



(51) International Patent Classification:
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/US2018/023683

(22) International Filing Date:
22 March 2018 (22.03.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/474,997 22 March 2017 (22.03.2017) US
62/481,257 04 April 2017 (04.04.2017) US

(71) Applicant: **THE ROCKEFELLER UNIVERSITY**
[US/US]; 1230 York Avenue, New York, NY 10021 (US).

(72) Inventors: **FUCHS, Elaine**; 430 East 63rd Street, Apartment 12a, New York, NY 10021 (US). **ADAM, Rene**; 303 East 89th, Apt. 3D, New York, NY 10128 (US). **GE, Yei-jing**; 504 E. 63rd St. Apt 29L, New York, NY 10065 (US).

(74) Agent: **NGUYEN, Quang, D.** et al.; Riverside Law, LLP, Glenhardie Corporate Center, 1285 Drummers Lane, Suite 202, Wayne, PA 19087 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

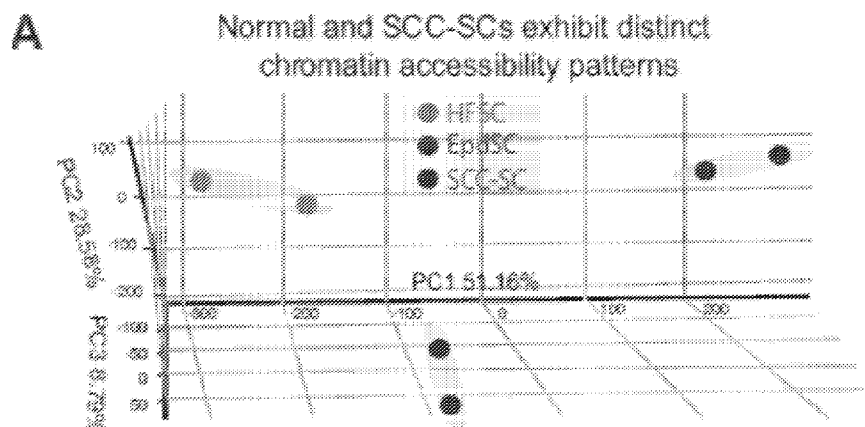
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: METHODS AND COMPOSITIONS FOR DETECTION AND TREATMENT OF CANCER



(57) Abstract: The present invention provides compositions and methods for diagnosing, treating and preventing cancer in an individual. The invention also provides methods for screening for a composition that modulates a desired epicenter.



TITLE OF THE INVENTION

Methods and Compositions for Detection and Treatment of Cancer

CROSS-REFERENCE TO RELATED APPLICATION

5 This application claims priority to U.S. Provisional Application Serial No. 62/474,997, filed March 22, 2017, and U.S. Provisional Application Serial No. 62/481,257, filed April 4, 2017, the contents of which are incorporated by reference herein in their entirety.

10 STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR
DEVELOPMENT

This invention was made with government support under NIH-AR31737 awarded by the National Institutes of Health. The government has certain rights in the invention.

15 BACKGROUND OF THE INVENTION

Human adult tissues harbor resident stem cells (SCs) responsible for homeostasis and wound repair. Tumorigenesis arises when normal SCs accumulate mutations that cause them to derail, shifting their homeostatic balance to favor tissue growth at the expense of differentiation. In contrast to wound-repair, where the growth:
20 differentiation imbalance is transient, cancers are refractory to tissue restoration cues, seemingly hijacking these normal cellular programs to fuel their molecular thirst for uncontrolled growth.

The notion that a “cancer is a wound that never heals” has its origins dating back to Rudolf Virchow in the mid 1800’s. Since then, tantalizing parallels
25 between cancer and wound have emerged in many contexts (Antsiferova and Werner, 2012; Arwert et al., 2012). For instance, it has long been recognized that human patients suffering from chronic wounds have increased susceptibility to cancers (Dunham, 1972; Haddow, 1972). Additionally, mice with gene mutations that enhance hair follicle SC activity heal their wounds faster, but they also exhibit enhanced susceptibility to
30 squamous cell carcinomas (Guasch et al., 2007; Hance et al., 2014). By contrast, mice whose skin possesses mutations that impede HFSC activation display reduced efficiency in wound closure, but also an increased resistance to cancers (Schober et al, 2007; Schober and Fuchs, 2011).

Intimate connections between wounds and tumors have also been drawn at the molecular level. Following serum stimulation, cultured fibroblasts elicit a robust wound repair signature resembling that of certain human carcinomas and predictive of poor patient prognosis (Chang et al., 2004; Iyer et al., 1999). Gene profiling studies in various wounded and tumorigenic epithelial tissues have further highlighted a concordant gene signature (Pedersen et al., 2003; Riss et al., 2006). Although intriguing, it remains unclear which of the normal SC remodeling pathways are exploited by tumor SCs and how cancers rewire pre-installed regulatory networks to support malignancy. The answers could be important in devising new and improved therapeutics for the treatments of chronic wounds as well as cancers.

Mouse skin offers an excellent genetically tractable model system to tackle these issues. Its epithelium has two distinct lineages, hair follicle (HF) and epidermis (Epd), each harboring their own resident SCs (Fuchs, 2016). HFSCs reside in a region of the follicle known as the bulge, and during normal homeostasis, their role is to fuel the cyclical bouts of HF regeneration and hair growth. By contrast, EpdSCs reside in the innermost (basal) layer of the epidermis, where they generate an upward flux of differentiating cells that produce the skin barrier.

Upon injury, both EpdSCs and HFSCs in the vicinity of the wound site are mobilized toward it, re-epithelializing the wound bed and restoring the skin barrier (Ito et al., 2005; Jensen et al., 2009; Levy et al., 2007; Tumber et al., 2004). Each lineage can also participate in cancer progression when its SCs acquire oncogenic HRAS mutations (Lapouge et al., 2011; White et al., 2011). At low levels, oncogenic HRAS drives the SCs to the hyper-proliferative and benign tumorigenic states; as RAS/MAPK levels rise, malignant, invasive squamous cell carcinomas (SCCs) develop (Rodriguez-Puebla et al., 1999). How stem cells acquire the plasticity that allows them to exit homeostasis and participate in wound-repair and malignant progression remains unknown.

Recent transcriptional and epigenetic chromatin landscaping of HFSCs and EpdSCs have suggested that chromatin dynamics and lineage commitment are governed by the signals emanating from the SC niche (Adam et al., 2015). Notably, HFSCs in the bulge niche are distinguished from EpdSCs in the basal epidermal layer by a cohort of transcription factors (TFs) that include SOX9, LHX2, TCF3/4, NFATc1, NFIB and FOXC1 (Blanpain et al., 2004; Morris et al., 2004; Tumber et al., 2004). SOX9 appears to be particularly important in governing the fate of skin SCs. Loss of function mutations in mice compromise HFSC function and result in the conversion of the SC niche into an

epidermal cyst (Kadaja et al., 2014). Conversely, ectopic SOX9 expression in the epidermis initiates the activation of other HFSC TFs (Adam et al., 2015).

During wound-repair, SOX9 is the only HFSC TF that remains expressed, albeit at reduced levels, as HFSCs mobilize and re-epithelialize the injured skin (Adam et al., 2015). Once HFSC progenies reach the epidermis and the wound is healed, SOX9 is no longer detected, indicative of a fate switch. Interestingly, SOX9 is also one of the few HFSC TFs that remain expressed in the tumor-initiating (stem) cells of SCCs (Lapouge et al., 2011; Schober and Fuchs, 2011). Despite shared expression of SOX9 in these states, the global gene expression patterns of HF-, Epd- and SCC-SCs remain strikingly different (Lapouge et al., 2011; Schober and Fuchs, 2011; Yang et al., 2015).

A major hurdle in digging deeper into SC plasticity in either wound-repair or cancer progression is the limiting supply of pure wound-induced or tumor-initiating SCs for transcriptome and chromatin landscaping. This is particularly the case for the wounded state, leaving us with a significant knowledge gap as to how cancers are wounds that do not heal and whether SC plasticity plays a role therein.

Therefore, there remains an unmet need for methods for identifying and regulating lineage transcription factors involved in the regulation of the transition of cells from homeostatic to cancerous states. The present invention satisfies these unmet needs.

SUMMARY OF THE INVENTION

The present invention describes a composition for treating cancer, wherein the composition comprises a modulator of a transcription epicenter, wherein the epicenter regulates cancer. In some embodiments, the modulator comprises an inhibitor of the activity of the transcription epicenter. In some embodiments, the modulator comprises an activator of the activity of the transcription epicenter. In some embodiments, the modulator modulates the activation of an epicenter regulator, wherein the regulator is selected from the group consisting of polymerases, acetyltransferases, histone deacetylases, methylases, histone demethylases, transcription factors, coactivators, corepressors, and enhancers. In some embodiments, the modulator modulates the activation of one or more transcription factors selected from the group consisting of SOX9, KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3.

In some embodiments, wherein the modulator comprises an inhibitor, the inhibitor is one or more molecules selected from the group consisting of a small interfering RNA (siRNA), a small guide RNA (gRNA), a microRNA, an antisense

nucleic acid, a ribozyme, an expression vector encoding a transdominant negative mutant, an antibody, a peptide, a chemical compound and a small molecule.

In some embodiments, wherein the modulator comprises an activator, the activator is one or more compound selected from the group consisting of a chemical
5 compound, a protein, a peptidomimetic, an antibody, a nucleic acid molecule.

In some embodiments, the composition comprises a modulator wherein the modulator binds a region of DNA, wherein the region of DNA comprising at least one binding transcription factor binding motif. In some embodiments, the transcription factor binding motif is selected from the group consisting of a SOX9 binding motif, a KLF5
10 binding motif, and a combination thereof. In some embodiments, the cancer is selected from the group consisting of squamous cell carcinoma and human head and neck squamous cell carcinoma.

In some embodiments, the invention relates to a method for treating cancer in a subject, the method comprising administering to subject in need thereof a modulator
15 of an epicenter, wherein the epicenter is accessible, and wherein the epicenter regulates cancer. In some embodiments, the modulator comprises an inhibitor of the activity of the transcription epicenter. In some embodiments, the modulator comprises an activator of the activity of the transcription epicenter. In some embodiments, the modulator modulates the activation of an epicenter regulator, wherein the regulator is selected from the group
20 consisting of polymerases, acetyltransferases, histone deacetylases, methylases, histone demethylases, transcription factors, coactivators, corepressors, and enhancers. In some embodiments, the modulator modulates the activation of one or more transcription factors selected from the group consisting of SOX9, KLF5, ETS, AP-1, GATA, GRHL, AP2, TCF, and NFI. In some embodiments, the inhibitor is one or more molecules selected
25 from the group consisting of a small interfering RNA (siRNA), a small guide RNA (gRNA), a microRNA, an antisense nucleic acid, a ribozyme, an expression vector encoding a transdominant negative mutant, an antibody, a peptide, a chemical compound and a small molecule. In some embodiments, the activator is one or more compound selected from the group consisting of a chemical compound, a protein, a peptidomimetic,
30 an antibody, a nucleic acid molecule. In some embodiments, the modulator binds a region of DNA, wherein the region of DNA comprising at least one binding transcription factor binding motif. In some embodiments, the transcription factor binding motif is selected from the group consisting of a SOX9 binding motif, a KLF5 binding motif, and a

combination thereof. In some embodiments, the cancer is squamous cell carcinoma. In some embodiments, the cancer is human head and neck squamous cell carcinoma.

In some embodiments, the invention relates to a method of diagnosing cancer in a subject, the method comprising the step of: (a) obtaining a sample from the
5 subject, (b) detecting epicenter activity relative to a comparator control in the sample, and (c) thereby diagnosing the subject with cancer. In some embodiments, the sample is a human sample. In some embodiments, the sample is skin tissue. In some embodiments, the cancer is squamous cell carcinoma. In some embodiments, the cancer is head and neck squamous cell carcinoma. In some embodiments, the comparator control is at least
10 one selected from the group consisting of: a positive control, a negative control, a historical control, a historical norm, or the level of a reference molecule in the biological sample.

In some embodiments, the invention relates to a method of identifying a cancer marker in a sample, the method comprising the steps of (a) classifying the
15 transcriptional state of the sample, (b) performing transcriptome analysis to identify modulated transcription factors, and (c) identifying a cancer marker. In some embodiments, the sample is a population of cells. In some embodiments, the cells are tumor stem cells. In some embodiments, the samples are sorted by transcriptional state wherein transcriptional state is characterized as significantly altered open chromatin state
20 relative to a comparator control. In some embodiments, transcriptionally sorted samples are statistically evaluated. In some embodiments, the sorted samples are mapped to genes. In some embodiments, the genes are classified as either induced or suppressed relative to a comparator control. In some embodiments, the comparator control is at least one selected from the group consisting of: a positive control, a negative control, a historical
25 control, a historical norm, or the level of a reference molecule in the biological sample. In some embodiments, the genes are subjected to pathway enrichment analysis. In some embodiments, the genes are screened for key regulatory elements. In some embodiments, the regulatory elements are mapped to transcription epicenters. In some embodiments, the epicenters are screened for genes corresponding to the regions containing the epicenters.
30 In some embodiments, the epicenters are linked to specific transcription factors. In some embodiments, the transcription factors are screened in databases.

In some embodiments, the invention relates to a composition for treating wounds comprising an inhibitor of SOX9.

In some embodiments, the invention relates to a method for treating a subject having a wound, the method comprising administering an inhibitor to a subject in need thereof and inhibitor of SOX9.

5 In some embodiments, the invention relates to a composition for detecting transcription epicenters comprising a plurality of probes and a plurality of promoters. In some embodiments, the plurality of probes comprises two probes. In some embodiments, the plurality of probes comprises a fluorescent protein. In some embodiments, the fluorescent protein is selected from the group consisting of RFP, iRFP and GFP. In some
10 embodiments, the plurality of promoters comprises two promoters. In some embodiments, the two promoters are selected from a group consisting of a H2B, SOX9, KLF5, PGK and miR21.

The invention also provides a method of screening for a composition that modulates a transcription epicenter, wherein the epicenter regulates cancer. In one embodiment, the method comprises: (a) administering an epicenter reporter construct into
15 cell, (b) detecting epicenter activity relative to a comparator control in the cell, and (c) thereby identifying a composition that modulate the epicenter.

BRIEF DESCRIPTION OF THE DRAWINGS

The following detailed description of preferred embodiments of the invention will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there are shown in the drawings embodiments which are presently preferred. It should be understood, however, that the invention is not limited to the precise arrangements and instrumentalities of the embodiments shown in the drawings.

