Restriction of Cellular Plasticity of Differentiated Cells Mediated by Chromatin Modifiers, Transcription Factors and Protein Kinases

Dylan P. Rahe and Oliver Hobert

Department of Biological Sciences, Howard Hughes Medical Institute, Columbia University, New York, NY

ORCID IDs: 0000-0001-7727-5561 (D.P.R.); 0000-0002-7634-2854 (O.H.)

ABSTRACT

Ectopic expression of master regulatory transcription factors can reprogram the identity of specific cell types. The effectiveness of such induced cellular reprogramming is generally greatly reduced if the cellular substrates are fully differentiated cells. For example, in the nematode C. elegans, the ectopic expression of a neuronal identity-inducing transcription factor, CHE-1, can effectively induce CHE-1 target genes in immature cells but not in fully mature non-neuronal cells. To understand the molecular basis of this progressive restriction of cellular plasticity, we screened for C. elegans mutants in which ectopically expressed CHE-1 is able to induce neuronal effector genes in epidermal cells. We identified a ubiquitin hydrolase, usp-48, that restricts cellular plasticity with a notable cellular specificity. Even though we find usp-48 to be very broadly expressed in all tissue types, usp-48 null mutants specifically make epidermal cells susceptible to CHE-1-mediated activation of neuronal target genes. We screened for additional genes that allow epidermal cells to be at least partially reprogrammed by ectopic che-1 expression and identified many additional proteins that restrict cellular plasticity of epidermal cells, including a chromatin-related factor (H3K79 methyltransferase, DOT-1.1), a transcription factor (nuclear hormone receptor NHR-48), two MAPK-type protein kinases (SEK-1 and PMK-1), a nuclear localized O-GlcNAc transferase (OGT-1) and a member of large family of nuclear proteins related to the Rb-associated LIN-8 chromatin factor. These findings provide novel insights into the control of cellular plasticity.

KEYWORDS

C. elegans plasticity transcription factor cellular fate

During the development of multicellular organisms, pluripotent cells differentiate into terminal fates through the adoption of various characteristics and features necessary for the proper functioning of the terminal cell type. Concomitant with differentiation, cells also progressively lose their original plasticity. Observations of the existence of cell fate restriction accompanied the first descriptions of pluripotency itself in the late 19th century (Lensch and Mummery 2013). Classical embryology experiments involving the division of early newt and sea urchin embryos, in conjunction with subsequent embryonic tissue transplantation experiments by Hans Spemann and Hilde Mangold, were key in demonstrating both the plasticity of early embryonic cells as well as the progressive restriction of their fate as cells progressed through critical stages of development (Lensch and Mummery 2013). Pioneering somatic nuclear transplantation experiments further demonstrated the progressive loss of cell fate plasticity in differentiated cells (Briggs and King 1952; Gurdon 1960; Gurdon 1962). In these experiments, nuclei were removed from the endoderm of progressively later stage frog embryos and transplanted into enucleated eggs, with the number of fully developed tadpoles scored from each experiment. As the frog endodermal cells progressed further in development, they became increasingly restricted in their ability to return to a plastic state when transplanted.

In addition to embryological manipulation, the ectopic expression of critical cell-fate-specifying transcription factors (TFs) has also been used to study cell fate plasticity. The master regulator of muscle fate, MyoD, can induce the differentiation of multiple differentiated cell types into muscle cells (Weintraub et al. 1989). However, this TF is only able to do so in some cell types, while others are refractory to MyoD activity. In C. elegans, ectopic expression of the MyoD homolog hlh-1
also results in the expression of muscle markers, but this activity drops off as development progresses (Fukushige and Krause 2005). Similarly, the endoderm specifying TFs END-1, END-3, ELT-2 and ELT-7 can drive expression of intestinal reporter constructs when ubiquitously expressed in early development (Sommermann et al. 2010). However, by the L1 stage, many fewer cells express the reporter. Using similar tools, the same lab later showed ELT-7 to be capable of disrupting tissues in much later stages of development (Riddle et al. 2016). The transformations described are extensive but are restricted to a small subset of tissues in the animal, consistent with developmental fate restriction in most cell types.

Multiple factors have been shown to play a role in developmental fate restriction. In particular, genes involved in the deposition and maintenance of histone marks associated with transcriptional repression, namely H3K37 and H3K9 methylation, have been implicated. Removal of the PRC2 component mes-2/E(z) from the early embryo can extend the time period in which hh-like and end-1 ectopic expression results in aberrant adoption of muscle or endoderm fates, respectively (Yuzuyk et al. 2009). Our lab has studied this problem by investigating the effect of ectopic expression of the transcription factor CHE-1 (Tursun et al. 2011). che-1 is normally expressed solely in the bilaterally symmetric ASE neuron pair (ASEL and ASER) and is required for the proper differentiation and function of this pair of chemosensory neurons (Chang et al. 2003; Uchida et al. 2003; Etchberger et al. 2007). In addition to wild-type animals, ectopic expression of che-1 is able to induce the expression of an ASER-reporter transgene (gfp-5:gfp) in a manner that is progressively restricted during development (Patel and Hobert 2017) (this work). Using this system, we found the histone chaperone lin-53 to be critical in maintaining the cell fate restriction of germ cells (Tursun et al. 2011). Subsequent analysis found its role in the germline to be as part of complex with PRC2, as removal of components of this complex (MES-2/E(z), MES-3, and MES-6/ESC) resulted in loss of fate restriction in the germline (Patel and Hobert 2017). In the heat shock screen, synchronized late L4 larvae were mutagenized for four hours in 50mM ethyl methanesulfonate (EMS) in M9 media. Five F1s were plated to each plate, and each plate was heat shocked at 37°C for 30min, and left overnight at 25°C. A list of strains and transgenes is provided as Supplemental Table S1.

