# Novel glycylated sugar alcohols protect ESC-specific microRNAs from degradation in iPS cells

Samantha Chang-Lin[1,2], Albert Hung[1,2], Donald C. Chang[1], Yi-Wen Lin[1], Shao-Yao Ying[2] and Shi-Lung Lin[1,*]

[1]WJWU & LYNN Institute for Stem Cell Research, 12145 Mora Drive, Santa Fe Springs, CA 90670, USA and
[2]Department of Cell & Neurobiology, Keck School of Medicine, University of Southern California, CA 90033, USA

## ABSTRACT

**Excessive accumulation of embryonic stem cell (ESC)-specific microRNAs occurs in both ESCs and induced pluripotent stem cells (iPSC); yet, the mechanism involved is unknown. In iPSCs, we for the first time found that novel glycylated sugar alcohols, particularly glycylglycerins, are tightly bound with ESC-specific microRNA precursors (pre-miRNA), such as pre-miR-302. Among these isolated glycylglycerins, we further identified that 1,3-diglycylglycerin and 1,2,3-triglycylglycerin are two major compounds bonded with negatively charged nucleic acids via electro-affinity and subsequently forming sugar-like coats in the hairpin-like double helix structures of pre-miRNAs. As a result, such glycylglycerin-formed coating serves as a protection layer against miRNA degradation. Moreover, we found that the pH value of iPSC cytosol determines the charges of these glycylglycerins. During iPSC differentiation, the cytosol pH is increased and hence neutralizes the charges of glycylglycerins, consequently leading to fast miRNA degradation. Therefore, the current findings not only explain how ESC-specific miRNAs are preserved and accumulated in iPSCs and ESCs but also demonstrate an important function of glycylglycerins in protecting the structural integrity of highly degradable miRNAs, providing a useful means for maintaining miRNA/siRNA function as well as developing the related RNA interference (RNAi) applications.**

## INTRODUCTION

Ribonucleic acids (RNAs) are highly degradable materials in nature; yet, both human embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) possess a unique set of highly abundant small non-coding RNAs (ncRNAs)—ESC-specific microRNAs (miRNAs). Among them, the miR-302 family is the most abundant miRNA species, taking up over 30% of the ncRNA populations in human ESCs and iPSCs. In humans, miR-302 is encoded in the intron 8 of the La Ribonucleoprotein Domain Family Member 7 (*LARP7*, *PIP7S*) gene and hence is co-expressed with the *LARP7* gene transcripts. Yet, miR-302 can only be found in undifferentiated pluripotent stem cells, whereas the expression of *LARP7* has been strongly detected in many somatic tissue cells, particularly in brain, heart, liver and testis (1,2), indicating that miR-302 is extremely degradable in somatic cells but not in pluripotent stem cells. Conversely, since *LARP7* is not the most abundantly expressed gene in either ESCs or iPSCs, it is further suggested that there must be a certain mechanism protecting miR-302 from degradation in these pluripotent stem cells.

During early embryogenesis around the 2–8 zygotic cell stage, maternal mRNAs are largely degraded while embryonic gene transcription begins. This maternal-zygotic transition also marks the beginning of zygotic reprogramming, in which excessive miR-302 is accumulated and then induces genome-wide DNA demethylation to reset the gene expression profiles of embryonic cells into a unique pluripotent state (3,4). Interestingly, the same miR-302 accumulation event also takes place in iPSCs during somatic cell reprogramming (5,6). In view of such a similar event occurring in both ESCs and iPSCs during the massive mRNA turnover period of reprogramming, we may use iPSCs as a comparable model for studying the mechanism underlying miR-302 preservation and accumulation in ESCs.

To collect sufficient iPSC extracts for molecular studies, we adopted a doxycycline-/tetracycline-on (*Tet-On*) inducible iPSC line as previously reported (6–8). This inducible iPSC line provides a unique advantage in that the reprogramming process of iPSCs can be synchronized at a certain stage easy for isolating and studying the cell contents of interest. Since the cell contents are changing from time to time during reprogramming, we specifically collected the cytosol extracts isolated from iPSCs at the 10–14 day post-reprogramming, which is relatively equivalent to the 4–16 cell stage of early zygotic development (3–5). After molecular screening and analyses with high performance liquid

*To whom correspondence should be addressed. Tel: +1 626 236 2885; Fax: +1 626 294 9950; Email: shilungl@mirps.org

chromatography and mass spectrometry (HPLC-MS), we found that these isolated cytosol extracts contain novel glycylated sugar alcohol compounds that protect hairpin-like small RNA molecules, in particular microRNA precursors (pre-miRNAs), from degradation in iPSCs.

Sugar alcohols are a generic kind of polyol alcohols derived from sugars and also frequently called polyhydric alcohol, polyalcohol or glycitol. As defined in polymer chemistry, polyols are compounds with multiple active hydroxyl groups available for organic reactions. Most sugar alcohols are water-soluble naturally occurring materials and are generally represented by a chemical formula $H(HCHO)_{n+1}H$, which is similar but different from sugars' $H(HCHO)_n HCO$. As sugar alcohols normally do not possess any charge, they are inert to negatively charged nucleic acids, such as deoxyribonucleic acids (DNAs) and RNAs. However, in iPSCs, we discovered a novel sugar/sugar alcohol modification—glycylation that replaces the hydroxyl groups (HO-) of a sugar or sugar alcohol with glycine's glycyl groups ($NH_3^+CH_2COO$-) and hence renders positive charges to the glycylated sugar or sugar alcohol, which then becomes highly interactive with DNAs and RNAs.

DNAs and RNAs possess negative charges due to their phosphodiester linkage (backbone) between consecutive nucleotides and hence tend to interact with positively charged compounds. Nevertheless, ions in the water often break down this linkage of DNA and RNA strands via hydrolysis. Water may also give ribonucleases (RNases) easy access to the phosphodiester backbone of RNAs, leading to fast RNA degradation. In a laboratory, alcohol-like materials are known to present strong polarity that can repel water molecules out of the DNA and RNA strands, a mechanism useful for protecting the structural integrity of DNAs and RNAs and thus preventing their degradation *in vitro*. Yet, scientists never know that pluripotent stem cells can also utilize sugar- or alcohol-like materials to preserve ESC-specific miRNAs in nature. To demonstrate this point, our present finding of glycylated sugar alcohols, in particular glycylglycerins, may provide an important insight into the molecular mechanism underlying ESC-specific miRNA preservation and accumulation in iPSCs and ESCs.

## MATERIALS AND METHODS

### Human iPSC derivation and iPSC cytosol extraction

Human iPSC derivation was performed as previously reported (6–8). An inducible *pTet-On-tTS-mir302s* lentiviral vector had been transgenically introduced into these iPSCs for modulating their reprogramming processes (6). After 2-week reprogramming induced by 10 μM of doxycycline (Dox)-mediated hsa-miR-302 expression and grown in feeder-free iPSC culture medium (6), the expressions of ESC-specific markers such as Oct4, Sox2 and Nanog on iPSCs were validated as previously reported (6,7). Then, iPSC cytosol was isolated by ultracentrifugation at 17500 *g* for 30 min at 4°C and further purified by passing it through a 0.01 μm ultrafilter column (30 kDa/100 nt-cutoff; Amicon Ultra-0.5 30K), following the manufacturer's suggestions (Millipore, Billerica, MA, USA). Approximately 0.8–1 ml of pure iPSC cytosol could be recovered from 1–1.2 billion

iPSCs. For isolating pre-miRNAs/miRNAs, the iPSC cytosol was further purified by a 0.001 μm nanofilter column (3 kDa/10 nt-cutoff; Amicon Ultra-0.5 3K) and recovered in the flow-through portion, while all small RNAs and their associated glycylglycerins/proteins were collected on the nanofilter and then dissolved in double-autoclaved DEPC-treated ddH$_2$O (pH 5.5–5.6) for further use in HPLC-MS, RT-qPCR and microarray analyses. The sizes of small RNAs so obtained were ranged from about 10 to 110 nt (or 3–30 kDa), including pre-miRNAs/miRNAs and tRNAs. For comparison between different types of pluripotent stem cells, we also used the same isolation method to collect cytosol and small RNA samples from human WA01/H1 ESCs (hESC H1; WiCell Research Institute, Madison, WI) (9) and 5-day-old *in-vitro*-fertilized mouse embryos (blastocyst stage IVF-E; C57BL/6 strain), respectively. All cytosol samples were freshly made and used in experiments immediately after extraction. Also, the pH value of cytosol was measured with a stainless steel micro pH probe and an AquaPro pH combination electrode probe (Thomas Scientific, Swedesboro, NJ, USA) immediately after extraction. Additionally, to prepare negative control RNAs for HPLC-MS and microinjection-mediated gene silencing assays, we also collected somatic small RNAs from human adult fibroblasts following the same isolation protocol.