Figure 1, comprising Figure 1A through Figure 1D, depicts results of example experiments demonstrating that tumor SCs undergo global changes in chromatin accessibility compared to their normal counterparts. Figure 1A depicts results from ATAC-seq performed on FACS sorted tumor (SCC-SCs) and normal (HFSCs and EpdSCs) stem cells. Genome-wide ATAC signals are plotted pair-wise on x- and y-axis to calculate their correlation coefficient of determination (R^2). Figure 1B depicts whole genome ATAC signals in SCC-, Epd- and HF-SCs are z-score-transformed and averaged across 100-bp genomic windows. Hierarchical clustering shows shared open chromatin regions between normal EpdSCs and HFSCs compared to SCC-SCs. Heatmap shows gain (yellow), loss (blue) or no change (black) of ATAC signals from normal to stress comparison. 'n' indicates number of genomic windows plotted. Figure 1C depicts cumulative density showing genes that are strongly upregulated in tumorigenic (T) versus normal (N) SCs are those that have gained ATAC peaks ($\log_2FC$ more positive, green curve right shift), whereas those that exhibit marked tumor-associated declines in expression are genes that have lost ATAC peaks ($\log_2FC$ more negative, red curve left shift). KS one-sided test was performed to compare genes that had gained or lost ATAC peaks relative to all genes. Figure 1D depicts SCC-SC, EpdSC and HFSC ATAC tracks of representative genes that are either constitutively active in all conditions, induced in tumor or suppressed in tumor.

Also see Figure 8.

Figure 2, comprising Figure 2A through Figure 2F, depicts the results of example experiments illustrating that SCC-SCs express both epidermal and HF lineage markers. Figure 2A depicts de novo motif analysis identifying enriched TF motifs associated with ATAC-peaks that are unique to HFSCs or EpdSCs compared to peaks that are common to both lineages. Figure 2B depicts immunofluorescence results

revealing that KLF5 is down-regulated in newly emerged hair follicles at embryonic day E17.5. Scale bar = 50 μ m. At least five biologically independent replicates were analyzed; shown is a representative image. Figure 2C depicts immunofluorescence results revealing lineage-specificity in SOX9 and KLF5 expression patterns under skin homeostasis. Bar = 50 μ m. At least five biologically independent replicates were analyzed; shown is a representative image. Figure 2D depicts ATAC tracks revealing enhanced chromatin accessibility within the KLF5 and SOX9 regulatory regions that are specific to SCC-SCs and not seen in the homeostatic SCs of the opposite lineage. Figure 2E depicts de novo motif analysis identifying enriched motifs for TFs associated with ATAC peaks that are unique to SCC-SCs compared to those that are shared between SCC-SCs and their normal counterparts. Figure 2F depicts immunofluorescence results revealing co-expression of SOX9 and KLF5 in SCCs. Bar = 50 μ m. At least five biologically independent replicates were analyzed; shown is a representative image.

Figure 3, comprising Figure 3A through Figure 3F, depicts the results of example experiments illustrating that lineage infidelity is a functionally obligatory hallmark of tumorigenesis. Figure 3A and Figure 3B depict that SCs of SCCs were freshly sorted by FACS and infected with lentivirus harboring CAS9 and small guide RNAs against KLF5 (red) or SOX9 (green). Scrambled (Scr) guide was used as control (black). Following two days of puromycin selection, half of each population was subjected to quantitative PCR to measure KLF5 and SOX9 expression levels (Figure 3A), and the other half was engrafted onto the backs of immunocompromised Nude mice (Figure 3B). Tumor volume was measured each week thereafter (mean $\pm$ std, n=5); Images were taken after three weeks (shown are representative images). At least two small guides for each gene were tested and results were consistent (also see Figure 10). Paired t test was performed for A. Two-way ANOVA with repeated measurement was performed for B. **P<0.01. Figure 3C depicts the results from genes associated with ATAC peaks in SCC-SCs containing SOX9 or KLF5 motifs (see Methods Details) that were analyzed by MolSigDB (Broad Institute) for their associated molecular pathways. Note preference of KLF5 for proliferation pathway genes and SOX9 for invasion pathway genes. Figure 3D depicts that overexpression (O/E) of KLF5 in vivo was achieved by lentiviral (LV) TRE-KLF5 infection of E9.5 K14-rtTA embryos. O/E of SOX9 in vivo was achieved by LV-rtTA infection of E9.5 TRESOX9 embryos. At P50, mice were subjected to Doxycycline treatment for three days and then pulsed with EdU for two hours prior to FACS analysis. Shown are % EdU+ of total basal cells. n= 9. Results are

shown as mean $\pm$ std. Paired t-test was performed. * $P < 0.05$, *** $P < 0.001$. Figure 3E depicts immunofluorescence data from engrafted SCC-SCs (transduced with GFP lentivirus to mark tumor versus stromal cells) that were ablated for KLF5 by CRISPR/CAS. Note decrease in progenitor marker K5 and increase in differentiation marker K10. Scale bar = 50 μ m. At least five biologically independent replicates were analyzed; shown are representative images. Figure 3F depicts keratinocytes from transgenic K14-rtTA embryos that were infected with lentiviral TRE-SOX9 or TRE-KLF5 (TRE-empty was used as control) and treated with doxycycline for three days, and then subjected to Boyden chamber assays. A layer of matrigel is placed on the lower filter. Keratinocytes in the upper chamber are assayed for their ability to respond to stimulatory dermal fibroblasts conditioned media in the lower chamber and degrade matrigel. One day after seeding, keratinocytes reaching the bottom chamber are counted and plotted as fold change compared to control. $n = 6$. Results are shown as mean $\pm$ std. Paired t-test was performed. *** $P < 0.001$. N.S., not significant.

Figure 4, comprising Figure 4A and Figure 4B, depicts the results of example experiments illustrating that epithelial wounding transiently inflicts and relies on lineage infidelity for repair. Figure 4A depicts immunofluorescence results revealing temporal changes of SOX9 and KLF5 expression patterns following a partial-thickness wound (see Methods Details). Scale bar = 50 μ m. At least five biologically independent replicates were analyzed; shown are representative images. Figure 4B depicts a schematic of inducible KLF5 and SOX9 knockout by in vivo CRISPR/CAS to analyze consequences to split-thickness wound-repair. Rosa26-Loxp-STOP-Loxp-Cas9-P2A-GFP E9.5 embryos were transduced with lentiviral (LV) CreER also carrying U6-sgKLF5 or U6-sgSOX9. At E18.5, tamoxifen was administered to pregnant females, and at P3, half of back skin of pups was directly frozen for sectioning and immune-labeling; the other half was treated with EDTA to remove and discard the epidermis. The dermis, including HFs, was then engrafted onto Nude mice. Grafts were analyzed two weeks later by either immunofluorescence (middle and bottom left panels) or FACS to quantify % of GFP+ cells out total epidermal basal cells (top right panel). Scale bar = 50 μ m. In SOX9 ablation staining, arrow points to an ablated HF (red is cytoplasmic rather than SOX9's nuclear staining) next to a non-transduced WT HF (arrow heads). For immunofluorescence, at least three biologically independent replicates were analyzed; shown are representative images. $n = 7$ for FACS quantification. Paired t test was performed. *** $P < 0.001$. N.S., not significant.

Figure 5, comprising Figure 5A through Figure 5G, depicts the results of example experiments illustrating that wounded and tumorigenic SCs display similar transcriptomes and genome-wide chromatin accessibilities. Figure 5A and Figure 5B depict that unsupervised hierarchical clustering and principle component analysis (PCA) of transcriptome data reveals similarities between SCC-SCs and Wound (Wd)-SCs relative to homeostatic SCs (HFSCs and EpdSCs). Figure 5C depicts that gene set enrichment analysis (GSEA) reveals striking parallels in transcriptome changes that occur in tumor and wound versus homeostatic SCs. Gene changes in tumor vs. normal are compared against pre-ranked changes in wound vs. normal. Left panel, up-regulated genes; right panel, down-regulated genes. Figure 5D depicts ATAC signals of SCC- and Wd-SC plotted on x- and y-axis, respectively, and that their correlation was calculated (coefficient of determination R^2). Figure 5E depicts results from genome-wide ATAC signals in Wd-, SCC-, Epd- and HF-SCs that are z-score-transformed and averaged across 100-bp genomic windows. Hierarchical clustering shows divergence between normal (Epd and HF) and stress-experienced (Wd and SCC) SCs, and heatmap shows gain (yellow), loss (blue) or no change (black) of ATAC signals in normal:stress comparisons. 'n' indicates number of genomic windows plotted. Figure 5F depicts ATAC tracks for Wd-, SCC-, Epd- and HF-SCs and H3K27Ac ChIPseq track for SCC-SCs which are shown for KLF5 (left) and mir21 (right) genes. Boxed peaks denote epicenters (ECs) cloned for testing enhancer activities with eGFP reporters in vivo (the tested KLF5 EC is the second boxed peak on the right). Grey shaded box indicates predicted TF motifs within ECs. Black bars indicated annotated super-enhancers (SE) in SCC-SCs for KLF5 and mir21. Figure 5G depicts that epifluorescence shows stress-induced EC driven reporter activity (eGFP) in the tumor and wound but not in normal skin, while all conditions show comparable LV transduction (H2BRFP). At least five biologically independent experiments were performed; shown are representative images.

Figure 6, comprising Figure 6A through Figure 6G, depicts the results of example experiments illustrating that activated epicenters comprised of lineage-infidelity and stress-induced TFs distinguish cancer from wound state. All immunofluorescence images in Figure 6 are representative of at least three biologically independent replicates. Scale bars = 50 μ m. Figure 6A depicts immunofluorescence results revealing pETS2 in tumor and wound but not in normal skin. Figure 6B depicts that forced activation of ETS2 in the skin epithelium is achieved by transducing K14rtTA embryos with lentiviral TRE-T72D-Ets2, and then treating neonatal mice with doxycycline for 4 weeks starting at P0

(Yang et al., 2015). Backskin was analyzed for pETS2, KLF5 and SOX9. Arrows point to the ectopic KLF5 expression in the HF and SOX9 in the Epd. Figure 6C depicts that immunofluorescence reveals KLF5 and SOX9 co-expression in EpdSCs at the wound edge five days after full-thickness punch wound. Figure 6D depicts that forced activation of KLF5 in the skin epithelium is achieved by transducing K14rtTA embryos with lentiviral TRE-KLF5, and inducing pups with doxycycline for two weeks starting at P0. Backskin was analyzed for KLF5 and SOX9 expression. Figure 6E depicts ATAC peaks at the SOX9 locus of Wd-, SCC-, Epd- and HF-SCs contrasting with SOX9 and LHX2 ChIPseq peaks at the SOX9 locus of HFSC-SCs. Boxed areas are magnified (HFSC-EC green, Tumor-EC red). Note silencing of SOX9's homeostatic enhancers and activation of new, tumor-specific enhancers in SCC-SCs. Note also a number of additional epicenters (red shaded) shared by wound and SCC. Figure 6F depicts epifluorescence results showing that SOX9's homeostatic enhancer drives GFP reporter activity only in HFSCs and not in tumor or wound states, while SOX9's SCC-specific enhancer drives GFP reporter activity only in the SCC and not in wound or normal homeostasis. Compare to Figure 3, where low stress ATAC peaks driver reporter expression in both SCC and wound but not normal homeostasis. All reporters have comparable lentiviral infectivity (H2BRFP). Figure 6G depicts that forced activation of SOX9 in the skin epithelium is achieved by transducing TRE-SOX9 embryos with lentiviral rtTA-H2BGFP, and then treating animals with doxycycline just prior to split-thickness engraftment (see Method Details). Skin grafts were then analyzed two weeks later, with half of the graft used for sectioning and immunofluorescence for SOX9, H2BGFP, KLF5 and pETS2, and the other half for EdU pulse and analyses.

Figure 7 depicts the results of example experiments illustrating a model for sustained lineage infidelity in diverging tumors from wounds. Under normal homeostasis, skin EpdSCs and HFSCs govern their own fates by activating lineage-specific homeostatic epicenters (ECs) for KLF5 and SOX9, respectively. In wound-repair, wound-ECs regulated by RAS/MAPK-induced, stress-responsive TFs ETS2 among others, become activated to drive expression of both KLF5 and SOX9 genes, irrespective of SC origin. This leads to transient lineage infidelity, which is resolved when RAS/MAPK and pETS2 activity wanes upon restoration of the skin barrier. Concomitantly, when KLF5 is expressed, SOX9's homeostatic HFSC-EC is repressed, and hence SOX9 becomes silenced when wound-EC lose their potency. In HFSCs that reach the epidermis or EpdSCs that had completed re-epithelialization, KLF5 and EpdSC fate prevail. In SCC

progression, RAS/MAPK and pETS2 activity are sustained, leading to elevated KLF5 and SOX9 gene expression driven by wound-ECs hijacked by the tumors. However, as SOX9 and KLF5 levels rise, new enhancer elements (tumor-ECs) are acquired, composed of lineage and stress-response TF motifs. The outcome is now a complete independence of KLF5 and SOX9 on their homeostatic enhancers, hence sustained lineage infidelity fueling malignancy.

Figure 8, comprising Figure 8A through Figure 8C, depicts the results of example experiments illustrating that attributes of ATAC-seq peaks in tumor and wound SCs confirm their quality and usefulness to identify gene regulatory regions. Figure 8A depicts results from averaging the genomic-wide distribution of ATAC signals for the transcription start site (TSS=0) or CCCTC-binding factor (CTCF) across SCC-, Epd- and HF-SC chromatin. Genome-wide data are plotted in kb relative to average tag count (average number of sequencing reads at each position). Note that the expected peaks and coverage are similar to that known for these chromatin features and which have been mapped previously by ChIPseq (ENCODE) (Buenestro et al., 2013). Figure 8B depicts the distribution of ATAC peaks (total numbers listed) over annotated genomic regions shown as pie charts for SCC-, Epd- and HF-SC chromatin. Note consistency with enriched signals over intergenic regulatory regions. Figure 8C depicts that the location of ATAC peak marks within an epicenter (EC) within the super-enhancer of the Cxcl14 gene that is known to be active in HFSCs, and bound by an entire suite of TFs that are flanked by H3K27Ac signals (Adam et al., 2015). Here, the current ATAC dataset for EpdSCs, SCC-SCs and HFSCs were plotted, to illustrate ATAC can be used to identify bona fide gene regulatory regions, in this case active in HFSC and silent in SCC-SC and EpdSC.

Figure 9 depicts the results of example experiments illustrating that S2KLF5 and SOX9 are co-expressed in hyperplastic and benign tumor stages. Immunofluorescence reveals co-expression of KLF5 and SOX9 early during hyperplastic and benign tumor (papilloma) stages that are generated by oncogenic HRasG12V expression. Scale bar = 50µm. Five independent replicates were processed; shown are representative images. Asterisk denotes autofluorescence, a frequent problem of stratum corneum and hair shaft. Bona fide SOX9 staining is nuclear.

Figure 10 depicts the results of example experiments illustrating efficient knockdown of KLF5 in SCC. Immunofluorescence for KLF5 and INTEGRIN β4 on tumors generated by subcutaneous injection of HRASG12V-transformed SCC cells into

host recipient Nude mice. Prior to injections, SCC cells were transduced with a GFP lentivirus (to mark the tumor cells as opposed to stroma) and also a KLF5 or Scramble control shRNA hairpin and puromycin selection marker, administered for 2d to obtain stable integration. Scale bar = 50 μ m. Two different KLF5 shRNA hairpins were tested and gave consistent results. Three biologically independent experiments were performed; shown are representative images.

