre1,*

1Genetics and Molecular Biology Branch, National Human Genome Research Institute, NIH, Bethesda, MD, USA and 2Department of Animal Biology, University of Illinois, Urbana, IL, USA

Received November 2, 2009; Revised December 28, 2009; Accepted January 18, 2010

Two common inflammatory skin disorders with impaired barrier, atopic dermatitis (AD) and psoriasis, share distinct genetic linkage to the Epidermal Differentiation Complex (EDC) locus on 1q21. The EDC is comprised of tandemly arrayed gene families encoding proteins involved in skin cell differentiation. Discovery of semidominant mutations in filaggrin (FLG) associated with AD and a copy number variation within the LCE genes associated with psoriasis provide compelling evidence for the role of EDC genes in the pathogenesis of these diseases. To date, little is known about the potentially complex regulatory landscape within the EDC. Here, we report a computational approach to identify conserved non-coding elements (CNEs) in the EDC queried for regulatory function. Coordinate expression of EDC genes during mouse embryonic skin development and a striking degree of synteny and linearity in the EDC locus across a wide range of mammalian (placental and marsupial) genomes suggests an evolutionary conserved regulatory milieu in the EDC. CNEs identified by comparative genomics exhibit dynamic regulatory activity (enhancer or repressor) in differentiating or proliferating conditions. We further demonstrate epidermal-specific, developmental in vivo enhancer activities (DNaseI and transgenic mouse assays) in CNEs, including one within the psoriasis-associated deletion, LCE3C_LCE3B-del. Together, our multidisciplinary study features a network of regulatory elements coordinating developmental EDC gene expression as an unexplored resource for genetic variants in skin diseases.

INTRODUCTION

The Epidermal Differentiation Complex (EDC) spanning 1.6 Mb on human 1q21 (mouse 3q) harbors four clusters of tandem gene families: Filaggrin (FLG)-like, Late Cornified Envelope (LCE), Small Proline Rich Region (SPRR) and the S100 genes (Fig. 1) (1–5). FLG-like, LCE and SPRR genes encode structural proteins which are cross-linked to form the essential epidermal barrier at the surface of the skin, although S100 genes encode chemoattractant proteins expressed when the barrier is impaired (6). Recently, the EDC has been implicated in two common inflammatory skin disorders with impaired barrier, atopic dermatitis (AD) and psoriasis, that both share distinct genetic linkage to the EDC suggesting a role for these genes in disease pathogenesis (7–9). Loss-of-function variants in the FLG gene that were first identified in ichthyosis vulgaris, a common dry, scaly skin disorder (10), are also strongly associated with AD with subsequent progression to asthma or allergic rhinitis known as the atopic march (8,11,12). However, only 50% of AD patients possess FLG null alleles (12) and exclusion of the common FLG alleles in familial AD studies continues to demonstrate linkage to the EDC suggesting additional genetic variants in the FLG-like genes and the EDC (9,13). Recent genome-wide association studies identified association of psoriasis to a 30-kb deletion spanning the LCE3C and LCE3B genes (LCE3C_LCE3B-del) (14,15). Although psoriatic skin samples from LCE3C_LCE3B-del genotyped patients demonstrated a decrease in LCE3C and LCE3B expression, it is possible that a regulatory element within the LCE3C_LCE3B-del allele could be a contributing factor as well.

The spatial and temporal expression of several genes in the EDC during developmental epidermal differentiation and their dense tandem genomic arrangement suggest a genomic mechanism to coordinate their expression. One intriguing model for coordinate expression is a network of cis-regulatory

*To whom correspondence should be addressed. Tel: +1 3014022314; Fax: +1 3014024929; Email: jsegre@nhgri.nih.gov

Published by Oxford University Press 2010.
elements in the EDC. Comparative genomics have greatly facilitated the identification of regulatory elements in Conserved Non-coding sequences or Elements (CNEs) (16). We apply this method to identify potential regulatory elements in the EDC locus as an untapped resource for functional genetics studies. Using this bioinformatics approach, we observe remarkable evolutionary conservation of the EDC locus across phylogenetically distinct mammalian genomes. Furthermore, we identified 48 CNEs in the EDC that exhibited dynamic regulatory activity during differentiating and proliferating conditions. We demonstrate epidermal-specific developmental in vivo enhancer activity in two CNEs, especially one in the psoriasis-associated \(LCE3C\_LCE3B\) allele. These results highlight a genomic mechanism to coordinate developmental expression of the EDC genes via cis-regulatory elements that could likely play a role in human skin disease.

