A Lentiviral Functional Proteomics Approach Identifies Chromatin Remodeling Complexes Important for the Induction of Pluripotency*

Anthony B. Mak†‡¶§, Zuyao Ni¶**, Johannes A. Hewel†¶§, Ginny I. Chen§‡‡, Guoqing Zhong†, Konstantina Karamboulas‡, Kim Blakely†¶§, Sandra Smiley†¶§, Edyta Marcon†, Deniza Roudeva‡§, Joyce Li‡, Jonathan B. Olsen‡§, Cuihong Wan‡§, Andrew Emili‡§ §§, Jack Greenblatt‡§¶¶, and Jason Moffat‡§

Protein complexes and protein-protein interactions are essential for almost all cellular processes. Here, we establish a mammalian affinity purification and lentiviral expression (MAPLE) system for characterizing the subunit compositions of protein complexes. The system is flexible (i.e., multiple N- and C-terminal tags and multiple promoters), is compatible with Gateway™ cloning, and incorporates a reference peptide. Its major advantage is that it permits efficient and stable delivery of affinity-tagged open reading frames into most mammalian cell types. We benchmarked MAPLE with a number of human protein complexes involved in transcription, including the RNA polymerase II-associated factor, negative elongation factor, positive transcription elongation factor b, SWI/SNF, and mixed lineage leukemia complexes. In addition, MAPLE was used to identify an interaction between the reprogramming factor Klf4 and the Swi/Snf chromatin remodeling complex in mouse embryonic stem cells. We show that the SWI/SNF catalytic subunit Smarca2/Brm is up-regulated during the process of induced pluripotency and demonstrate a role for the catalytic subunits of the SWI/SNF complex during somatic cell reprogramming. Our data suggest that the transcription factor Klf4 facilitates chromatin remodeling during reprogramming. Molecular & Cellular Proteomics 9:811–823, 2010.

The analysis of protein-protein interactions (PPIs)** and protein complexes is of central importance to biological research and facilitates our understanding of how molecular events drive phenotypic outcomes. Moreover, large scale protein interaction data can be used to generate protein interaction networks, which can then be used to predict disease genes and model biology in any living organism.

A number of methods (e.g., yeast two-hybrid) have been developed to examine binary protein interactions in a systematic format and applied to model systems (1–8). However, affinity purification (AP) coupled with tandem MS has become the method of choice for the identification of protein complexes (9, 10). Large scale PPI studies using a high throughput and systematic AP-MS approach have been performed for Escherichia coli (11, 12) and Saccharomyces cerevisiae (13–15). In fact, large scale efforts using AP-MS have connected an estimated 60% of the yeast proteome, demonstrating the power of coupling systematic biochemical purifications with mass spectrometry (13–16).

AP-MS has also been used extensively for purification of mammalian protein complexes (17), but this has been mostly restricted to small scale studies and the use of either cell lines that are easy to transfect or highly validated antibodies against specific targets. For example, Glatter et al. (18) recently developed an integrated workflow where a high density interactome was developed for the protein phosphatase 2A complex. This workflow relies on “flip-in” technology to introduce transgenes into a common genomic site in HEK293 cells and, similar to work by other groups (19, 20), utilizes an
human embryonic kidney; MEF, mouse embryonic fibroblast; shRNA, short hairpin RNA; TEV, tobacco etch virus; CMV, cytomegalovirus; hPGK, human phosphoglycerate kinase; mPGK, mouse phosphoglycerate kinase; Ef1α, elongation factor 1α; Gw, Gateway; qRT-PCR, quantitative RT-PCR; puro, puromycin; hygro, hygromycin; rTA, reverse tetracycline-controlled transactivator; pLd, plasmid lentiviral destination; pLx, plasmid lentiviral expression; PAF, RNA polymerase II-associated factor; NELF, negative elongation factor; P-TEFb, positive transcription elongation factor b; SWI/SNF, switch/sucrose non-fermentable; MLL, mixed lineage leukemia; VM, V5-Myc; Bia-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxyethyl)propane-1,3-diol; AQUA, absolute quantification. See supplemental Tables 1–7 for gene names.

This is an open access article under the CC BY license.

Molecular & Cellular Proteomics 9.5 811
inducible promoter to control expression levels of bait proteins (18). Unfortunately, the utility of these approaches is not easily extended to multiple cell types, including primary cells, and a few selected cell types are almost certainly insufficient to recapitulate all biologically relevant protein interactions in mammals. Many protein interactions occur dynamically in distinct cellular contexts and vary with a multitude of factors (e.g., embryonic development, tissue type, cell cycle phase, nutrient availability, etc.) that affect epigenetic regulation. Therefore, an efficient strategy for systematic identification of PPI by AP-MS in multiple mammalian cell types (e.g., primary diploid and diseased cells) with the potential for integration into a high throughput workflow would be valuable for mapping mammalian protein interaction networks.

Deciphering the chromatin code is arguably the next important milestone in biology. Understanding how all genes are transcribed and regulated in an epigenetic manner will generate cell- and tissue-specific genomic profiles that connect genotype to phenotype. This applies particularly to stem cell biology where somatic cells can be converted to pluripotent cells in a patient-specific manner, providing the raw materials for regenerative medicine. The rapid advances in stem cell research motivated us to develop a system to identify PPI in virtually any mammalian cell type. To this end, we developed an integrated strategy for mammalian functional proteomics with the following features in mind: 1) applicability to most mammalian cell types, 2) compatibility with publicly and commercially available cDNA libraries, and 3) versatility with regard to various affinity purification schemes. To accommodate these features, we combined lentiviral technology (21–23), Gateway™ cloning technology (24), and a unique affinity purification tag including a built-in reference peptide. We then established a functional proteomics workflow for AP-MS. We leveraged this workflow for more than 20 target proteins and multiple cell types, including human cells and primary mouse cells, and benchmarked its utility for identifying PPI and protein complexes related to transcription and chromatin modification.