Figure 11, comprising Figure 11A and Figure 11B, depicts the results of example experiments illustrating efficient CRISPR/CAS-mediated SOX9 and KLF5 gene ablation in vivo. Amniotic sacs of living Rosa26-Loxp-STOP-Loxp-Cas9-P2A-GFP E9.5 embryos were injected in utero with lentivirus (LV) harboring PGK-CreER and a U6-sgRNA against KLF5 or SOX9 or Scramble control. Embryos were removed for immunofluorescence analysis of either sagittal skin sections (Figure 11A, E18.5, KLF5 ablated; Scr) or whole mount (Figure 11B, E16.5, SOX9 ablated; Scr). Note that KLF5 and GFP are mutually exclusive in the epidermis, reflective of efficient KLF5 knockout in LV-transduced cells. Note that for SOX9, green GFP (boxed insets in lower left corners of mainframes) show high transductions across all conditions, as that HF formation is abrogated in SOX9 sg-transduced regions and not in Scr sg-transduced regions. This is consistent with the essential role of SOX9 in HFSCs (Nowak et al., 2008). P-cadherin (PCAD) and LHX2 served as controls to mark HFs and were not affected by SOX9 ablation. Scale bar = 50 μ m for sagittal sections, 200 μ m for whole mount images. At least two distinct sgRNAs were tested for each gene and both gave consistent results. At least 5 biologically independent experiments were performed; shown are representative images.

Figure 12, comprising Figure 12A through Figure 12D, depicts the results of example experiments illustrating that wound and tumor transcriptomes and chromatin accessibility profiles share similarities distinct from proliferative progenies of stem cells. Figure 12A depicts Venn diagrams showing how transcripts up or downregulated by $\geq 4X$ ($p < 0.05$) in SCC-SCs relative to their WT counterparts (numbers of transcripts indicated) compare to those up or downregulated in Wd-SCs relative to unwounded SCs. Note that 98.5% of all changes—up or down—are concordant (Fisher's exact test, p value = 0) while only 1.5% are discordant. Figure 12B depicts gene set enrichment analysis (GSEA) compares the transcripts up or downregulated by $\geq 4X$ ($p < 0.05$) in stressed conditions vs. normal compared against pre-ranked changes in proliferative TAC (transient amplifying cells, derived from HFSCs) vs. homeostatic HFSCs. Comparisons

of signatures of TACs and Wd-SCs are shown in the top panels; TACs and SCC-SCs in the bottom panels. Note that these signatures do not significantly overlap, indicating that the similarities seen between Wd-SCs and SCC-SCs go beyond merely similarities in proliferative features. Figure 12C depicts that strong ATAC signals are centered throughout the transcription start sites (TSS, left) and CTCF motif sites (right) of the wounded (Wd) SC chromatin. Plots of the genome-wide data illustrate the expected peaks (numbers on X axis in kb; Y axis, average number of sequencing reads at each position). Figure 12D depicts the distribution of ATAC peaks over annotated genomic regions of Wd-SC chromatin, shown as a pie chart, and highlights the enrichment of ATAC signals over intergenic regulatory regions.

Figure 13, comprising Figure 13A through Figure 13D, depicts the results of example experiments illustrating that stress-induced enhancers drive stem cell lineage infidelity in skin malignancy. For all immunofluorescence in Figure 13, at least 3 biologically independent replicates were performed and representative images are shown. Scale bars = 50µm. Figure 13A depicts ATAC peaks at the *Sca1* and *Hes1* locus (examples of epidermal genes) showing stress-related enhancer epicenters (ECs) (red shades) displaying specific signals in SCC and Wd but not in normal homeostatic SCs. These ECs are enriched for binding motifs for ETS, AP1 and often STAT factors, all of which are induced in stress states. Figure 13B depicts immunofluorescence results revealing the presence of AP1 factors JUN and FOS as well as pSTAT3 in tumor and wound. FOS is absent in homeostatic skin and JUN, pSTAT3 are weak. Figure 13C depicts that forced activation of ETS2 in the skin epithelium is achieved by transducing E9.5 K14rtTA embryos in utero with lentivirus harboring TRE-ETS2 (T72D). Doxycycline was administered starting at P0 and mice were analyzed at 4 weeks (Yang et al., 2015). Sagittal sections of back skins were analyzed for JUN, FOS and pSTAT3 expression. Figure 13D depicts that ATAC peaks at the *SOX9* locus show stress-induced ECs (red shades) with specific signals in SCC and WD but not in normal homeostatic SCs.

Figure 14, comprising Figure 14A through Figure 14F, depicts the results of example experiments illustrating that KLF5 reprograms HF into epidermis and sustained SOX9 pushes wound into tumor. Figure 14A depicts that forced activation of KLF5 in the skin epithelium is achieved by transducing K14rtTA embryos with lentiviral TRE-KLF5, and inducing pups with doxycycline for 2 weeks starting at P0. Back skin was analyzed for K10, LORICRIN, and TENASCIN C. Inset show in the TRE-KLF5

group magnified images for ectopic K10, LORICRIN in the HF and a transduced (H2BRFP) follicle showing reduced TENASCIN C compared to a WT follicle. Figure 14B and Figure 14C depict that forced activation of SOX9 in the skin epithelium was achieved by transducing E9.5 TRESOX9 embryos in utero with lentivirus harboring PGK-rtTA-H2BGFP. Doxycycline was administered starting at E18.5 and at P3, split-thickness grafts were performed on host Nude mice (see Method Details). Skin grafts were processed 2 weeks later, and either a) sagittal sectioned and stained for K5, K10, LORICRIN, shown in Figure 14E, or b) EdU pulsed for 2 hours and subjected to FACS analysis for % of EdU+ cells out of total basal cells, shown in Figure 14F. n = 8 for Figure 14F. Results are shown as mean +/- std. Paired t-test was performed. ***P <0.001.

Figure 15 is an image demonstrating that epicenter activity in cultured mouse and human SCC lines.

Figure 16 is an image demonstrating epicenter activity in human and neck squamous cell carcinoma models.

DETAILED DESCRIPTION

The present invention provides compositions and methods for detecting and treating cancer. The present invention is partly based upon the observation that human patients suffering from chronic wounds have increased susceptibility to cancers. In addition, the present invention is partly based on the observation that mice with gene mutations that enhance hair follicle stem cell activity heal their wounds faster but also exhibit susceptibility to squamous cell carcinoma. In contrast, mice whose skin possesses mutations that impede hair follicle stem cell activation display reduced efficiency in wound closure, but an increased resistance to cancers.

In one embodiment, the present invention provides methods for detecting and modulating transcription epicenters and super-enhancer regions in DNA that are specifically accessible in cells transitioning to a cancerous state. The present invention provides compositions and methods for preventing and treating cancer through use of modulators of epicenters and super-enhancer regions. The present invention also provides a composition for detecting accessible promoter regions for the purpose of identifying new epicenters and relevant transcription factors that bind thereto that may be new targets for cancer therapeutics.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the present invention, 5 the preferred materials and methods are described herein. In describing and claiming the present invention, the following terminology will be used.

It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

The articles “a” and “an” are used herein to refer to one or to more than 10 one (*i.e.*, to at least one) of the grammatical object of the article. By way of example, “an element” means one element or more than one element.

“About” as used herein when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass non-limiting variations of $\pm 40\%$ or $\pm 20\%$ or $\pm 10\%$, $\pm 5\%$, $\pm 1\%$, or $\pm 0.1\%$ from the specified value, as such 15 variations are appropriate.

The term “abnormal” when used in the context of organisms, tissues, cells or components thereof, refers to those organisms, tissues, cells or components thereof that differ in at least one observable or detectable characteristic (e.g., age, treatment, time of day, etc.) from those organisms, tissues, cells or components thereof that display the 20 “normal” (expected) respective characteristic. Characteristics which are normal or expected for one cell or tissue type, might be abnormal for a different cell or tissue type.

As used herein the terms “alteration,” “defect,” “variation,” or “mutation,” refers to a mutation in a gene in a cell that affects the function, activity, expression (transcription or translation) or conformation of the polypeptide that it encodes. Mutations 25 encompassed by the present invention can be any mutation of a gene in a cell that results in the enhancement or disruption of the function, activity, expression or conformation of the encoded polypeptide, including the complete absence of expression of the encoded protein and can include, for example, missense and nonsense mutations, insertions, deletions, frameshifts and premature terminations. Without being so limited, mutations 30 encompassed by the present invention may alter splicing the mRNA (splice site mutation) or cause a shift in the reading frame (frameshift).

A disease or disorder is “alleviated” if the severity of a sign or symptom of the disease or disorder, the frequency with which such a sign or symptom is experienced by a patient, or both, is reduced. “Alleviating” specific cancers and/or their pathology

includes degrading a tumor, for example, breaking down the structural integrity or connective tissue of a tumor, such that the tumor size is reduced when compared to the tumor size before treatment. "Alleviating" cancer includes reducing the rate at which the cancer spreads to other organs.

5 The term "amplification" refers to the operation by which the number of copies of a target nucleotide sequence present in a sample is multiplied.

By the term "applicator," as the term is used herein, is meant any device including, but not limited to, a hypodermic syringe, a pipette, an iontophoresis device, a patch, and the like, for administering the compositions of the invention to a subject.

10 The term "antibody," as used herein, refers to an immunoglobulin molecule which specifically binds with an antigen. Antibodies can be intact immunoglobulins derived from natural sources or from recombinant sources and can be immuno-reactive portions of intact immunoglobulins. Antibodies are typically tetramers of immunoglobulin molecules. The antibodies in the present invention may exist in a
15 variety of forms including, for example, polyclonal antibodies, monoclonal antibodies, Fv, Fab and F(ab)₂, as well as single chain antibodies and humanized antibodies (Harlow et al., 1999, In: Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY; Harlow et al., 1989, In: Antibodies: A Laboratory Manual, Cold Spring Harbor, New York; Houston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; Bird
20 et al., 1988, Science 242:423-426).

 The term "antibody fragment" refers to at least one portion of an intact antibody and refers to the antigenic determining variable regions of an intact antibody. Examples of antibody fragments include, but are not limited to, Fab, Fab', F(ab')₂, and Fv fragments, linear antibodies, sdAb (either V_L or V_H), camelid V_{HH} domains, scFv
25 antibodies, and multi-specific antibodies formed from antibody fragments. The term "scFv" refers to a fusion protein comprising at least one antibody fragment comprising a variable region of a light chain and at least one antibody fragment comprising a variable region of a heavy chain, wherein the light and heavy chain variable regions are contiguously linked via a short flexible polypeptide linker, and capable of being
30 expressed as a single chain polypeptide, and wherein the scFv retains the specificity of the intact antibody from which it was derived. Unless specified, as used herein an scFv may have the V_L and V_H variable regions in either order, e.g., with respect to the N-terminal and C-terminal ends of the polypeptide, the scFv may comprise V_L-linker-V_H or may comprise V_H-linker-V_L.

An “antibody heavy chain,” as used herein, refers to the larger of the two types of polypeptide chains present in all antibody molecules in their naturally occurring conformations.

5 An “antibody light chain,” as used herein, refers to the smaller of the two types of polypeptide chains present in all antibody molecules in their naturally occurring conformations. κ and λ light chains refer to the two major antibody light chain isotypes.

By the term “synthetic antibody” as used herein, is meant an antibody which is generated using recombinant DNA technology, such as, for example, an antibody expressed by a bacteriophage as described herein. The term should also be
10 construed to mean an antibody which has been generated by the synthesis of a DNA molecule encoding the antibody and which DNA molecule expresses an antibody protein, or an amino acid sequence specifying the antibody, wherein the DNA or amino acid sequence has been obtained using synthetic DNA or amino acid sequence technology which is available and well known in the art.

15 By the term “specifically binds,” as used herein with respect to an antibody, is meant an antibody which recognizes a specific antigen, but does not substantially recognize or bind other molecules in a sample. For example, an antibody that specifically binds to an antigen from one species may also bind to that antigen from one or more species. But, such cross-species reactivity does not itself alter the
20 classification of an antibody as specific. In another example, an antibody that specifically binds to an antigen may also bind to different allelic forms of the antigen. However, such cross reactivity does not itself alter the classification of an antibody as specific. In some instances, the terms “specific binding” or “specifically binding,” can be used in reference to the interaction of an antibody, a protein, or a peptide with a second chemical species, to
25 mean that the interaction is dependent upon the presence of a particular structure (e.g., an antigenic determinant or epitope) on the chemical species; for example, an antibody recognizes and binds to a specific protein structure rather than to proteins generally. If an antibody is specific for epitope “A”, the presence of a molecule containing epitope A (or free, unlabeled A), in a reaction containing labeled “A” and the antibody, will reduce the
30 amount of labeled A bound to the antibody.

The term “antigen” or “Ag” as used herein is defined as a molecule that provokes an immune response. This immune response may involve either antibody production, or the activation of specific immunologically-competent cells, or both. The

skilled artisan will understand that any macromolecule, including virtually all proteins or peptides, can serve as an antigen. Furthermore, antigens can be derived from recombinant or genomic DNA. A skilled artisan will understand that any DNA, which comprises a nucleotide sequences or a partial nucleotide sequence encoding a protein that elicits an immune response therefore encodes an “antigen” as that term is used herein. Furthermore, one skilled in the art will understand that an antigen need not be encoded solely by a full length nucleotide sequence of a gene. It is readily apparent that the present invention includes, but is not limited to, the use of partial nucleotide sequences of more than one gene and that these nucleotide sequences are arranged in various combinations to encode polypeptides that elicit the desired immune response. Moreover, a skilled artisan will understand that an antigen need not be encoded by a “gene” at all. It is readily apparent that an antigen can be generated synthesized or can be derived from a biological sample. Such a biological sample can include, but is not limited to a tissue sample, a tumor sample, a cell or a biological fluid.

The term “anti-tumor effect” as used herein, refers to a biological effect which can be manifested by a decrease in tumor volume, a decrease in the number of tumor cells, a decrease in the number of metastases, an increase in life expectancy, or amelioration of various physiological symptoms associated with the cancerous condition. An “anti-tumor effect” can also be manifested by the ability of the peptides, polynucleotides, cells and antibodies of the invention in prevention of the occurrence of tumor in the first place.

The term “cancer” as used herein is defined as disease characterized by the abnormal growth of aberrant cells. Cancer cells can spread locally or through the bloodstream and lymphatic system to other parts of the body. Examples of various cancers include but are not limited to, breast cancer, prostate cancer, ovarian cancer, cervical cancer, skin cancer, pancreatic cancer, colorectal cancer, renal cancer, liver cancer, brain cancer, lymphoma, leukemia, lung cancer, squamous cell carcinoma, sarcoma and the like.

As used herein, the term “marker” or “biomarker” is meant to include a parameter which is useful according to this invention for determining the presence and/or severity of disease.

The level of a marker or biomarker “significantly” differs from the level of the marker or biomarker in a reference sample if the level of the marker in a sample from the patient differs from the level in a sample from the reference subject by an amount

greater than the standard error of the assay employed to assess the marker, and preferably at least 10%, and more preferably 25%, 50%, 75%, or 100%.