### Analyses of glycylglycerins using high performance liquid chromatography—mass spectrometry (HPLC-MS)

HPLC programs were run by an Ultimate 3000 HPLC machine (Thermo Scientific, Waltham, MA, USA) with a DNAPac PA-100 column (BioLC Semi-Prep 9 × 250 mm) at a flow rate of 3.6 ml/min. Starting buffer was 40 mM Tris–HCl (pH 6.5–7.6) and mobile buffer was 40 mM Tris–HCl (pH 6.5–7.6) with 500 mM sodium perchlorate. HPLC separation is based on the charge affinity of samples to ion-exchange chromatography resins during buffer elution. Signals of charged molecules, such as nucleic acids and glycylated sugar alcohols, were detected and measured with both ultra RS charged aerosol and ultraviolet scan detectors (Thermo Scientific). Fragments of HPLC-purified samples were separately collected at detected peaks and further analyzed with an LTQ-Orbitrap electrospray ionization mass spectrometer (ESI-MS; Thermo Scientific). After HPLC purification, each collected sample fragment was desalted and washed twice with double-autoclaved DEPC-treated ddH$_2$O on a 0.001 micron nanofilter (Amicon Ultra-0.5 3K) and then dissolved in an acetonitrile/isopropanol/HPLC-ddH$_2$O (40/20/40) solution containing 20 μM piperidine and 20 μM imidazole. ESI-MS analyses were performed by the chemical core facility at University of Southern California, following a previously reported protocol (10). MS separation was based on the mass-to-charge ($m/z$) ratios of the sample molecules. MS spectrums were collected from 5–300 and 2100–3300 $m/z$, respectively, and further analyzed with a Thermo Scientific Xcalibar 4.0 software.

### Analyses of miRNA/siRNA levels using microarrays and RT-qPCR

The purity and quantity of miRNA/siRNA were first assessed with 4% low-melting-point agarose gel electrophore-

sis and spectrophotometer at UV 260/280 nm (Bio-Rad, Hercules, CA, USA). Then, microarray analyses were performed by LC Sciences (San Diego, CA, USA), using ∼10 µg of purified small RNAs from each sample, respectively. Each microarray chip was hybridized with a single sample labeled with either Cy3 or Cy5. Background subtraction, data normalization and statistic calculation were performed following manufacturer's protocols. For RT-qPCR, we used a set of TaqMan primers directed against hsa-miR-302a or mmu-miR-295, respectively, and the related Real-Time PCR kit (Life Technologies, Grand Island, NY, USA), following the manufacturer's instructions. Signals were detected with an ABI7300 Real-Time PCR System (Applied Biosystems, Life Technologies).

### Synthesis of siRNAs and pre-miRNAs

A synthetic siRNA mimic of hsa-miR-302a, namely siR-302, was purchased from Sigma-Genosys (St. Louis, MO, USA), while pre-miR-302a and pre-miR-295 were synthesized by Bio-Synthesis (Lewisville, TX, USA). Duplexes of siR-302 were formed by hybridization of a pair of perfectly matched RNA sequences: 5′-UAAGUGCUUC CAUGUUUUAG UGA-3′ and 5′-UCACUAAAAC AUG-GAAGCAC UUA-3′ in an annealing buffer provided by the manufacturer. For serving as standard markers in ESI-MS analyses, the synthetic sequences of pre-miR-302a and pre-miR-295 used were 5′-pCCACCACUUA AACGUGGAUG UACUUGCUUU GAAACUAAAG AAGUAAGUGC UUCCAUGUUU UGGUGAUGG-3′ and 5′-pGGUGAGACUC AAAUGUGGGG CACACU-UCUG GACUGUACAU AGAAAGUGCU ACUACUU-UUG AGUCUCUCC-3′, respectively.

### Analysis of RNA degradation with gel shift electrophoresis

RNA samples (15 µg) with or without DGG/TGG treatments were mixed with 1× RNA gel loading dye (ThermoFisher Scientific) and then run on a 4% low-melting-point agarose gel in 1× Tris-Acetate-EDTA (TAE) buffer (pH 8.0) at 4°C.

### Microinjection and western blot analyses

Microinjection was performed with an Eppendorf Inject-Man NI2 micromanipulator and a FemtoJet microinjector under a Nikon TE2000 inverted microscope. We injected 0.2 pg of DGG/TGG-mixed siR-302 (in 0.2 pL of 1× phosphate buffered saline (PBS) buffer at pH 6.5) into the cytoplasm of each single keratinocyte, grew the injected cells in feeder-free iPSC culture medium at 37°C, 5% $CO_2$ for 5 days and then measured the changes of miR-302-targeted gene expressions using western blot assays. The feeder-free iPSC culture medium was consisted of MEF-conditioned hESC medium supplemented with 20% KO serum replacement (Invitrogen, Carlsbad, CA, USA) and 1% Minimum Essential Medium (MEM) non-essential amino acids, 100 µM ß-mercaptoethanol, 1 mM GlutaMax, 1 mM sodium pyruvate, 10 ng/ml bFGF, 5 ng/ml LIF, 10 ng/ml FGF4, 50 IU/ml penicillin, 50 mg/ml streptomycin, 0.1 µM A83-01, 1 mM valproic acid and 5 µM Y-27632. Western blot analysis was performed as previously reported (5,6).

### Statistical analyses

All data were shown as averages and standard deviations (SD). Mean of each test group was calculated by AVERAGE of Microsoft Excel. SD was performed by STDEV. Statistical analysis of data was performed by One-Way ANOVA. Tukey and Dunnett's *t* post hoc test were used to identify the significance of data difference in each group. $P < 0.05$ was considered significant (SPSS v12.0, Claritas Inc.).
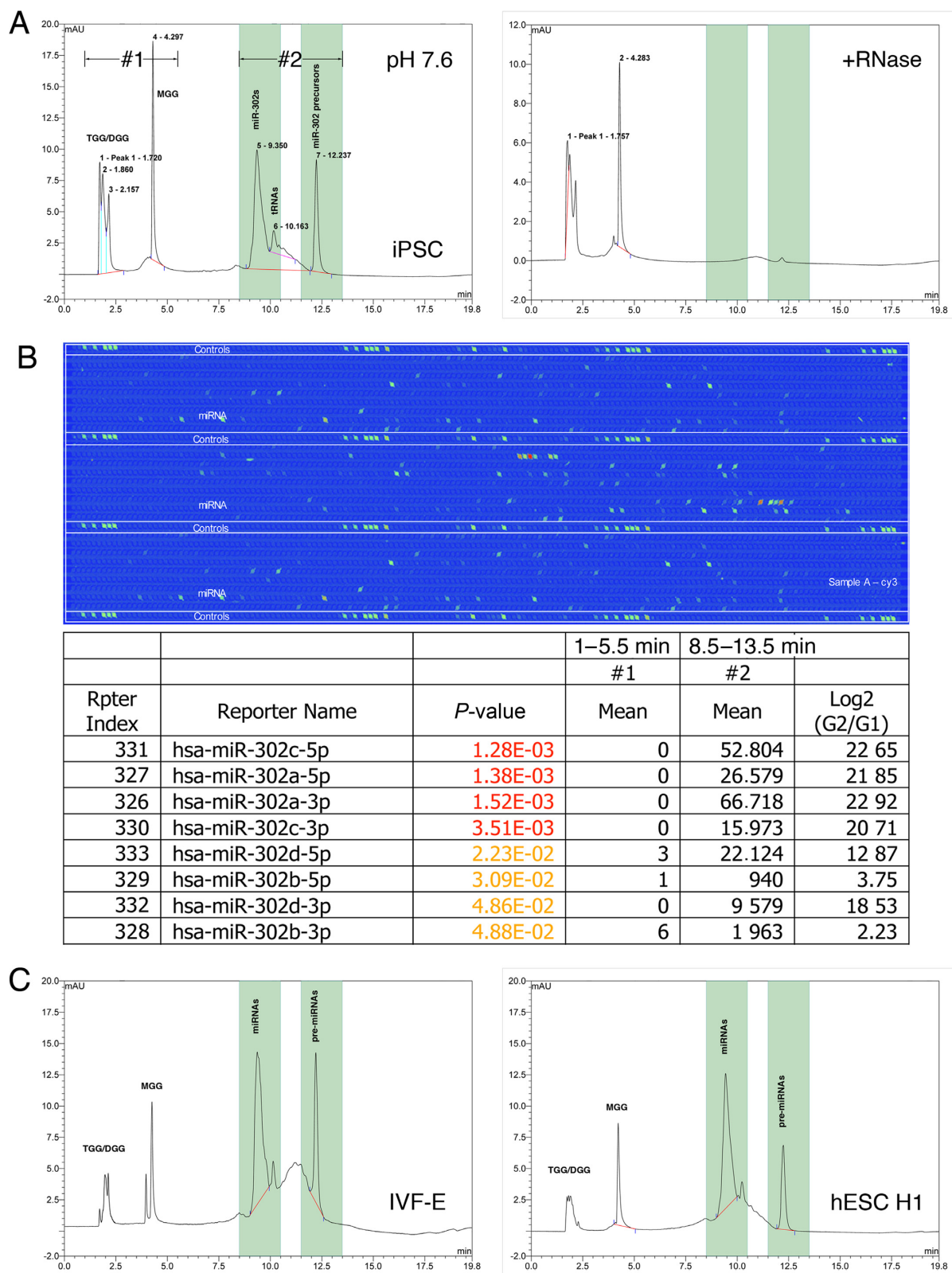
## RESULTS