### EXPERIMENTAL PROCEDURES

#### Cell Culture

HEK293 and HEK293T cells were cultured in Dulbecco’s modified Eagle’s medium with 10% fetal bovine serum and antibiotics as described previously (22, 25). Mouse R1 embryonic stem cells were maintained on feeder cells or expanded on gelatin-coated tissue culture plates as described previously (26).

#### Plasmid Construction

Please refer to supplemental Table 4 for a complete list of plasmids used in this study and supplemental Table 5 for primer sequences used to generate the plasmids in this study.

**Gateway-compatible Entry Clones**—Gateway-compatible entry clones (supplemental Table 4) were obtained by: 1) human ORFeome library (Open Biosystems), 2) UltimateORF collection (Invitrogen), 3) PCR amplification from Mammalian Gene Collection clones to create entry clones into the pDONR223 construct using Gateway BP Clonase enzyme mixture (Invitrogen) according to the manufacturer’s protocol, and 4) BP reaction with pMXs (Addgene) expression clones. All entry clones were sequence-verified in full.

**Plasmid Lentiviral Destination (pLD) Vectors**—pLD vectors were all constructed by cloning using T4 DNA ligase (New England Biolabs) according to the manufacturer’s protocol. The EF1α promoter was PCR-amplified from pEF-myc-mito (Invitrogen) and cloned into pGateway 5’CMV5’triple FLAG and pGateway 5’CMV3’triple FLAG to generate pGateway 5’EF1α5’triple FLAG and pGateway 5’EF1α3’triple FLAG. The EF1α promoter, the Gateway cassette, and either the N- or the C-terminal 3×FLAG tag were PCR-amplified with primers Ndel-EF1α-Gw and EF1a-Ntag-Gw-BstB1_2 or Ndel-EF1α-Gw and EF1a-Gw-Tag-C_tag_BstB1_2 from pGateway 5’EF1α5’triple FLAG and pGateway 5’EF1α3’triple FLAG, respectively, and cloned into the pLKO.1 vector (22) using NdeI/BstBI to generate pLD-puromycin resistance-EF1α, N-terminal triple Flag (pLD-puro-EnF) and the corresponding C-terminal triple FLAG construct (pLD-puro-EcF). To efficiently clone in various affinity tags in replacement of the N terminus (5’) 3×FLAG tag flanking Nhel/Agel was introduced by first PCR amplifying the 3×FLAG-His6 (FH) tag2 with primers Pmll_Nhel_FH_F and Hindlll_Agel_FH_R. Then, the PCR-amplified 3×FLAG-His6 tag was digested with Pmll/Hindlll and cloned into a pGateway 5’EF1α3’ triple FLAG first digested by Kpnll followed by blunting with DNA polymerase I large (Klenow) fragment (New England Biolabs) according to the manufacturer’s protocol and then digested with Hindlll to generate pLD-puro-EnF. N- and C-terminal versatile affinity (VA) tags were gene-synthesized (Bio Basic Inc.) with flanking Nhel/Agel and Xbal/BstBI sites, respectively, to allow for subcloning from the pUC57 host vector into pLD-puro-EnF and pLD-puro-EcF to generate pLD-puro-EnVA and pLD-puro-EcVA, respectively. The CMV promoter from the pLJm1 vector (Addgene) was PCR-amplified using Ndel_hCMV_F and either Nhel_hCMV_R or Mll_hCMV_R and cloned into pLD-puro-EnVA and pLD-puro-EcVA using NdeI/Nhel or NdeI/MluI to generate pLDPuro-EnF. N- and C-terminal versatile affinity (VA) tags were gene-synthesized (Bio Basic Inc.) with flanking Nhel/Agel and Xbal/BstBI sites, respectively, to allow for subcloning from the pUC57 host vector into pLD-puro-EnF and pLD-puro-EcF to generate pLD-puro-EnVA and pLD-puro-EcVA, respectively. The CMV promoter from the pLJm1 vector (Addgene) was PCR-amplified using Ndel_hCMV_F and either Nhel_hCMV_R or Mll_hCMV_R and cloned into pLD-puro-EnVA and pLD-puro-EcVA using NdeI/Nhel and NdeI/MluI to generate pLD-puro-CnVA and pLD-puro-CcVA, respectively. pLD-hygro-EnVA was generated by subcloning the hygromycin resistance gene from the pLJm6 vector into the pLD-puro-CnVA and pLD-puro-CcVA, respectively.

For generation of the pLD-puro-TcVA and pLD-puro-TrVA vectors, the rtTA2 vector was PCR-amplified with primers Nhel_rTAF_F and Pmll_rTA_R and cloned into pLKO_TRC901 vector (Broad Institute of the Massachusetts Institute of Technology and Harvard University) using Nhel/Kpnll. PCR-amplified hPGK-pac-2A-rtTA2 was then PCR-amplified using the primers hPGK-5′ and NsiI_rTA and cloned into pLDPuro-TrVA and pLDPuro-TcVA using SfiI/NsiI to generate pLD-puro-2A-rtTA-TrVA and pLD-puro-2A-rtTA-TcVA, respectively. All DNA fragments cloned into the pLKO.1 backbone vector were sequence-verified in full.

