Dynamic transcriptional and epigenomic reprogramming from pediatric nasal epithelial cells to induced pluripotent stem cells

Hong Ji, PhD,[a] Xue Zhang, PhD,[b] Sunghee Oh, PhD,[c] Christopher N. Mayhew, PhD,[d] Ashley Ulm, BA,[a] Hari K. Somineni, MS,[a] Mark Ericksen, BS,[a] James M. Wells, PhD,[d,e] and Gurjit K. Khurana Hershey, MD, PhD[a]

Cincinnati, Ohio, and Cheongju, South Korea

Background: Induced pluripotent stem cells (iPSCs) hold tremendous potential, both as a biological tool to uncover the pathophysiology of disease by creating relevant human cell models and as a source of cells for cell-based therapeutic applications. Studying the reprogramming process will also provide significant insight into tissue development.

Objective: We sought to characterize the derivation of iPSC lines from nasal epithelial cells (NECs) isolated from nasal mucosa samples of children, a highly relevant and easily accessible tissue for pediatric populations.

Methods: We performed detailed comparative analysis on the transcriptomes and methylomes of NECs, iPSCs derived from NECs (NEC-iPSCs), and embryonic stem cells (ESCs).

Results: NEC-iPSCs express pluripotent cell markers, can differentiate into all 3 germ layers *in vivo* and *in vitro*, and have a transcriptome and methylome remarkably similar to those of ESCs. However, residual DNA methylation marks exist, which are differentially methylated between NEC-iPSCs and ESCs. A subset of these methylation markers related to epithelium development and asthma and specific to NEC-iPSCs persisted after several passages *in vitro*, suggesting the retention of an epigenetic memory of their tissue of origin. Our analysis also identified novel candidate genes with dynamic gene expression and DNA methylation changes during reprogramming, which are indicative of possible roles in airway epithelium development.

Conclusion: NECs are an excellent tissue source to generate iPSCs in pediatric asthmatic patients, and detailed characterization of the resulting iPSC lines would help us better understand the reprogramming process and retention of epigenetic memory. (J Allergy Clin Immunol 2015;135:236-44.)

Key words: *Induced pluripotent stem cells, nasal epithelial cells, DNA methylation, gene expression, epigenetic memory, asthma*

Generation of induced pluripotent stem cells (iPSCs) from somatic cells offers enormous potential for modeling diseases, generating cells for therapeutic purposes, and elucidating developmental processes.[1-6] iPSCs are generated from a variety of somatic cells through ectopic expression of specific factors in these cells, which induces a state of pluripotency that closely resembles that of embryonic stem cells (ESCs).[7,8] Since initial success with integrating viral vectors in 2006,[9] many groups have used nonintegrating vectors,[10,11] the delivery of reprogramming factor RNAs or proteins, and small molecules[8,12,13] or chemicals[14] to increase reprogramming efficiency, reduce random footprint mutations, and minimize the tumorigenicity of the resulting iPSCs.

The reprogramming process through which a somatic cell acquires pluripotency is an epigenetic transformation. Although there are reports suggesting that these iPSCs are indistinguishable from embryonic stem cells (ESCs) in terms of DNA methylation and gene expression profiles,[15-19] other reports, including nucleotide-resolution DNA methylation mapping, suggest that iPSCs have different epigenomic and gene expression profiles compared with ESCs and that these differences are mitotically transmittable.[20-28] Furthermore, functional differences have been noted between iPSCs and ESCs in some differentiation assays.[23,24,28-30] Differences in induction, culture conditions, and methods of assessing variation could explain these inconsistencies. Additional complexity results from the source of the tissue and the age of the donor, which affect the efficiency of reprogramming and the differentiation capacity of iPSCs because of the retention of epigenetic memory.[23,24,28,31] Therefore iPSCs derived from adult somatic tissues might harbor safety risks for therapeutic applications.[32] Nevertheless, reprogramming cells that are readily accessible might be more broadly applicable for modeling diseases and generating autologous cells for therapeutics. Thus it is important to evaluate newly generated human somatic cell–derived iPSCs and select those iPSC clones with minimal residual markers from the cells of origin for medical applications.

In this article we generated iPSCs from nasal epithelial cells (NECs) isolated from the nasal mucosa of asthmatic children (NEC-iPSCs). Nasal mucosa was chosen because it can be easily sampled, even in pediatric populations, and this tissue is relevant to asthma because the upper airway shares many similarities with the lower airway epithelium.[33] We characterized the epigenetic and gene expression profiles of these NEC-iPSCs and compared them with those of ESCs and the nasal cells from which they originated. We found that dynamic DNA methylation changes occur during reprogramming and that the transcriptomes and

From the Divisions of [a]Asthma Research, [b]Human Genetics, [d]Developmental Biology, and [e]Endocrinology, Cincinnati Children's Hospital Medical Center, and the Department of Pediatrics, University of Cincinnati, Cincinnati, and [c]the Division of Human Genetics, Kim Sook Za Children's Hospital Medical Center Research Foundation, Cheongju.

Supported by R21AI101375 (to H.J.), National Institutes of Health/National Center for Advancing Translational Sciences 8UL1TR000077-04 (to H.J.), U19 AI070412 (to H.J.), and 2U19AI70235 (to G.K.K.H.).

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication May 15, 2014; revised July 24, 2014; accepted for publication August 27, 2014.

Available online October 14, 2014.

Corresponding author: Hong Ji, PhD, Division of Asthma Research, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229. E-mail: Hong.Ji@cchmc.org.

0091-6749/$36.00

http://dx.doi.org/10.1016/j.jaci.2014.08.038

Abbreviations used
CCHMC: Cincinnati Children's Hospital Medical Center
cNEC: Cultured nasal epithelial cell
DMP: Differentially methylated point
ESC: Embryonic stem cell
FDR: False discovery rate
hESC: Human embryonic stem cell
HFF: Human foreskin fibroblast
iPSC: Induced pluripotent stem cell
NEC: Nasal epithelial cell

methylomes of NEC-iPSCs are remarkably similar to those of ESCs. Some DNA methylation marks of parental tissue origin existed in NEC-iPSCs, even after 15 passages, but the expression levels of nearby genes were indistinguishable from those in ESCs. In addition, bioinformatic analysis on transcriptional and DNA methylation profiles during reprogramming from NECs to iPSCs revealed novel genes and pathways that might be involved in the development of airway epithelium.