RESULTS

Developmental coordination of EDC gene expression during epidermal differentiation and skin barrier formation

To investigate the degree to which the EDC genes are coordinately regulated during skin epidermal development, we employed a semi-quantitative analysis of pan-EDC gene locus temporal expression in murine embryonic skin (Fig. 2A). In mice, epidermal differentiation commences at embryonic day (E)15.5 followed by acquisition of skin barrier formation at E16.5 (17). Our results indicate initial expression of a majority of the EDC genes (44/61) genes, at E15.5 compared with control E13.5 (Fig. 2B). By E16.5, 11 additional genes were induced, most notably the \(LCE\) genes and increased expression of other EDC genes. The \(LCE\) genes at the extreme 5' and 3' ends of the \(LCE\) gene family were induced at E15.5 followed by induction of the more internal \(LCE\) genes at E16.5. Although several of the \(S100\) genes demonstrated similar expression levels both at E15.5 and E16.5, not all \(S100\) genes were expressed suggesting a discrete mechanism of regulation. These data confirm and expand the previous SPRR and \(LCE\) gene family-centric analyses, which revealed cluster-specific gene expression in various adult human epithelia (18,19). Together, these data support a coordination of EDC temporal expression during epidermal development that can be facilitated by cis-regulatory elements.

The EDC represents an ultraconserved microsyntenic genomic block in mammals

To identify CNEs, we aligned orthologous EDC genes from multiple, phylogenetically distinct mammalian species: human, chimp, rhesus, mouse, rat, dog and opossum (Fig. 3). The metatherian (or marsupial) opossum, representing one of the furthest mammalian phylogenetic branches (diverged 180 million years ago from the human), provides unique insights in ascertaining mammalian biological processes in comparison to these mammals (20–22). For example, opossums exhibit the same patterning of epidermal barrier acquisition observed in other mammals (23,24) yet at an accelerated pace given the shortened gestational age in utero (~13 days) (Supplementary Material, Fig. 1). Given the observation of conserved patterning of epidermal barrier formation and its unique phylogenetic position in the mammalian lineage, we, therefore, incorporated the opossum genome to empower our comparative genomics studies. We find that the alignment of orthologous EDC genes in multiple mammalian species demonstrated a striking degree of linearity (order or arrangement of the genes) and synteny (genes occurring on the same chromosome) of orthologous EDC genes. Consistent with a shared genome-wide sequence identity to the human (93%) (25), the chimp and rhesus EDC loci are highly conserved (linear and syntenic). This conservation extended to the mammalian order rodentia where mouse and rat EDC loci differed from the human EDC locus in size (3.1 and 3.9 Mb, respectively) owing to a large

---

**Figure 1.** The EDC is comprised of tandemly arrayed gene families. Human (hg18) EDC on chromosome 1q21 (1.6 Mb).
1.3 Mb ‘gene desert’ insertion (telomeric to HRN and centromeric to the FLG genes). Analysis of orthologous genes in the opossum also revealed conserved linearity with a large 309 Mb insertion separating the FLG-like and SPRR gene families from the LCE and S100 gene families (21). Analysis of the EDC orthologous genes in the dog revealed conserved linearity among the FLG-like, LCE and SPRR gene families with limited EDC synteny as the S100 gene family mapped to another chromosome, suggesting independent regulation of the S100 genes in the dog. Although the evolutionary time between the radiation of mammalian species is too short to produce the null hypothesis of randomly distributed genes in the genomes, the existence of orthologous EDC genes as linear and syntenic loci across mammalian genomes suggests a genetic mechanism to maintain the EDC as a regulatory block.