**Plasmid Lentiviral Expression (pLex) Vectors**—pLex vectors were constructed using the Gateway LR Clonase II enzyme mixture (Invitrogen catalog number 11791-020) according to the manufacturer’s protocol between Gateway-compatible entry clones and pLD vectors.

plX vectors were sequence-verified.
Stable Cell Lines

Lentivirus was produced and used to infect either HEK293 or R1 cells at a multiplicity of infection <1 as described previously (22). Transduced cells were selected with puromycin (Sigma) at a concentration of 1 μg/ml for R1 cells and 2 μg/ml for HEK293 for a minimum of 48 h.

Western Blots and Antibodies

Cells were subjected to high salt lysis buffer (10 mM Tris-HCl, pH 7.9, 10% glycerol, 420 mM NaCl, 0.1% Tris-HEPES, 1% DTT, 10 mM NaF, 0.25 mM Na2VO4, and 1× protease inhibitor mixture (Sigma)) by three freeze-thaw cycles as described previously (27) or by radioimmune precipitation assay buffer (50 mM Tris-HCl, pH 7.4, 1% Nonidet P-50, 0.25% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA, 10 mM Na2VO4, 10 mM sodium pyrophosphate, 25 mM NaF, and 1× protease inhibitor mixture (Sigma)) followed by centrifugation at 14,000 rpm for 1 h at 4 °C to remove insoluble material. 20–100 μg was separated by either a 10% SDS-PAGE gel or a NuPage® Novex® Bis-Tris 4–12% SDS-PAGE gel (Invitrogen) and transferred to nitrocellulose or PVDF membranes. Transferred samples were immunoblotted with primary antibodies (supplemental Table 6), followed by incubation with horseradish peroxidase-conjugated goat anti-mouse or goat anti-rabbit secondary antibodies (Santa Cruz Biotechnology, Inc.). Western blot detection was performed using enhanced chemiluminescence (GE Healthcare). The intensities of protein bands were quantified by Fluor-S™ (Bio-Rad).

Affinity Purifications

Lysates for affinity purifications were prepared from 5 × 15-cm plates (HEK293 samples) or 2.5 × 15-cm plates (mouse R1 cells) from stable transgenic cells generated by mammalian affinity purification and lentiviral expression (MAPLE). HEK293 stable transgenic cells were lysed in high salt lysis buffer (described above), and R1 stable transgenic cells were lysed using 50 mM HEPES-KOH, pH 8.0, 0.1% protease inhibitor mixture (Sigma) by three freeze-thaw cycles. Cells were incubated on ice for a minimum of 30 min and centrifuged at 14,000 rpm at 4 °C for 15 min to 1 h to remove insoluble material. All MAPLE-generated HEK293 samples were purified with FLAG followed by His purifications. MAPLE-generated R1 samples were purified by a sole FLAG purification and by Strept-Tactin followed by FLAG purifications. FLAG purifications were performed as described previously (25). For the FLAG-His purifications, cell lysates were incubated with FLAG M2-agarose beads (Sigma) at 4 °C for 4 h and washed with a low salt lysis buffer followed by tobacco etch virus (TEV) protease cleavage buffer (20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 0.1% Nonidet P-40, and 0.1 mM EDTA) followed by incubation with 0.1 mg/ml TEV protease combined with 2 μg of 3×FLAG peptide (Sigma) at 4 °C overnight. The TEV protease-cleaved products were further incubated with nickel-nitrilotriacetic acid-agarose (Qiagen) at 4 °C for 4 h. After washing with nickel-nitrilotriacetic acid buffer (20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 5 mM imidazole, and 0.1 mM EDTA), proteins were eluted with 500 mM ammonium hydroxide (pH >11). Strept-Tactin-FLAG purifications were performed by incubating cell lysates with Strept-Tactin-Sepharose resin (IBA) at 4 °C for 4 h followed by washes with a low salt wash buffer (10 mM HEPES-KOH, pH 8.0, 100 mM KCl, 0.1% Nonidet P-40, 0.5 mM, and 1 mM DTT). VA-tagged baits were eluted from the Strept-Tactin-Sepharose with 10 mM d-biotin (Sigma) for 30 min and then subjected to the same protocol as FLAG purifications.

Mass Spectrometry (LC-MS/MS)

For affinity-purified baits from MAPLE-generated cell lines in HEK293, half of the affinity-purified sample (equivalent to 2.5 × 15-cm plates of HEK293 cells) was precipitated by TCA (final concentration, 20%; Sigma) at 4 °C for overnight followed by cold acetone washing. Samples were then subjected to reduction reaction with 2 mM tris(2-carboxyethyl)phosphine HCl at room temperature for 45 min followed by alkylation reaction with 10 mM iodoacetamide in the dark for 40 min. After the addition of CaCl2 (final concentration, 1 mM), proteins were tryptically digested by using a Sigma Singles kit (T5757) according to the manufacturer’s instructions at 37 °C with gentle shaking (110 rpm) overnight. Digestion was terminated by 1% formic acid (Fluka). 18 μl of 100 μl of digested sample was loaded on a microcolumn using the EASY-nLC system (Proxenon, Odense, Denmark). The microchromatography column was constructed in a 120-mm × 75-μm tip pulled with a column puller (Sutter Instrument, Novato, CA) and packed with 3-μm Luna C18(2) stationary phase (Phenomenex, Torrance, CA). The organic gradient was driven by the EASY-nLC system over 105 min using buffers A and B (98% buffer A (95% water, 5% acetonitrile, and 0.1% formic acid) to 90% buffer B (95% acetonitrile and 0.1% formic acid in water) over 45 min) at a flow rate of 300 nl/min. The gradient was held at 2% B for 1 min followed by a 2-min increase to 6% B, 76-min increase to 26% B, 5-min increase to 90% B, 5-min hold at 90% B, 1-min decrease to 2% B, and 8-min hold at 2% B. Eluted peptides were directly sprayed into an LTQ linear ion trap mass spectrometer (ThermoFisher Scientific, San Jose, CA) using a nanospray ion source (Proxeon). A spray voltage of +2.5 kV was applied. The mass spectrometer was programmed with Xcalibur 2.0 software such that one precursor survey scan was performed for a mass range of m/z 400–2000 followed by three data-dependent MS/MS scans selected based on the three most abundant precursor ions and a precursor signal threshold of 500 counts. The exclusion list was enabled to exclude a maximum of 500 ions for 60 s. Overall, there were three biological replicate samples as well as two technical replicate samples to yield six data sets for each bait. Samples were randomized in their analysis so that cross-contamination would be filtered out during data analysis. Moreover, a 30-min wash step was applied between each sample to reduce cross-contamination.