METHODS

Human subjects and nasal mucosal cell sampling/processing

Nasal epithelial samples were collected from participants of the Exposure Sibling Study, a case-control study of asthmatic patients and their nonasthmatic siblings living in the Cincinnati Metropolitan area. Four children (age, 13-17 years) with asthma were included in this study. A trained clinical research coordinator obtained informed consent from participants and their parents/guardians by using a protocol approved by the Cincinnati Children's Hospital Medical Center (CCHMC) Institutional Review Board. Asthma diagnosis was confirmed with the diagnosing allergist/pulmonologist at the CCHMC or from the child's community pediatrician according to American Thoracic Society criteria. These children did not show any signs or symptoms of atopy at the time of recruitment and sample collection, according to parent questionnaires. Two samples were used for induction to pluripotency, and 2 were used for fresh and cultured sample comparison. Nasal mucosa sampling was performed with a CytoSoft Brush (Medical Packaging Corp, Camarillo, Calif), and the samples were immediately taken to the laboratory for processing, as previously described.[34] The NECs were cultured in bronchial epithelial cell growth medium (Lonza, Basel, Switzerland) before they were subject to reprogramming. DNA and RNA were extracted immediately from a portion of the NECs with the AllPrep DNA/RNA Micro kit (Qiagen, Hilden, Germany), according to the manufacturer's protocols.

Generation of iPSC lines from nasal mucosal samples, foreskin, and blood

Nasal epithelial samples obtained from children as described above were cultured in bronchial epithelial cell growth medium until they reached confluence.[35] Primary human foreskin fibroblasts (HFFs) from 3 healthy neonates were cultured from foreskin tissue. Blood from a healthy donor was subjected to Ficoll centrifugation to enrich for PBMCs. For lentiviral-mediated reprogramming factor delivery, cells were transduced with a polycistronic lentivirus expressing Oct4, Sox2, Klf4, c-Myc, and dTomato in the presence of polybrene (Santa Cruz Biotechnology, Dallas, Tex).[36] For integration-free reprogramming, HFFs were nucleofected (program U20) with the EBNA1/OriP-based episomal plasmids pCLXE-hOct3/4-shp53, pCLXE-hSox2-Klf4, pCLXE-hLmyc-Lin28, and pCLXE-GFP obtained from Addgene (Cambridge, Mass; ID nos.: 27077, 27078, 27080, and 27082).[37] Cells were transferred to mouse embryonic fibroblasts and cultured in standard human embryonic stem cell (hESC) media containing 4 ng/mL basic fibroblast growth factor with or without SPT cocktail (2 μmol/L SB431542 [Stemgent, Cambridge, Mass], 0.5 μmol/L PD0325901 [Stemgent], and 0.5 μmol/L thiazovivin [Stemgent]) for 10 days, followed by culture in hESC media without SPT cocktail. For some experiments, cells were transferred to Matrigel (Corning, NY) 6 days after lentiviral transduction and cultured in mTeSR1 media. Cultures were fed daily until hESC-like colonies appeared. Colonies with similar morphology to hESCs were excised, transferred to feeder-free culture conditions consisting of Matrigel and mTeSR1, and expanded in culture similar to National Institutes of Health–approved ESCs (WA09 or H9; Fig 1, *B*).[38] PBMCs were reprogrammed, as described previously.[39] Materials and methods for further characterization of these iPSCs are included in the Methods section in this article's Online Repository at www.jacionline.org.

RNA sequencing and gene expression analysis

RNA (approximately 1 μg) was used for Illumina sequencing (Illumina, San Diego, Calif). Read alignment, splice identification, expression-level quantification, and identification of differentially expressed genes were performed by using previously described methods.[40-47] The ggplot2 and reshape2 library in R software were applied to draw the heat map of the correlation matrix among samples based on $\log_2$-scaled expression levels.

DNA methylation microarray processing and analysis

Genomic DNA was bisulfite treated and assayed with the Illumina Infinium HumanMethylation450 BeadChip (Illumina). Array quality was assessed by using sample-independent and sample-dependent internal control probes included on the array for staining, extension, hybridization, specificity, and bisulfite conversion. One NEC-iPSC sample exhibited low intensity for all sample-dependent controls, suggesting a problematic quality of the sample, and was excluded from subsequent analyses. The remaining 11 samples had greater than 98% CpG sites detected at a *P* value of .01 and approximately 95% bisulfite conversion. The signal intensities were then background adjusted and normalized by using the methylation module and used to calculate β values as follows:

$$\beta = \frac{Signal_{methylation}}{Signal_{methylation} + Signal_{unmethylation} + 100}$$

The following CpG sites were excluded from analysis: (1) CpG sites that were not detected in all samples at a *P* value of .01; (2) CpG sites on X and Y chromosomes; (3) CpG sites with 1 or more bead number smaller than 5; and (4) CpG sites with SNPs present nearby (>10 or ≤10 bp from the query site). These procedures resulted in 11 samples and 350,950 CpG sites.

The difference in β values for each of the CpG sites was tested. NECs and cNECs were paired by subject, and therefore paired *t* tests were performed. For other comparisons, 2-sample *t* tests were conducted. A false discovery rate (FDR) was calculated by using the Benjamini-Hochberg procedure or with the q value package to enhance the power in the identification of differential methylation. CpG sites with FDRs of 0.05 or less and absolute β differences of 0.1 or greater were selected as DMPs. Among the DMPs between NEC-iPSCs and ESCs, we further separated sites with aberrant reprogramming and those with parental memory. CpG sites with DNA methylation in NEC-iPSCs significantly outside the range of cNECs and ESCs were considered sites with aberrant reprogramming. The remaining sites with methylation in NEC-iPSCs either between cNECs and ESCs or no significant difference from cNECs were considered sites with potential memory.

Figs E1 to E3, Tables E1 to E8, and Video E1 can be found in this article's Online Repository at www.jacionline.org. Supplementary Methods for standard pluripotent stem cell maintenance and characterization, gene ontology analysis, association of genome-wide DNA methylation with gene expression, quantitative RT-PCR, and bisulfite pyrosequencing can also be found in this article's Online Repository.