Regulatory activities in the CNEs of the EDC

In our search for regulatory elements, we identified 48 CNEs (range of 115–1403 bp; mean = 431 bp) using a conservative alignment of orthologous EDC loci from multiple mammalian species (Fig. 4A, Supplementary Material, Table S1, Methods). Totaling 19,847 bp, CNEs represent 1.2% of the human EDC, a majority of which are located intergenically. To determine the regulatory potential of the CNEs, we performed luciferase reporter assays on cultured epidermal cells (keratinocytes) under terminally differentiating and proliferating conditions (1.2 and 0.05 mM Ca\(^{2+}\), respectively). Fourteen random Non-Conserved Non-coding Elements (NCNE) in the EDC (range of 115–1109 bp; mean = 496 bp) were used as a negative control. Nine CNEs exhibit enhancer activity (mean >2-fold increased luciferase activity) and six CNEs exhibit repressor activity (mean >2-fold decreased luciferase activity) under differentiating conditions (Fig. 4B). Under proliferating conditions, four CNE continued to demonstrate enhancer activity (Fig. 4C) and three CNE continued to demonstrate repressor activity. Three additional CNEs exhibited repressor activity exclusively under proliferating conditions. By comparison, all NCNE failed to exhibit enhancer or repressor activity (data not shown). In summary, these data demonstrate an enrichment of regulatory activity (30%, 15/48) of the EDC CNEs during differentiation compared with 20% (10/48) of the EDC CNEs during proliferation.

h923 and h621 in LCE3C_LCE3B-del function as epidermal-specific developmental enhancers in vivo

We further characterized the in vivo regulatory activity of two CNEs, selected either by in vitro enhancer activity or by disease relevance, using DNaseI hypersensitivity assays and transgenic mice. h923 is a 657-bp CNE located 2.6 kb upstream of the involucrin (IVL) transcriptional start site and demonstrated the highest enhancer activity in vitro. h621 is a 333-bp CNE that maps within the psoriasis-associated LCE3C_LCE3B-del region. Regulatory elements, typically devoid of core nucleosome structure, are hypersensitive to DNaseI treatment (26). We quantified DNaseI sensitivity of human primary epidermal cells by real-time PCR and overlapping amplicons that tile across the targeted CNEs (27). In vivo tissue-specific enhancer activity was assayed using transgenic

Figure 2. Coordinated expression of EDC genes during epidermal differentiation and barrier formation. (A) Mouse (mm9) EDC (chr3), 3.1 Mb, are comprised of 4 gene families (FLG-like [II], LCE [III], SPRR [IV] and S100 [I and V]). Group I represents a cluster of 2 S100 genes (S100A10, S100A11). (B) Heatmap reflecting a semiquantitative real-time PCR analysis of EDC gene expression from mouse dorsal skin at E15.5 (epidermal differentiation) and E16.5 (barrier formation). Experiments were done in triplicate (n = 2 per embryonic stage) and normalized to β-2-microglobulin. Gray scale legend, fold-change over E13.5 expression.
**DISCUSSION**

Our multidisciplinary study demonstrates a network of inter-spersed cis-regulatory elements in the EDC to coordinate gene expression during mammalian epidermal differentiation. Furthermore, we report a striking degree of synteny and linearity in EDC across eutherian and metatherian mammalian phylogenetic species. Discovery of extant metatherian evolutionary origins in the development of a unique mammalian tissue empowered our comparative genomics studies ascertaining CNEs as regulatory elements. We demonstrate the dynamic nature of regulatory activity of EDC CNEs with enhancer and repressor activity in differentiating versus proliferating conditions. Moreover, we show that two CNEs (h621 and h923) demonstrate in vivo epidermal-specific, developmental enhancer activity using DNaseI hypersensitivity assays and transgenic mice.

**h923 enhancer is sufficient to direct epidermal tissue specificity**

h923 demonstrated consistent high enhancer activity in our cell reporter assays suggesting a dual regulatory role in proliferating and differentiating conditions. Although h923 partially overlaps with a previously described distal regulatory region in the human IVL promoter required for tissue-specificity and expression when tested with its endogenous promoter in transgenic mice (29), our results demonstrate that h923 is sufficient to direct epidermal tissue specificity without its native IVL promoter.