For MS analysis of affinity-purified samples from MAPLE-generated R1 lines, samples were lyophilized in a SpeedVac and trypsin-digested in 50 mM ammonium bicarbonate, pH 8 (0.75 μg for 16 h followed by 0.25 μg for 2 h). The ammonium bicarbonate was removed by SpeedVac, and the samples were resuspended in buffer A (2% acetonitrile and 0.1% formic acid). Then, samples were individually and directly loaded onto capillary columns packed in house with Magic C18 AQ (5 μm, 100 Å). MS/MS data were acquired from a ThermoFinnigan LTQ equipped with a Proxenon NanoSource and an Agilent 1100 capillary pump via data-dependent mode (over a 2-h–20% acetonitrile gradient).

Analysis of LC-MS/MS Data

For MAPLE-generated data from HEK293 cells, RAW files were extracted with the extractms program and submitted to database.
search using SEQUEST v2.7 and a modified IPI_HUMAN database (version 3.53; 73,748 entries). The modification consisted of adding BSA (Swiss-Prot accession number P02769), GFP (Swiss-Prot accession number P42212), TEV (Swiss-Prot accession number P04517), streptavidin (Swiss-Prot accession number P22629), the beacon peptide (amino acid sequence, ELFNLLGENQPPVVK), and the reverse sequences of all entries, resulting in a total database size of 147,506 entries. Search parameters were set to allow for one missed cleavage site and one fixed modification of +57 for cysteine using precursor and fragment ion tolerances of 3 and 0 m/z, respectively. Protein hits were filtered using the StatQuest program with a confidence level of 99%. Spectral counts were normalized using normalized spectral abundance factors (28) based on protein length and sum of the spectral counts to enable comparison of protein levels across different runs or within a single run. There were 21 independent cell lines (i.e., 19 VA-tagged baits and enhanced green fluorescent protein (eGFP) and no tag controls) from which 1916 prey proteins were identified. A two-tailed t test was performed on the set of 1916 prey proteins identified by 19 independent baits versus the eGFP and no tag controls to filter background contaminants. Significant p values (p < 0.05) highlighted bait enrichment, and thus these proteins were included in the list. This resulted in a list of 222 confident prey proteins. To add further stringency, prey identified in all three purifications were included in the final analysis set. Lastly, isoforms were removed, and 62 high confidence prey proteins remained. The normalized spectral counts were averaged over the replicate runs to produce an average normalized spectral count for hierarchical clustering. A matrix containing one column for each bait with all its associated prey was produced for all 19 baits. The matrix was clustered using hierarchical clustering with average linkage distance and visualized in Treeview.

For MAPLE-generated data from R1 cells, RAW files were converted into mgf format and were searched using the Mascot search engine (Matrix Sciences) against the Mouse_RefseqV32 database (version 32) in which 35,188 entries were searched. The search parameters included a precursor ion mass tolerance of 3.0 Da and a fragment ion mass tolerance of 0.6 Da. Search parameters were set to allow two missed cleavages and methionine oxidation as a variable modification (fixed modifications were not applicable). Common contaminants associated withFLAG purifications, frequent flyers in mass spectrometry analyses (27), and protein hits found in the VA-tagged GFP samples were removed from the protein hit list. Only protein hits that were detected in all biological replicates, that had a Mascot score greater than 60, and with at least one unique peptide are reported in supplemental Table 3.

**RNA Interference Experiments**

R1 cells were used to validate lentiviral shRNAs and were infected as described (22). Validation of shRNAs was performed for three biological replicates. The following RNAi Consortium (Sigma-Aldrich) clones were used: shGFP;shGFP-1, shKlf4;shKLF4-1826, shBrg1-1; shBrg1-2;shBrg1-3089, shBrm-1;shBrm-3320, and shBrg1-2;shBrg1-3089. Knockdowns were allowed to occur for 5 days down to occur before the addition of doxycycline. After 7 days, cells were fixed with 4% paraformaldehyde (electron microscopy grade; Electron Microscopy Sciences) and stained with the alkaline phosphatase substrate kit I (Vector Laboratories) according to the manufacturer’s protocol. Reprogramming efficiency was scored by counting alkaline phosphatase-positive colonies for three independent assays.

**RESULTS**

**MAPLE System**—The MAPLE workflow is outlined in supplemental Fig. 1, and its key features are illustrated in Fig. 1a. Briefly, a custom lentiviral plasmid is used to introduce an affinity purification tag onto a given ORF using Gateway recombinant cloning technology (see “Experimental Procedures”). The resulting lentiviral expression constructs are then packaged into lentivirus expression particles, which are used to transduce various types of target cells to create stably expressing cell lines. Protein lysates derived from expanded cell lines are utilized for affinity purifications. Affinity-purified samples are then processed for “gel-free” peptide shotgun sequencing by LC-MS/MS (27). The resulting spectra are used to search sequence databases to identify co-purifying proteins. Because of the broad host range of lentiviruses (22), this procedure can be used to identify PPI and protein complexes in any mammalian cell type that can be amplified to obtain a sufficient amount of cell lysate.