The term “control or reference standard” describes a material comprising none, or a normal, low, or high level of one of more of the marker (or biomarker) expression products of one or more the markers (or biomarkers) of the invention, such that the control or reference standard may serve as a comparator against which a sample can be compared.

By the phrase “determining the level of marker (or biomarker) expression” is meant an assessment of the degree of expression of a marker in a sample at the nucleic acid or protein level, using technology available to the skilled artisan to detect a sufficient portion of any marker expression product.

“Differentially increased expression” or “up regulation” refers to biomarker product levels which are at least 10% or more, for example, 20%, 30%, 40%, or 50%, 60%, 70%, 80%, 90% higher or more, and/or 1.1 fold, 1.2 fold, 1.4 fold, 1.6 fold, 1.8 fold, 2.0 fold higher or more, and any and all whole or partial increments therebetween than a control.

“Differentially decreased expression” or “down regulation” refers to biomarker product levels which are at least 10% or more, for example, 20%, 30%, 40%, or 50%, 60%, 70%, 80%, 90% lower or less, and/or 2.0 fold, 1.8 fold, 1.6 fold, 1.4 fold, 1.2 fold, 1.1 fold or less lower, and any and all whole or partial increments therebetween than a control.

A “disease” is a state of health of an animal wherein the animal cannot maintain homeostasis, and wherein if the disease is not ameliorated then the animal’s health continues to deteriorate. In contrast, a “disorder” in an animal is a state of health in which the animal is able to maintain homeostasis, but in which the animal’s state of health is less favorable than it would be in the absence of the disorder. Left untreated, a disorder does not necessarily cause a further decrease in the animal’s state of health.

“Encoding” refers to the inherent property of specific sequences of nucleotides in a polynucleotide, such as a gene, a cDNA, or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having either a defined sequence of nucleotides (e.g., DNA, cDNA, rRNA, tRNA and mRNA) or a defined sequence of amino acids and the biological properties resulting therefrom. Thus, a gene encodes a protein if transcription and translation of mRNA corresponding to that gene produces the protein in a cell or other biological system. Both

the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings, and the non-coding strand, used as the template for transcription of a gene or cDNA, can be referred to as encoding the protein or other product of that gene or cDNA.

5 Unless otherwise specified, a “nucleotide sequence encoding an amino acid sequence” includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. Nucleotide sequences that encode proteins and RNA may include introns.

 “Effective amount” or “therapeutically effective amount” are used
10 interchangeably herein, and refer to an amount of a compound, formulation, material, or composition, as described herein effective to achieve a particular biological result. Such results may include, but are not limited to, the inhibition of virus infection as determined by any means suitable in the art.

 As used herein “endogenous” refers to any material from or produced
15 inside an organism, cell, tissue or system.

 As used herein, the term “exogenous” refers to any material introduced from or produced outside an organism, cell, tissue or system.

 The term “expression” as used herein is defined as the transcription and/or translation of a particular nucleotide sequence driven by its promoter.

20 As used herein, an “immunoassay” refers to a biochemical test that measures the presence or concentration of a substance in a sample, such as a biological sample, using the reaction of an antibody to its cognate antigen, for example the specific binding of an antibody to a protein. Both the presence of the antigen or the amount of the antigen present can be measured.

25 The term “inhibit,” as used herein, means to suppress or block an activity or function by at least about ten percent relative to a control value. Preferably, the activity is suppressed or blocked by 50% compared to a control value, more preferably by 75%, and even more preferably by 95%.

 As used herein, an “instructional material” includes a publication, a
30 recording, a diagram, or any other medium of expression which can be used to communicate the usefulness of a component of the invention in a kit for detecting biomarkers disclosed herein. The instructional material of the kit of the invention can, for example, be affixed to a container which contains the component of the invention or be shipped together with a container which contains the component. Alternatively, the

instructional material can be shipped separately from the container with the intention that the instructional material and the component be used cooperatively by the recipient.

“Isolated” means altered or removed from the natural state. For example, a nucleic acid or a peptide naturally present in a living animal is not “isolated,” but the same nucleic acid or peptide partially or completely separated from the coexisting materials of its natural state is “isolated.” An isolated nucleic acid or protein can exist in substantially purified form, or can exist in a non-native environment such as, for example, a host cell.

In the context of the present invention, the following abbreviations for the commonly occurring nucleic acid bases are used. “A” refers to adenosine, “C” refers to cytosine, “G” refers to guanosine, “T” refers to thymidine, and “U” refers to uridine.

The term “label” when used herein refers to a detectable compound or composition that is conjugated directly or indirectly to a probe to generate a “labeled” probe. The label may be detectable by itself (e.g. radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, may catalyze chemical alteration of a substrate compound or composition that is detectable (e.g., avidin-biotin). In some instances, primers can be labeled to detect a PCR product.

The “level” of one or more biomarkers means the absolute or relative amount or concentration of the biomarker in the sample.

The term “marker (or biomarker) expression” as used herein, encompasses the transcription, translation, post-translation modification, and phenotypic manifestation of a gene, including all aspects of the transformation of information encoded in a gene into RNA or protein. By way of non-limiting example, marker expression includes transcription into messenger RNA (mRNA) and translation into protein, as well as transcription into types of RNA such as transfer RNA (tRNA) and ribosomal RNA (rRNA) that are not translated into protein.

The terms “microarray” and “array” refers broadly to both “DNA microarrays” and “DNA chip(s),” and encompasses all art-recognized solid supports, and all art-recognized methods for affixing nucleic acid molecules thereto or for synthesis of nucleic acids thereon. Preferred arrays typically comprise a plurality of different nucleic acid probes that are coupled to a surface of a substrate in different, known locations. These arrays, also described as “microarrays” or colloquially “chips” have been generally described in the art, for example, U.S. Pat. Nos. 5,143,854, 5,445,934, 5,744,305, 5,677,195, 5,800,992, 6,040,193, 5,424,186 and Fodor et al., 1991, Science, 251:767-777,

each of which is incorporated by reference in its entirety for all purposes. Arrays may generally be produced using a variety of techniques, such as mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase synthesis methods. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, e.g., U.S. Pat. Nos. 5,384,261, and 6,040,193, which are incorporated herein by reference in their entirety for all purposes. Although a planar array surface is preferred, the array may be fabricated on a surface of virtually any shape or even a multiplicity of surfaces. Arrays may be nucleic acids on beads, gels, polymeric surfaces, fibers such as fiber optics, glass or any other appropriate substrate. (See U.S. Pat. Nos. 5,770,358, 5,789,162, 5,708,153, 6,040,193 and 5,800,992, which are hereby incorporated by reference in their entirety for all purposes.) Arrays may be packaged in such a manner as to allow for diagnostic use or can be an all-inclusive device; e.g., U.S. Pat. Nos. 5,856,174 and 5,922,591 incorporated in their entirety by reference for all purposes. Arrays are commercially available from, for example, Affymetrix (Santa Clara, Calif.) and Applied Biosystems (Foster City, Calif.), and are directed to a variety of purposes, including genotyping, diagnostics, mutation analysis, marker expression, and gene expression monitoring for a variety of eukaryotic and prokaryotic organisms. The number of probes on a solid support may be varied by changing the size of the individual features. In one embodiment, the feature size is 20 by 25 microns square, in other embodiments features may be, for example, 8 by 8, 5 by 5 or 3 by 3 microns square, resulting in about 2,600,000, 6,600,000 or 18,000,000 individual probe features.

“Measuring” or “measurement,” or alternatively “detecting” or “detection,” means assessing the presence, absence, quantity or amount (which can be an effective amount) of either a given substance within a clinical or subject-derived sample, including the derivation of qualitative or quantitative concentration levels of such substances, or otherwise evaluating the values or categorization of a subject’s clinical parameters.

By the term “modulating,” as used herein, is meant mediating a detectable increase or decrease in the activity and/or level of a mRNA, polypeptide, or a response in a subject compared with the activity and/or level of a mRNA, polypeptide or a response in the subject in the absence of a treatment or compound, and/or compared with the activity and/or level of a mRNA, polypeptide, or a response in an otherwise identical but untreated subject.

Unless otherwise specified, a “nucleotide sequence encoding an amino acid sequence” includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. The phrase nucleotide sequence that encodes a protein or an RNA may also include introns to the extent that the nucleotide
5 sequence encoding the protein may in some version contain an intron(s).

The term “operably linked” refers to functional linkage between a regulatory sequence and a heterologous nucleic acid sequence resulting in expression of the latter. For example, a first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional
10 relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in the same reading frame.

“Parenteral” administration of an immunogenic composition includes, e.g.,
15 subcutaneous (s.c.), intravenous (i.v.), intramuscular (i.m.), or intrasternal injection, or infusion techniques.

The terms “patient,” “subject,” “individual,” and the like are used interchangeably herein, and refer to any animal, or cells thereof whether *in vitro* or *in situ*, amenable to the methods described herein. In certain non-limiting embodiments, the
20 patient, subject or individual is a human.

The term “polynucleotide” as used herein is defined as a chain of nucleotides. Furthermore, nucleic acids are polymers of nucleotides. Thus, nucleic acids and polynucleotides as used herein are interchangeable. One skilled in the art has the general knowledge that nucleic acids are polynucleotides, which can be hydrolyzed into
25 the monomeric “nucleotides.” The monomeric nucleotides can be hydrolyzed into nucleosides. As used herein polynucleotides include, but are not limited to, all nucleic acid sequences which are obtained by any means available in the art, including, without limitation, recombinant means, i.e., the cloning of nucleic acid sequences from a recombinant library or a cell genome, using ordinary cloning technology and PCR™, and
30 the like, and by synthetic means.

As used herein, the terms “peptide,” “polypeptide,” and “protein” are used interchangeably, and refer to a compound comprised of amino acid residues covalently linked by peptide bonds. A protein or peptide must contain at least two amino acids, and no limitation is placed on the maximum number of amino acids that can comprise a

protein's or peptide's sequence. Polypeptides include any peptide or protein comprising two or more amino acids joined to each other by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. "Polypeptides" include, for example, biologically active fragments, substantially homologous polypeptides, oligopeptides, homodimers, heterodimers, variants of polypeptides, modified polypeptides, derivatives, analogs, fusion proteins, among others. The polypeptides include natural peptides, recombinant peptides, synthetic peptides, or a combination thereof.

As used herein, the term "providing a prognosis" refers to providing a prediction of the probable course and outcome of disease, including prediction of severity, duration, chances of recovery, etc. The methods can also be used to devise a suitable therapeutic plan, e.g., by indicating whether or not the condition is still at an early stage or if the condition has advanced to a stage where aggressive therapy would be ineffective.

A "reference level" of a biomarker means a level of the biomarker that is indicative of a particular disease state, phenotype, or lack thereof, as well as combinations of disease states, phenotypes, or lack thereof. A "positive" reference level of a biomarker means a level that is indicative of a particular disease state or phenotype. A "negative" reference level of a biomarker means a level that is indicative of a lack of a particular disease state or phenotype.

"Sample" or "biological sample" as used herein means a biological material isolated from an individual. The biological sample may contain any biological material suitable for detecting the desired biomarkers, and may comprise cellular and/or non-cellular material obtained from the individual.

"Standard control value" as used herein refers to a predetermined amount of a particular protein or nucleic acid that is detectable in a sample. The standard control value is suitable for the use of a method of the present invention, in order for comparing the amount of a protein or nucleic acid of interest that is present in a sample. An established sample serving as a standard control provides an average amount of the protein or nucleic acid of interest in the sample that is typical for an average, healthy person of reasonably matched background, e.g., gender, age, ethnicity, and medical history. A standard control value may vary depending on the protein or nucleic acid of

interest and the nature of the sample.

The terms “subject,” “patient,” “individual,” and the like are used interchangeably herein, and refer to any animal, or cells thereof whether *in vitro* or *in situ*, amenable to the methods described herein. In certain non-limiting embodiments, the
5 patient, subject or individual is a human.

The term “therapeutic” as used herein means a treatment and/or prophylaxis. A therapeutic effect is obtained by suppression, remission, or eradication of a disease state.

A “vector” is a composition of matter which comprises an isolated nucleic
10 acid and which can be used to deliver the isolated nucleic acid to the interior of a cell. Numerous vectors are known in the art including, but not limited to, linear polynucleotides, polynucleotides associated with ionic or amphiphilic compounds, plasmids, and viruses. Thus, the term “vector” includes an autonomously replicating plasmid or a virus. The term should also be construed to include non-plasmid and non-
15 viral compounds which facilitate transfer of nucleic acid into cells, such as, for example, polylysine compounds, liposomes, and the like. Examples of viral vectors include, but are not limited to, adenoviral vectors, adeno-associated virus vectors, retroviral vectors, lentiviral vectors, and the like.

Ranges: throughout this disclosure, various aspects of the invention can be
20 presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to
25 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 2.7, 3, 4, 5, 5.3, and 6. This applies regardless of the breadth of the range.

30 Description

The present invention relates to the identification of transcription epicenters and super-enhancer regions, and using modulators of the activity of transcription epicenters and super-enhancer regions. Epicenters are short regions of DNA that can be bound by proteins, for example transcription factors, to increase the likelihood

that transcription of a particular gene or genes can occur. Generally, epicenters are regulatory regions of DNA where one or more transcription factors bind. Generally, epicenters are shouldered by active histone markers (e.g. H3K27Ac). Generally, super-enhancers are broad open chromatin domains marked with active histone markers,
5 comprising one or more epicenters.

In one aspect, the present invention provides compositions and methods for detecting and treating cancer. For example, in one embodiment, the composition comprises a modulator of transcription epicenter activity or super-enhancer activity. For example, in one embodiment, the composition modulates the accessibility of the
10 transcription epicenter or super-enhancer. In one embodiment, the transcription epicenter is a short (around one kilobase of nucleotides in length) active subdomain within super-enhancers, which is particularly enriched for transcription factor binding sites and allows for cooperative binding of a cohort of cell type- and state- specific transcription factors. Epicenters are tissue-, lineage- and cell-stage specific, change dynamically depending on
15 the microenvironment and thus drive gene expression in a highly spatio-temporally selective manner. In one embodiment, the super-enhancer region is a densely spaced cluster of active enhancers (often larger than 10 kilobases of nucleotides in length) with unusually strong enrichment for the binding of cell-type specific transcription factors and transcriptional coactivators, including Mediator (MED1). Super-enhancers are also
20 marked by highly concentrated active histone markers (including H3K27ac), and harbor multiple clusters of epicenters to drive cell type- and state- specific gene expression. Super-enhancers associate with critical cell-identity genes and thus dictate cellular behavior and fate.

In one embodiment, the composition comprises a modulator of one or
25 more epicenter regulator proteins or other components associated with transcription epicenter or super-enhancer accessibility. For example, in one embodiment, the modulator modulates the expression or activity of the one or more epicenter regulator proteins. Exemplary epicenter regulatory proteins associated with transcription epicenter or super-enhancer accessibility include, but is not limited to, polymerases,
30 acetyltransferases, histone deacetylases, methyltransferases, and histone demethylases. In one embodiment, the composition modulates the assembly of transcription machinery specific to the assembly of active transcription complexes within epicenters or super-enhancer regions. In some embodiments, the composition comprises a modulator of one or more factors that bind to the transcription epicenter or super-enhancer. Exemplary

factors include, but are not limited to, transcription factors, coactivators, corepressors, and enhancers. For example, in certain embodiments, the modulator modulates the expression or activity of the one or more factors. In one embodiment, the modulator modulates the formation of a complex of the one or more factors.