### Identification of glycylglycerins

Our principle of searching natural miRNA protecting agents in iPSCs is based on the interaction of electro-affinity charges between miRNA and its binding/protecting agents. Since miRNAs are negatively charged molecules that tend to interact with positively charged materials, it is necessary to maintain such an electro-affinity interaction in order to co-purify the complexes of miRNA and its binding/protecting agents. However, due to the use of alkaline and acidic chemicals, conventional miRNA isolation methods often neutralize the electro-affinity of positively charged materials, leading to a failure in detection of the binding/protecting agents after miRNA isolation. To overcome this problem, we used physical methods, such as ultracentrifugation and filtration, to isolate ESC-specific miRNAs from human iPSCs.
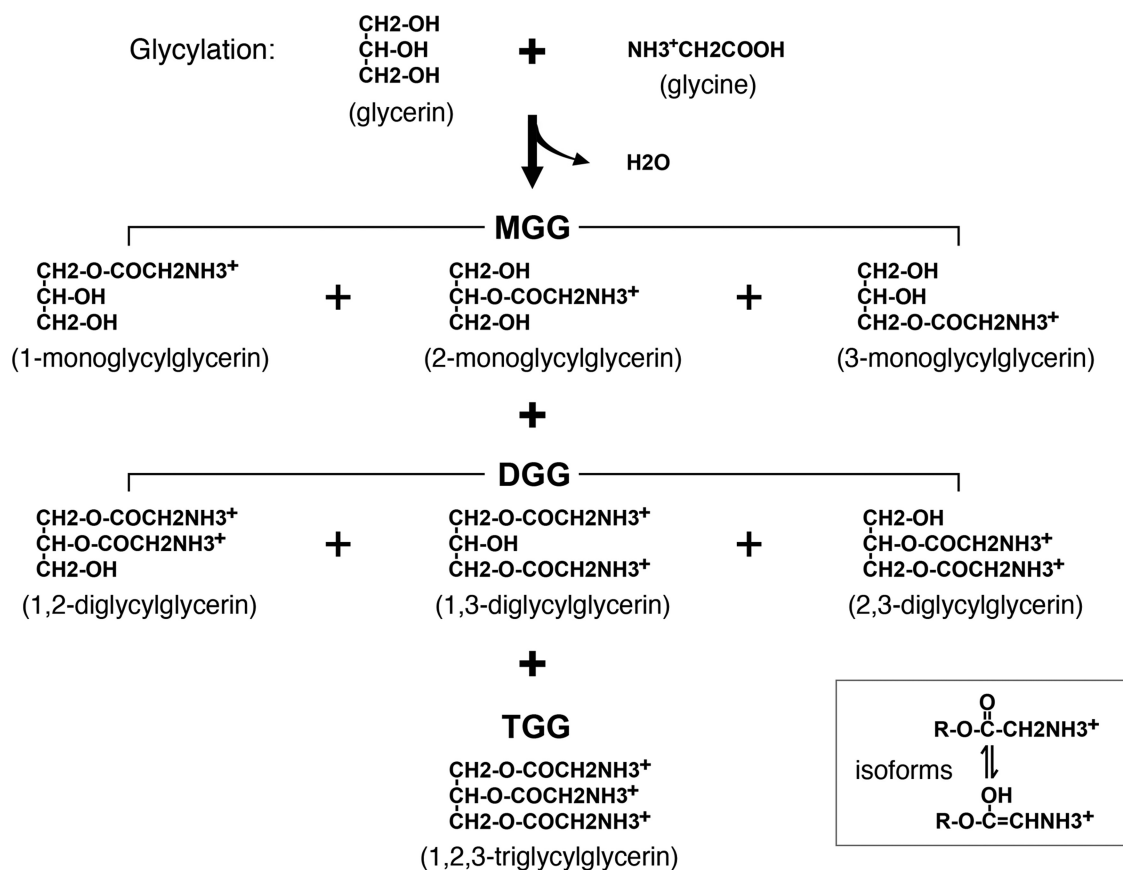
The isolated ESC-specific miRNAs were screened by HPLC-MS and microarrays to determine the major RNA contents in iPSC cytosol, as shown in Figure 1A and B. As previously reported (2), both of HPLC (1A) and microarray data (1B) indicated that miR-302 familial miRNAs and their precursors were the most abundant small RNA species in human iPSCs (group #2 of Figure 1A and B; retention time = 8.8–13 min). Accompanying the isolated miRNAs, we further detected non-RNA materials that were separated from the miRNAs/pre-miRNAs by HPLC at pH 7.6 (group #1 of Figure 1A; retention time = 1.6–4.7 min). These non-RNA materials cannot be digested by RNase and is positively charged in view of their short retention time in an anion-exchange HPLC resin column, of which the average retention time (3.1 min) is even less than a single nucleotide (4.3 min). Further MS analyses revealed that these non-RNA materials are glycylated sugar alcohols in particular, glycylglycerins (Supplementary Figure S1A–E).

Glycylation of glycerin (or glycerol) is a novel sugar/sugar alcohol modification that has never been reported in living cells before. As depicted in Figure 2, glycylation generates three kinds of glycylglycerins in iPSCs, including 1-, 2-, 3-monoglycylglycerins (MGG), 1,2-, 2,3-, 1,3-diglycylglycerins (DGG) and 1,2,3-triglycylglycerin (TGG). All of these identified glycylglycerins form isoforms, of which the chemical structures can be detected by spectrophotometry at wavelength 260 nm, as shown in the right-bottom corner of Figure 2. In chemical modeling with MS analyses, the structures of MGG resemble monosaccharides such as fructose, while DGG and TGG are similar to disaccharides. Most noteworthily, we discovered that DGG and TGG were only detected in undifferentiated

**Figure 1.** Analysis of small charged molecules (3–30 kDa) extracted from iPSC cytosol. (**A**) Results of HPLC analyses at pH 7.6 showed two major groups of extracted small molecules, including group #1 isolated from retention time 1.0–5.5 min containing RNase resistant factors and group #2 isolated from retention time 8.5–13.5 min containing small RNAs. (**B**) Microarray analyses of group #2 (G2) extractions showed the majority of isolated small RNAs were ESC-specific miR-302 familial miRNAs and their pre-miRNAs. (**C**) Analyses of samples extracted from mouse IVF-Es (left chart) and human H1 ESCs (right chart) also showed similar HPLC peak patterns, indicating the conservation of these charged molecules in human iPSCs, ESCs and mouse blastocyst-stage embryos.

Glycylation:

$$CH_2-OH$$
$$CH-OH$$
$$CH_2-OH$$
(glycerin)

$+$

$NH_3^+CH_2COOH$
(glycine)

$\searrow H_2O$

**MGG**

$CH_2-O-COCH_2NH_3^+$
$CH-OH$
$CH_2-OH$
(1-monoglycylglycerin)

$+$

$CH_2-OH$
$CH-O-COCH_2NH_3^+$
$CH_2-OH$
(2-monoglycylglycerin)

$+$

$CH_2-OH$
$CH-OH$
$CH_2-O-COCH_2NH_3^+$
(3-monoglycylglycerin)

$+$

**DGG**

$CH_2-O-COCH_2NH_3^+$
$CH-O-COCH_2NH_3^+$
$CH_2-OH$
(1,2-diglycylglycerin)

$+$

$CH_2-O-COCH_2NH_3^+$
$CH-OH$
$CH_2-O-COCH_2NH_3^+$
(1,3-diglycylglycerin)

$+$

$CH_2-OH$
$CH-O-COCH_2NH_3^+$
$CH_2-O-COCH_2NH_3^+$
(2,3-diglycylglycerin)

$+$

**TGG**

$CH_2-O-COCH_2NH_3^+$
$CH-O-COCH_2NH_3^+$
$CH_2-O-COCH_2NH_3^+$
(1,2,3-triglycylglycerin)

isoforms
$$R-O-\overset{O}{\overset{\|}{C}}-CH_2NH_3^+$$
$$\updownarrow$$
$$R-O-\overset{OH}{\overset{|}{C}}=CHNH_3^+$$

**Figure 2.** Category of HPLC-MS-identified glycylglycerins. In human iPSCs, the glycylation of glycerin forms three different kinds of glycylglycerins, including monoglycylglycerins (MGG), diglycylglycerins (DGG) and triglycylglycerin (TGG). All these glycylglycerins carry positive charges in their amino groups at $\leq$pH 7.0 and can form isoforms as shown in the right bottom corner box.

rather than differentiated iPSCs, suggesting that they are specifically bound with ESC-specific miRNAs in undifferentiated iPSCs and thus may be served as a novel marker for undifferentiated iPSCs.