Mutagenesis

In the heat shock screen, synchronized late L4 larvae were mutagenized for four hours in 50mM ethyl methanesulfonate (EMS) in M9 media. Five F1s were plated to each plate, and each plate was heat shocked at 37°C for one hour when F2s reached the L4/Young Adult stage. In the epidermis-specific screen, animals were mutagenized as before, but each generation was synchronized by alkaline bleach treatment. After bleaching, F1 embryos were distributed among 10 150mm NGM plates. When synchronized F2s reached adulthood, animals were washed extensively and sorted using a COPAS Biosorter for animals with high GFP expression. >300,000 F2 animals were sorted from each of the ten F1 plates, and Ecta animals were singled for further analysis.

RNAi

Appropriate RNAi clones from Ahringer library (Kamath and Ahringer 2003) were grown O/N in LB+Carbenicillin, then 350 µL was plated onto NGM plates supplemented carbenicillin to 25 µg/ml and IPTG to 1mM. Once dried, 5-10 L4s were added. Phenotype was assessed in the following generation.

CRISPR/Cas9 genome engineering

GFP insertions were engineered using co-CRISPR, as described in Kim et al. 2014. Homologous repair template for usp-48 was a TOPO-cloned PCR-product from the GFP-tagged fosmid from the Transgeneomics Project (Sarov et al. 2012). GFP was inserted between the last codon and the stop codon of usp-48. The ogt-1 repair template was generated by PCR fusion using the codon-optimized GFP sequence from Dickinson et al. 2015 and cloned into the pMiniT2 vector (NEB). The dot-1::gfp strain (ot899) was generated using the SEC method (Dickinson et al. 2015).
Single-molecule FISH

smFISH was performed using Custom Stellaris™ FISH probes, purchased from Biosearch Technologies and staining was done according to the manufacturer’s protocol, with slight modifications. Briefly, animals were fixed in 10% formaldehyde in M9 for 30 min., then washed into 70% ethanol for 1-3 days at 4°C. Animals were resuspended in 200 μl hybridization buffer (10% Formamide, 10% Dextran Sulfate, 1 mg/mL E. coli RNA, 2 mg/mL Vanadyl ribonucleoside complex, 200 μg/mL BSA) to which probes were added to 5-10μM and incubated O/N at 37°C. After hybridization, worms were washed in 2x SSC, 10% Formamidine, incubated with 5 μg/mL DAPI, and mounted with antifade buffer for microscopy. Probes correspond to coding regions of gcy-5 or unc-10 cDNA (excluding UTRs).

Microscopy

μManager, a plugin for ImageJ (Fiji) was used for processing on Figure 1F, Figure 3, Figure 6B&C, and Figure 9. Fig.S2 (Stuurman et al. 2007). All other images were taken on a LSM 880 Confocal microscope and processed using Zeiss Zen software.

Tail epidermis ploidy measurements

Ploidy was assessed using DAPI staining (Lozano et al. 2006), with modifications. Specifically, day 1 adult hermaphrodite N2 animals were fixed in -20°C DMF and washed immediately for four times with PBS-0.1% Triton-X100. During the second wash, 5 ng/mL DAPI was added and incubated for 30 min. at 4°C, after which the final two washes were completed before mounting with Vectashield (Vector Labs). Animals were imaged using a Zeiss LSM 880 confocal microscope. Fluorescence intensity of individual nuclei was measured using ImageJ 2.0 software: Z-projection by sum-slices method, followed by freehand selection of nucleus, then Integrated Density using the measurement tool. Ploidy was estimated by normalizing to multiple neuronal nuclei from the tail region of each image.

Wormsorter operation

COPAS Biosort (Union Biometrica) was used according to Doitsidou et al. 2008. Briefly, the Wormsorter was washed, then unmutagenized animals were recorded using green fluorescence (488nm exc. wavelength, GFP filter) and time-of-flight, a measure of the size of the sorted object. Objects of similar size as wild-type but with much higher levels
of green fluorescence were sorted to plates, and a fluorescence dissecting scope was used to isolate individual Ecta mutants to new plates.

**Mapping-by-Sequencing**

Mutants were mapped according to Doitsidou et al. 2010. Briefly, mutants are crossed to a strain (CB4856) that is polymorphic at known loci. F2 animals from this cross are then screened for the Ecta phenotype, and several (>15) are single and allowed to grow to a large population size. These plates are then pooled together and genomic DNA extracted for sequencing.

**WGS library preparation, sequencing and bioinformatic analysis**

_mutants_ were sequenced using 150 bp paired end reads on an Illumina HiSeq 2500 at BGI (Beijing Genomics Institute). All other sequencing libraries were prepared using Nextera tagmentation kits (Illumina) and sequenced using an Illumina NextSeq machine using either 75bp or 150bp read length. For the sequence analysis, FASTQ files were aligned to WS220 genome using _bwa-aln_ (Li and Durbin 2009). Resulting bam files were sorted and indexed using _samtools_ (Li et al. 2009). For SNV mapping, variant SNV alleles were called using the GATK “Best Practices” pipeline: duplicate reads are called with _picard_ (http://broadinstitute.github.io/picard/), and GATK package was used both to recalibrate base scores according to known mapping variants (-T BaseRecalibrator) and call variants (-T HaplotypeCaller), both for known variant alleles (for mapping,-genotyping_mode GENOTYPE_GIVEN_ALLELES) and novel variants (for mutant analysis,-genotyping_mode DISCOVERY) (Depristo et al. 2011). A custom R script was used to plot the observed frequency of known SNVs vs. genome position to determine linkage. Allele frequency for each known variant was plotted independently for each chromosome vs. physical chromosome position, and a Loess regression line was fitted to the data. WGS without SNV mapping was performed using GATK without base recalibration. After a HaplotypeCaller run, variants were subtracted from other variants to create exclusive variant lists for each mutant using _bcftools isec_ (Danecék et al. 2011; Narasimhan et al. 2016). With or without mapping, characterization of the effects of variants was performed using _ReSeq_ gene structures and _SnpEff_ (Cingolani et al. 2012). Scripts used in bioinformatics are available upon request.