**h621 enhancer maps to the psoriasis-associated LCE3C_LCE3B-del**

Despite our observation of negative cell-based reporter results in h621, we were able to demonstrate in vivo epidermal-specific enhancer activity in our transgenic mice. Our findings are consistent with previous reports of regulatory polymorphisms that fail to recapitulate their in vivo effects in cell-based assays (30). Identification of in vivo developmental enhancer activity in h621 mapping within the psoriasis-associated LCE3C_LCE3B-del allele suggests an alternative disease mechanism in LCE3C_LCE3B-del-psoriasis, a loss of regulatory activity affecting global transcription of the EDC.
Figure 4. Regulatory activities in the CNEs of the EDC. (A) CNEs (grouped into clusters I–V, as depicted in Fig. 1) are labeled as distance (kb) from the S100A10 transcriptional start site. Identification of CNEs span (hg18) chr1:150,202,011–151,891,137 including −20 kb of the transcriptional start site of S100A10 and +20 kb downstream from S100A1 transcript. EDC CNEs were tested for in vitro enhancer and repressor activity (luciferase reporter assays) in keratinocytes under (B) differentiating and (C) proliferating conditions. CNEs exhibiting >2-fold increased luciferase activity demonstrate enhancer activity and those that exhibit >2-fold decreased luciferase activity demonstrate repressor activity. Columns represent an average of two independent experiments performed in duplicate. Numbered CNEs are highlighted in designated boxes where Rectangle = enhancers (differentiating and proliferating), Hexagon = repressors (differentiating and proliferating), Triangle = enhancers (differentiating only), Oval = repressors (proliferating only), Bars, standard error.
Cluster of regulatory elements (Group II): association with AD?

Interestingly, the first genome-wide association study for AD recently identified a tagging SNP in linkage disequilibrium (LD) with the FLG-like gene family within the EDC even when excluding individuals with the two common mutations in the FLG locus (R501X and 2282del4) (9). Another study for AD also demonstrated linkage to the EDC even after excluding the FLG common mutations (31). Since these studies only genotyped subjects for the two common FLG mutations, it is possible that either other FLG variants or alternatively, other genetic variants within the LD block account for the residual association in these studies (12). To that end, we observe a cluster of CNEs with some of the highest regulatory activity under differentiating conditions in this LD block (Group II, Fig. 5A), suggesting possible FLG-like gene regulatory regions to analyze in AD etiology.

Developmental enhancers in disease

This study augments the evidence that developmental enhancers may be implicated and causative in disease (32). In addition to their role in development, they may also play an extensive role in regulating adult tissue repair and in the case of skin diseases, response to barrier impairment or environmental stress. As we learn more about the role of lincRNAs and miRNAs that are also evolutionarily conserved (33,34), it would be intriguing to investigate whether our CNEs encode these small RNA molecules as well. In summary, our study not only provides a much-anticipated source of potential genetic variants in AD and psoriasis but also underscores the importance of extensive genomic and complementary functional studies in conjunction with genetic studies.

MATERIALS AND METHODS

Semiquantitative real-time PCR

RNA was extracted from the dorsal skin of E13.5, E15.5 and E16.5 mice with TRIzol/chloroform (Invitrogen) and RNeasy purification (Qiagen). cDNA was generated from the RNA (5 \mu g) using Superscript VILO cDNA kit (Invitrogen). Semiquantitative real-time PCR on cDNA was performed in triplicate (ABI 7300) using primers specific for target sequence (35) (Supplementary Material, Table S2) and SYBR Green (Invitrogen) to measure increased fluorescence of targeted amplicon. Primers used in this study amplify across intron boundaries where available and were confirmed for specificity using UCSC’s ‘in silico PCR’ feature. Real-time PCR analysis (using the \text{ΔΔC}_T method per manufacturer’s instructions) included experimental data with single-peak dissociation curves and normalization to β2-microglobulin expression.

Comparative genomic sequence analysis

The EDC loci from human (hg18), chimp (panTro2), rhesus (rheMac2), mouse (mm8), rat (rn4), dog (canFam2) and opossum (monDom4) sequences were obtained from UCSC using the ‘convert’ option from the human reference sequence and aligned using Mutispecies Percent Identity Plot Maker, MultiPipMaker (36) to obtain CNEs. ‘Single coverage’ and ‘repeat masked sequence’ options were selected in MultiPipMaker to exclude false positive regions that could be generated from paralogous gene sequences represented in the
EDC. Sequences identified by MultiPipMaker as alignable with at least 100 bp and >50% nucleotide identity across all seven mammalian genomes in non-coding regions were designated as CNEs. CNEs immediately adjacent to exon and introns were excluded. For synteny block analysis, the EDC loci were aligned using Mauve (version 2.3.0) (37) with the default settings and seed families enabled.