RESULTS

Nasal mucosa samples were obtained from 2 children with asthma (aged 13 and 17 years, respectively), and primary cultured

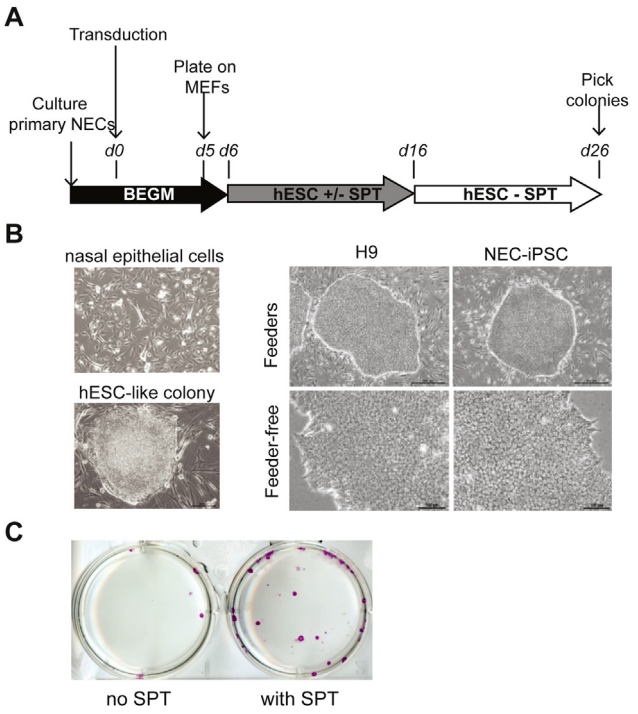

FIG 1. Reprogramming of pediatric NECs. **A,** Timeline and key steps for reprogramming. **B,** Representative morphology of NECs before transduction and ESC-like colonies after transduction. hESCs and NEC-iPSCs before (*top*; *scale bar* = 500 μm) and after (*bottom*; *scale bar* = 100 μm) transition to feeder-free culture are also shown. **C,** SPT cocktail enhances reprogramming efficiency. Purple color indicates alkaline phosphatase staining.

nasal epithelial cells (cNECs) were subjected to transduction with a polycistronic lentivirus (Fig 1, *A*).[48] Transduced cells were plated on mouse embryonic fibroblasts and cultured in standard hESC media containing 4 ng/mL basic fibroblast growth factor with or without SPT cocktail.[14] Colonies with similar morphology to ESCs were picked and expanded for several passages on mouse embryonic fibroblasts before transition to feeder-free culture conditions consisting of Matrigel and mTeSR1 (Fig 1, *B*). One iPSC line was generated from 1 donor, and 3 iPSC lines were generated from the other donor. The addition of SPT cocktail greatly enhanced the reprogramming efficiency from 0.0044% to 0.022% (Fig 1, *C*), which is consistent with a previous report.[14]

Using immunocytochemistry, we analyzed NEC-iPSC lines for expression of markers shared with ESCs. Consistent with their hESC-like morphology (Fig 1, *B*), the iPSCs were positive for OCT4, Tra-1-60, and alkaline phosphatase staining (Figs 1, *C*, and 2, *A*). Additional analysis demonstrated that, when compared with NECs, homeobox transcription factor (*NANOG*) expression in NEC-iPSCs was increased to a comparable level as seen in ESCs, and the expression of keratin 19 (*CK19*), a marker for epithelial cells, was decreased (Fig 2, *B*). Consistent with the activation of endogenous pluripotency-associated gene expression, reprogramming of NECs was accompanied by demethylation of CpG sites at the *OCT4* and *NANOG* promoters (Fig 2, *C*).

Next, we evaluated the differentiation potential of the NEC-iPSCs using *in vitro* embryonic body formation and *in vivo* teratoma induction. NEC-iPSCs readily formed embryonic bodies *in vitro*, and genes specific to each of the 3 embryonic germ layers were expressed (Fig 3, *A*). In addition, NEC-iPSCs differentiated into beating cardiomyocytes *in vitro* (see Video E1 in this article's Online Repository). When NEC-iPSCs were injected into NOD/SCID $\gamma C^{-/-}$ mice, they formed well-differentiated cystic teratomas containing tissues derived from all 3 germ layers (Fig 3, *B*). Cytogenetic analysis showed normal karyotypes (Fig 3, *C*), indicating that reprogramming did not introduce gross chromosomal rearrangements. Collectively, our analyses indicate the successful reprogramming of human primary NECs into pluripotent iPSCs.

To broadly investigate the molecular similarity between NEC-iPSCs and ESCs, we generated genome-wide gene expression and DNA methylation profiles. Comprehensive RNA sequencing analysis from 13.9 to 18.2M reads per sample revealed that NEC-iPSCs are indistinguishable from ESCs in terms of gene expression (Fig 4, *A*). Furthermore, NEC-iPSCs and ESCs showed great similarity, as evidenced by more than 0.9 correlation coefficients between them (Fig 4, *B*), which was greater than observed between ESC lines.[49] However, 54 genes (16 upregulated and 38 downregulated genes in iPSCs) were differentially expressed with an FDR of less than 0.05 after multiple testing corrections (edgeR R package, version 2.13). Nineteen of these 54 genes were differentially expressed among ESCs, NEC-iPSCs, and cNECs (Fig 4, *C*). Fourteen of 54 differentially expressed genes were further visualized after filtering out genes with high variability within their own group (see Fig E1). Interestingly, 7 of these 14 genes in NEC-iPSCs are different from all other cell types, suggesting that NEC-iPSCs might acquire these distinct expression patterns during reprogramming (Fig E1). The expression levels of 3 genes (*FAAH*, *TRPC4*, and *RP11-455F5.3*) in NEC-iPSCs are similar to those in cNECs, suggesting that these expression signatures are footprints of their parental tissue (Fig E1).

We also elucidated the changes in gene expression during reprogramming from cNECs to NEC-iPSCs. In total, 4944 genes (2096 upregulated and 2848 downregulated genes in NEC-iPSCs) are significantly differentially expressed (FDR < 0.05; Fig 4, *C-E*, and see Table E1). As expected, the reprogramming process resulted in loss of the nasal epithelial profile and acquisition of a pluripotent stem cell profile (Table E2).

Along with the dynamic expression changes from cNECs to iPSCs, DNA methylation profiles of NEC-iPSCs became distinct from NECs and cNECs and similar to ESCs (Fig 5, *A*). DNA methylation changes at many CpG sites were negatively correlated with expression alterations at genes located within 1500 bp (see Fig E2). To identify methylation patterns important to maintaining an airway epithelial phenotype, we compared primary NECs to iPSCs. CpG sites (84,864 [24.2%]) underwent significant DNA methylation changes (FDR ≤ 0.05, difference in β ≥ 0.10; see Table E3). Seven thousand two hundred thirty-four (8.5%) of these CpG sites were previously identified differentially methylated regions, many related to cancer or reprogramming. Among the 91.5% newly identified differentially methylated points (DMPs), 34.4% (26,717) are located within enhancers, and 9.3% (7,198) are associated with promoters. Gene ontology and pathway analysis of the top 3,000 most significant hits reveal that many of them are located close to or within genes involved in transcriptional regulation and organ development (see Table E4). Specifically, there is an enrichment of genes that are important for gene ontology terms, such as respiratory system development, lung development, and epithelial tube formation, including 102 CpG sites in *MUC* genes, which encode proteins coating the epithelia of the airways, intestines, and other mucous membrane–containing organs. Many other airway-specific