**Data availability**

Strains will be deposited at the Caenorhabditis Genetics Center (CGC) and are available upon request. Supplemental material available at FigShare: https://doi.org/10.25387/g3.8114237.

**RESULTS AND DISCUSSION**

A genetic screen for cell fate plasticity reveals the ubiquitin hydrolase usp-48

We have previously established a powerful assay system to study the restriction of cellular plasticity (Tursun et al. 2011; Patel and Hobert 2017). Misexpression of the _CHE-1_ transcription factor, a master regulator of ASE neuron identity, under control of a ubiquitously expressed heat-shock promoter in embryonic or early larval stages is capable of inducing ASE identity features in other cell types, while misexpression in late larval or adult stages does not result in broad ectopic activation of ASE identity markers (Patel and Hobert 2017). We recapitulated these previously reported results by demonstrating that a short induction of _CHE-1_ expression from the _otIs305_ array at early larval stages results in many tissues inducing ectopic expression of _gcy-5_, a direct target of _che-1_ (Figure 1A,B). Heat shock at later stages show a progressive restriction of the induction of _gcy-5:gfP_ in other tissues (Figure 1A,B). We sought to understand the underlying genetic basis of the progressive restriction of cellular plasticity by moving beyond our previous, candidate RNAi screens and instead screened in an unbiased manner for EMS-induced mutants in which ectopic induction of _che-1_ expression in adult animals will permit induction of _che-1_ target gene expression (schematized in Figure 1C).

A pilot, non-clonal mutagenesis screen identified several sterile mutants with ectopic _gcy-5:gfP_ expression after _che-1_ induction at adult stages, but we were not able isolate heterozygous parents to stably propagate these mutants. We therefore switched to a semi-clonal screen in which 5 F1s from a mutagenized P0 population were single into each screening plate. A total of 7302 haploid genomes were screened and 36 mutants were recovered. All mutants were sterile, and were maintained as heterozygotes. 35 of the 36 mutants displayed ectopic _gcy-5:gfP_ in the germline. Since several such “germline-reprogrammed” mutants had been described before (Tursun et al. 2011; Patel et al. 2012; Kolundzic et al. 2018; Hajduska et al. 2019), no further analysis of this class of mutants was performed. We rather focused on the only mutant that exhibited a phenotype in somatic tissue, _ot674_. In _ot674_ mutant animals, heat shock induction of _che-1_ at any stage resulted in expression of _gcy-5:gfP_ specifically in the epidermis of the animals (Figure 1E). We refer to this phenotype as the Ecta (Epidermal CHE-1 Target Activation) phenotype throughout this paper. The Ecta phenotype is specific to _CHE-1_ activity, as neither heat shock in the absence of _hsp:che-1_ nor heat shock without the transgene results in ectopic expression of _gcy-5:gfP_ in _ot674_ mutants.

In an RNAi screen for chromatin factors whose depletion may permit _che-1_ target gene induction (Tursun et al. 2011), we had observed, but not previously reported, that RNAi of one predicted chromatin factor, _usp-48_, resulted in an Ecta phenotype. We did not follow up this result due to the highly variable nature of the RNAi phenotype. However, after retrieval of the _ot674_ mutant, we decided to sequence the _usp-48_ locus in these animals through conventional Sanger sequencing and indeed found a nonsense mutation in the third exon of _usp-48_. The mutant was backcrossed and balanced with _ht2_, a chromosomal translocation between the first and third chromosome (McKim et al. 1988). In addition, a nonsense allele (mgI57) previously isolated in the Ruv-kun Lab for a different phenotype was obtained and was found to phenocopy _ot674_ (Figure 1F). Yet another allele of _usp-48_ was retrieved from another screen for the Ecta mutant phenotype, described in more detail below. The _usp-48_ Ecta phenotype could be rescued with a wild-type copy of the gene, expressed under control of an epidermis-specific promoter. We conclude that _usp-48_ is the causative mutation for the Ecta phenotype.

_usp-48_ encodes the previously uncharacterized _C. elegans_ ortholog of the human ubiquitin hydrolase USP48 (Komander et al. 2009) (Figure 2). Its function is not well understood, but recent studies on the mammalian homolog have suggested multiple roles. Two independent studies performing Chromatin Immunoprecipitation-Mass Spectrometry (ChIP-MS) on multiple histone marks coprecipitated USP48 with anti-H3K4me3, H3K4me1, H3K27ac, H3K79me2, and H3K36me3 antibodies (Engelen et al. 2015; Ji et al. 2015). All of these histone marks are associated with active transcription or promoter/enhancer activity (reviewed in (Gates et al. 2017)). This is notable because previously identified mutants with cell fate restriction phenotypes were mostly implicated in gene repression (Tursun et al. 2011; Patel et al. 2012; Kolundzic et al. 2018; Hajduska et al. 2019). One exception is the
Assessing the effect of che-1 and usp-48 on endogenous gene expression

We sought to assess the expression of CHE-1 targets in the epidermis of usp-48 mutants by measuring endogenous mRNA using single-molecule fluorescence in situ hybridization (smFISH) (Raj et al. 2008), rather than transgenic reporters. In this method, fluorescence-labeled DNA oligonucleotides tiling an individual mRNA are hybridized in fixed animals, and individual mRNAs are thus observed as individual puncta. We generated smFISH probes to two ASE-expressed genes, *gcy-5* (a direct transcriptional target of CHE-1 protein) and *unc-10*, a pan-neuronally expressed synaptic protein (Koushika et al. 2001). Both genes are ectopically activated by che-1 in a germ cell reprogramming paradigm (Tursun et al. 2011). We confirmed specificity of the smFISH probes by showing that fluorescent signals are not observed in *gcy-5* and *unc-10* deletion mutants, respectively, but detectable in wild-type animals with expected expression patterns (Figure 3A,B). In usp-48(ot674) mutants, but not heterozygous controls, punctate signals of the *unc-10* and *gcy-5* mRNAs were clearly visible within the epidermis after ectopic CHE-1 induction (Figure 3C). Thus, not only is a *gcy-5*-promoter-driven transgene activated, but also the endogenous locus itself, along with another marker of broad neuronal fate.