5 Accordingly, in one embodiment, the invention provides compositions and methods for modulating epicenters and/or super-enhancer regions. As referred elsewhere herein, modulating epicenters and/or super-enhancer regions include but is not limited to modulating the accessibility of the transcription epicenter and/or super-enhancer, modulating proteins or other components associated with transcription epicenter or super-
10 enhancer accessibility, desired transcription factors, and the likes.

 In one embodiment, the modulator modulates binding of the one or more factors to the transcription epicenter or super-enhancer.

 In one embodiment, the composition modulates one or more transcription factors identified to be oncogenic. In one embodiment, the composition modulates one or
15 more transcription factors identified to regulate distinct cell phenotypes. In one embodiment, the composition modulates one or more transcription factors identified to regulate cellular communication. In one embodiment, the composition modulates one or more transcription factors identified to regulate cellular movements and adhesion. In one embodiment, the one or more transcription factors include, but are not limited to, SOX9,
20 KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NF1, ETS2, and STAT3. In one embodiment, the composition modulates one or more of SOX9 and KLF5.

 In some embodiments, the present invention provides methods for treating and preventing cancer. For example, in one embodiment, the method comprises modulating epicenters and super-enhancers that are specifically accessible in cells
25 transitioning to a cancerous state. In one embodiment, the method comprises administering a modulator of the activity of a transcription epicenter or super-enhancer. In some embodiments, the method comprises modulating epicenters or super-enhancer regions by administering compositions described herein. For example, in one embodiment, the method comprises administering a modulator of one or more
30 transcription factors, as described herein.

 In some embodiments, the present invention also provides methods for detecting new epicenters and relevant transcription factors that bind thereto that may be new targets for cancer diagnostics and therapeutics. In some embodiments, the present invention provides transcription probes indicating the accessibility of certain promoter

regions, transcription factors, epicenters, super-enhancers, and/or epicenter regulating proteins.

In one aspect, the present invention provides a method of diagnosing cancer in a subject by detecting the activity of a transcription epicenter or super-enhancer.

5 For example, in certain embodiments, the method comprises detecting the accessibility of the transcription epicenter or super-enhancer. In one embodiment, the method comprises detecting the presence or abundance of one or more components associated with accessibility of the transcription epicenter or super-enhancer. In one embodiment, the method comprises detecting the presence or abundance of one or more factors that bind to
10 the transcription epicenter or super-enhancer. In one embodiment, the method comprises detecting the presence or abundance of one or more transcripts associated with the transcription epicenter or super-enhancer.

In some embodiments, the invention provides a composition for treating wounds. In certain embodiments, the composition comprises a modulator epicenter
15 activity such that when administered, wound healing is improved. In some embodiments, the composition comprises an activator of epicenter activity and/ or transcription factor activity that promote wound healing, cell proliferation, cell migration, differentiation, remodeling and homeostasis. In some embodiments, the composition comprises an inhibitor of epicenter activity and/or transcription factors that prevent wound healing, cell
20 proliferation, cell migration, differentiation, remodeling and homeostasis.

In some embodiments, the present invention provides methods for treating wounds comprising administering to a subject a composition that modulates epicenter activity. In some embodiments, the method comprises administering to a subject in need thereof, a composition that modulated epicenter activity, wherein when the composition is
25 administered, wound healing is improved. In some embodiments, the method comprises administering a composition modulates the activity of one or more epicenters. In some embodiments, the method comprises administering a composition modulates the activity of one or more transcription factors.

30 Compositions

In one embodiment, the composition provides a modulator (e.g., an inhibitor or activator) of transcription epicenter activity or super-enhancer activity. In some embodiments, the invention relates to compositions that modulate chromatin remodeling, thereby modulating the accessibility of epicenters or super-enhancers. In

some embodiments, the invention relates to modulators of histone modifications (e.g., mono-methylation, di-methylation, tri-methylation, acetylation). In some embodiments, the present invention relates to modulators of specific enzymes that modify chromatin (e.g., acetyltransferases, deacetylases, methyltransferases, and kinases). In some
5 embodiments, the present invention relates to modulators of ATP-dependent chromatin remodeling (e.g., modulators of remodelers of chromatin including but not limited to SWI/SNF, ISWI, NuRD/Mi-2/CHD, INO80 and SWR1).

In one embodiment, the composition provides a modulator of one or more transcription factors that bind transcription epicenters. In some embodiments, the
10 composition provides a modulator of one or more transcription factors that modulate epicenter activity. In some embodiments, the transcription factors include but are not limited to SOX9, KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3.

15 Modulators: Inhibitors and Activators

In one embodiment, the invention provides compositions and methods for modulating epicenters and/or super-enhancer regions. As referred elsewhere herein, modulating epicenters and/or super-enhancer regions include but is not limited to modulating the accessibility of the transcription epicenter and/or super-enhancer,
20 modulating proteins or other components associated with transcription epicenter or super-enhancer accessibility, desired transcription factors, and the likes.

Inhibitors

In one embodiment, inhibitors are used to prevent expression of or activity
25 of one or more transcription epicenters. In some embodiments, the inhibitors are used to prevent binding of proteins to transcription epicenters. In some embodiments, the proteins are transcription factors, regulators of transcription factor expression or activity, or regulators of transcription. In some embodiments, the inhibitors of transcription factor expression are regulators of transcription and translation of transcription factors,
30 including but are not limited to siRNA, antisense nucleic acids, ribozymes, small molecules, and antagonists. In some embodiments, the regulators of transcription factor activity are enzymes including but not limited to kinases. In some embodiments, regulators of transcription include by are not limited to polymerases, acetyltransferases, histone deacetylases, and methylases.

siRNA

In one embodiment, siRNA is used to decrease the activity of one or more epicenters. In one embodiment, the siRNA is used to decrease the level of one or more of transcription factor (e.g., SOX9, KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3) binding to an epicenter. In one embodiment, the siRNA is used to decrease the level of one or more epicenter regulator protein. RNA interference (RNAi) is a phenomenon in which the introduction of double-stranded RNA (dsRNA) into a diverse range of organisms and cell types causes degradation of the complementary mRNA. In the cell, long dsRNAs are cleaved into short 21-25 nucleotide small interfering RNAs, or siRNAs, by a ribonuclease known as Dicer. The siRNAs subsequently assemble with protein components into an RNA-induced silencing complex (RISC), unwinding in the process. Activated RISC then binds to complementary transcript by base pairing interactions between the siRNA antisense strand and the mRNA. The bound mRNA is cleaved and sequence specific degradation of mRNA results in gene silencing. See, for example, U.S. Patent No. 6,506,559; Fire et al., 1998, Nature 391(19):306-311; Timmons et al., 1998, Nature 395:854; Montgomery et al., 1998, TIG 14 (7):255-258; David R. Engelke, Ed., RNA Interference (RNAi) Nuts & Bolts of RNAi Technology, DNA Press, Eagleville, PA (2003); and Gregory J. Hannon, Ed., RNAi A Guide to Gene Silencing, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2003). Soutschek et al. (2004, Nature 432:173-178) describe a chemical modification to siRNAs that aids in intravenous systemic delivery. Optimizing siRNAs involves consideration of overall G/C content, C/T content at the termini, T_m and the nucleotide content of the 3' overhang. See, for instance, Schwartz et al., 2003, Cell, 115:199-208 and Khvorova et al., 2003, Cell 115:209-216. Therefore, the present invention also includes methods of decreasing levels of the desired transcription factor at the protein level using RNAi technology. In doing so, the present invention includes methods of decreasing the activity of one or more epicenters.

In other related aspects, the invention includes an isolated nucleic acid encoding an inhibitor, wherein an inhibitor such as an siRNA or antisense molecule, inhibits the desired one or more epicenters, one or more transcription factors binding thereto, a derivative thereof, a regulator thereof, or a downstream effector, operably linked to a nucleic acid comprising a promoter/regulatory sequence such that the nucleic acid is preferably capable of directing expression of the protein encoded by the nucleic

acid. Thus, the invention encompasses expression vectors and methods for the introduction of exogenous DNA into cells with concomitant expression of the exogenous DNA in the cells such as those described, for example, in Sambrook et al. (2001, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York),
5 and in Ausubel et al. (1997, Current Protocols in Molecular Biology, John Wiley & Sons, New York) and as described elsewhere herein. In another aspect of the invention, the desired epicenter, one or more transcription factors binding thereto, or a regulator thereof, can be inhibited by way of inactivating and/or sequestering one or more of the epicenters, transcription factors, or a regulator thereof. As such, inhibiting the effects of the epicenter
10 or one or more transcription factors binding thereto can be accomplished by using a transdominant negative mutant.

In another aspect, the invention includes a vector comprising an siRNA or antisense polynucleotide. Preferably, the siRNA or antisense polynucleotide is capable of inhibiting the expression of the one or more transcription factors or other proteins
15 involved in the regulation of the epicenter. The incorporation of a desired polynucleotide into a vector and the choice of vectors is well-known in the art as described in, for example, Sambrook et al., *supra*, and Ausubel et al., *supra*, and elsewhere herein.

The siRNA or antisense polynucleotide can be cloned into a number of types of vectors as described elsewhere herein. For expression of the siRNA or antisense
20 polynucleotide, at least one module in each promoter functions to position the start site for RNA synthesis.

In order to assess the expression of the siRNA or antisense polynucleotide, the expression vector to be introduced into a cell can also contain either a selectable marker gene or a reporter gene or both to facilitate identification and selection of
25 expressing cells from the population of cells sought to be transfected or infected through viral vectors. In other embodiments, the selectable marker may be carried on a separate piece of DNA and used in a co-transfection procedure. Both selectable markers and reporter genes may be flanked with appropriate regulatory sequences to enable expression in the host cells. Useful selectable markers are known in the art and include, for example,
30 antibiotic-resistance genes, such as neomycin resistance and the like.

Antisense nucleic acids

In one embodiment of the invention, an antisense nucleic acid sequence which is expressed by a plasmid vector is used to inhibit a desired epicenter or one or

more transcription factors binding thereto. The antisense expressing vector is used to transfect a mammalian cell or the mammal itself, thereby causing reduced epicenter activity, endogenous expression of the one or more transcription factors or epicenter regulator proteins.

5 Antisense molecules and their use for inhibiting gene expression are well known in the art (*see, e.g.*, Cohen, 1989, In: Oligodeoxyribonucleotides, Antisense Inhibitors of Gene Expression, CRC Press). Antisense nucleic acids are DNA or RNA molecules that are complementary, as that term is defined elsewhere herein, to at least a portion of a specific mRNA molecule (Weintraub, 1990, Scientific American 262:40). In
10 the cell, antisense nucleic acids hybridize to the corresponding mRNA, forming a double-stranded molecule thereby inhibiting the translation of genes.

 The use of antisense methods to inhibit the translation of genes is known in the art, and is described, for example, in Marcus-Sakura (1988, Anal. Biochem. 172:289). Such antisense molecules may be provided to the cell via genetic expression
15 using DNA encoding the antisense molecule as taught by Inoue, 1993, U.S. Patent No. 5,190,931.

 Alternatively, antisense molecules of the invention may be made synthetically and then provided to the cell. Antisense oligomers of between about 10 to about 30, and more preferably about 15 nucleotides, are preferred, since they are easily
20 synthesized and introduced into a target cell. Synthetic antisense molecules contemplated by the invention include oligonucleotide derivatives known in the art which have improved biological activity compared to unmodified oligonucleotides (*see* U.S. Patent No. 5,023,243).

 Compositions and methods for the synthesis and expression of antisense
25 nucleic acids are as described elsewhere herein.

Ribozymes

 Ribozymes and their use for inhibiting gene expression are also well known in the art (*see, e.g.*, Cech et al., 1992, J. Biol. Chem. 267:17479-17482; Hampel et al., 1989, Biochemistry 28:4929-4933; Eckstein et al., International Publication No. WO
30 92/07065; Altman et al., U.S. Patent No. 5,168,053). Ribozymes are RNA molecules possessing the ability to specifically cleave other single-stranded RNA in a manner analogous to DNA restriction endonucleases. Through the modification of nucleotide sequences encoding these RNAs, molecules can be engineered to recognize specific

nucleotide sequences in an RNA molecule and cleave it (Cech, 1988, J. Amer. Med. Assn. 260:3030). A major advantage of this approach is the fact that ribozymes are sequence-specific.

There are two basic types of ribozymes, namely, tetrahymena-type
5 (Hasselhoff, 1988, Nature 334:585) and hammerhead-type. Tetrahymena-type ribozymes recognize sequences which are four bases in length, while hammerhead-type ribozymes recognize base sequences 11-18 bases in length. The longer the sequence, the greater the likelihood that the sequence will occur exclusively in the target mRNA species. Consequently, hammerhead-type ribozymes are preferable to tetrahymena-type ribozymes
10 for inactivating specific mRNA species, and 18-base recognition sequences are preferable to shorter recognition sequences which may occur randomly within various unrelated mRNA molecules.

In one embodiment of the invention, a ribozyme is used to inhibit a desired epicenter, one or more regulators of epicenter activity, or one or more transcription
15 factors binding thereto. Ribozymes useful for inhibiting the expression of a target molecule may be designed by incorporating target sequences into the basic ribozyme structure which are complementary, for example, to the mRNA sequence of the collagen modifying enzyme of the present invention. Ribozymes targeting a desired collagen modifying enzyme may be synthesized using commercially available reagents (Applied
20 Biosystems, Inc., Foster City, CA) or they may be genetically expressed from DNA encoding them.

Small Molecules

When the inhibitor of the invention is a small molecule, a small molecule
25 agonist may be obtained using standard methods known to the skilled artisan. Such methods include chemical organic synthesis or biological means. Biological means include purification from a biological source, recombinant synthesis and in vitro translation systems, using methods well known in the art.

Combinatorial libraries of molecularly diverse chemical compounds
30 potentially useful in treating a variety of diseases and conditions are well known in the art as are method of making the libraries. The method may use a variety of techniques well-known to the skilled artisan including solid phase synthesis, solution methods, parallel synthesis of single compounds, synthesis of chemical mixtures, rigid core structures, flexible linear sequences, deconvolution strategies, tagging techniques, and generating

unbiased molecular landscapes for lead discovery vs. biased structures for lead development.

In a general method for small library synthesis, an activated core molecule is condensed with a number of building blocks, resulting in a combinatorial library of covalently linked, core-building block ensembles. The shape and rigidity of the core determines the orientation of the building blocks in shape space. The libraries can be biased by changing the core, linkage, or building blocks to target a characterized biological structure ("focused libraries") or synthesized with less structural bias using flexible cores.