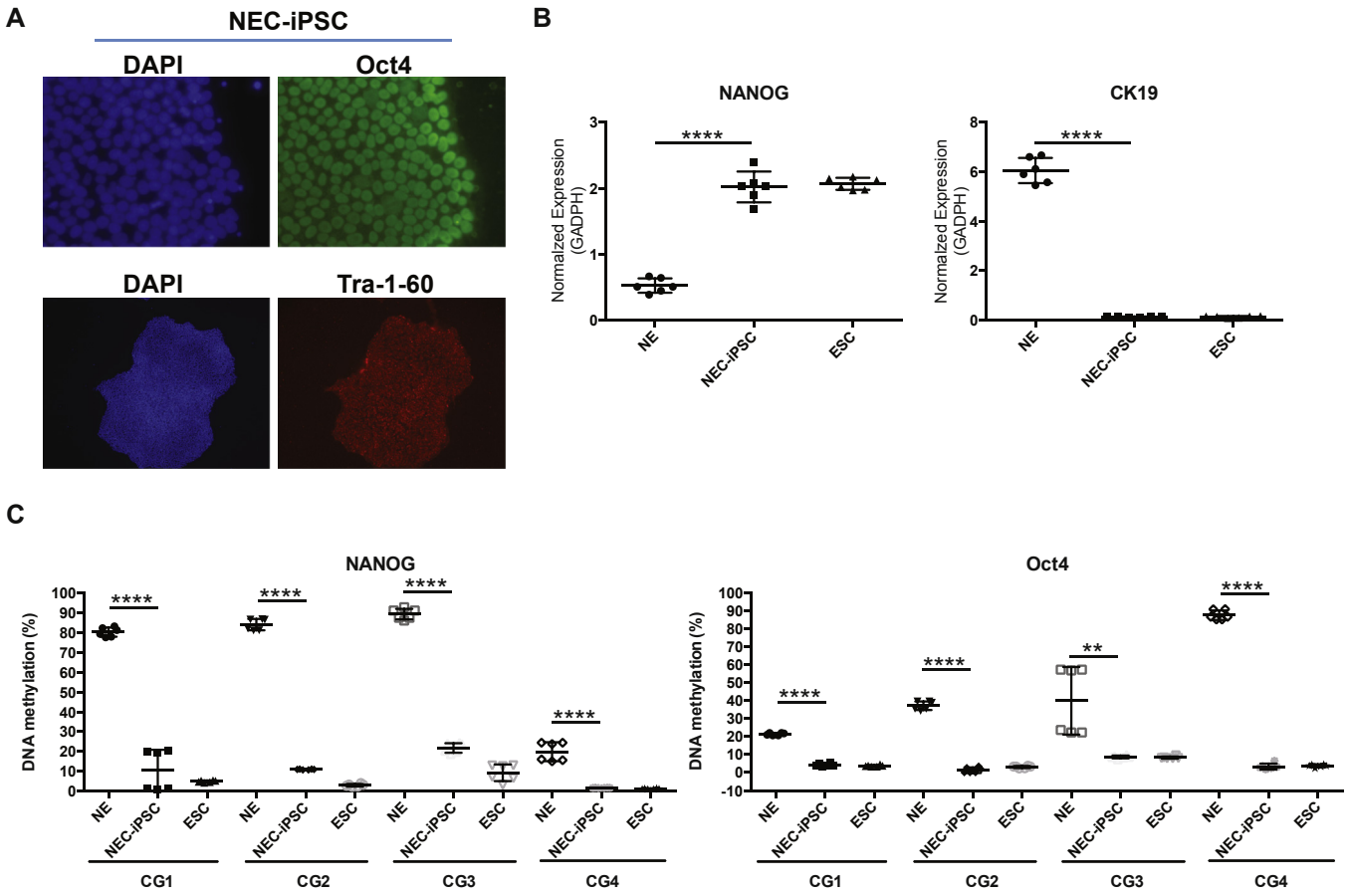

FIG 2. NEC-iPSCs express pluripotency markers through downregulation of promoter methylation. **A,** Nuclear expression of Oct4 and Tra-1-60 in NEC-iPSCs. *DAPI*, 4′-6-Diamidino-2-phenylindole dihydrochloride. **B,** Expression of *NANOG* and *CK19* in ESCs, NEC-iPSCs, and nasal epithelial samples *(NE)*. *GAPDH*, Glyceraldehyde-3-phosphate dehydrogenase. **C,** Promoter methylation of *OCT4* and *NANOG* in samples from Fig 2, *B*. Data represent means ± SDs from 3 independent experiments of 2 biological samples. One-way ANOVA: $^{**}P < .01$ and $^{****}P < .001$.

markers, including *KRT5*, *NGFR*, *ARG2*, *KRT16*, and *CFTR*, are also less methylated in NECs compared with NEC-iPSCs (Fig 5, *B*, and see Table E3). Transcription factors and pathways known to direct airway development, including *WNT3A*, *FGF2*, *SHH*, *FGF7*, *FGF10*, and *BMP4*,[48,50] undergo dynamic DNA methylation changes during reprogramming (see Table E3). There are also significant differences in DNA methylation comparing cNECs and NECs, suggesting that culturing primary cells from tissues alters DNA methylation profiles of functionally important genes (Fig 5).

When NEC-iPSCs were compared with ESCs, 99.5% (349,219/350,950) of the CpG sites were similarly methylated. Such similarity with ESCs in DNA methylation is superior to iPSC lines generated from 6 other sources (with differences from ESCs varying between 0.92% and 3.82%),[51,52] suggesting that NECs are an excellent resource for iPSC generation. Despite the large similarity in methylation patterns, differential methylation was still detected in 1731 CpG sites ($q \leq 0.05$, absolute difference in $\beta \geq 0.10$; see Table E5, *A*). These differences could either be due to aberrant DNA methylation profiles introduced by reprogramming[51,52] or memory of the tissue of origin, as documented in other iPSC lines.[23,24,28] We identified 458 CG sites with potential aberrant DNA methylation introduced by reprogramming (see Fig E3, *A*, and Table E5, *B*), including 14 CpG sites located in 3 previously reported genes (*TMEM132C*, *FAM19A5*, and *DPP6*). The remaining 1273 CpG sites can process memory from NECs (see Fig E3, *B*, and Table E5, *C*). From this, gene ontology analysis revealed that 20 CpG sites are close to or within 15 genes involved in epithelial cell differentiation and morphogenesis (cluster 4 in Table E6, *A*), supporting the existence of epigenetic memory from epithelial cells. Indeed, these CpG sites have similar DNA methylation in NEC-iPSCs compared with that seen in their parental NECs (Fig 6, *A*). A CpG site located in the promoter of *RPTN* is differentially methylated between NEC-iPSCs and ESCs, with a similar methylation level in NEC-iPSCs compared with that seen in their parental tissue (Fig 6, *B*). This difference in DNA methylation persisted for 15 passages, suggesting the retention of this memory. *RPTN* encodes reptin, a protein involved in cornified cell envelope formation.[53,54] Similarly, we observed differential methylation at a CpG site located in the *SPRR2A* promoter; however, this difference disappeared after 15 passages (Fig 6, *C*), which is consistent with the previous observation that epigenetic memory at selected loci disappears after extensive passaging *in vitro*.[23,28] In addition to the memory related to epithelial lineage, we also observed significant lower DNA methylation in NEC-iPSCs compared with ESCs at a CpG site located within