Postdevelopmental requirement of usp-48 for restricting cellular plasticity

A temperature-sensitive allele of usp-48 that we retrieved from another screen, described below, allowed us to assess whether usp-48 activity is required during the initial differentiation of the hypoderms or whether it is required in later, postdevelopmental stages to prevent che-1 from inducing its targets genes. This temperature-sensitive allele, ot872, contained a point mutation in the second exon (Figure 2). At 25°C, these animals exhibit some sterility and display 100% penetrant Ecta phenotype. At 20°C, ot872 animals are fertile but still display 100% penetrant Ecta phenotype. At 15°C, these animals are fertile and show a significantly reduced penetrance of the Ecta phenotype (Figure 4). To access the postdevelopmental requirement of usp-48, we grew ot872 animals at 15°C and shifted them to 20°C or 25°C at the L4 stage. This temperature resulted in a 100% Ecta phenotype, reminiscent of animals raised continuously at 20 or 25°C (Figure 4). These data suggest that USP-48 is required postdevelopmentally and acutely to restrict the plasticity of epidermal cells and prevent ectopic CHE-1 from being able to act on its targets.

USP-48 protein is nuclear localized and very broadly expressed

In order to further investigate possible roles of usp-48, an endogenous translational C-terminal GFP reporter was built using CRISPR/Cas9-mediated genome engineering (Figure 5A). The resultant usp-48 allele exhibited nuclear GFP expression very broadly at all stages of development (Figure 5B-D). In pachytene germline mes-4 gene, which restricts CHE-1 activity in the germline, mes-4 deposits methylation on H3K36, a mark correlated with actively transcribed genes (Patel et al. 2012). H3K36 methylation plays a role in the proper distribution of the repressive H3K27 methylation mark (Gaydos et al. 2012; Patel et al. 2012).

**Figure 2** usp-48 encodes a conserved ubiquitin hydrolase. A: USP-48 contains a conserved C19 peptidase domain as well as a ubiquitin-like (Ubl) domain at the C-terminus. Multiple alleles (ot872, ot674, and mg457) have been isolated and exhibit the Ecta phenotype. B: The C19 domain contains four catalytic residues, all of which are conserved in mouse and human orthologs (blue boxes), but the C-terminal Ubl domain is less well conserved in *C. elegans*, such that it is not recognized by domain searching algorithms (SMART, InterProScan).
Figure 3 Induction of endogenous neuronal genes gcy-5 and unc-10 upon ectopic che-1 expression in a usp-48 mutant background. A,B: gcy-5 (panel A) and unc-10 (panel B) mRNA smFISH probes are gene specific, since staining of the respective deletion mutant animals revealed no puncta. Scoring of punctae is quantified in the right panels. C: Ectopic expression of che-1 induces endogenous gcy-5 and unc-10 mRNA expression in the epidermis of usp-48 mutants. After heat-shock induction of che-1 (hs > che-1), gcy-5 and unc-10 mRNA are induced in usp-48 mutant epidermis (as assessed by smFISH), but not heterozygous controls.
we observe other tissue types. However, we can rule out this possibility, because *usp-48* and *endoreduplication* (Hedgecock and White 1985). One could envision during larval development and adulthood, a process known as worms is that epidermal nuclei continue to replicate their DNA that make their nuclear biology different from other cells in the animals are nearly completely penetrant Ecta.

The expression of *usp-48* with a temperature sensitive allele. At 15°C, *usp-48* is ~60% Ecta, but at 20°C or higher the penetrance is ~100%. When shifted at the L4 stage, before che-1 expression begins, to either 20°C or 25°C, the animals are nearly completely penetrant Ecta.

cells and oocytes, clusters of brighter expression can be observed, consistent with chromatin binding (Figure 5C,D'). The only observed cells not exhibiting nuclear expression were newly fertilized zygotes, until the 4- or 8-cell embryo, when nuclear expression can be observed again (Figure 5D). The absence of GFP signal in early embryos prompted us to ask if *usp-48* mRNA was maternally inherited. To this end, progeny from a *trans*-heterozygote of this CRISPR allele and *ot674* were generated. If *usp-48* is maternally inherited, 100% of resultant embryos should display GFP expression. This was the case, with fluorescence persisting to a lesser extent into the L1 stage. Thus, there exists maternal contribution of the mRNA despite loss of GFP signal in the early embryonic nuclei (Figure 5E).

The expression of *usp-48* in all cell types, combined with the ubiquitous induction of che-1 in our protocol, raises the intriguing question as to why it is only the epidermis that becomes susceptible to che-1-misexpression. *usp-48* does not appear to display selective defects in epidermal differentiation. At the young adult stage, when *usp-48* mutants display the Ecta phenotype, the epidermis of *usp-48* mutants looks superficially normal by light microscopy and animals are not Dumpy or Blistered (common phenotypes associated with epidermal defects) and have normal adult alae. Moreover, we examined the expression of two collagens, terminal markers of epidermal identity, *dpy-7* and *col-93*, and found no effects of *usp-48* on the expression of these genes. One feature of many *C. elegans* epidermal cells that make their nuclear biology different from other cells in the worms is that epidermal nuclei continue to replicate their DNA during larval development and adulthood, a process known as endoreduplication (Hedgecock and White 1985). One could envision that this could lead epidermal chromatin to be more “susceptible” to *usp-48* and che-1-dependent gene expression changes, compared to other tissue types. However, we can rule out this possibility, because we observe che-1 target activation in *usp-48* mutants upon che-1 misexpression in non-syncytial epidermal cells of the tail, which do not endoreduplicate (Figure 6A,B). Other cell types that endoreduplicate (e.g., the intestine) also show no *gcy-5* induction, corroborating that endoreduplication alone cannot explain the phenotype. We furthermore note that since tail epidermal cells do not fuse with other epidermal cells, as do the epidermal cells derived from dividing seam cells during larval development (Hedgecock and White 1985), we can also rule out that processes related to cell fusion may explain the particular sensitivity of epidermal cells to che-1-dependent, ectopic gene activation in *usp-48* mutants. Consistent with this notion, in *eff-1* mutants, where all epidermal cells fail to fuse (Mohler et al. 2002), che-1 is still able to activate *gcy-5::gfp* in *usp-48* mutants (Figure 6C). One possibility that could explain the epidermal specificity and which we cannot completely rule out is that the expression levels achieved by the heat shock promoter may be higher in the epidermis than in other tissues.