In one embodiment, the small molecule is able to inhibit one or more epicenters or transcription factor binding domains. In one embodiment, the small molecule induces the expression of the intracellular enzyme procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2). In one embodiment, the small molecule is minoxidil or a salt or chemical analog thereof.

Antagonist

In another aspect of the invention, one or more epicenters, one or more epicenter regulator proteins, or one or more transcription factors binding thereto, can be inhibited by way of inactivating and/or sequestering the epicenter, regulator of epicenter activity, or one or more transcription factors binding thereto. As such, inhibiting the effects of an epicenter, regulator of epicenter activity, or one or more transcription factors binding thereto can be accomplished by using a transdominant negative mutant. Alternatively, an antibody specific for the epicenter, regulator, or one or more transcription factor, otherwise known as an antagonist to the epicenter, regulator, or one or more transcription factor may be used. In one embodiment, the antagonist is a protein and/or compound having the desirable property of interacting with a binding partner of the epicenter, regulator, or one or more transcription factor, and thereby competing with the corresponding protein. In another embodiment, the antagonist is a protein and/or compound having the desirable property of interacting with the epicenter, regular, or one or more transcription factor and thereby sequestering the epicenter, regulator, or one or more transcription factor.

As will be understood by one skilled in the art, any antibody that can recognize and bind to an antigen of interest is useful in the present invention. Methods of making and using antibodies are well known in the art. For example, polyclonal

antibodies useful in the present invention are generated by immunizing rabbits according to standard immunological techniques well-known in the art (*see, e.g.*, Harlow et al., 1988, In: *Antibodies, A Laboratory Manual*, Cold Spring Harbor, NY). Such techniques include immunizing an animal with a chimeric protein comprising a portion of another protein such as a maltose binding protein or glutathione (GSH) tag polypeptide portion, and/or a moiety such that the antigenic protein of interest is rendered immunogenic (*e.g.*, an antigen of interest conjugated with keyhole limpet hemocyanin, KLH) and a portion comprising the respective antigenic protein amino acid residues. The chimeric proteins are produced by cloning the appropriate nucleic acids encoding the marker protein into a plasmid vector suitable for this purpose, such as but not limited to, pMAL-2 or pCMX.

However, the invention should not be construed as being limited solely to methods and compositions including these antibodies or to these portions of the antigens. Rather, the invention should be construed to include other antibodies, as that term is defined elsewhere herein, to antigens, or portions thereof. Further, the present invention should be construed to encompass antibodies, *inter alia*, bind to the specific antigens of interest, and they are able to bind the antigen present on Western blots, in solution in enzyme linked immunoassays, in fluorescence activated cells sorting (FACS) assays, in magnetic-activated cell sorting (MACS) assays, and in immunofluorescence microscopy of a cell transiently transfected with a nucleic acid encoding at least a portion of the antigenic protein, for example.

One skilled in the art would appreciate, based upon the disclosure provided herein, that the antibody can specifically bind with any portion of the antigen and the full-length protein can be used to generate antibodies specific therefor. However, the present invention is not limited to using the full-length protein as an immunogen. Rather, the present invention includes using an immunogenic portion of the protein to produce an antibody that specifically binds with a specific antigen. That is, the invention includes immunizing an animal using an immunogenic portion, or antigenic determinant, of the antigen.

Once armed with the sequence of a specific antigen of interest and the detailed analysis localizing the various conserved and non-conserved domains of the protein, the skilled artisan would understand, based upon the disclosure provided herein, how to obtain antibodies specific for the various portions of the antigen using methods well-known in the art or to be developed.

The skilled artisan would appreciate, based upon the disclosure provided herein, that that present invention includes use of a single antibody recognizing a single antigenic epitope but that the invention is not limited to use of a single antibody. Instead, the invention encompasses use of at least one antibody where the antibodies can be
5 directed to the same or different antigenic protein epitopes.

The generation of polyclonal antibodies is accomplished by inoculating the desired animal with the antigen and isolating antibodies which specifically bind the antigen therefrom using standard antibody production methods such as those described in, for example, Harlow et al. (1988, In: Antibodies, A Laboratory Manual, Cold Spring
10 Harbor, NY).

Monoclonal antibodies directed against full length or peptide fragments of a protein or peptide may be prepared using any well-known monoclonal antibody preparation procedures, such as those described, for example, in Harlow et al. (1988, In: Antibodies, A Laboratory Manual, Cold Spring Harbor, NY) and in Tuszynski et al.
15 (1988, Blood, 72:109-115). Quantities of the desired peptide may also be synthesized using chemical synthesis technology. Alternatively, DNA encoding the desired peptide may be cloned and expressed from an appropriate promoter sequence in cells suitable for the generation of large quantities of peptide. Monoclonal antibodies directed against the peptide are generated from mice immunized with the peptide using standard procedures
20 as referenced herein.

Nucleic acid encoding the monoclonal antibody obtained using the procedures described herein may be cloned and sequenced using technology which is available in the art, and is described, for example, in Wright et al. (1992, Critical Rev. Immunol. 12:125-168), and the references cited therein. Further, the antibody of the
25 invention may be "humanized" using the technology described in, for example, Wright et al., and in the references cited therein, and in Gu et al. (1997, Thrombosis and Hematocyst 77:755-759), and other methods of humanizing antibodies well-known in the art or to be developed.

The present invention also includes the use of humanized antibodies
30 specifically reactive with epitopes of an antigen of interest. The humanized antibodies of the invention have a human framework and have one or more complementarity determining regions (CDRs) from an antibody, typically a mouse antibody, specifically reactive with an antigen of interest. When the antibody used in the invention is humanized, the antibody may be generated as described in Queen, et al. (U.S. Patent No.

6, 180,370), Wright et al., (supra) and in the references cited therein, or in Gu et al. (1997, Thrombosis and Hematocyst 77(4):755-759). The method disclosed in Queen et al. is directed in part toward designing humanized immunoglobulins that are produced by expressing recombinant DNA segments encoding the heavy and light chain

5 complementarity determining regions (CDRs) from a donor immunoglobulin capable of binding to a desired antigen, such as an epitope on an antigen of interest, attached to DNA segments encoding acceptor human framework regions. Generally speaking, the invention in the Queen patent has applicability toward the design of substantially any humanized immunoglobulin. Queen explains that the DNA segments will typically include an

10 expression control DNA sequence operably linked to the humanized immunoglobulin coding sequences, including naturally-associated or heterologous promoter regions. The expression control sequences can be eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells or the expression control sequences can be prokaryotic promoter systems in vectors capable of transforming or transfecting

15 prokaryotic host cells. Once the vector has been incorporated into the appropriate host, the host is maintained under conditions suitable for high level expression of the introduced nucleotide sequences and as desired the collection and purification of the humanized light chains, heavy chains, light/heavy chain dimers or intact antibodies, binding fragments or other immunoglobulin forms may follow (Beychok, Cells of

20 Immunoglobulin Synthesis, Academic Press, New York, (1979), which is incorporated herein by reference).

The invention also includes functional equivalents of the antibodies described herein. Functional equivalents have binding characteristics comparable to those of the antibodies, and include, for example, hybridized and single chain antibodies, as

25 well as fragments thereof. Methods of producing such functional equivalents are disclosed in PCT Application WO 93/21319 and PCT Application WO 89/09622.

Functional equivalents include polypeptides with amino acid sequences substantially the same as the amino acid sequence of the variable or hypervariable regions of the antibodies. "Substantially the same" amino acid sequence is defined herein as a

30 sequence with at least 70%, preferably at least about 80%, more preferably at least about 90%, even more preferably at least about 95%, and most preferably at least 99% homology to another amino acid sequence (or any integer in between 70 and 99), as determined by the FASTA search method in accordance with Pearson and Lipman, 1988 Proc. Nat'l. Acad. Sci. USA 85: 2444-2448. Chimeric or other hybrid antibodies have

constant regions derived substantially or exclusively from human antibody constant regions and variable regions derived substantially or exclusively from the sequence of the variable region of a monoclonal antibody from each stable hybridoma.

Single chain antibodies (scFv) or Fv fragments are polypeptides that
5 consist of the variable region of the heavy chain of the antibody linked to the variable region of the light chain, with or without an interconnecting linker. Thus, the Fv comprises an antibody combining site.

Functional equivalents of the antibodies of the invention further include fragments of antibodies that have the same, or substantially the same, binding
10 characteristics to those of the whole antibody. Such fragments may contain one or both Fab fragments or the F(ab')₂ fragment. The antibody fragments contain all six complement determining regions of the whole antibody, although fragments containing fewer than all of such regions, such as three, four or five complement determining regions, are also functional. The functional equivalents are members of the IgG
15 immunoglobulin class and subclasses thereof, but may be or may combine with any one of the following immunoglobulin classes: IgM, IgA, IgD, or IgE, and subclasses thereof. Heavy chains of various subclasses, such as the IgG subclasses, are responsible for different effector functions and thus, by choosing the desired heavy chain constant region, hybrid antibodies with desired effector function are produced. Exemplary constant
20 regions are gamma 1 (IgG1), gamma 2 (IgG2), gamma 3 (IgG3), and gamma 4 (IgG4). The light chain constant region can be of the kappa or lambda type.

The immunoglobulins of the present invention can be monovalent, divalent or polyvalent. Monovalent immunoglobulins are dimers (HL) formed of a hybrid heavy chain associated through disulfide bridges with a hybrid light chain. Divalent
25 immunoglobulins are tetramers (H₂L₂) formed of two dimers associated through at least one disulfide bridge.

Activators

In various embodiments, the present invention includes compositions and
30 methods of treating wounds in a subject. In various embodiments, the present invention includes compositions and methods of diagnosing and treating cancer. Examples of wounds include skin wounds, skin ulcers, skin punctures, incision wounds, lacerations, avulsions, and the like.

In various embodiments, the composition for treating a wound comprises an activator of one or more epicenters or one or more transcription factor binding thereto. In one embodiment, the activator of the invention increases epicenter activity. In one embodiment, the activator of the invention increases the amount of transcription factor polypeptide, the amount of transcription factor mRNA, the amount of transcription factor activity, or a combination thereof. In some embodiments, the transcription factor is SOX9, KLF (for example KLF5), GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3.

It will be understood by one skilled in the art, based upon the disclosure provided herein, that an increase in the level of epicenter activity or one more transcription factors encompasses the increase in transcription factor expression, including transcription, translation, or both. The skilled artisan will also appreciate, once armed with the teachings of the present invention, that an increase in the accessibility or activity of the epicenter or level of one or more transcription factors includes an increase in epicenter or transcription factor activity. Thus, increasing the epicenter activity or level or activity of the transcription factor includes, but is not limited to, increasing the amount of transcription factor polypeptide, increasing transcription, translation, or both, of a nucleic acid encoding transcription factor; and it also includes increasing any activity of a transcription factor polypeptide as well. It additionally includes transcriptional activity localized to the epicenter or regulated by the activity of the epicenter.

Thus, the present invention relates to the prevention and treatment of a disease (e.g., cancer) or disorder by administration of a transcription factor polypeptide, a recombinant transcription factor polypeptide, an active transcription factor polypeptide fragment, an activator of transcription factor expression or activity, or activator of epicenter activity.

It is understood by one skilled in the art, that an increase in the activity of the epicenter or level of epicenter regulator protein or transcription factor binding thereto encompasses the increase of regulator protein or transcription factor protein expression. Additionally, the skilled artisan would appreciate, that an increase in the level of regulator protein or transcription factor includes an increase in epicenter or transcription factor activity. Thus, increasing the level or activity of regulator protein or transcription factor includes, but is not limited to, increasing transcription, translation, or both, of a nucleic acid encoding regulator protein or transcription factor; and it also includes increasing any activity of one or more regulator proteins or transcription factors as well.

Activation of transcription factors can be assessed using a wide variety of methods, including those disclosed herein, as well as methods well-known in the art or to be developed in the future. That is, the routineer would appreciate, based upon the disclosure provided herein, that increasing the level or activity of epicenter, regulator protein or one or more transcription factor can be readily assessed using methods that assess the level of a nucleic acid encoding regulator protein or transcription factor (e.g., mRNA) and/or the level of regulator protein or transcription factor polypeptide in a biological sample obtained from a subject.

An epicenter, regulator protein or transcription factor activator can include, but should not be construed as being limited to, a chemical compound, a protein, a peptidomimetic, an antibody, a nucleic acid molecule. One of skill in the art would readily appreciate, based on the disclosure provided herein, that an epicenter activator, regulator protein activator, or transcription factor activator encompasses a chemical compound, a protein, a peptidomimetic, an antibody, and/or a nucleic acid molecule that increases the level, enzymatic activity, or the like of epicenter, regulator protein, or transcription factor. In some embodiments, the enzymatic activity is post-translational modification (e.g., phosphorylation, acetylation, SUMOylation, ubiquitylation). Additionally, an epicenter, regulator protein, or transcription factor activator encompasses a chemically modified compound, and derivatives, as is well known to one of skill in the chemical arts.

It will be understood by one skilled in the art, based upon the disclosure provided herein, that an increase in the level of transcription factor or regulator protein encompasses the increase in regulator protein or transcription factor expression, including transcription, translation, or both. The skilled artisan will also appreciate, once armed with the teachings of the present invention, that an increase in the level of transcription factor includes an increase in regulator protein or transcription factor activity (e.g., enzymatic activity, receptor binding activity, etc.). Thus, increasing the level or activity of regulator protein or transcription factor includes, but is not limited to, increasing the amount of regulator protein or transcription factor polypeptide, increasing transcription, translation, or both, of a nucleic acid encoding regulator protein or transcription factor; and it also includes increasing any activity of a regulator protein or transcription factor polypeptide as well. The epicenter, regulator protein, and one or more transcription factor activator compositions and methods of the invention can selectively activate epicenter, regulator protein, or transcription factor. Thus, the present invention relates to disease

treatment (i.e., cancer treatment) by administration of a regulator protein or transcription factor polypeptide, a recombinant regulator protein or transcription factor polypeptide, an active regulator protein or transcription factor polypeptide fragment, or an activator of regulator protein or transcription factor expression or activity.

5 Further, one of skill in the art would, when equipped with this disclosure and the methods exemplified herein, appreciate that an epicenter activator, regulator protein activator, or transcription factor activator includes such activators as discovered in the future, as can be identified by well-known criteria in the art of pharmacology, such as the physiological results of activation of epicenters, regulator proteins, or transcription
10 factors as described in detail herein and/or as known in the art. Therefore, the present invention is not limited in any way to any particular epicenter, regulator protein, or transcription factor activator as exemplified or disclosed herein; rather, the invention encompasses those activators that would be understood by the routineer to be useful as are known in the art and as are discovered in the future.

15 Further methods of identifying and producing an epicenter activator, regulator protein activator, or transcription factor activator are well known to those of ordinary skill in the art, including, but not limited to, obtaining an activator from a naturally occurring source. Alternatively, an epicenter, regulator protein, or transcription factor activator can be synthesized chemically. Further, the routineer would appreciate,
20 based upon the teachings provided herein, that an epicenter, regulator protein, or transcription factor activator can be obtained from a recombinant organism. Compositions and methods for chemically synthesizing epicenter, regulator protein, or transcription factor activators and for obtaining them from natural sources are well known in the art and are described in the art.