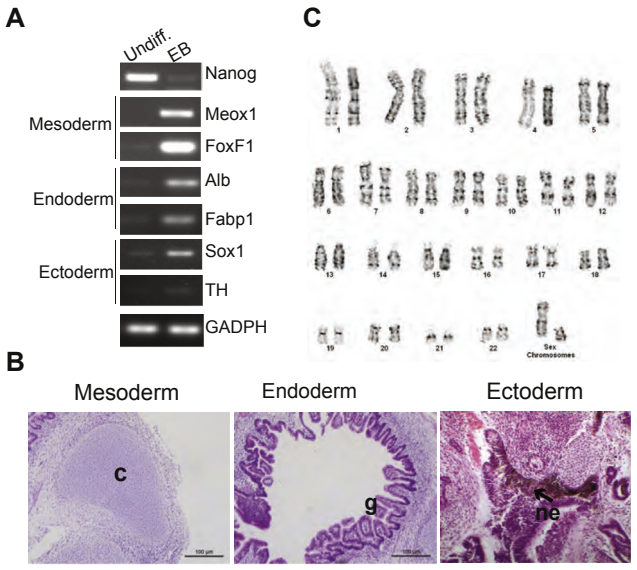

FIG 3. NEC-iPSCs can differentiate into 3 germ layers *in vitro* and *in vivo*. **A,** Expression of pluripotency and germ layer markers in NEC-iPSCs *(Undiff.)* and embryonic bodies *(EB)*. *GAPDH,* Glyceraldehyde-3-phosphate dehydrogenase. **B,** *In vivo* differentiation of NEC-iPSCs in teratomas. Tissues were derived from 3 embryonic germ layers: glandular epithelia (*g*; endoderm), cartilage (*c*; mesoderm), and pigmented neuroepithelium (*ne*; ectoderm). **C,** G-banded karyotype analysis of NEC-iPSCs.

the promoter of the *CAT* gene, even after 15 passages (Fig 6, *D*, and see Table E5, *C*). *CAT* encodes catalase, a key antioxidant enzyme in defense against oxidative stress, and contributes to asthma.[55-57] Importantly, residual DNA methylation marks in *SPRR2A* and *CAT* are specific to the NEC-iPSCs we generated because iPSCs derived from HFFs and PBMCs have significantly different DNA methylation levels (Fig 6, *C* and *D*). No significant gene expression differences were associated with these DNA methylation differences between NEC-iPSCs and ESCs (Fig E3, *C* and *D*). Collectively, our data demonstrated the persistence of epigenetic memory in NEC-iPSCs, particularly in genes related to epithelial function and asthma.

DISCUSSION

In the present study, for the first time, we report the generation of iPSCs from NECs of asthmatic children. NEC-iPSCs generated in the present study are functionally similar to hESCs and comparable with the iPSCs generated from airway epithelial cells from a healthy donor.[19] The transcriptome and methylome of the NEC-iPSCs were also remarkably similar to those of the hESCs, the gold standard of pluripotent stem cells. However, several previously uncovered DNA methylation markers in epithelium-specific and disease-related genes persist in our NEC-iPSCs, even after multiple passages, which is suggestive of a stable epigenetic memory of their tissue of origin. Comparative analysis between NEC-iPSCs and NECs also identified novel candidates that can contribute to normal development of airway epithelium. In addition to providing a novel and readily accessible somatic cell source for iPSC generation from pediatric populations, these NEC-iPSCs offer new innovative methods to model disease and can be used for drug screening and regeneration.

Because the tissues of origin have a significant effect on iPSC reprogramming and differentiation caused by the epigenetic memories they might harbor, the somatic cells used for iPSC reprogramming and the methods used to reprogram need to be carefully evaluated considering their downstream clinical applications. The reprogramming of iPSCs is an epigenetic process, and it has been previously reported that residual memory of the tissue of origin exists in iPSCs and that some persisted even after extensive passaging.[21,23,24,58-60] Such differences in genomic methylation might not alter the behavior of undifferentiated stem cells but could become apparent when the stem cells are differentiated.[23,24,28] We used NECs, an easily accessible tissue for the pediatric population, and the resulting iPSCs seem to harbor fewer epigenetic differences (0.5%) from ESCs compared with other tissues (0.92% to 3.82%).[51,52] Cultured NECs that we reprogrammed have altered molecular profiles compared with those of patients' nasal cells (Figs 4, *A* and *D*, and 5). The progenitor-like phenotype of these proliferating cells might be beneficial given that previous studies have found that progenitor cells are easier to reprogram than terminally differentiated cells.[61] We observed epithelial lineage-specific DNA methylation marks in all NEC-iPSCs, and some persisted even after multiple passages, supporting the notion that there are persistent lineage-specific epigenetic markers in iPSCs. However, our RNA sequencing data demonstrated that these markers did not correlate with any detectable gene expression differences between ESCs and NEC-iPSCs (see Fig E3, *C* and *D*). Interestingly, the expression levels of several genes located adjacent to these residual marks are different between NEC-iPSCs and cNECs (see Fig E3, *D*), even though their methylation status is similar. This implies that transcriptional reprogramming occurred at these genes, possibly through mechanisms other than DNA methylation at these loci, thus leaving footprints of DNA methylation from their origin. In addition to memory of epithelial lineages, we also observed persistent residual DNA methylation marks at other genomic locations, such as a CpG site in the promoter of *CAT*, a gene involved in response to oxidative stress and asthma pathogenesis. *CAT* expression is significantly downregulated in NEC-iPSCs during reprogramming (see Table E1), yet its DNA methylation remains similar to that seen in NECs (Fig 6, *D*, and see Table E5, *C*). Our observations support the use of DNA methylation profiling in addition to gene expression and functional analysis to carefully characterize newly generated iPSC lines, although the functional effect of such memory and a mechanistic basis for escaping reprogramming warrant further investigation.