**A different screen that probes specifically for epidermal cell fate plasticity reveals a large set of plasticity restriction mutants**

In order to perhaps better understand *usp-48* function, we set out to screen for mutants that phenocopy the Ecta phenotype of *usp-48*. The isolation of so many sterile mutants (that we did not further pursue, as mentioned above), made us concerned that our heat-shock based approach to induce che-1 expression may make animals unnecessarily sick. To avoid this issue, we designed a different screening approach. We constructed a new transgene (*ot15642*) that expresses che-1 under the control of the promoter of *col-19*, which drives expression from the L4/Adult molt onwards specifically in the epidermis (Cox and Hirsh 1985; Liu et al. 1995). In contrast to a transgene that expresses che-1 throughout all developmental stages in the epidermis (*dpy-7prom::che-1*) and which results in the ectopic induction of *gcy-5::gfp* expression in the epidermis (Patel and Hobert 2017), this transgene does not induce *gcy-5::gfp* expression (Figure 7A, C). The different effects of these two transgenes corroborate the phenomenon of plasticity restriction in the adult epidermis. Using this *col-19prom::che-1* transgene, we conducted a non-clonal screen with the aid of a Wormsorter, a FACs machine specialized for sorting of entire animals (Doitsidou et al. 2008). In order to reduce the possibility of isolating siblings from the same mutant F1, we separated F1s onto ten different plates. Thus, individual Ecta animals from different screening plates are assured to be independent alleles (Figure 7B). This approach indeed yielded a large number of completely fertile mutants. All but one plate contained multiple Ecta mutants that were fertile and bred true. Of those that bred true, we were able to maintain 23 strains that were both highly penetrant and expressive.

The 23 mutant strains were first screened for mutations in the *usp-48* locus by Sanger sequencing. One of them turned out to be the temperature-sensitive *usp-48* allele, *ot872*, that we already described above. 14 strains that harbored no mutation in *usp-48* were selected for whole genome sequencing (WGS) in combination with SNP-based mapping, using a procedure we previously described (Doitsidou et al. 2010). Each allele was mapped to a specific region through such linkage analysis, and through a combination of complementation testing, transgene rescue, and/or RNAi phenocopy experiments we were able to map nine of these alleles to five distinct additional genes, as described below. The remaining four alleles have been mapped to a chromosomal region but remain uncloned. The remaining eight mutant strains retrieved from the screen were subsequently whole genome-sequenced without any mapping, which revealed (a) two additional alleles of the genes that we mapped in parallel,
(b) four strains were revealed to be siblings of already identified mutations, and (c) two alleles of a gene (sek-1) whose function is intimately related to one of the genes that we first cloned (pmk-1; explained further below). In summary, 23 strains were analyzed, of which 19 were confirmed to be independent alleles, 15 of which correspond to seven genes that we molecularly identified, and four mutants remain uncloned. Relevant alleles are summarized in Table 1 and representative micrographs of these mutants are shown in Figure 7C. Cloning information is summarized in Table 2, mapping information and gene structures and locations of mutations are shown in Figure 8. The following genes were identified by this screen.

**dot-1.1/DOT1L**, an H3K79 methyltransferase, restricts cellular plasticity in the epidermis

The ot885 allele was mapped to the left arm of chromosome I and found to contain a missense mutation in *dot-1.1* which codes for one
of five *C. elegans* orthologs of the phylogenetically conserved H3K79 histone methyltransferase DOT1 (Feng et al. 2002). An available nonsense mutation and RNAi against *dot-1.1* phenocopied the defects observed in *ot885* animals (Table 2). As with *usp-48*, we find that the ectopic expression of the *gcy-5* reporter transgene after *che-1* induction in *dot-1.1* mutants is an accurate reflection of endogenous *gcy-5* transcription since we observed endogenous *gcy-5* induction by smFISH analysis (Figure 9).

The retrieval of *dot-1.1* from this screen and its phenotypic similarity to *usp-48* mutants is notable because in vertebrates, chromatin immunoprecipitation with antibodies raised against the H3K79 mark deposited by DOT-1.1 pulls down vertebrate USP48 protein (Engelen et al. 2015; Ji et al. 2015). *dot-1.1/DOT1L* is a unique methyltransferase both in that it does not methylate a histone tail, but rather a residue residing within the core globular domain of the histone H3, and that it uses a domain other than SET to catalyze this activity (Wood et al. 2018). Despite its original characterization as being required for silencing of telomeres, H3K79 methylation by Dot1 and its homologs is most closely associated with actively transcribed genes (Wood et al. 2018), as is USP48 by virtue of its physical association with histone marks associated with active genes (Engelen et al. 2015; Ji et al. 2015).