25 One of skill in the art will appreciate that an activator can be administered as a small molecule chemical, a protein, a nucleic acid construct encoding a protein, or combinations thereof. Numerous vectors and other compositions and methods are well known for administering a protein or a nucleic acid construct encoding a protein to cells or tissues. Therefore, the invention includes a method of administering a protein or a
30 nucleic acid encoding a protein that is an activator of an epicenter, regulator protein, or one or more transcription factor.

One of skill in the art will realize that diminishing the amount or activity of a molecule that itself diminishes the amount or activity of epicenter, regulator protein, or transcription factor can serve to increase the amount or activity of epicenter, regulator

protein, or transcription factor. Any inhibitor of a regulator of epicenter, regulator protein, or transcription factor is encompassed in the invention. As a non-limiting example, antisense is described as a form of inhibiting a regulator of epicenter, regulator protein or transcription factor in order to increase the amount or activity of epicenter, regulator

5 protein, or transcription factor. Antisense oligonucleotides are DNA or RNA molecules that are complementary to some portion of a mRNA molecule. When present in a cell, antisense oligonucleotides hybridize to an existing mRNA molecule and inhibit translation into a gene product. Inhibiting the expression of a gene using an antisense oligonucleotide is well known in the art (Marcus-Sekura, 1988, Anal. Biochem. 172:289),

10 as are methods of expressing an antisense oligonucleotide in a cell (Inoue, U.S. Pat. No. 5,190,931). The methods of the invention include the use of antisense oligonucleotide to diminish the amount of a molecule that causes a decrease in the amount or activity or epicenter, regulator protein, or transcription factor, thereby increasing the amount or activity of epicenter, regulator protein, or transcription factor. Contemplated in the

15 present invention are antisense oligonucleotides that are synthesized and provided to the cell by way of methods well known to those of ordinary skill in the art. As an example, an antisense oligonucleotide can be synthesized to be between about 10 and about 100, more preferably between about 15 and about 50 nucleotides long. The synthesis of nucleic acid molecules is well known in the art, as is the synthesis of modified antisense

20 oligonucleotides to improve biological activity in comparison to unmodified antisense oligonucleotides (Tullis, 1991, U.S. Pat. No. 5,023,243).

Similarly, the expression of a gene may be inhibited by the hybridization of an antisense molecule to a promoter or other regulatory element of a gene, thereby affecting the transcription of the gene. Methods for the identification of a promoter or

25 other regulatory element that interacts with a gene of interest are well known in the art, and include such methods as the yeast two hybrid system (Bartel and Fields, eds., In: The Yeast Two Hybrid System, Oxford University Press, Cary, N.C.).

Alternatively, inhibition of a gene expressing a protein that diminishes the level or activity of an epicenter, regulator protein or one or more transcription factors can

30 be accomplished through the use of a ribozyme. Using ribozymes for inhibiting gene expression is well known to those of skill in the art (see, e.g., Cech et al., 1992, J. Biol. Chem. 267:17479; Hampel et al., 1989, Biochemistry 28: 4929; Altman et al., U.S. Pat. No. 5,168,053). Ribozymes are catalytic RNA molecules with the ability to cleave other single-stranded RNA molecules. Ribozymes are known to be sequence specific, and can

therefore be modified to recognize a specific nucleotide sequence (Cech, 1988, J. Amer. Med. Assn. 260:3030), allowing the selective cleavage of specific mRNA molecules. Given the nucleotide sequence of the molecule, one of ordinary skill in the art could synthesize an antisense oligonucleotide or ribozyme without undue experimentation, provided with the disclosure and references incorporated herein.

One of skill in the art will appreciate that a regulator protein or transcription factor polypeptide, a recombinant regulator protein or transcription factor polypeptide, or an active regulator protein or transcription factor polypeptide fragment can be administered singly or in any combination thereof. Further, a regulator protein or transcription factor polypeptide, a recombinant regulator protein or transcription factor polypeptide, or an active regulator protein or transcription factor polypeptide fragment can be administered singly or in any combination thereof in a temporal sense, in that they may be administered simultaneously, before, and/or after each other. One of ordinary skill in the art will appreciate, based on the disclosure provided herein, that a regulator protein or transcription factor polypeptide, a recombinant regulator protein or transcription factor polypeptide, or an active regulator protein or transcription factor polypeptide fragment can be used to prevent or treat a disease or disorder such as cancer, and that an activator can be used alone or in any combination with another regulator protein or transcription factor polypeptide, recombinant regulator protein or transcription factor polypeptide, active regulator protein or transcription factor polypeptide fragment, or regulator protein or transcription factor activator to effect a therapeutic result.

One of skill in the art, when armed with the disclosure herein, would appreciate that the treating a disease or disorder such as cancer encompasses administering to a subject a polypeptide, a recombinant polypeptide, an active polypeptide fragment, or activator as a preventative measure against a disease or disorder such as cancer. As more fully discussed elsewhere herein, methods of increasing the level or activity of an epicenter, regulator protein, or transcription factor encompass a wide plethora of techniques for increasing not only epicenter, regulator protein, or transcription factor activity, but also for increasing expression of a nucleic acid encoding regulator protein or transcription factor. Additionally, as disclosed elsewhere herein, one skilled in the art would understand, once armed with the teaching provided herein, that the present invention encompasses a method of preventing a wide variety of diseases where increased expression and/or activity of regulator protein or transcription factor mediates, treats or

prevents the disease. Further, the invention encompasses treatment or prevention of such diseases discovered in the future.

The invention encompasses administration of a polypeptide, a recombinant polypeptide, an active polypeptide fragment, or an activator to practice the methods of the invention; the skilled artisan would understand, based on the disclosure provided herein, how to formulate and administer the appropriate polypeptide, recombinant polypeptide, active polypeptide fragment, or activator to a subject. However, the present invention is not limited to any particular method of administration or treatment regimen. This is especially true where it would be appreciated by one skilled in the art, equipped with the disclosure provided herein, including the reduction to practice using an art-recognized model of a neurodegenerative disease, that methods of administering a polypeptide, a recombinant polypeptide, an active polypeptide fragment, or activator can be determined by one of skill in the pharmacological arts.

As used herein, the term “pharmaceutically-acceptable carrier” means a chemical composition with which an appropriate polypeptide, recombinant polypeptide, active polypeptide fragment, or activator, may be combined and which, following the combination, can be used to administer the appropriate polypeptide, recombinant polypeptide, active polypeptide fragment, or activator to a subject.

Probes

In one embodiment, the invention relates to probes used to detect epicenters, for example cancer-related epicenters. In one embodiment, the probes are transcription reporter constructs comprising promoters of genes determined to be regulated by transcription factors associated with epicenters of interest. In one embodiment, the construct comprises promoters or genes that are involved in signaling pathways of interest. In some embodiments, the pathways of interest include lineage infidelity, cell differentiation, stem cell plasticity in different microenvironments, epithelial-mesenchymal transition, cancer, and wound healing. In some embodiments, the probes comprise an internal H2BRFP control and eGFP tagged to a promoter of a transcription factor of interest (e.g., SOX9, KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3, either individually or as a cohort). In some embodiments, the reporter is a fluorescent protein. In some embodiments, the reporter is a luciferase reporter. In some embodiments, the reporter is a CRE recombinase. In some embodiments, the reporter is a CAS9 enzyme.

Methods

In one aspect, the invention provides methods of treating or preventing cancer. In some embodiments, the invention provides methods of treating cancer
5 comprise modulating epicenters or super-enhancers determined to be differentially regulated in cancer. In some embodiments, the method comprises modulating epicenter activity. In some embodiments, the invention provides methods for modulating the assembly of transcriptional machinery specific to one or more epicenters.

In some embodiments, the method comprises administering to a subject in
10 need thereof, one or more compositions of the invention described herein.

In some embodiments, the invention provides methods for detecting or diagnosing cancer in a subject or biological sample obtained from a subject. For example, in certain embodiments, the method comprises detecting the accessibility of the transcription epicenter or super-enhancer. In one embodiment, the method comprises
15 detecting the presence or abundance of one or more components associated with accessibility of the transcription epicenter or super-enhancer. In one embodiment, the method comprises detecting the presence or abundance of one or more factors that bind to the transcription epicenter or super-enhancer. In one embodiment, the method comprises detecting the presence or abundance of one or more transcripts associated with
20 transcription epicenter or super-enhancer.

Treating Cancer

One aspect of the invention provides a method of treating cancer in an individual, the method comprising administering to the individual an effective cancer-
25 inhibiting amount of an inhibitor of epicenter activity. The invention further provides a method of inhibiting cancer in an individual in need thereof, the method comprising administering to the individual an effective cancer-inhibiting amount of any one of the compositions described herein.

The disclosed compounds can be used to prevent, abate, minimize, control,
30 and/or lessen cancer in humans and animals. The disclosed compounds can also be used to slow the rate of primary cancerous growth. The disclosed compounds when administered to a subject in need of treatment can be used to stop the spread of cancer cells. As such, the compounds disclosed herein can be administered as part of a combination therapy with one or more drugs or other pharmaceutical agents. When used

as part of the combination therapy, the decrease in cancer and reduction in primary cancerous growth afforded by the disclosed compounds allows for a more effective and efficient use of any pharmaceutical or drug therapy being used to treat the patient. In addition, control of cancer by the disclosed compound affords the subject a greater ability to concentrate the disease in one location.

In one embodiment, the invention provides methods for preventing cancer or other cancerous cells as well as to reduce the rate of tumor growth. The methods comprise administering an effective amount of one or more of the disclosed compounds to a subject diagnosed with a cancer or cancerous cells or to a subject having cancer or cancerous cells.

The following are non-limiting examples of cancers that can be treated by the disclosed methods and compositions: Acute Lymphoblastic; Acute Myeloid Leukemia; Adrenocortical Carcinoma; Adrenocortical Carcinoma, Childhood; Appendix Cancer; Basal Cell Carcinoma; Bile Duct Cancer, Extrahepatic; Bladder Cancer; Bone Cancer; Osteosarcoma and Malignant Fibrous Histiocytoma; Brain Stem Glioma, Childhood; Brain Tumor, Adult; Brain Tumor, Brain Stem Glioma, Childhood; Brain Tumor, Central Nervous System Atypical Teratoid/Rhabdoid Tumor, Childhood; Central Nervous System Embryonal Tumors; Cerebellar Astrocytoma; Cerebral Astrocytoma/Malignant Glioma; Craniopharyngioma; Ependymoblastoma; Ependymoma; Medulloblastoma; Medulloepithelioma; Pineal Parenchymal Tumors of intermediate Differentiation; Supratentorial Primitive Neuroectodermal Tumors and Pineoblastoma; Visual Pathway and Hypothalamic Glioma; Brain and Spinal Cord Tumors; Breast Cancer; Bronchial Tumors; Burkitt Lymphoma; Carcinoid Tumor; Carcinoid Tumor, Gastrointestinal; Central Nervous System Atypical Teratoid/Rhabdoid Tumor; Central Nervous System Embryonal Tumors; Central Nervous System Lymphoma; Cerebellar Astrocytoma Cerebral Astrocytoma/Malignant Glioma, Childhood; Cervical Cancer; Chordoma, Childhood; Chronic Lymphocytic Leukemia; Chronic Myelogenous Leukemia; Chronic Myeloproliferative Disorders; Colon Cancer; Colorectal Cancer; Craniopharyngioma; Cutaneous T-Cell Lymphoma; Esophageal Cancer; Ewing Family of Tumors; Extragonadal Germ Cell Tumor; Extrahepatic Bile Duct Cancer; Eye Cancer, intraocular Melanoma; Eye Cancer, Retinoblastoma; Gallbladder Cancer; Gastric (Stomach) Cancer; Gastrointestinal Carcinoid Tumor; Gastrointestinal Stromal Tumor (GIST); Germ Cell Tumor, Extracranial; Germ Cell Tumor, Extragonadal; Germ Cell Tumor, Ovarian; Gestational Trophoblastic Tumor;

- Glioma; Glioma, Childhood Brain Stem; Glioma, Childhood Cerebral Astrocytoma; Glioma, Childhood Visual Pathway and Hypothalamic; Hairy Cell Leukemia; Head and Neck Cancer; Hepatocellular (Liver) Cancer; Histiocytosis, Langerhans Cell; Hodgkin Lymphoma; Hypopharyngeal Cancer; Hypothalamic and Visual Pathway Glioma;
- 5 intraocular Melanoma; Islet Cell Tumors; Kidney (Renal Cell) Cancer; Langerhans Cell Histiocytosis; Laryngeal Cancer; Leukemia, Acute Lymphoblastic; Leukemia, Acute Myeloid; Leukemia, Chronic Lymphocytic; Leukemia, Chronic Myelogenous; Leukemia, Hairy Cell; Lip and Oral Cavity Cancer; Liver Cancer; Lung Cancer, Non-Small Cell; Lung Cancer, Small Cell; Lymphoma, AIDS-Related; Lymphoma, Burkitt; Lymphoma,
- 10 Cutaneous T-Cell; Lymphoma, Hodgkin; Lymphoma, Non-Hodgkin; Lymphoma, Primary Central Nervous System; Macroglobulinemia, Waldenstrom; Malignant Fibrous Histiocytoma of Bone and Osteosarcoma; Medulloblastoma; Melanoma; Melanoma, intraocular (Eye); Merkel Cell Carcinoma; Mesothelioma; Metastatic Squamous Neck Cancer with Occult Primary; Mouth Cancer; Multiple Endocrine Neoplasia Syndrome,
- 15 (Childhood); Multiple Myeloma/Plasma Cell Neoplasm; Mycosis; Fungoides; Myelodysplastic Syndromes; Myelodysplastic/Myeloproliferative Diseases; Myelogenous Leukemia, Chronic; Myeloid Leukemia, Adult Acute; Myeloid Leukemia, Childhood Acute; Myeloma, Multiple; Myeloproliferative Disorders, Chronic; Nasal Cavity and Paranasal Sinus Cancer; Nasopharyngeal Cancer; Neuroblastoma; Non-Small Cell Lung
- 20 Cancer; Oral Cancer; Oral Cavity Cancer; Oropharyngeal Cancer; Osteosarcoma and Malignant Fibrous Histiocytoma of Bone; Ovarian Cancer; Ovarian Epithelial Cancer; Ovarian Germ Cell Tumor; Ovarian Low Malignant Potential Tumor; Pancreatic Cancer; Pancreatic Cancer, Islet Cell Tumors; Papillomatosis; Parathyroid Cancer; Penile Cancer; Pharyngeal Cancer; Pheochromocytoma; Pineal Parenchymal Tumors of Intermediate
- 25 Differentiation; Pineoblastoma and Supratentorial Primitive Neuroectodermal Tumors; Pituitary Tumor; Plasma Cell Neoplasm/Multiple Myeloma; Pleuropulmonary Blastoma; Primary Central Nervous System Lymphoma; Prostate Cancer; Rectal Cancer; Renal Cell (Kidney) Cancer; Renal Pelvis and Ureter, Transitional Cell Cancer; Respiratory Tract Carcinoma Involving the NUT Gene on Chromosome 15; Retinoblastoma;
- 30 Rhabdomyosarcoma; Salivary Gland Cancer; Sarcoma, Ewing Family of Tumors; Sarcoma, Kaposi; Sarcoma, Soft Tissue; Sarcoma, Uterine; Sezary Syndrome; Skin Cancer (Nonmelanoma); Skin Cancer (Melanoma); Skin Carcinoma, Merkel Cell; Small Cell Lung Cancer; Small Intestine Cancer; Soft Tissue Sarcoma; Squamous Cell Carcinoma, Head and Neck Squamous Cell Carcinoma, Squamous Neck Cancer with

Occult Primary, Metastatic; Stomach (Gastric) Cancer; Supratentorial Primitive Neuroectodermal Tumors; T-Cell Lymphoma, Cutaneous; Testicular Cancer; Throat Cancer; Thymoma and Thymic Carcinoma; Thyroid Cancer; Transitional Cell Cancer of the Renal Pelvis and Ureter; Trophoblastic Tumor, Gestational; Urethral Cancer; Uterine
 5 Cancer, Endometrial; Uterine Sarcoma; Vaginal Cancer; Vulvar Cancer; Waldenstrom Macroglobulinemia; and Wilms Tumor.