To confirm the existence of glycylglycerins in ESCs and blastocyst stage embryos, we performed HPLC analyses using small RNAs isolated from cytosol of 5-day-old *in-vitro*-fertilized mouse blastocysts (IVF-E) and human H1 ESCs (hESC H1), respectively, of which the small RNAs were obtained by the same isolation methods as described in 'Materials and Methods' section. As a result, Figure 1C showed that all three kinds of glycylglycerins were also presented in the IVF-Es and hESCs and could be co-purified with ESC-specific miRNAs/pre-miRNAs to demonstrate similar HPLC peak patterns. Based on these HPLC results and further spectrophotometer measurement, the concentrations of detected glycylglycerins in human iPSCs, hESCs and mouse IVF-Es were estimated to be $\sim$0.73–0.95 ($n = 4$), 0.31–0.37 ($n = 2$) and 0.49 mM, respectively. Meanwhile, we also measured the concentrations of isolated ESC-specific miRNAs/pre-miRNAs in human iPSCs, hESCs and mouse IVF-Es, showing $\sim$16.5–17.3, 14.6–15.2 and 19.7 $\mu$M, respectively. However, because our physical isolation methods could only collect glycylglycerins that were bound with miRNAs and pre-miRNAs, not unbound glycylglycerins,

the concentrations of total glycylglycerins in stem cell cytosol may be higher than the currently estimated numbers.

**Glycylglycerins protect ESC-specific miRNAs from degradation in iPSCs**

After HPLC separation of glycylglycerins from ESC-specific miRNAs, the purified miRNAs and their pre-miRNAs became very unstable in cytosol. As shown in Figure 3, when 100 $\mu$g of the purified miRNAs/pre-miRNAs were mixed with 200 $\mu$l of iPSC cytosol in the absence or presence of 0.1 M DGG/TGG and then incubated at 20°C, we found that without any protection (miRNA only) over 46 and 73% of the miRNA/pre-miRNA population was quickly degraded within 12 and 24 h, respectively, whereas the miRNAs/pre-miRNAs mixing with DGG/TGG (miR + D/TGG) showed no detectable degradation during the same incubation time. This result demonstrated that these iPSC-derived glycylglycerins, particularly DGG and TGG, likely can facilitate the protection and preservation of ESC-specific miRNAs/pre-miRNAs in iPSCs.

**Application of glycylglycerins in miRNA and siRNA preservation**

Given that siRNA is often used as a synthetic mimic of natural pre-miRNAs, we may apply glycylglycerins to pro-
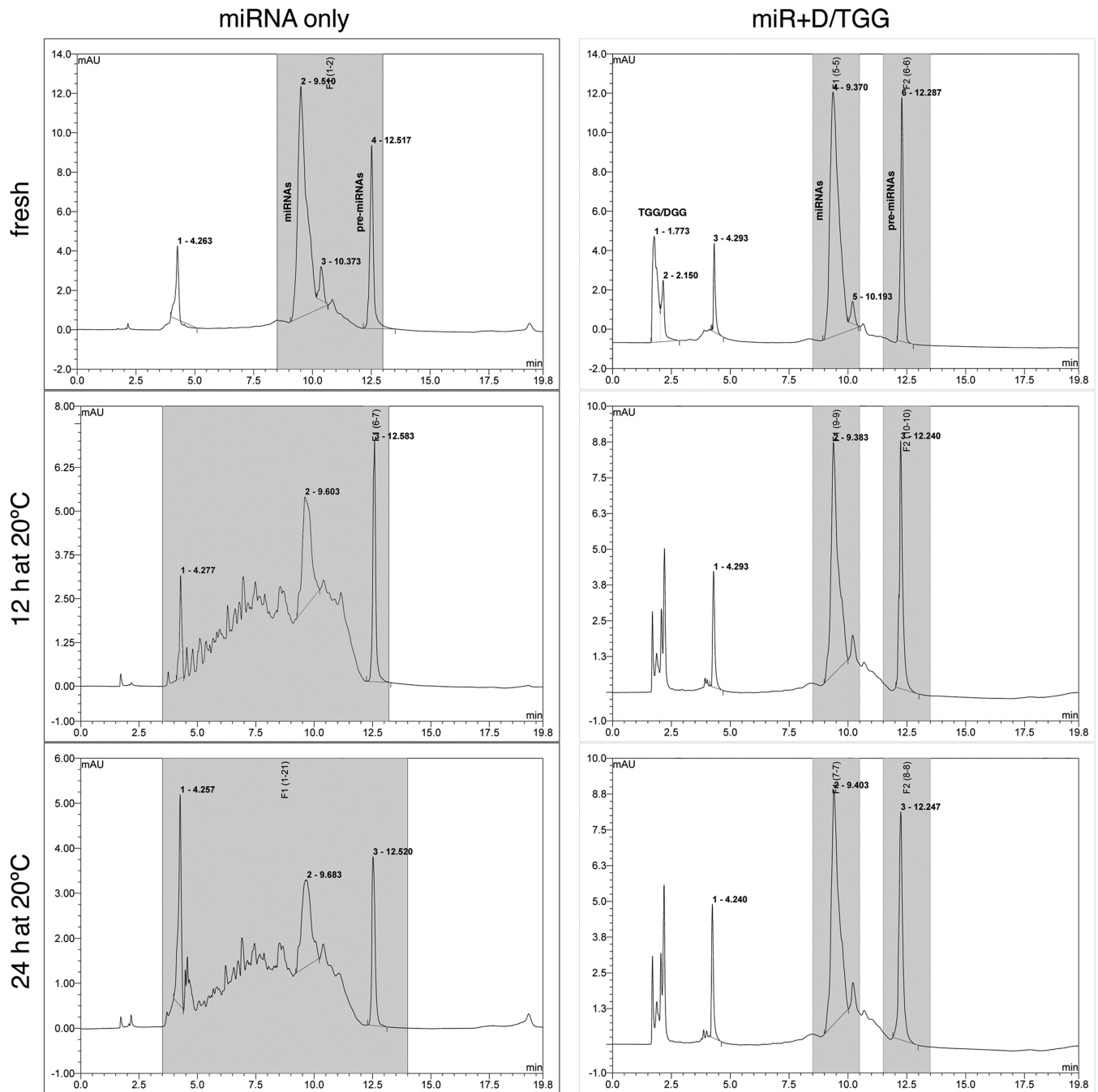
**Figure 3.** HPLC analysis of ESC-specific miRNA/pre-miRNA degradation in iPSC cytosol in the absence (left panels; miRNA only) or presence of DGG and TGG (right panels; miR + D/TGG). In the absence of glycylglycein protection, >46% and >73% of the whole miRNA/pre-miRNA population was quickly degraded (smeared down into smaller charges/sizes toward the left side of the HPLC chart) after incubating in iPSC cytosol for 12 and 24 h, respectively, whereas no degradation was detected when DGG/TGG and miRNAs/pre-miRNAs were mixed together in the cytosol.