**VA Tag**—Several individual and dual affinity tag combinations incorporating 3×FLAG, His6, Protein G, or Strep III were assessed in a lentiviral context by fusing each tag with the eGFP at either its N or C terminus and determining performance by Western blot analyses, fluorescence microscopy, and binding to appropriate resins (data not shown). Although all the tags that we constructed were functional (as tested by the above mentioned assays), we wanted a flexible solution that could accommodate multiple purification schemes. This motivated us to construct a novel ~12-kDa triple affinity tag.

**Reprogramming Assay**

Secondary mouse embryonic fibroblast (MEF) lines 1B and 6C were maintained and induced to reprogram as described previously (29). Briefly, 2.5 × 10^4 MEFs were plated into 12-well plates for the reprogramming assay when overexpressing GFP, Brg1, and Brm by MAPLE. 1.5 × 10^4 MEFs were plated into 12-well plates for reprogramming assays after infection with lentiviral shRNAs. We allowed 72 h after lentivirus removal for expression of transgenes or knockdown to occur before the addition of doxycycline. After 7 days, cells were fixed with 4% paraformaldehyde (electron microscopy grade; Electron Microscopy Sciences) and stained with the alkaline phosphatase substrate kit I (Vector Laboratories) according to the manufacturer’s protocol. Reprogramming efficiency was scored by counting alkaline phosphatase-positive colonies for three independent assays.

**cDNAs were produced by first strand synthesis from 2 μg of total RNA according to the manufacturer’s instructions (Invitrogen catalog number 11754). Real time PCR using primers for each gene (supplemental Table 7) was performed on a 2-μl aliquot from a total of 400 μl of cDNA with the SYBR Green kit (Fermentas catalog number K0221) using the 7300 Real Time OGR System (Applied Biosystems) in a 10-μl volume in duplicate. PCR consisted of 40 cycles of 95 °C for 15 s and 55 °C for 30 s. A final cycle (95 °C for 15 s and then 60 °C) generated a dissociation curve to confirm a single product. The cycle number required to reach a threshold in the linear range (Qf) was determined and compared with a standard curve for each primer set generated by five 3-fold dilutions of genomic DNA samples of known concentration. Values were normalized to β-actin. The copy number was determined based on the standard curve generated by running known concentrations of the genomic DNA (1 ng of DNA = 300 copies).**
termed the VA tag, that includes 3×FLAG, His6, and Strep III (30) epitopes for the following reasons. First, 3×FLAG has been widely used, is small in size, and is amenable to immunofluorescence. Second, His6 is widely used and allows for protein purification under denaturing conditions. Third, Strep III is highly selective with little nonspecific binding and efficiently binds to desthiobiotin and biotin for elution (30). The VA tag also contains a dual TEV protease cleavage site (31) and a unique, yeast-derived, high responding proteotypic peptide (ELFNLLGENQPPVVIK) that serves as a molecular “beacon” (32) during mass spectrometry (Fig. 1a).

Importantly, all three epitopes of the VA tag were easily detected on either the N or C terminus of eGFP by immunoblot (Fig. 1b), and VA-GFP localized predominantly to the cytoplasm by GFP fluorescence and anti-FLAG immunofluorescence (Fig. 1c). Furthermore, all three epitopes in the VA tag were capable of being captured on appropriate resins with little to no nonspecific adsorption to Protein G beads (Fig. 1d).

To confirm bait retrieval, we monitored the abundance of the bait-derived beacon peptide by LC-MS/MS using a heavy
stable isotope-labeled synthetic (i.e., AQUA-labeled) reference peptide spiked into the sample digest as an internal standard (supplemental Fig. 2, a and b). As expected, both the unlabeled (m/z 905.4) and AQUA-labeled (m/z 909.4) beacon peptides co-eluted during liquid chromatography (Fig. 1z). The corresponding MS/MS spectra of these precursors revealed a characteristic shift in mass of reporter y-ions (e.g., m/z 652.5 → 660.5), demonstrating the potential utility of the AQUA-labeled standard for identifying bait proteins (Fig. 1f).

**MAPLE Is an Effective System for Identifying Specific PPIs and Protein Complexes**—To examine an entire complex by reciprocal tagging and validate MAPLE for the characterization of human complexes, we tested the evolutionarily conserved RNA polymerase II-associated factor (PAF) complex as it has been purified previously from human cells and serves as a good control to assess our MAPLE workflow (33, 34). The PAF complex is involved in mediating efficient transcription elongation by RNA polymerase II, mRNA quality control, and chromatin modification that is coupled to transcription elongation (35). The human and yeast PAF complexes contain PAF1, CDC73, CTR9, and LEO1, although the human complex additionally contains SKI8/WDR61 and appears to lack the RTF1 subunit stably associated with the yeast complex at least in some cell types (33, 34, 36). Each of the subunits of the PAF core complex, PAF1, CDC73, CTR9, and LEO1, and RTF1 were VA-tagged and subjected to MAPLE followed by LC-MS/MS to identify a high confidence network of reciprocal protein interactions (18). Importantly, all of these tagged baits were localized to previously reported subcellular compartments, migrated at their predicted molecular weights, and were present at levels comparable with or below the corresponding endogenous proteins (Fig. 2, a and b). Baits were purified on anti-FLAG and nickel resins and analyzed either by SDS-PAGE followed by silver staining or by trypsin digestion and tandem mass spectrometry to identify potential interacting protein partners. Untagged CDC73 and VA-tagged eGFP were used as negative controls to generate nonspecific background profiles necessary for filtering out common contaminants. Each bait protein was purified from three independent cultures and analyzed twice by data-dependent LC-MS/MS, producing a total of six sample runs per bait.