Luciferase assay
A library of CNE and NCNE from the EDC was generated by PCR amplification from human BAC clones using Fast Start Fidelity Taq (Roche) and subsequent cloning into Gateway (Invitrogen) compatible pGL3 (Promega) plasmids upstream of a mouse Sprr1a promoter driving firefly luciferase expression. NCNE were selected using a sliding window view of the human EDC locus to identify non-conserved (no shared alignment between mouse, rat, dog and opossum) DNA sequences across the EDC locus in intergenic regions that represented a range from 115 to 1109 bp and a median size of 496 bp similar to the CNEs. All CNEs and NCNEs were sequence verified. Dual luciferase assays (Promega) were performed in duplicate in two independent experiments on mouse SP-1 keratinocytes plated on 6-well plates under proliferating (0.05 mM Ca\(^{2+}\)) or terminal differentiating (1.3 mM Ca\(^{2+}\)) conditions in S-MEM/10% chelex-treated fetal bovine serum (Lifeblood Medical) and measured 24 and 72 h post-transfection (Fluorskan Ascent FL fluorimeter, Thermo Scientific), respectively (35). For each well, firefly luciferase activity was normalized to the co-transfected Renilla luciferase activity and empty vector control.

DNasel hypersensitivity assay
Human skin samples were obtained with appropriate informed consent and reviewed by the NIH Office of Human Subject Review. Only human newborn foreskin samples that screened positive for LCE3CF/LCE3C_LCE3B-del were used. DNasel assay. Foreskin samples were cut into 1 cm pieces and incubated on dispase:HBSS (1:1) (BD Biosciences) with frequent shaking and pipetting to obtain single cell isolation. Single cells were lysed with 0.03% NP-40/10 mM Tris–HCl (pH 7.5)/10 mM NaCl/3 mM MgCl\(_2\) buffer and digested with increasing amounts of DNasel (Roche) at 37°C to obtain DNasel-treated nuclei. After removal of proteins with overnight proteinase K digestion, DNasel-digested DNA was subsequently isolated via phenol/chloroform extraction. Double-stranded DNA was quantitated using PicoGreen (Invitrogen) according to the manufacturer’s instructions. DNasel sensitivity was assayed by real-time PCR (ABI 7300) using 9.5 ng of DNasel-digested DNA, SYBR Green (Qiagen) and tiling primers (200 bp amplicons with 50 bp overlap) to amplify targeted sequence (Supplementary Material, Table S3) (27). A non-conserved region (~8.7 kb downstream of SPRR3) was designed as a negative DNasel control.

In vivo mouse enhancer assay
All animals were maintained in an AAALAC accredited, murine pathogen-free facility at the National Institutes of Health (Bethesda, MD, USA) in accordance with institutional protocols and the Guide for the Care and Use of Laboratory Animals. All animal procedures were approved by the NHGRI Animal Care and Use Committee. CNEs were cloned into the Gateway-compatible hsp68-lacZ reporter vector (gift from Marcelo Nobrega, University of Chicago), sequence-verified, purified by cesium chloride gradient (Loftstrand) and linearized with SalI (NEB) prior to injection into fertilized eggs that were implanted into pseudopregnant females. Analysis was done on transgenic founder embryos (F0) at the corresponding embryonic stage post-transplantation (where transplantation day is designated as embryonic day [E] 0.5) and were genotyped using lacZ primers. Whole-mount embryos were stained overnight for β-galactosidase staining activity using X-gal (Fermentas) after cold fixation as previously described (38). β-galactosidase stained tissue sections were obtained from paraffin blocks (Histoserv).

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS
We thank Cherry Yang, Yicong Liu and Roli Mandana for technical expertise, Carole Yee for primary keratinocyte samples, Peter Chines for primer tile design, Elsa Escobar and Lauretine Sop for mouse husbandry and Julia Fekecs for assistance in preparing the figures. We thank Greg Crawford and David Sankoff for helpful discussions with the genome-wide DNase hypersensitivity data and genomic synteny, respectively. We also thank Anne Bowcock, Elizabeth Grice and Stacie Loftus for manuscript critiques.

Conflict of Interest statement. None declared.

FUNDING
This work was supported by the National Human Genome Research Institute Intramural Research Program and the National Institute of Health K99AR055948 to C.G.S.

REFERENCES