**ogt-1/OGT, the sole intracellular O-GlcNAc Transferase, also restricts cellular plasticity in the epidermis**

Another mutant allele causing an Ecta phenotype, *ot878* (Figure 7C), mapped to chromosome III and contained a missense mutation in *ogt-1* (Figure 8). A null allele of this gene as well as RNAi resulted in the same Ecta phenotype. Two additional missense alleles of *ogt-1* (Figure 8, Table 1) were identified by sequencing two independently isolated mutants in the screen. As with *usp-48* and *dot-1.1*, we find that the ectopic expression of the *gcy-5* reporter transgene upon *che-1* expression in *ogt-1* mutants is an accurate reflection of endogenous *gcy-5* transcription since we observed endogenous *gcy-5* induction by smFISH analysis (Figure 9).

*ogt-1* encodes the sole *C. elegans* ortholog of O-GlcNAc Transferase (OGT), an enzyme responsible for the addition of a single N-acetyl
Glucosamine (GlcNAc) carbohydrate moiety to a serine or threonine hydroxyl group (reviewed in Levine and Walker 2016). While the vast majority of glycosylation occurs in the endoplasmic reticulum or Golgi, OGT is the sole example of intracellular glycosylation known. OGT has thousands of cytoplasmic and nuclear targets, and is thought to achieve specificity by protein-protein interactions with non-catalytic tetra-tripeptide repeats that are present in the N-terminal half of the protein. OGT mutants are lethal in mammals, zebrafish, and flies, but, surprisingly, ogt-1 deletion alleles are viable in C. elegans (Hanover et al. 2005) (this study). OGT in Drosophila was originally identified as a Polycomb-group gene super sex combs, linking its activity to transcriptional regulation (Sinclair et al. 2009). Some evidence has been reported for the existence of O-GlcNAc modifications of histones H2A, H2B, H3 and H4 (Sakabe et al. 2010; Zhang et al. 2011). However, other studies have found no evidence for this modification on histones (Gagnon et al. 2015). What is less controversial is the existence of multiple TFs and even Pol II as substrates of OGT (Nagel and Ball 2014; Lewis et al. 2016). Either indirectly through TFs or directly, OGT and O-GlcNAc has been found to associate with chromatin, mainly upstream of highly active promoters in both worms and mammals (Love et al. 2010; Deplus et al. 2013). Hence, to the extent that OGT has been shown to associate with chromatin, it is mainly correlated with active transcription, as observed for USP48 and DOT1L.

pmk-1/p38\(\alpha\) and sek-1/MMK3/6, members of a conserved stress-signaling pathway, restrict cellular plasticity in the epidermis

Mapping to the same region of chromosome IV, the Ecta mutant alleles \(ot886\) and \(ot887\) contained missense mutations in the pmk-1
One of the mutant allele was mapped to the right arm of LGX and found to harbor a mutation in a gene, C08A9.6, which we named lido-1 (for LIN-8 domain containing). The causative nature of the mutant allele is suggested by phenocopying the effect of this mutation by RNAi against lido-1 (Table 2). lido-1 is one of 16 genes in the C. elegans genome whose protein shares a central domain of homology (Interpro accession number IPR005020) with the previously described LIN-8 protein (Davison et al. 2005)(Figure 10). We have named all other genes that contain this domain of homology lido as well. In most of these proteins this domain encompasses almost the entire protein. Two additional C. elegans genes are likely pseudogenes with premature stop codons; without such codon their protein products would also contain a LIN-8 domain. Most lido genes can be found in two clusters, one on the X chromosome (containing 6 directly adjacent lido genes) and one on the second chromosome (containing 7 lido genes plus lin-8) (Figure 10A).

Intriguingly, by inspecting other genome sequences, we find that other Caenorhabditis species have significantly expanded or retracted their complement of lido genes. C. remanei encodes 88 lido genes, while C. japonica only contains only 5 lido genes (Figure 10B). Equally intriguingly, in other Caenorhabditis species, the encoded proteins have acquired additional domains, most of them involved directly in DNA or chromatin-related activities. For example, 21 of the C. remanei LIDO proteins have MADF DNA binding domains and a total of 8 directly adjacent lido genes and one on the second chromosome (containing 7 lido genes plus lin-8) (Figure 10A).

Outside the Caenorhabditidae, only two species contain this domain of homology, a protein from a single Cyanobacteria species (Moorea producens) and a single protein from the ciliate Pseudocochliometrum persalinus. According to the Interpro database, the LIN-8 domain of the ciliate protein displays also some homology to a SET domain, suggesting that the LIN-8 domain and the SET domain may share some evolutionary relationship. The homology of LIDO-1 to LIN-8, a class A SynMuv gene, is a further indication of a function of LIDO-1 in chromatin biology. Synthetic multivulva (SynMuv) genes were identified as being redundantly required (Synthetic) for the inhibition of let-60/Ras signaling during the formation of the vulva (Fay and Yochem 2007). SynMuv mutants in each class correspond to groups of interacting genes, and of the two major classes, the SynMuvB class...
contains homologs of DP, E2F, and Rb, several of which have been suggested to form various transcriptionally repressive complexes, including the DRM/DREAM and NuRD complexes (Fay and Yochem 2007). SynMuvA class genes, however, which include lin-15A, lin-8, lin-38, and lin-56, encode novel nuclear proteins with no non-nematode orthologs (Davison et al. 2005; Fay and Yochem 2007; Davison et al. 2011). LIN-8 has been shown to physically interact with LIN-35/Rb transcriptional repressor (a SynMuvB gene) (Davison et al. 2005).

While the specific function of the LIN-8 domain is not known, it is possible that LIDO-1, similar to LIN-8, interacts with other nuclear genes involved in transcriptional repression, or possibly LIN-35/Rb itself, given the extensive homology between LIDO-1 and LIN-8 in the region deemed minimally necessary for LIN-8-LIN-35 binding (Davison et al. 2005). It is possible that LIDO-1 interacts with some of the other genes identified in this screen. A link between LIDO-1 and SynMuv genes is particularly intriguing because usp-48 mutants, which phenocopy lido-1 mutants, have previously been shown to genetically interact with the SynMuv machinery (Cui et al. 2006).