To score protein-protein interactions, we considered reproducibility and background, and an enrichment score was determined using a combination of protein length-normalized spectral counts and filtering criteria (see below and “Experimental Procedures”). The resulting values were scaled between 0 and 0.01 where 0 represents no prey detected and values over 0.005 were deemed highly significant. The results show that all core members of the human PAF complex consistently co-purified with each of the PAF baits with high enrichment scores (Fig. 2c, supplemental Fig. 5, and supplemental Tables 1 and 2). Consistent with the observation by Zhu et al. (34), SKI8/WDR61 consistently co-purified with all core members of the PAF complex including PAF1, LEO1, CTR9, and CDC73 and served as a good positive control for our scoring schema and the overall workflow (Fig. 2c). Also as expected (34), none of the PAF core components were observed in RTF1 affinity purifications (Fig. 2c). As expected, several preys were weakly captured in common between RTF1 and two PAF-related baits, LEO1 and PAF1, including C17orf79, an uncharacterized human open reading frame, suggesting that there may be some association among these proteins in human cells (supplemental Tables 1 and 2). Overall, the MAPLE workflow permitted the efficient identification of authentic protein complexes in human cells and their rapid validation by reciprocal tagging.

**Systematic Analysis of Protein Complexes Involved in Transcription and Chromatin Modification**—As further validation of the MAPLE workflow, systematic examination of protein complexes involved in various aspects of transcription was performed. The reciprocal tagging strategy used for the PAF complex was again applied to several previously documented multiprotein complexes linked to positive and negative regulation of transcription elongation and chromatin remodeling, including the positive transcription elongation factor b (P-TEFb) (37), negative elongation factor (NELF) (38, 39), mixed lineage leukemia (MLL) (40, 41), and SWI/SNF (42) complexes (Fig. 2c, supplemental Fig. 5, and supplemental Tables 1 and 2). A total of 17 proteins representing subunits of the PAF complex and four other complexes as well as PTEN and JUNB were built as baits. Each stably expressing cell line was analyzed by immunoblot and immunofluorescence to confirm that each tagged protein migrated at its predicted molecular weight and localized to the correct subcellular compartment. Each of the baits was purified from HEK293 cells, and its putative interaction partners were identified by LC-MS/MS as described above.

For the NELF complex, all four members of the complex (COBRA1, WHSC2, TH1L, and RDBP) were readily detected as baits and consistently interacted with each other as preys in reciprocal affinity purifications except that TH1L and RDBP did not co-purify when COBRA1 was used as bait (Fig. 2c). This suggested that the tag on COBRA1 interfered with its association with TH1L and RDBP and highlighted the value of performing reciprocal experiments with MAPLE to maximize subunit coverage for multiprotein complexes.

The P-TEFb complex, containing the cyclin-dependent kinase CDK9 and its cyclin partner (37), also performed well in reciprocal purifications (Fig. 2c). In addition, CDK9 affinity purifications revealed novel potential interactors with obvious functional consequences. One was the CDC37 cochaperone that promotes the association of HSP90 with its protein kinase subset of client proteins to maintain their stability and signaling functions (43). CDC37, HSP90AB1, and HSP90AA2 all co-purified with CDK9 as bait, suggesting that CDK9 could be another client of the CDC37-HSP90 complex.
Likewise, the three members of the MLL histone methyltransferase complex that served as baits (ASH2L, WDR5, and RBBP5) reciprocally co-purified with each other (Fig. 2c). The MLL proto-oncogene is a recurrent site of genetic rearrangements in acute leukemias (40). The MLL gene is the founding member of the mammalian SET family of histone-lysine methyltransferases that are responsible for regulating gene expression patterns during development (40). The evolutionarily conserved protein DPY30 also co-purified with each of these baits, consistent with previous evidence that it forms a subcomplex with the ASH2L, RBBP5, and WDR5 proteins that are shared by all human Set1-like histone methyltransferase complexes (44, 45). CXXC1, a protein that recognizes CpG sequences, also co-purified with ASH2L, WDR5, and RBBP5 and is also known to interact with the Set1 histone H3K4-specific methyltransferases in the regulation of MLL target genes (46). Consistent with the notion that ASH2L, WDR5, and RBBP5 form a subcomplex consistently present in MLL complexes, MAPLE purifications revealed interactions with other members of the Set1/COMPASS and MLL complexes, including SETD1A, SETD1B, HCFC1, HCFC2, UTX/KDM6A, MLL, MLL3, and MLL4 (Fig. 2c). Further investigation into the nature of these interactions could help refine the composition of these complexes.

The last complex examined by MAPLE in HEK293 cells was the highly conserved SWI/SNF chromatin remodeling complex composed of seven core subunits and one of three different catalytic subunits (BRG1/BAF, BRM/BAF, or PBAF).
Three subunits common to BRG1/BAF, BRM/BAF, and PBAF, namely SMARCC2, SMARCE1, and SMARCD1, were subjected to the MAPLE workflow, and affinity-purified proteins were identified by LC-MS/MS. Affinity capture of all known SWI/SNF subunits was achieved for all three SWI/SNF baits (Fig. 2c), and some potential novel interactions were identified as well (Fig. 2c and supplemental Tables 1 and 2).

For example, DPF2 was identified with all three SWI/SNF baits (Fig. 2c). DPF2, also known as REQuiem or UBID4, is a member of the d4 domain family characterized by a zinc finger-like structural motif and may function as a transcription factor important for the apoptotic response (47). This suggests that DPF2 may regulate the role of SWI/SNF during apoptosis.