tecting the structural integrity of siRNA and so as to maintain its full function for developing RNA interference (RNAi)-related applications. Small ncRNAs like siRNAs and shRNAs are known to be highly unstable and degradable *in vitro* as well as *in vivo*. To overcome this problem, we tested the feasibility of using DGG/TGG to preserve siRNA integrity in iPSC cytosol at 37°C. As shown in Figure 4A, both results of gel electrophoresis and RT-qPCR demonstrated that >94% of the siRNAs (siR-302)

pre-mixed with 0.1 M DGG/TGG remained intact up to 5 days in iPSC cytosol (pH 6.7–6.9), whereas most (>90%) of the pure siR-302 without any glycylglycerin protection was quickly degraded within 48 h. Since the effect of RNAi usually takes several days to become fully effective, of which the onset timing is dependent on the concentration of siRNAs/miRNAs, the expression rates of siRNA/miRNA-targeted genes and the half-life of the target proteins, this finding suggests that glycylglycerins such as DGG and
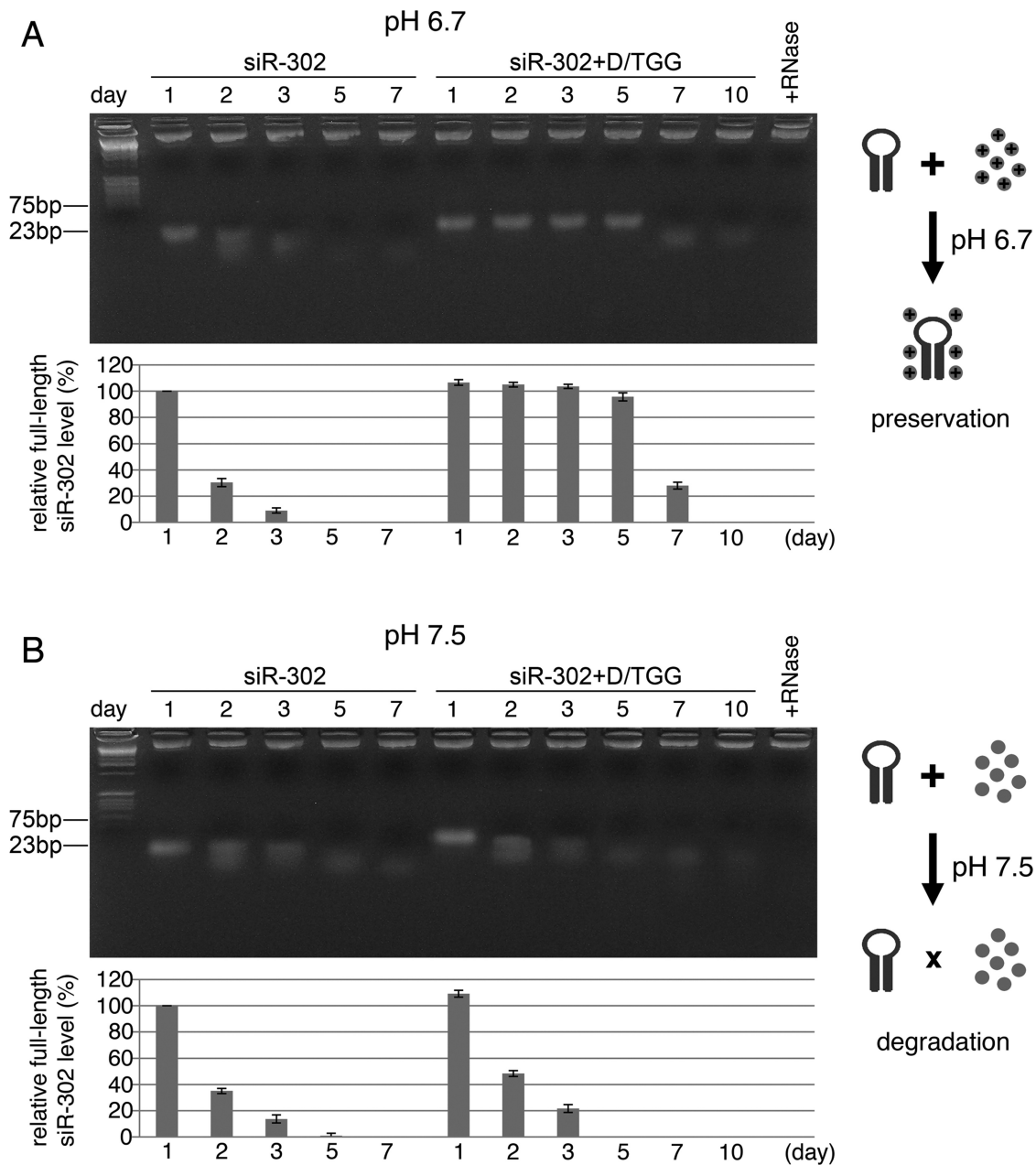
**Figure 4.** Effects of pH alteration on DGG/TGG-protected siRNA siR-302 ($n = 4$, $P < 0.05$). Comparisons between unprotected and DGG/TGG-protected siR-302 degradation in iPSC cytosol at (**A**) pH 6.7 versus (**B**) pH 7.5 were performed, using 4% low-melting-point agarose gel electrophoresis (photo shown on top) and RT-qPCR with a miR-302a-specific primer (chart listed below). (**A**) DGG/TGG-protected siR-302 remained intact up to 5 days in iPSC cytosol at pH 6.7, while most (>90%) of unprotected siR-302 was degraded within 48 h. (**B**) Both DGG/TGG-mixed and unprotected siR-302 molecules were largely degraded within 48 h in iPSC cytosol at pH 7.5.

TGG are useful for maintaining siRNA/miRNA integrity long enough to show the full effects of siRNA/miRNA-mediated RNAi.

### pH value determines the protection effects of glycylglycerins

During iPSC cytosol collection, we measured the pH value of undifferentiated iPSC cytosol to be about 6.7–6.9 ($n = 9$), whereas previous studies have reported that the cytoplasmic pH of differentiated mammalian cells is ranged between 7.0–7.4 (11,12). Particularly, the cytosolic pH of iPSCs dif-

ferentiated under an atmospheric condition containing 22–23% $O_2$ and 0.5% $CO_2$ at 37°C could be measured as high as 7.3–7.8, depending on the primordial cell types of the differentiated cells ($n = 6$). To understand this effect of pH alteration on miRNA/siRNA stability, we adjusted iPSC cytosol to pH 7.5 using DEPC-treated Tris–HCl buffer (pH 8.0) and then repeated the same experiments as used in Figure 4A. Yet, the result of Figure 4B showed that the addition of DGG and TGG at pH 7.5 failed to protect siR-302 from degradation, indicating that the positive charges of glycylglycerins likely have been neutralized at this higher pH point

and hence could not interact with negatively charged siR-NAs. This result may also reveal one of the possibilities that explain why ESC-specific miRNAs are highly preserved in undifferentiated iPSCs but quickly degraded prior to stem cell differentiation.

## Glycylglycerins preserve miRNA/siRNA function

To determine how glycylglycerins affect miRNA/siRNA function, we micro-injected DGG/TGG-mixed siR-302 into the cytoplasm of single keratinocytes (0.2 pg per cell) and then used western blot assays to determine the resulting effects on the protein expression levels of miR-302-targeted genes in these injected cells (Figure 5A). Previous studies have reported that miR-302 functions to silence multiple epigenetic regulators, such as AOF1, MECP1-p66, MECP2, MBD2 and DNMT1, so as to induce global genomic DNA demethylation during the initiation stage of somatic cell reprogramming (3,5,6,13). Because in somatic cells the promoters of many ESC-specific marker genes, in particular Oct4, Sox2 and Nanog (OSN), are inhibited by DNA methylation, the miR-302-induced DNA demethylation can therefore release such promoter inhibition and then activates OSN expression, consequently leading to somatic cell reprogramming to form iPSCs (3–6). Additionally, miR-302 can further silence an Oct4 inhibitor, NR2F2, to enhance Oct4 expression (14). As predicted by these previous reports, Figure 5B showed that the DGG/TGG-bound siR-302 successfully knocked down all the reported target genes and induced concurrent OSN expression in the injected cells, while the injection of unprotected siR-302 failed to elicit any overt effect, which may resemble a similar condition occurring in *LARP7*-expressing somatic cells.

To further validate the target specificity of DGG/TGG-bound siR-302, we repeated the same microinjection experiment using DGG/TGG-mixed small RNAs isolated from human fibroblasts, showing no effect on the protein levels of the reported miR-302 target genes (Figure 5B, the most right lane). This negative result of control RNAs further confirmed that the target specificity of DGG/TGG-bound siR-302 is determined by the siR-302 sequence rather than DGG/TGG. Nevertheless, since a complete somatic cell reprogramming event takes 10–14 days, which is twice longer than the effective time of DGG/TGG protection (5–6 days), one microinjection of DGG/TGG-bound siR-302 may not be sufficient to support the full reprogramming process of iPSC derivation. In view of this obstacle, how to apply a second dosage of DGG/TGG-bound siR-302 into these partially reprogrammed cells will be the key challenge to achieve the final result of full iPSC derivation.