**nhr-48, a nuclear hormone receptor gene, restricts cellular plasticity**

Two other mutants with the Ecta phenotype, ot880 and ot881, were mapped to the same region on the X chromosome, and both contained different missense mutations in the gene nhr-48 (Figure 7). A genetic complementation cross showed they were alleles of the same gene. In addition, another mutant strain in our collection of Ecta mutants harbored a premature stop codon in the nhr-48 locus. All nhr-48 mutant strains are viable and fertile with no obvious pleiotropies. As with the other genes described above, we find that the ectopic expression of the gcy-5 reporter transgene upon che-1 expression in nhr-48 mutants is an accurate reflection of endogenous gcy-5 transcription since we observed endogenous gcy-5 induction by smFISH analysis (Figure 9).

*nhr-48* is a member of the family of nuclear hormone receptors (NHR), a class of TFs that contain an N-terminal DNA-binding
domain (DBD) and a C-terminal ligand-binding domain (LBD). In *C. elegans*, the NHR gene family has undergone a dramatic expansion, with 284 NHR genes compared to 48 and 21 for humans and flies, respectively (Sluder et al. 1999; Maglich et al. 2001). While most of these expanded genes have no homologs in humans or other non-nematode species, *nhr-48* is a member of the well-conserved NR1 family of NHRs, which also contain the DAF-12 and NHR-8 gene. This family includes the mammalian Vitamin D Receptor (VDR) and Drosophila Ecdysone Receptor (EcR) (Maglich et al. 2001). Despite these similarities, *nhr-48* lacks significant conservation in the LBD region, and may not contain a functional ligand binding domain (Lu et al. 2016). The only known role for *NHR-48* in *C. elegans* is to act as a transcriptional repressor in a subset of pharyngeal gland cells (Ghai et al. 2012).

**Genes that restrict epidermal cellular fate plasticity are broadly expressed**

We generated translational C-terminal GFP fusion constructs for a subset of the genes described above. For *dot-1.1* and *ogt-1*, we tagged the endogenous locus with *gfp*, using CRISPR/Cas9-mediated genome engineering, while *nhr-48* expression was examined using a fosmid-based transgene from the TransgeneOme project (Figure 11A) (Sarov et al. 2012). In all three cases, each reporter showed very broad if not ubiquitous, nuclear expression throughout development (Figure 11B), as we had also observed with *usp-48* (Figure 4). Similarly, the *pmk-1* gene had previously been shown to be very broadly, if not ubiquitously expressed (Berman et al. 2001; Pagano et al. 2015). Not only do all these proteins appear to be ubiquitously expressed, but they also all localize to the nucleus (Figure 2, Figure 11). In the case of *pmk-1*, its gene product is reported to translocate to the nucleus upon activation by certain types of stress (Berman et al. 2001; Pagano et al. 2015).

**CONCLUSIONS**

We have described here a collection of mutants that permit the ectopically expressed master regulator of ASE neuron identity, *che-1*, to induce ASE-like features in other cell types that are normally refractory to ectopic *che-1* expression. These genes therefore can be considered to normally restrict the plasticity of cellular identity. Several of the genes identified here encode proteins involved in transcription and chromatin modification and based on studies in other organisms, some functional and/or physical links appear to exist between these proteins. We propose that these proteins either directly or indirectly establish a specific chromatin landscape that makes epidermal cell refractory to aberrant gene activation. Hence *che-1* is normally not able to activate its target genes, but can do so upon removal of these genes.

---

**Figure 10** Lido protein family. A: Two genomic clusters of *C. elegans* *lido* genes on LGII and LGX. B: Domain organization of LIDO proteins. C: Phylogenetic tree of nematode LIDO proteins. Gray and black triangles indicate *lin-8* and *lido-1*, respectively.
A striking aspect of our findings is that all proteins that we have analyzed here are very broadly, if not ubiquitously expressed nuclear proteins, but the loss of at least one these proteins, USP-48, appears to confer cellular plasticity only to epidermal cells. Using the same che-1 misexpression assay that we employ here, ogt-1 mutants were very recently shown to confer cellular plasticity also to germ cells, in which OGT-1 functionally interacts with the chromodomain protein MRG-1 (Hajduskova et al. 2019). The plasticity phenotype of mrg-1 mutants is, in turn, restricted to germ cells (Hajduskova et al. 2019). Given that we have previously shown that terminal differentiation of specific neuron types renders cells refractory to ectopic transcription factor-mediated reprogramming, the simplest explanation for the cellular specificity of these plasticity mutants could be that these factors, in spite of their broad expression, have a role in specifying the terminally differentiated state of only specific cell types, such as epidermal cells in the case of usp-48. However, none of the mutants that we analyzed appears to have any obvious defects in epidermal cell differentiation, as deduced by overall intact morphology and marker gene expression. Nevertheless, loss of the genes described here must transform the gene regulatory state of epidermal cells in a manner that makes these cells more "plastic". The normal function of these genes may therefore lie in ensure robustness of the regulatory state of these cells.

ACKNOWLEDGMENTS

We thank Chi Chen for injections, Gabriel Hayes and Gary Ruvkun for mg157, Iden Sapse for help with smFISH and Tulsi Patel for both help with the original screen and comments on the manuscript. This work was supported by the National Institutes of Health (NRSA Kirschstein predoctoral fellowship F31NS080435 to DR and R21NS076191-01 to OH). OH is an Investigator of the Howard Hughes Medical Institute.
LITERATURE CITED

Berman, K., J. McKay, L. Avery, and M. Cobb, 2001 Isolation and character- ization of pmk-1(33); three p38 homologs in Caenorhabditis elegans. Mol. Cell Biol. Res. Commun. 4: 337–344. https://doi.org/10.1006/ mcrb.2001.0300

Brenner, S. 1974 The genetics of Caenorhabditis elegans. Genetics 77: 71–94.