**MAPLE Synopsis**—In our initial assessment of 19 bait proteins by MAPLE, 1916 prey proteins were identified through LC-MS/MS with one or more spectral counts with a confidence of 99% (corresponding to an false discovery rate of 0.01). Enrichment scores were calculated for each potential prey, yielding a total of 62 high confidence preys representing 148 interactions (see supplemental Fig. 3 for filtering criteria). Unsupervised hierarchical clustering of the baits and high confidence preys indicates that the known subunits of the various complexes share a high measure of similarity and cluster together (Fig. 2c). This representation is also a good visualization tool to identify baits that may interact nonspecifically with multiple preys. As described above, our results show good concordance with what has been published in the literature and some overlap with database resources like CORUM, the comprehensive resource of mammalian protein complexes (Ref. 48) and see supplemental Fig. 4). Based on these analyses, we conclude that using MAPLE for reciprocal tagging of known components of large complexes tends to capture the biological diversity of each complex. To examine whether the MAPLE workflow could be applied in a similar parallel followed by LC-MS/MS to identify candidate interacting partners.

Several novel KLF4 protein interactors were identified, including the catalytic subunits SMARCA4/BRG1 and SMARCA2/BRM of the SWI/SNF chromatin remodeling complex (supplemental Table 3 and supplemental Fig. 6). Using a co-immunoprecipitation assay, the interaction between the KLF4 bait and endogenous SMARCA4/BRG1 was validated in mouse embryonic stem cells (Fig. 3c). The interaction between KLF4 and SMARCA2/BRM was validated by co-immunoprecipitation in HEK293 cells (Fig. 3d). In addition, the c-Myc bait was able to co-immunoprecipitate endogenous SMARCA4/BRG1 in mouse ES cells (Fig. 3c), validating a KLF4-SMARCA4/BRG1 interaction previously reported in a high throughput AP-MS study in human cancer cells (17). Taken together, these data indicate that KLF4 interacts, directly or indirectly, with the catalytic subunits SMARCA4/BRG1 and SMARCA2/BRM of the SWI/SNF chromatin remodeling complex in pluripotent stem cells.

**Requirement for SWI/SNF during Induction of Pluripotency**—A direct link between the SWI/SNF chromatin remodeling complex and somatic cell reprogramming has not been clearly established. Because Klf4 is important for induced pluripotency (50, 51) and we observed an association between KLF4 and the SWI/SNF complex in mouse ES cells, the
question is raised whether the SWI/SNF-mediated chromatin remodeling contributes to the process of induced pluripotency. To test this idea, we used an established reprogramming assay comprising secondary MEFs that can be converted into induced pluripotent stem cells by addition of doxycycline (29). To examine the efficiency and kinetics of reprogramming in the secondary MEF lines 1B and 6C after doxycycline was added to induce reprogramming, samples were taken to examine expression of pluripotency markers. In parallel, high resolution video microscopy was performed on 6C cells to observe induced pluripotent stem colony formation (supplemental Fig. 7). The expression of the pluripotency markers Oct4, Nanog, Eras, Zfp42, Nrob1, and Fbx15 was evident 7 days following addition of doxycycline, whereas the expression of Sox2 and Foxd3 was delayed but established by day 15 (Fig. 4a and data not shown). The number of alkaline phosphatase-positive colonies was manually counted by three independent observers, and results were averaged. Error bars represent standard error (n = 3).
To assess the requirement of Smarca4/Brg1 and Smarca2/Brm during reprogramming, we perturbed their expression levels in either 6C or 1B cells by using lentiviral cDNAs (overexpression) or short hairpin RNAs (knockdown) (supplemental Fig. 8). Cells were induced to reprogram, and the efficiency was scored based on the number of alkaline phosphatase-positive colonies after 7 days. Lentivirus-mediated expression of GFP or the pLKO.1 vector did not significantly alter the number of alkaline phosphatase-positive colonies formed by 6C cells by 2-fold (p < 0.05) and the number of alkaline phosphatase-positive colonies formed by 1B cells by 7-fold (p < 0.01) (Fig. 4c). Furthermore, knockdown of Smarca4/Brg1 or Smarca2/Brm with either of two independent shRNAs reduced the number of alkaline phosphatase-positive colonies formed by 6C cells by 2-fold (p < 0.02) (Fig. 4d). Taken together, these results indicate that the catalytic subunits of the SWI/SNF chromatin remodeling complex, likely in association with KLF4, are important for somatic cell reprogramming.

**DISCUSSION**

The goal of this study was to design a system to facilitate identification of protein complexes in a wide variety of mammalian cell types. To this end, we developed MAPLE and coupled it to tandem mass spectrometry to identify protein complexes. MAPLE works efficiently with most cell types, thereby granting access to cell types that may better recapitulate the natural environment of a protein complex. This could be particularly important in cells and tissues that undergo substantive epigenetic regulation such as stem cells. MAPLE is based on a custom-built lentiviral plasmid that is Gateway-compatible and relies on commercially available cDNA libraries. A unique and versatile affinity tag (i.e., VA) was created to accommodate different purification schemes and to contain a novel yeast proteotypic peptide (i.e., beacon) that acts as a reference for the bait during affinity purifications. As such, the beacon serves as a “digital Western.”

A major disadvantage of mammalian AP-MS approaches is the amount of starting material (i.e., cells) required to achieve high protein purification yields. Although the original tandem affinity purification tag (52) is still routinely in use in mammalian cells for protein complex characterization, other affinity tags like the GS tag have helped reduce the amount of starting material required by 10–100-fold (53). With the MAPLE system, we typically used ~5e7 cells for starting material, similar to Bürststümmer et al. (53). The VA tag is compatible with single, dual, or even triple affinity purification schemes so the choice of method will impact how much starting material is required. That is, more purification steps require more starting material. The downfall of tandem purification methods is that transient or weak interactions are generally lost.