## L-ascorbic acid promotes glycylglycerin formation *invitro*

Glycine and glycerin were never known to interact with each other before; yet, we found that glycylation of glycerin can naturally occur under a ≤pH 6.8 condition in the presence of L-ascorbic acid. As we had found that the pH of iPSC cytosol is ranged about 6.7–6.9, it is conceivable that iPSC cytosol may contain certain acidic factors to maintain this relatively low pH. During HPLC-MS analyses of iPSC and IVF-E cytosol, we also observed that both samples contained about 2–4 mM of L-ascorbic acid, which may lead

to the relatively low pH of iPSC cytosol. Based on this finding, we further tested a pure chemical reaction by mixing 80 mM glycine, 80 mM glycerin and 4 mM L-ascorbic acid together in $1\times$ PBS solution (final pH 6.8) and then incubated the mixture at 37°C for 8 h. Surprisingly, the result of Figure 6A showed that ∼0.54 ± 0.11 mM of glycylglycerins was generated in the test tubes ($n = 6$) without addition of any enzyme, indicating that glycylglycerins, particularly DGG and TGG, can be naturally produced in iPSCs and IVF-Es under a cytosolic pH 6.7–6.9 condition in the presence of L-ascorbic acid at 37°C. Given that this glycylation of glycerin could not occur in the same reaction at pH 7.3, it further suggested that most somatic cells with a cytosolic pH around 7.3 may not be able to produce glycylglycerins in particular, DGG and TGG. However, although our current data had identified a possible source of glycylglycerins in iPSCs and IVF-Es, this finding did not rule out the possible involvement of another enzymatic or transporter-/receptor-mediated mechanism in glycylglycerin production, of which the details remain to be determined.

## Analyses of glycylglycerin-bound ESC-specific miRNAs/pre-miRNAs in IVF-Es

After understanding the pH effect on the binding interaction between DGG/TGG and miRNAs/pre-miRNAs, we further performed HPLC-MS analyses of DGG/TGG-bound pre-miR-302a under various buffered conditions to see the changes of pre-miR-302a charges, of which the result showed a marked shift of the pre-miR-302a peak from 12.23 min at pH 7.3 to 1.76 min at pH 6.8 during HPLC elution (Figure 6B). This result indicated that the binding of positively charged DGG/TGG molecules indeed neutralized the negative charges of pre-miRNAs and hence significantly reduced the retention time of DGG/TGG-bound pre-miR-302a in the anion-exchange HPLC column. According to this finding, we learned the exact timing for collecting unbound glycylglycerins and glycylglycerin-bound pre-miRNAs, respectively, during HPLC separation under different pH conditions. In order to further investigate the presence of DGG/TGG-bound pre-miRNAs in IVF-E cytosol, we collected the sample isolated from the 1.7–1.8 min section of HPLC fragments at pH 6.8 (Figure 6B, right panel) and then used it for ESI-MS analysis (Figure 7A–D).