Gene ontology and pathway analysis of genes that are differentially expressed and epigenetically modified during reprogramming revealed many known genes related to epithelial differentiation from stem cells, supporting the successful epigenetic reprogramming from epithelial lineages to pluripotent cells. Our study also identified many novel candidates that might play important roles in airway epithelium development; however, because we started reprogramming from proliferating primary nasal cells in culture, our study might not reveal genes involved in the terminal airway differentiation steps, such as tight junction formation or cilia development. For example, we identified a significantly demethylated CpG site within the *MUC15* promoter and within *ANO3* in NECs compared with NEC-iPSCs and ESCs. *ANO3* encodes anoctamin-3, which might function as a calcium-activated chloride channel and is associated with asthma and eczema.[62] The expression of *KRT5* and *SPRR2A* is turned off during reprogramming (Fig 4, *D* and *E*). *KRT5* encodes type II cytokeratin, which is specifically expressed in the basal layer of the

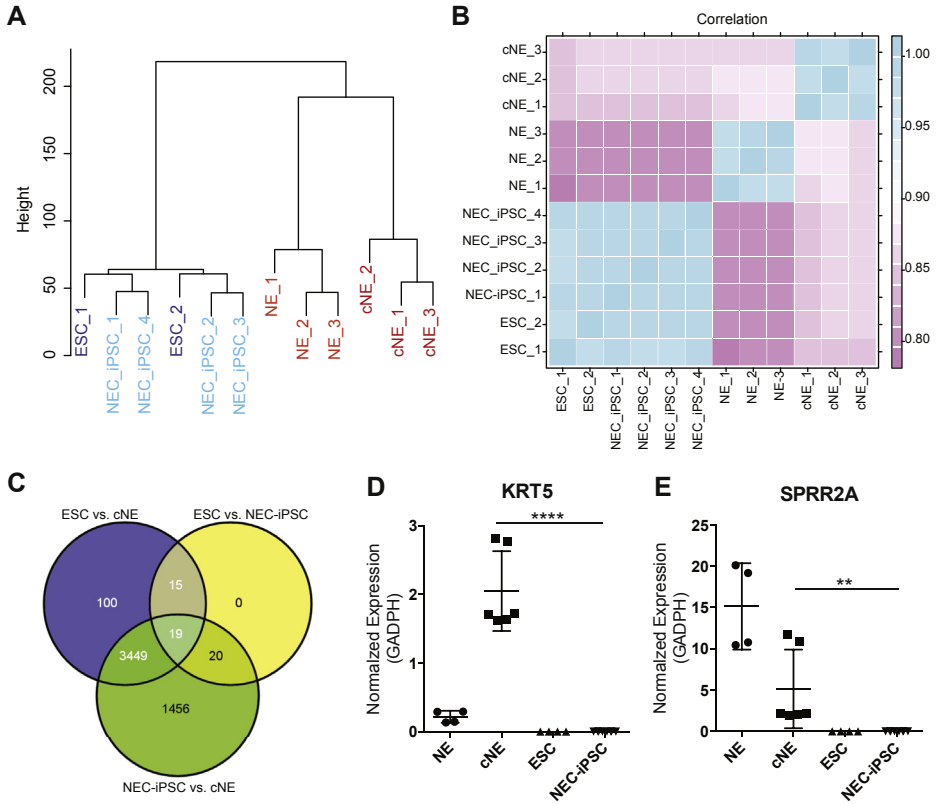

FIG 4. NEC-iPSCs have similar transcriptomes compared with ESCs. **A,** Hierarchic clustering of ESCs, NEC-iPSCs, cNECs, and nasal epithelial samples *(NEs)*. **B,** Correlation map of samples shown in Fig 4, *A*. Pearson coefficients between samples were plotted. **C,** Overlap between the differentially expressed genes from indicated comparisons. **D** and **E,** Quantitative RT-PCR of *KRT5* and *SPRR2A*. Data represent means ± SDs of indicated samples. *GAPDH*, Glyceraldehyde-3-phosphate dehydrogenase. One-way ANOVA: $^{**}P < .01$ and $^{****}P < .0001$.

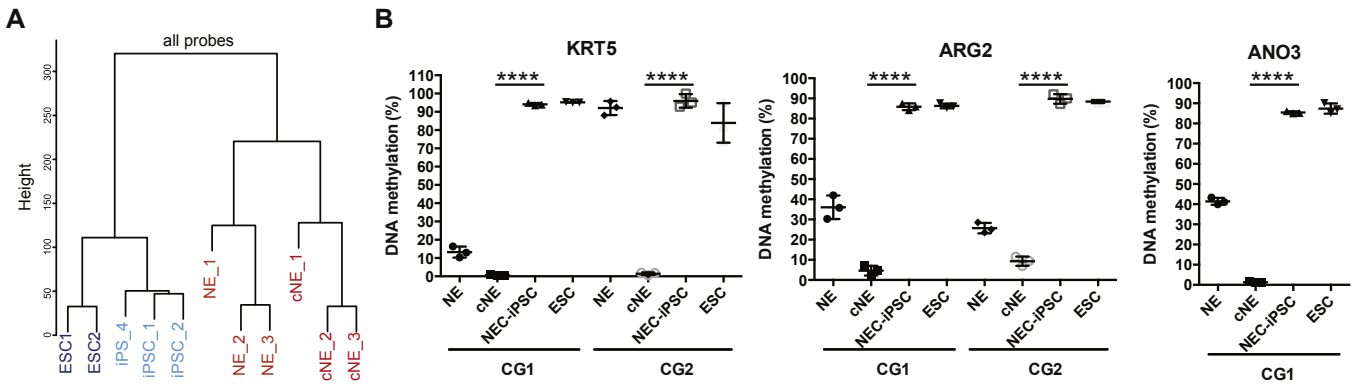

FIG 5. NEC-iPSCs have a similar methylome compared with that of ESCs. **A,** Hierarchic clustering of the same samples shown in Fig 4. **B,** Reprogramming of airway-specific markers *KRT5*, *ARG2*, and *ANO3*. Data represent means ± SDs of indicated samples. *NE*, Nasal epithelial samples. One-way ANOVA: $^{****}P < .0001$.

epidermis with its family member KRT14 by all stratified squamous epithelia.[63] *SPRR2A* is one of the small proline-rich protein genes (SPRRs) that encode precursors of the cornified cell envelope, which are specifically expressed during keratinocyte terminal differentiation.[64] How these genes contribute to airway epithelium development is currently unknown, and they are novel candidates for follow-up. Recently, ESCs and patient-specific iPSCs were progressively differentiated into Nkx2.1-expressing lung progenitors and proximal lung epithelial cells, providing a useful platform for disease modeling and *in vitro* drug testing.[48,50,65] Despite this recent success, the regulation of airway epithelium development is poorly understood compared with that of other systems. Therefore the identification of novel regulators of these processes in our study will facilitate the