Furthermore, we introduced a constitutive promoter in our MAPLE system and demonstrated that the CMV and PGK promoters driving cDNAs in HEK293 and R1 cells, respectively, resulted in bait expression levels comparable with endogenous. However, some genes or cells may be sensitive to bait overexpression, resulting in false outcomes. Similar to other studies, we demonstrate that MAPLE can be adapted to be tetracycline-regulatable and -inducible (supplemental Fig. 9). Although determining an optimal level of bait expression would be ideal, it is not practical for a systematic and high throughput approach.

To benchmark the MAPLE workflow in cells that are easy to manipulate, we examined 19 bait proteins involved in different aspects of chromatin biology or disease in HEK293 cells. Except for the PTEN phosphatase and the JUNB proto-oncogene, each of the baits was part of an evolutionarily conserved multiprotein complex. PAF1, CDC73, CTR9, and LEO1 represent evolutionarily conserved core components of the human PAF complex involved in transcription elongation. CDK9 and CCNT1 constitute the P-TEFb complex. COBRA1, WHSC2, TH1L, and RDBP represent the core components of the NELF complex. WDR5, RBBP5, and ASH2L make up a key subcomplex that likely regulates all the SET-like histone methyltransferase complexes, including SET1 and all MLL proteins. Lastly, SMARCD1, SMARCC2, and SMARCE1 are common components of all three characterized SWI/SNF chromatin remodeling complexes, including BRG/BAF, BRM/BAF, and PBAF. Importantly, reciprocal tagging validated all the known protein complexes in our benchmarking study, and thus it represents a good strategy to capture new complex members. The application of MAPLE and LC-MS/MS across a broad range of baits in an unbiased manner will yield a high density protein interaction data set with a wealth of new biological information.

A key feature of the MAPLE system is that lentiviruses are efficient at stably transducing most cell types. By transducing cells at a low multiplicity of infection, one can limit the number of integrants per cell and keep cell-to-cell expression levels more constant. The fact that the resulting populations of cells are heterogeneous because of variability in integration is advantageous as this mitigates the possibility of bias due to clonal expansion. To put this to the test, we generated stable mouse R1 ES cells expressing VA-tagged KLF4, one of the classical reprogramming factors that convert somatic cells into induced pluripotent stem cells (51) and performed AP-MS using lysates from this stable cell line to identify protein interactions. We were able to show that KLF4 interacts with the SWI/SNF chromatin remodeling complexes containing SMARCA4/BRG1 and SMARCA2/BRM.

Consistent with this, recent studies have found that ES cells contain a functionally and structurally specialized chromatin remodeling complex, esBAF, that is critical for the self-renewal of ES cells and maintenance of the stem cell fate (54–57). In fact, Crabtree and co-workers (56) proposed that
the surface of esBAF complexes is tailored for interactions with factors found specifically in ES cells and that through these functional interactions esBAF maintains the pluripotent chromatin landscape. Using chromatin immunoprecipitation coupled with high throughput sequencing technology, Ho et al. (55) were able to demonstrate that esBAF is enriched at transcription start sites, occupies genes of the core pluripotency network (i.e. Oct4, Sox2, and Nanog), represses developmental genes, and opposes Polycomb complexes by direct repression of subunits of the PRC1 complex. Nevertheless, a direct role for esBAF or any of the SWI/SNF complexes in reprogramming or dedifferentiation had not been shown.

Because KLF4 has an important role during somatic cell reprogramming to induced pluripotent stem cells and our data uncovered an interaction between KLF4 and the catalytic subunits of the SWI/SNF chromatin remodeling complexes, an important remaining question was to examine the requirement of SWI/SNF chromatin remodeling complexes during somatic cell reprogramming. Thus, we turned to a model where secondary MEFs can be induced to pluripotency by the addition of doxycycline (29). The first clue that somatic cell reprogramming may have a slightly different SWI/SNF requirement than maintenance of the pluripotent state (55, 56) came from the observation that SMARCA2/BRM expression is induced during reprogramming/dedifferentiation (Fig. 4b). This observation may have important consequences for how SWI/SNF subunits are distributed during reprogramming versus maintenance of pluripotency or self-renewal. To further investigate this, the onset of the expression of pluripotency markers was examined in two independent clones (1B and 6C) that were induced into pluripotency following knockdown or overexpression of SWI/SNF catalytic subunits (i.e. SMARCA4/BRG1 or SMARCA2/BRM). These results indicate that slight changes in the concentrations of the catalytic subunits of the esBAF (or any SWI/SNF complex involved in reprogramming) complex have drastic consequences for reprogramming to the pluripotent state. The interaction between KLF4 and SWI/SNF complexes may help to establish the pluripotent transcriptional circuitry.

Acknowledgments—We thank Dr. András Nagy for kindly providing the R1 mouse ES cell line, the 1B and 6C secondary MEFs, and the pDONR221-Klf4 construct and Dr. Tony Pawson for the pGateway vectors. We thank Drs. Anthony Gramolini, Stéphane Angers, Janet Rossant, and Brian Cox for helpful discussions.*

* This work was supported in part by grants from the Canadian Institutes of Health Research (to J. M. (Grant 178975), J. G., and A. E.) and in part by a Network Centers of Excellence Stem Cell Network grant (to J. M.).

†† This article contains supplemental Figs. 1–9 and Tables 1–7.

¶¶ These authors contributed equally to this work.

§§ Supported by a Natural Sciences and Engineering Research Council Canada graduate scholarship.

** Supported by a Canadian Institutes of Health Research postdoctoral fellowship.