Since miR-302a and miR-295 are well known markers for human and mouse ESCs, respectively, and their HPLC-MS patterns were still unclear at the time of this study, we then used synthetic pre-miR-302a (MW = 22 251 g) and pre-miR-295 (MW = 22 241 g) mixed with purified DGG/TGG (MW = 206/263) to set up the standard HPLC-MS spectrums of these ESC markers bound with DGG/TGG. As shown in Figure 7A and B, due to their similar molecular weights (MW), the MS spectrums of both pre-miR-302a (7A) and pre-miR-295 (7B) showed in a proximity area around 2330∼3150 $m/z$ with a charge ($z$) number ranged between 9–12, while some detached DGG and TGG molecules were also detected at 206 and 263 $m/z$, respectively. After further software analysis and calculation, the measured MWs of DGG/TGG-bound pre-miR-302a and pre-miR-295 were approximately 28 164 and 28 154 g, respectively. Compared to their original unbound
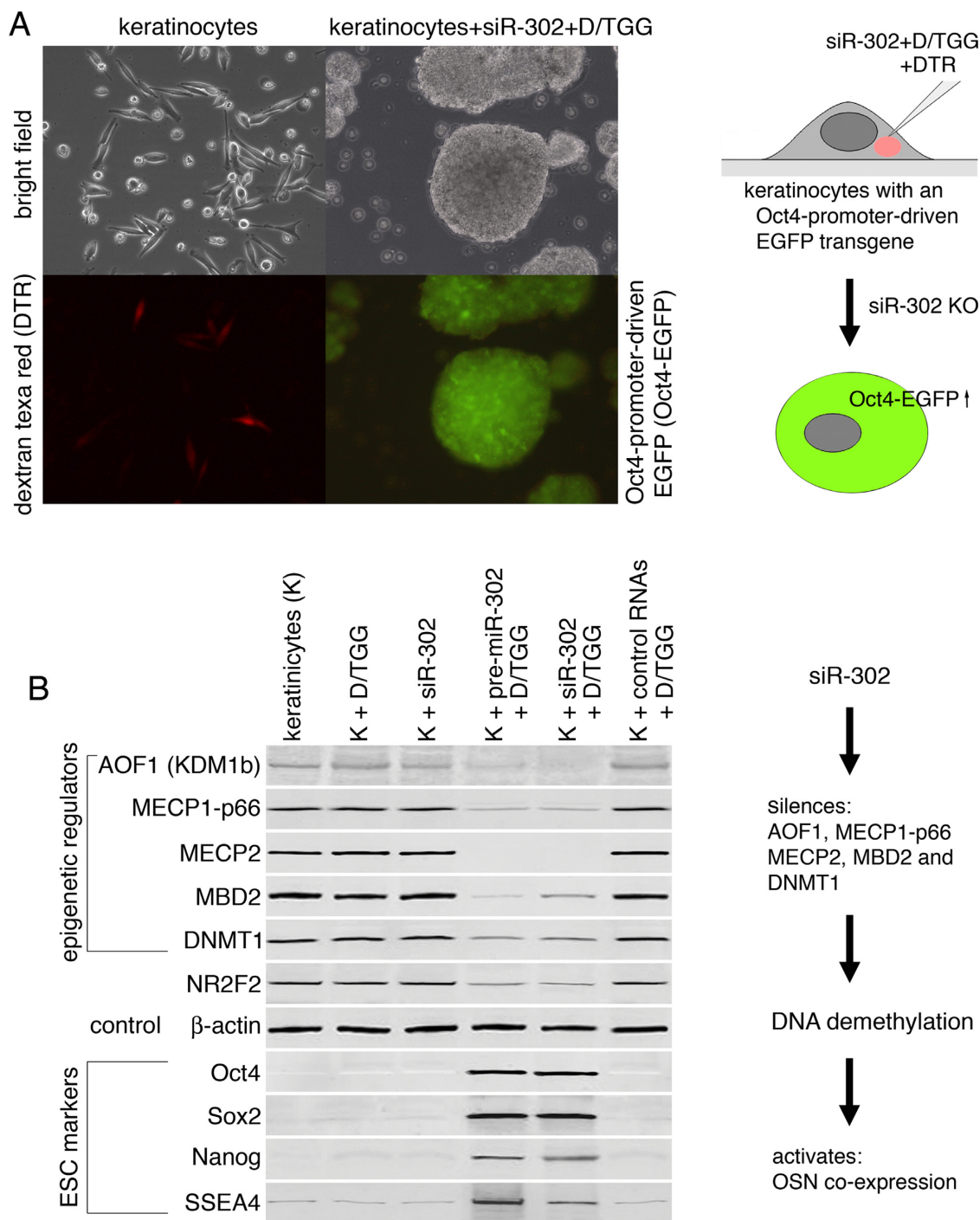
**Figure 5.** Functionality of DGG/TGG-protected siR-302. (**A**) Microinjection of DGG/TGG-mixed siR-302 into the cytoplasm of single keratinocytes that had been pre-transfected with an Oct4-promoter-driven green fluorescent protein transgene (Oct4-EGFP). The injected cells were labeled by co-injection of a red fluorescent dye—dextran texa red (DTR) with the DGG/TGG-mixed siR-302. (**B**) Western blot assays showed the gene silencing effects of siR-302 on reported miR-302-targeted genes, including lysine (K)-specific demethylase 1B (AOF1, KDM1b), methyl-CpG-binding protein 1, subunit p66 (MECP1-p66), methyl-CpG-binding protein 2 (MECP2), methyl-CpG-binding domain protein 2 (MBD2), DNA (cytosine-5-)-methyltransferase 1 (DNMT1) and nuclear receptor subfamily 2, group F, member 2 (NR2F2). In addition, silencing of these target genes led to the elevated expression of ESC-specific reprogramming factors, such as Oct4, Sox2 and Nanog.
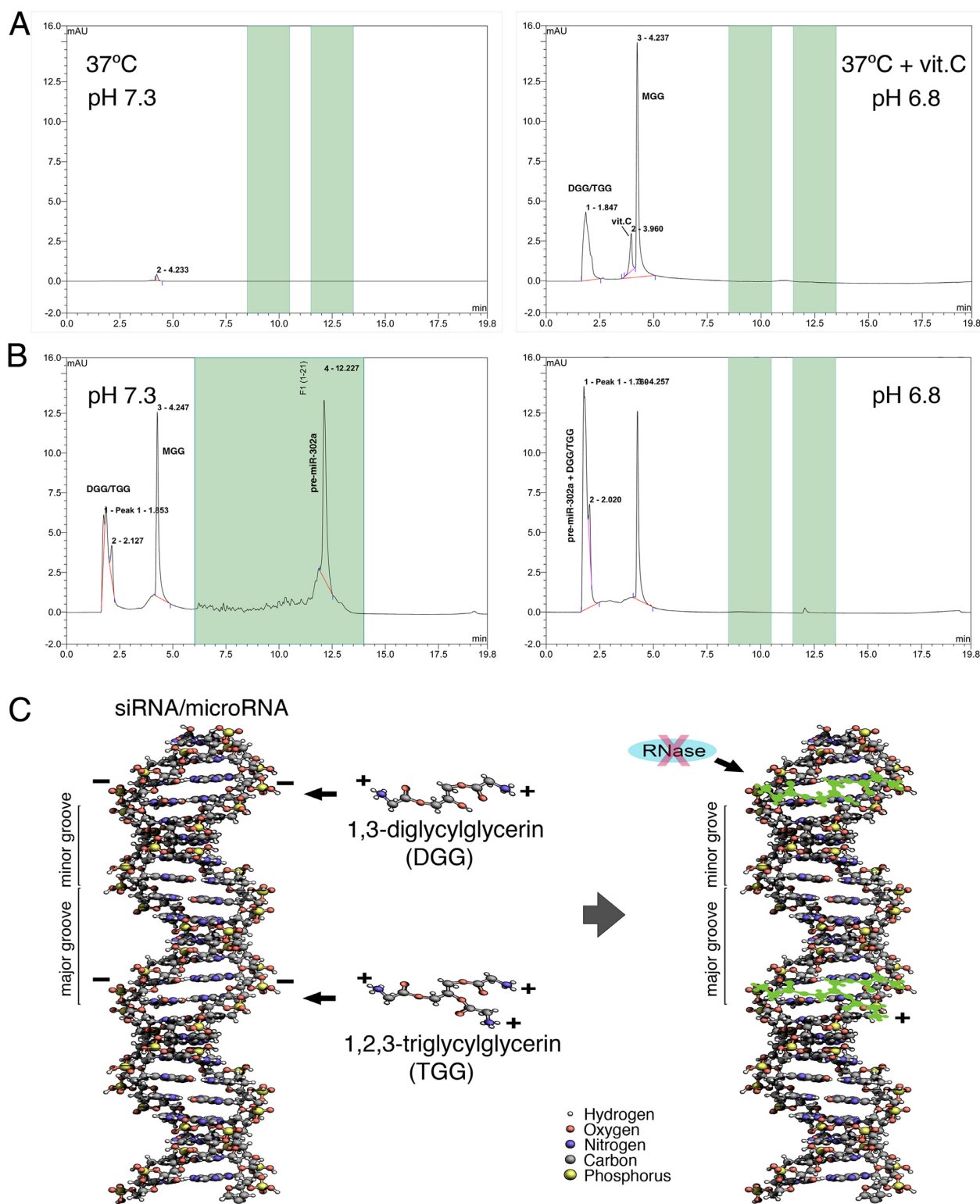
**Figure 6.** HPLC detection of glycylglycerin production and glycylglycerin-bound pre-miRNA. (**A**) HPLC analyses showed that production of glycyl-glycerins can naturally occur in a pure chemical reaction between glycine and glycerin at pH 6.8 in the presence of L-ascorbic acid (vit.C), but not in the same reaction at pH 7.3. (**B**) HPLC peak of DGG/TGG-mixed pre-miR-302a was shifted from 12.23 min at pH 7.3 (unbound) to 1.76 min at pH 6.8 (DGG/TGG-bound), indicating a marked loss of pre-miRNA negative charges after binding with DGG/TGG. After HPLC separation at pH 6.8, the DGG/TGG-bound pre-miR-302a was collected from the 1.7–1.8 min section of HPLC fragments and further used for MS analyses, as shown in Figure 7A. (**C**) Chemical 3D modeling of electro-binding between DGG/TGG and pre-miRNA/siRNA suggested that the long strand structures of 1,3-diglycylglycerin and 1,2,3-triglycylglycerin may fit into the minor grooves of the pre-miRNA/siRNA double helix structures and hence bind to the negatively charged phosphodiester-linkage backbones (green shadow) located in the minor grooves of pre-miRNA/siRNA via the positively charged amino groups of DGG/TGG, so as to form a layer of protective coating for preserving pre-miRNA/siRNA integrity.
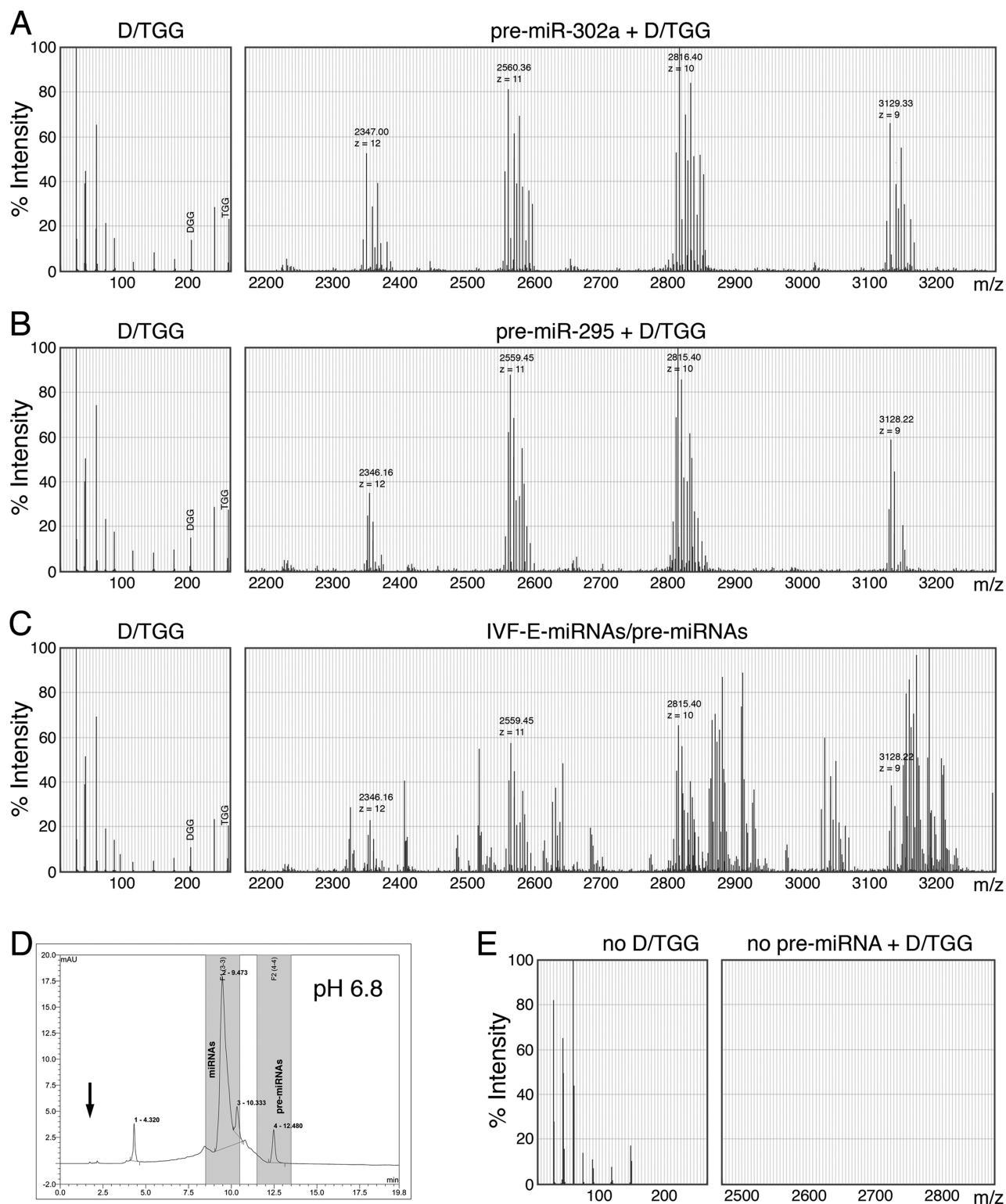
**Figure 7.** ESI-MS spectrums of (**A**) DGG/TGG-bound pre-miR-302a, (**B**) DGG/TGG-bound pre-miR-295, (**C**) ESC-specific miRNAs/pre-miRNAs isolated from mouse IVF-Es (5-day-old blastocysts) and (**E**) somatic small RNAs isolated from human fibroblasts. All tested samples were separately collected from the 1.7–1.8 min section (indicated by an arrow) of HPLC fragments run in a pH 6.8 buffered condition, as shown in (**D**) and Figure 6B. MS spectrums of DGG/TGG-related peaks were detected between 5–300 $m/z$, while those of ESC-specific pre-miRNAs were ranged from 2100 to 3800 $m/z$ with a charge ($z$) number of 8–12; yet, due to chart limitation, the MS spectrums shown here represented only the size range between 2100–3300 $m/z$ with a $z$ number of 9–12.