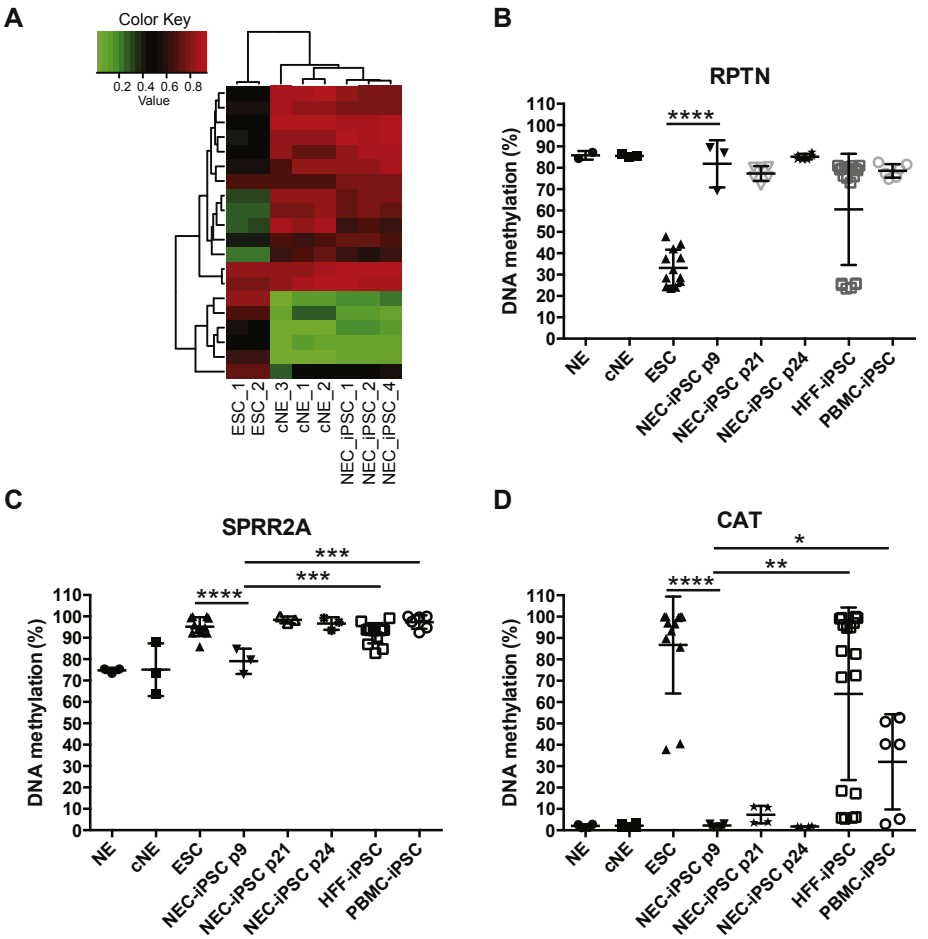

FIG 6. Epigenetic memory of parental tissue persists in NEC-iPSCs. **A,** DNA methylation at 20 CpG sites with epigenetic memory related to epithelial function. **B-D,** DNA methylation at CpG sites located within the *RPTN*, *SPRR2A*, and *CAT* promoters. Data represent means ± SDs of duplicate experiments for 2 or more biological samples for each indicated cell type. One-way ANOVA: **P* < .05, ***P* < .01, ****P* < .001, and *****P* < .0001.

optimization of such directed differentiation and establishment of functional airway epithelium for *in vitro* modeling of diseases.

We thank the CCHMC Genetic Variation and Gene Discovery Core Facility for RNA sequencing, the Genomics and Sequencing core at the University of Cincinnati for Illumina Infinium 450K array processing, the Pluripotent Stem Cell Facility for iPSC generation, and the Pyrosequencing Laboratory for locus-specific DNA methylation analysis. HFF cells used for reprogramming were obtained through the Department of Dermatology, University of Cincinnati, and were a kind gift from Dr Susanne Wells. PBMCs used for reprogramming were obtained from the Stem Cell Processing Core at Cincinnati Children's Hospital.

Key messages

- **iPSCs from NECs are remarkably similar to ESCs, and NECs are excellent tissue sources to generate iPSCs.**
- **Residual DNA methylation markers from parental tissue persist in iPSCs, which are located in genes related to epithelial function and asthma.**

REFERENCES

1. Batista LF, Pech MF, Zhong FL, Nguyen HN, Xie KT, Zaug AJ, et al. Telomere shortening and loss of self-renewal in dyskeratosis congenita induced pluripotent stem cells. Nature 2011;474:399-402.
2. Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, et al. Modelling schizophrenia using human induced pluripotent stem cells. Nature 2011;473:221-5.
3. Soldner F, Hockemeyer D, Beard C, Gao Q, Bell GW, Cook EG, et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell 2009;136:964-77.
4. Pasca SP, Portmann T, Voineagu I, Yazawa M, Shcheglovitov A, Pasca AM, et al. Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. Nat Med 2011;17:1657-62.
5. Yoshida Y, Yamanaka S. Recent stem cell advances: induced pluripotent stem cells for disease modeling and stem cell-based regeneration. Circulation 2010; 122:80-7.
6. Spence JR, Mayhew CN, Rankin SA, Kuhar MF, Vallance JE, Tolle K, et al. Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. Nature 2011;470:105-9.
7. Yamanaka S. Induced pluripotent stem cells: past, present, and future. Cell Stem Cell 2012;10:678-84.
8. Hou P, Li Y, Zhang X, Liu C, Guan J, Li H, et al. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science 2013;341: 651-4.

9. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006;126:663-76.
10. Stadtfeld M, Nagaya M, Utikal J, Weir G, Hochedlinger K. Induced pluripotent stem cells generated without viral integration. Science 2008;322:945-9.
11. Zhou W, Freed CR. Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells. Stem Cells 2009;27:2667-74.
12. Shi Y, Do JT, Desponts C, Hahm HS, Scholer HR, Ding S. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell 2008;2:525-8.
13. Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen AE, et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol 2008;26:795-7.
14. Lin T, Ambasudhan R, Yuan X, Li W, Hilcove S, Abujarour R, et al. A chemical platform for improved induction of human iPSCs. Nat Methods 2009;6:805-8.
15. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, et al. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell 2011;144:439-52.
16. Guenther MG, Frampton GM, Soldner F, Hockemeyer D, Mitalipova M, Jaenisch R, et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell 2010;7:249-57.
17. Newman AM, Cooper JB. Lab-specific gene expression signatures in pluripotent stem cells. Cell Stem Cell 2010;7:258-62.
18. Boulting GL, Kiskinis E, Croft GF, Amoroso MW, Oakley DH, Wainger BJ, et al. A functionally characterized test set of human induced pluripotent stem cells. Nat Biotechnol 2011;29:279-86.
19. Ono M, Hamada Y, Horiuchi Y, Matsuo-Takasaki M, Imoto Y, Satomi K, et al. Generation of induced pluripotent stem cells from human nasal epithelial cells using a Sendai virus vector. PLoS One 2012;7:e42855.
20. Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell 2009;5:111-23.
21. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 2009;41:1350-3.
22. Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. PLoS One 2010;5:e8975.
23. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, et al. Epigenetic memory in induced pluripotent stem cells. Nature 2010;467:285-90.
24. Kim K, Zhao R, Doi A, Ng K, Unternaehrer J, Cahan P, et al. Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. Nat Biotechnol 2011;29:1117-9.
25. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 2011;471:68-73.
26. Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR. Transcriptional signature and memory retention of human-induced pluripotent stem cells. PLoS One 2009;4:e7076.
27. Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, et al. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat Cell Biol 2011;13:541-9.
28. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol 2010;28:848-55.
29. Bar-Nur O, Russ HA, Efrat S, Benvenisty N. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. Cell Stem Cell 2011;9:17-23.
30. Hu BY, Weick JP, Yu J, Ma LX, Zhang XQ, Thomson JA, et al. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. Proc Natl Acad Sci U S A 2010;107:4335-40.
31. Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, et al. The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature 2009;460:1136-9.
32. Miura K, Okada Y, Aoi T, Okada A, Takahashi K, Okita K, et al. Variation in the safety of induced pluripotent stem cell lines. Nat Biotechnol 2009;27:743-5.
33. Poole A, Urbanek C, Eng C, Schageman J, Jacobson S, O'Connor BP, et al. Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. J Allergy Clin Immunol 2014;133:670-8.e12.
34. Guajardo JR, Schleifer KW, Daines MO, Ruddy RM, Aronow BJ, Wills-Karp M, et al. Altered gene expression profiles in nasal respiratory epithelium reflect stable versus acute childhood asthma. J Allergy Clin Immunol 2005;115:243-51.
35. Karp PH, Moninger TO, Weber SP, Nesselhauf TS, Launspach JL, Zabner J, et al. An in vitro model of differentiated human airway epithelia. Methods for establishing primary cultures. Methods Mol Biol 2002;188:115-37.
36. Voelkel C, Galla M, Maetzig T, Warlich E, Kuehle J, Zychlinski D, et al. Protein transduction from retroviral Gag precursors. Proc Natl Acad Sci U S A 2010;107:7805-10.
37. Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, et al. A more efficient method to generate integration-free human iPS cells. Nat Methods 2011;8:409-12.
38. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic stem cell lines derived from human blastocysts. Science 1998;282:1145-7.
39. Suzuki T, Mayhew C, Sallese A, Chalk C, Carey BC, Malik P, et al. Use of induced pluripotent stem cells to recapitulate pulmonary alveolar proteinosis pathogenesis. Am J Respir Crit Care Med 2014;189:183-93.
40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.
41. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat Methods Med Res 2013;22:519-36.
42. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.
43. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 2010;11:94.
44. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.
45. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. Nat Biotechnol 2009;27:455-7.
46. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511-5.
47. Xu G, Deng N, Zhao Z, Judeh T, Flemington E, Zhu D. SAMMate: a GUI tool for processing short read alignments in SAM/BAM format. Source Code Biol Med 2011;6:2.
48. Longmire TA, Ikonomou L, Hawkins F, Christodoulou C, Cao Y, Jean JC, et al. Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. Cell Stem Cell 2012;10:398-411.
49. Allegrucci C, Young LE. Differences between human embryonic stem cell lines. Hum Reprod Update 2007;13:103-20.
50. Wong AP, Bear CE, Chin S, Pasceri P, Thompson TO, Huan LJ, et al. Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. Nat Biotechnol 2012;30:876-82.
51. Ruiz S, Diep D, Gore A, Panopoulos AD, Montserrat N, Plongthongkum N, et al. Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. Proc Natl Acad Sci U S A 2012;109:16196-201.
52. Planello ACJ, Sharma V, Singhania R, Mbabaali F, Müller F, Alfaro JA, et al. Aberrant DNA methylation reprogramming during induced pluripotent stem cell generation is dependent on the choice of reprogramming factors. Cell Regen 2014;3:4.
53. Krieg P, Schuppler M, Koesters R, Mincheva A, Lichter P, Marks F. Repetin (Rptn), a new member of the "fused gene" subgroup within the S100 gene family encoding a murine epidermal differentiation protein. Genomics 1997;43:339-48.
54. Huber M, Siegenthaler G, Mirancea N, Marenholz I, Nizetic D, Breitkreutz D, et al. Isolation and characterization of human repetin, a member of the fused gene family of the epidermal differentiation complex. J Invest Dermatol 2005;124:998-1007.
55. Islam T, McConnell R, Gauderman WJ, Avol E, Peters JM, Gilliland FD. Ozone, oxidant defense genes, and risk of asthma during adolescence. Am J Respir Crit Care Med 2008;177:388-95.
56. Fitzpatrick AM, Jones DP, Brown LA. Glutathione redox control of asthma: from molecular mechanisms to therapeutic opportunities. Antioxid Redox Signal 2012;17:375-408.
57. Comhair SA, Erzurum SC. Redox control of asthma: molecular mechanisms and therapeutic opportunities. Antioxid Redox Signal 2010;12:93-124.
58. Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, et al. Dissecting direct reprogramming through integrative genomic analysis. Nature 2008;454:49-55.
59. Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 2009;27:361-8.

60. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 2009;27:353-60.
61. Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, et al. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell 2008;133:250-64.
62. Dizier MH, Margaritte-Jeannin P, Madore AM, Esparza-Gordillo J, Moffatt M, Corda E, et al. The ANO3/MUC15 locus is associated with eczema in families ascertained through asthma. J Allergy Clin Immunol 2012;129:1547-53.e3.
63. Nelson WG, Sun TT. The 50- and 58-kdalton keratin classes as molecular markers for stratified squamous epithelia: cell culture studies. J Cell Biol 1983; 97:244-51.
64. Tong L, Corrales RM, Chen Z, Villarreal AL, De Paiva CS, Beuerman R, et al. Expression and regulation of cornified envelope proteins in human corneal epithelium. Invest Ophthalmol Vis Sci 2006;47:1938-46.
65. Mou H, Zhao R, Sherwood R, Ahfeldt T, Lapey A, Wain J, et al. Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs. Cell Stem Cell 2012;10:385-97.