Briggs, R. and T. J. King, 1952 Transplantation of Living Nuclei From Blastula Cells into Encuclated Frogs’ Eggs. Proc. Natl. Acad. Sci. USA 38: 455–463. https://doi.org/10.1073/pnas.38.5.455

Chang, S., R. J. Johnston Jr., and O. Hobert, 2003 A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of C. elegans. Genes Dev. 17: 2123–2137. https://doi.org/10.1101/ gadt.1117903

Cheung, W. D., and G. W. Hart, 2008 AMP-activated protein kinase and glucose deprivation. J. Biol. Chem. 283: 13009–13020. https:// doi.org/10.1074/jbc.M801222200

Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen et al., 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnPeff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6: 80–92. https://doi.org/10.4161/ fly.19695

Cox, G. N., and D. Hirsh, 1985 Stage-specific patterns of collagen gene expression during development of Caenorhabditis elegans. Mol. Cell. Biol. 5: 363–372. https://doi.org/10.1107/MCB.5.2.363

Cui, M., E. B. Kim, and M. Han, 2006 Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in C. elegans. PLoS Genet. 2: e74. https://doi.org/10.1371/journal.pgen.0020074

Danecek, P., A. Auton, G. Abecasis, A. C. Albers, E. Banks et al., 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Engelen, E., J. H. Brandsma, M. J. Moen, L. Sognirle, D. H. Dekkers et al., 2015 Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nat. Commun. 6: 7155. https://doi.org/10.1038/ncomms8155

Engelen, E., J. H. Brandsma, M. J. Moen, L. Signorile, D. H. Dekkers et al., 2015 Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nat. Commun. 6: 7155. https://doi.org/10.1038/ncomms8155

Etchberger, J. F., A. Lorch, M. C. Sleumer, R. Zapf, S. J. Jones et al., 2007 The molecular signature and cis-regulatory architecture of a C. elegans gustatory neuron. Genes Dev. 21: 1653–1674. https://doi.org/10.1101/gad.1560107

Fay, D. S., and J. Yochem, 2007 The SynMuv genes of Caenorhabditis elegans in vulval development and beyond. Dev. Biol. 306: 1–9. https://doi.org/10.1016/j.ydbio.2007.03.016

Feng, Q., H. Wang, H. H. Ng, H. Erdjument-Bromage, P. Tempst et al., 2002 Methylation of H3-Klysine 79 is mediated by a new family of HMTases without a SET domain. Curr. Biol. 12: 1052–1058.

Ferreiro, L., M. Barragan, A. Gubern, E. Ballestar, M. Joaquin et al., 2010 The p38 SAPK is recruited to chromatin via its interaction with transcription factors. J. Biol. Chem. 285: 31819–31828. https:// doi.org/10.1074/jbc.M110.155846

Fukusige, T., and M. Krause, 2005 The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early C. elegans embryos. Development 132: 1795–1805. https://doi.org/10.1242/dev.01774

Gagnon, J., S. S. Duou, N. Zamorano, N. V. Jannantuomo, I. Hammond-Martel et al., 2015 Undetectable histone O-GlcNAcylation in mammalian cells. Epigenetics 10: 677–691. https://doi.org/10.1080/ 15592294.2015.1060387

Gates, L. A., C. E. Foulds, and B. W. O’Malley, 2017 Histone Marks in the ‘Driver’s Seat’: Functional Roles in Steering the Transcription Cycle. Trends Biochem. Sci. 42: 977–989. https://doi.org/10.1016/ j.tibs.2017.10.004

Gaydos, L. J., A. Rechtsteiner, T. A. Egelhofer, C. R. Carroll, and S. Strome, 2012 Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. Cell Reports 2: 1169–1177. https://doi.org/10.1016/j. celprep.2012.09.019

Ghai, V., R. B. Smit, and J. Gaudet, 2012 Transcriptional regulation of HLH-6-independent and subtype-specific genes expressed in the Caenorhabditis elegans pharyngeal glands. Mech. Dev. 129: 284–297. https://doi.org/10.1016/j.med.2012.06.005

Gurdon, J. B., 1960 The developmental capacity of nuclei taken from differentiating endoderm cells of Xenopus laevis. J. Embryol. Exp. Morphol. 8: 505–526.

Gurdon, J. B., 1962 The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J. Embryol. Exp. Morphol. 10: 622–640.

Hajduska, M., G. Baytek, E. Kolundzic, A. Goolschan, M. Kazmierczak et al., 2019 MRG-1/MRG15 Is a Barrier for Germ Cell to Nerve Reprogramming in Caenorhabditis elegans. Genetics 211: 121–139. https://doi.org/10.1534/genetics.118.301674

Hanover, J. A., M. E. Forsythe, P. T. Hennessey, T. M. Brodigan, D. C. Love et al., 2005 A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout. Proc. Natl. Acad. Sci. USA 102: 11266–11271. https://doi.org/10.1073/pnas.0408771102

Hedgecock, E. M., and J. G. White, 1985 Polyploid tissues in the nematode Caenorhabditis elegans. Dev. Biol. 107: 128–133. https://doi.org/10.1016/0012-1606(85)90381-1

Hobert, O., 2016 Terminal Selectors of Neuronal Identity. Curr. Top. Dev. Biol. 116: 455–475. https://doi.org/10.1016/bs.ctdbh.2015.12.007

Jia, X., D. B. Dadon, B. J. Abraham, T. I. Lee, R. Jaenisch et al., 2015 Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. Proc. Natl. Acad. Sci. USA 112: 3841–3846. https://doi.org/10.1073/pnas.1502971112

Kim, D. H., R. Feinbaum, G. Alloing, F. E. Emerson, D. A. Garsin et al., 2002 A conserved p38 MAP kinase pathway in Caenorhabditis elegans innate immunity. Science 297: 623–626. https://doi.org/10.1126/ science.1073759

Kim, H., T. Ishidate, K. S. Ghanta, M. Seth, D. Conte Jr. et al., 2014 A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. Genetics 197: 1069–1080. https://doi.org/10.1038/ genetics.2014.166839