MWs, we estimated that one pre-miR-302a or pre-miR-295 could be bound by over 22–29 DGG/TGG molecules, which markedly neutralized the negative changes of a pre-miRNA. Yet, as we also observed some DGG/TGG molecules detached from pre-miR-302a and pre-miR-295 during MS analyses, it suggested that the actual binding number of DGG/TGG to a pre-miRNA may be higher than the currently estimated numbers.

Following the establishment of DGG/TGG-bound pre-miR-302a and pre-miR-295 spectrums, we then used these standard MS peak patterns to investigate the presence of DGG/TGG-bound ESC-specific miRNAs/pre-miRNAs in cytosol of mouse IVF-Es. As shown in Figure 7C, the peaks of DGG/TGG-bound ESC-specific pre-miRNAs were measured ranged between 2300 and >3300 $m/z$ with a charge ($z$) number of 9–12. Among these many detected molecular peaks, we could clearly identify all the major peaks of DGG/TGG-bound pre-miR295 at 2346.16, 2559.45, 2815.40 and 3128.22 $m/z$, indicating the existence of DGG/TGG-bound pre-miR-295 in mouse IVF-Es *in vivo*. Additionally, we also repeated the same experiment using ESC-specific miRNAs/pre-miRNAs isolated from cytosol of human ESCs (hESC H1), showing different major peaks of DGG/TGG-bound pre-miR-302a at 2347.00, 2560.36, 2816.40 and 3129.33 $m/z$ (Supplementary Figure S2). These results successfully demonstrated a potential use of the established pre-miRNA MS spectrums as standard markers for screening and identifying the expression of certain specific miRNAs in biological samples. In contrast, when performing the same experiment using somatic small RNAs isolated from human fibroblasts, neither DGG/TGG nor any DGG/TGG-bound pre-miRNA was detected in the corresponding peak locations of HPLC fragments (Figure 7D) or MS spectrums (Figure 7E), further confirming the unique presence of these DGG/TGG-bound ESC-specific miRNAs/pre-miRNAs in human ESCs and mouse blastocysts, respectively.

## DISCUSSION

Pluripotent stem cells like iPSCs can be a treasure box for us to discover and develop novel medicine. They contain numerous unique and effective ingredients useful for stimulating cell/tissue/organ regeneration, repairing damaged or aged tissues, treating degenerative diseases (i.e. diabetes, osteoporosis and Alzheimer's diseases etc.) and preventing tumor formation and cancer progression/metastasis. To reveal these potentials, we have used iPSCs as a tool for screening, identifying, isolating and producing novel drug candidates as well as studying the mechanism underlying pluripotent stem cell derivation (15). As a result, the current study disclosed for the first time that glycylated sugar alcohols in particular, glycylglycerins, protect hairpin-like small RNA molecules, such as pre-miRNAs, from degradation in human iPSCs. Due to the structural similarity between pre-miRNAs and siRNAs, these glycylglycerins also provide protection against siRNA degradation, leading to a very useful means for developing miRNA/siRNA-related RNAi applications.

Understanding the function of sugar and sugar-like materials in ESCs and iPSCs is crucial for the progress of stem cell research; yet, this field is rarely studied due to the limited availability of iPSCs and ESCs. With the advantage of inducible iPSC derivation technologies, we were able to complete this study and understand the important role of glycylglycerins in preservation of ESC-specific miRNAs/pre-miRNAs in human iPSCs, ESCs and mouse blastocysts. Our data suggest that in iPSCs (at pH 6.7–6.9) these positively charged glycylglycerins, particularly DGG and TGG, bind to negatively charged miRNAs/pre-miRNAs via electro-affinity and so as to protect them from degradation. To demonstrate this point, an illustration of electro-affinity between pre-miRNAs/siRNAs and DGG/TGG molecules was shown in Figure 6C. However, prior to differentiation, it was also noted that the cellular pH value is raised to ≥7.3 and hence neutralizes the positive charges of glycylglycerins, leading to quick degradation of ESC-specific miRNAs/pre-miRNAs. This finding may explain why ESC-specific miRNAs are largely degraded before stem cell differentiation.

Glycylation of glycerin is the first sugar/sugar alcohol modification known to have a valid function related to ESC biology. Due to the structural similarity between glycylglycerins and mono-/di-saccharides, we believe that these glycylglycerins may possess other functionality in iPSCs and ESCs. Moreover, since there may exist more varieties of novel sugar and sugar alcohol modifications in iPSCs and ESCs, it is important to understand the functions of these differently modified sugars and sugar alcohols. The knowledge so obtained may facilitate the advance of stem cell research into real applications. In view of such abundant knowledge and novel materials found in iPSCs, how to explore their usefulness will be the next challenge.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## REFERENCES

1. Alazami,A.M., Al-Owain,M., Alzahrani,F., Shuaib,T., Al-Shamrani,H., Al-Falki,Y.H., Al-Qahtani,S.M., Alsheddi,T., Colak,D. and Alkuraya,F.S. (2012) Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a syndrome of facial dysmorphism, intellectual disability, and primordial dwarfism. *Hum. Mutat.*, **33**, 1429–1434.
2. Suh,M.R., Lee,Y., Kim,J.Y., Kim,S.K., Moon,S.H., Lee,J.Y., Cha,K.Y., Chung,H.M., Yoon,H.S., Moon,S.Y. *et al.* (2004) Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.*, **270**, 488–498.
3. Lin,S.L. (2011) Deciphering the mechanism behind induced pluripotent stem cell generation. *Stem Cells*, **29**, 1645–1649.

4. Lin,S.L. and Chen,J. (2013) Mechanism of miR-302-mediated iPS cell generation. In: Sell,S (ed). *Stem Cells Handbook*. Springer Publishers press, NY, pp. 119–127.

5. Lin,S.L., Chang,D., Chang-Lin,S., Lin,C.H., Wu,D.T.S., Chen,D.T. and Ying,S.Y. (2008) Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. *RNA*, **14**, 2115–2124.

6. Lin,S.L., Chang,D., Lin,C.H., Ying,S.Y., Leu,D. and Wu,D.T.S. (2011) Regulation of somatic cell reprogramming through inducible mir-302 expression. *Nucleic Acids Res.*, **39**, 1054–1065.

7. Lin,S.L., Chang,D., Ying,S.Y., Leu,D. and Wu,D.T.S. (2010) MicroRNA miR-302 inhibits the tumorigenecity of human pluripotent stem cells by coordinate suppression of CDK2 and CDK4/6 cell cycle pathways. *Cancer Res.*, **70**, 9473–9482.

8. Lin,S.L. and Ying,S.Y. (2012) Mechanism and method for generating tumor-free iPS cells using intronic microRNA miR302 induction. In: Ying,SY (ed). *MicroRNA Protocols*. 2nd edn. Springer Publishers press, NY, pp. 295–324.

9. Thomson,J.A., Itskovitz-Eldor,J., Shapiro,S.S., Waknitz,M.A., Swiergiel,J.J., Marshall,V.S. and Jones,J.M. (1998) Embryonic stem cell lines derived from human blastocysts. *Science*, **282**, 1145–1147.

10. Kullolli,M., Knouf,E., Arampatzidou,M., Tewari,M. and Pitteri,S.J. (2014) Intact microRNA analysis using high resolution mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **25**, 80–87.

11. Roos,A. and Boron,W.F. (1981) Intraellular pH. *Physiol. Rev.*, **61**, 296–434.

12 Bright,G.R., Fisher,G.W., Rogowska,J. and Taylor,D.L. (1987) Fluorescence ratio imaging microscopy: temporal and spatial measurements of cytoplasmic pH. *J. Cell Biol.*, **104**, 1019–1033.

13. Lee,M.R., Prasain,N., Chae,H.D., Kim,Y.J., Mantel,C., Yoder,M.C. and Broxmeyer,H.E. (2013) Epigenetic regulation of NANOG by miR-302 cluster-MBD2 completes induced pluripotent stem cell reprogramming. *Stem Cells*, **31**, 666–681.

14. Rosa,A. and Brivanlou,A.H. (2011) A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation. *EMBO J.*, **30**, 237–248.

15. Chen,S.K.J. and Lin,S.L. (2013) Recent patents on microRNA-induced pluripotent stem cell generation. *Recent Patents Regenerative Med.*, **3**, 5–16.