Chemotherapeutic agents include cytotoxic agents (e.g., 5-fluorouracil, cisplatin, carboplatin, methotrexate, daunorubicin, doxorubicin, vincristine, vinblastine, oxorubicin, carmustine (BCNU), lomustine (CCNU), cytarabine USP, cyclophosphamide,
 10 estramucine phosphate sodium, altretamine, hydroxyurea, ifosfamide, procarbazine, mitomycin, busulfan, cyclophosphamide, mitoxantrone, carboplatin, cisplatin, interferon alfa-2a recombinant, paclitaxel, teniposide, and streptozocin), cytotoxic alkylating agents (e.g., busulfan, chlorambucil, cyclophosphamide, melphalan, or ethylesulfonic acid), alkylating agents (e.g., asaley, AZQ, BCNU, busulfan, bisulphan,
 15 carboxyphthalatoplatinum, CBDCA, CCNU, CHIP, chlorambucil, chlorozotocin, cis-platinum, clomesone, cyanomorpholinodoxorubicin, cyclodisone, cyclophosphamide, dianhydrogalactitol, fluorodopan, hepsulfam, hycanthone, iphosphamide, melphalan, methyl CCNU, mitomycin C, mitozolamide, nitrogen mustard, PCNU, piperazine, piperazinedione, pipobroman, porfiromycin, spirohydantoin mustard, streptozotocin,
 20 teroxirone, tetraplatin, thiotepa, triethylenemelamine, uracil nitrogen mustard, and Yoshi-864), antimetabolic agents (e.g., allicolchicine, Halichondrin M, colchicine, colchicine derivatives, dolastatin 10, maytansine, rhizoxin, paclitaxel derivatives, paclitaxel, thiocolchicine, trityl cysteine, vinblastine sulfate, and vincristine sulfate), plant alkaloids (e.g., actinomycin D, bleomycin, L-asparaginase, idarubicin, vinblastine sulfate,
 25 vincristine sulfate, mitramycin, mitomycin, daunorubicin, VP-16-213, VM-26, navelbine and taxotere), biologicals (e.g., alpha interferon, BCG, G-CSF, GM-CSF, and interleukin-2), topoisomerase I inhibitors (e.g., camptothecin, camptothecin derivatives, and morpholinodoxorubicin), topoisomerase II inhibitors (e.g., mitoxantron, amonafide, m-AMSA, anthrapyrazole derivatives, pyrazoloacridine, bisantrene HCL, daunorubicin,
 30 deoxydoxorubicin, menogaril, N,N-dibenzyl daunomycin, oxanthrazole, rubidazone, VM-26 and VP-16), and synthetics (e.g., hydroxyurea, procarbazine, o,p'-DDD, dacarbazine, CCNU, BCNU, cis-diamminedichloroplatinum, mitoxantrone, CBDCA, levamisole, hexamethylmelamine, all-trans retinoic acid, gliadel and porfimer sodium).

Antiproliferative agents are compounds that decrease the proliferation of cells. Antiproliferative agents include alkylating agents, antimetabolites, enzymes, biological response modifiers, miscellaneous agents, hormones and antagonists, androgen inhibitors (e.g., flutamide and leuprolide acetate), antiestrogens (e.g., tamoxifen citrate and analogs thereof, toremifene, droloxifene and roloxifene), Additional examples of
5 specific antiproliferative agents include, but are not limited to levamisole, gallium nitrate, granisetron, sargramostim strontium-89 chloride, filgrastim, pilocarpine, dexrazoxane, and ondansetron.

The inhibitors of the invention can be administered alone or in
10 combination with other anti-cancer agents, including cytotoxic/antineoplastic agents and anti-angiogenic agents. Cytotoxic/anti-neoplastic agents are defined as agents which attack and kill cancer cells. Some cytotoxic/anti-neoplastic agents are alkylating agents, which alkylate the genetic material in tumor cells, e.g., cis-platin, cyclophosphamide, nitrogen mustard, trimethylene thiophosphoramide, carmustine, busulfan, chlorambucil,
15 belustine, uracil mustard, chlomaphazin, and dacabazine. Other cytotoxic/anti-neoplastic agents are antimetabolites for tumor cells, e.g., cytosine arabinoside, fluorouracil, methotrexate, mercaptopurine, azathioprine, and procarbazine. Other cytotoxic/anti-neoplastic agents are antibiotics, e.g., doxorubicin, bleomycin, dactinomycin, daunorubicin, mithramycin, mitomycin, mytomycin C, and daunomycin. There are
20 numerous liposomal formulations commercially available for these compounds. Still other cytotoxic/anti-neoplastic agents are mitotic inhibitors (vinca alkaloids). These include vincristine, vinblastine and etoposide. Miscellaneous cytotoxic/anti-neoplastic agents include taxol and its derivatives, L-asparaginase, anti-tumor antibodies, dacarbazine, azacytidine, amsacrine, melphalan, VM-26, ifosfamide, mitoxantrone, and
25 vindesine.

Anti-angiogenic agents are well known to those of skill in the art. Suitable anti-angiogenic agents for use in the methods and compositions of the present disclosure include anti-VEGF antibodies, including humanized and chimeric antibodies, anti-VEGF aptamers and antisense oligonucleotides. Other known inhibitors of angiogenesis include
30 angiostatin, endostatin, interferons, interleukin 1 (including alpha and beta) interleukin 12, retinoic acid, and tissue inhibitors of metalloproteinase-1 and -2. (TIMP-1 and -2). Small molecules, including topoisomerases such as razoxane, a topoisomerase II inhibitor with anti-angiogenic activity, can also be used.

Other anti-cancer agents that can be used in combination with the disclosed compounds include, but are not limited to: acivicin; aclarubicin; acodazole hydrochloride; acronine; adozelesin; aldesleukin; altretamine; ambomycin; ametantrone acetate; aminoglutethimide; amsacrine; anastrozole; anthramycin; asparaginase; asperlin; azacitidine; azetepa; azotomycin; batimastat; benzodepa; bicalutamide; bisantrene hydrochloride; bisnafide dimesylate; bizelesin; bleomycin sulfate; brequinar sodium; bropiramine; busulfan; cactinomycin; calusterone; caracemide; carbetimer; carboplatin; carmustine; carubicin hydrochloride; carzelesin; cedefingol; chlorambucil; cirolemycin; cisplatin; cladribine; crisnatol mesylate; cyclophosphamide; cytarabine; dacarbazine; dactinomycin; daunorubicin hydrochloride; decitabine; dexormaplatin; dezaguanine; dezaguanine mesylate; diaziquone; docetaxel; doxorubicin; doxorubicin hydrochloride; droloxifene; droloxifene citrate; dromostanolone propionate; duazomycin; edatrexate; eflornithine hydrochloride; elsamitrucin; enloplatin; enpromate; epipropidine; epirubicin hydrochloride; erbulozole; esorubicin hydrochloride; estramustine; estramustine phosphate sodium; etanidazole; etoposide; etoposide phosphate; etoprine; fadrozole hydrochloride; fazarabine; fenretinide; floxuridine; fludarabine phosphate; fluorouracil; fluorocitabine; fosquidone; fostriecin sodium; gemcitabine; gemcitabine hydrochloride; hydroxyurea; idarubicin hydrochloride; ifosfamide; ilmofofosine; interleukin II (including recombinant interleukin II, or rIL2), interferon alfa-2a; interferon alfa-2b; interferon alfa-n1; interferon alfa-n3; interferon beta-I a; interferon gamma-I b; iproplatin; irinotecan hydrochloride; lanreotide acetate; letrozole; leuprolide acetate; liarozole hydrochloride; lometrexol sodium; lomustine; losoxantrone hydrochloride; masoprocol; maytansine; mechlorethamine hydrochloride; megestrol acetate; melengestrol acetate; melphalan; menogaril; mercaptopurine; methotrexate; methotrexate sodium; metoprine; meturedopa; mitindomide; mitocarcin; mitocromin; mitogillin; mitomalcin; mitomycin; mitosper; mitotane; mitoxantrone hydrochloride; mycophenolic acid; nocodazole; nogalamycin; ormaplatin; oxisuran; paclitaxel; pegaspargase; peliomycin; pentamustine; peplomycin sulfate; perfosfamide; pipobroman; piposulfan; piroxantrone hydrochloride; plicamycin; plomestane; porfimer sodium; porfiriomycin; prednimustine; procarbazine hydrochloride; puromycin; puromycin hydrochloride; pyrazofurin; riboprine; rogletimide; safingol; safingol hydrochloride; semustine; simtrazene; sparfosate sodium; sparsomycin; spirogermanium hydrochloride; spiromustine; spiroplatin; streptonigrin; streptozocin; sulofenur; talisomycin; tecogalan sodium; tegafur; teloxantrone hydrochloride; temoporfin; teniposide; teroxirone; testolactone; thiamiprine; thioguanine; thiotepa;

tiazofurin; tirapazamine; toremifene citrate; trestolone acetate; tricyriline phosphate; trimetrexate; trimetrexate glucuronate; triptorelin; tubulozole hydrochloride; uracil mustard; uredepa; vapreotide; verteporfin; vinblastine sulfate; vincristine sulfate; vindesine; vindesine sulfate; vinepidine sulfate; vinglycinate sulfate; vinleurosine sulfate; 5 vinorelbine tartrate; vinrosidine sulfate; vinzolidine sulfate; vorozole; zeniplatin; zinostatin; zorubicin hydrochloride. Other anti-cancer drugs include, but are not limited to: 20-epi-1,25 dihydroxyvitamin D3; 5-ethynyluracil; abiraterone; aclarubicin; acylfulvene; adecypenol; adozelesin; aldesleukin; ALL-TK antagonists; altretamine; ambamustine; amidox; amifostine; aminolevulinic acid; amrubicin; amsacrine; 10 anagrelide; anastrozole; andrographolide; angiogenesis inhibitors; antagonist D; antagonist G; antarelix; anti-dorsalizing morphogenetic protein-1; antiandrogen, prostatic carcinoma; antiestrogen; antineoplaston; antisense oligonucleotides; aphidicolin glycinate; apoptosis gene modulators; apoptosis regulators; apurinic acid; ara-CDP-DL-PTBA; arginine deaminase; asulacrine; atamestane; atrimustine; axinastatin 1; axinastatin 15 2; axinastatin 3; azasetron; azatoxin; azatyrosine; baccatin III derivatives; balanol; batimastat; BCR/ABL antagonists; benzochlorins; benzoylstauroporine; beta lactam derivatives; beta-alethine; betaclamycin B; betulinic acid; bFGF inhibitor; bicalutamide; bisantrene; bisaziridinylspermine; bisnafide; bistratene A; bizelesin; breflate; bropirimine; budotitane; buthionine sulfoximine; calcipotriol; calphostin C; camptothecin derivatives; 20 canarypox IL-2; capecitabine; carboxamide-amino-triazole; carboxyamidotriazole; CaRest M3; CARN 700; cartilage derived inhibitor; carzelesin; casein kinase inhibitors (ICOS); castanospermine; cecropin B; cetorelix; chlorins; chloroquinoxaline sulfonamide; cicaprost; cis-porphyrin; cladribine; clomifene analogues; clotrimazole; collismycin A; collismycin B; combretastatin A4; combretastatin analogue; conagenin; 25 crambescidin 816; crisnatol; cryptophycin 8; cryptophycin A derivatives; curacin A; cyclopentantraquinones; cycloplatam; cypemycin; cytarabine ocfosfate; cytolytic factor; cytostatin; dacliximab; decitabine; dehydrotidemnin B; deslorelin; dexamethasone; dexifosfamide; dexrazoxane; dexverapamil; diaziqune; didemnin B; didox; diethylnorspermine; dihydro-5-azacytidine; dihydrotaxol, 9-; dioxamycin; diphenyl 30 spiromustine; docetaxel; docosanol; dolasetron; doxifluridine; droloxifene; dronabinol; duocarmycin SA; ebselen; ecomustine; edelfosine; edrecolomab; eflornithine; elemene; emitefur; epirubicin; epristeride; estramustine analogue; estrogen agonists; estrogen antagonists; etanidazole; etoposide phosphate; exemestane; fadrozole; fazarabine; fenretinide; filgrastim; finasteride; flavopiridol; flezelastine; fluasterone; fludarabine;

- fluorodaunorubicin hydrochloride; forfenimex; formestane; fostriecin; fotemustine; gadolinium texaphyrin; gallium nitrate; galocitabine; ganirelix; gelatinase inhibitors; gemcitabine; glutathione inhibitors; hepsulfam; heregulin; hexamethylene bisacetamide; hypericin; ibandronic acid; idarubicin; idoxifene; idramantone; ilmofofosine; ilomastat;
- 5 imidazoacridones; imiquimod; immunostimulant peptides; insulin-like growth factor-1 receptor inhibitor; interferon agonists; interferons; interleukins; iobenguane; iododoxorubicin; ipomeanol, 4-; iroplact; irsogladine; isobengazole; isohomohalicondrin B; itasetron; jasplakinolide; kahalalide F; lamellarin-N triacetate; lanreotide; leinamycin; lenograstim; lentinan sulfate; leptolstatin; letrozole; leukemia inhibiting factor; leukocyte
- 10 alpha interferon; leuprolide+estrogen+progesterone; leuprorelin; levamisole; liarozole; linear polyamine analogue; lipophilic disaccharide peptide; lipophilic platinum compounds; lissoclinamide 7; lobaplatin; lombricine; lometrexol; lonidamine; losoxantrone; lovastatin; loxoribine; lurtotecan; lutetium texaphyrin; lysofylline; lytic peptides; maitansine; mannostatin A; marimastat; masoprocol; maspin; matrilysin
- 15 inhibitors; matrix metalloproteinase inhibitors; menogaril; merbarone; meterelin; methioninase; metoclopramide; MIF inhibitor; mifepristone; miltefosine; mirimostim; mismatched double stranded RNA; mitoguazone; mitolactol; mitomycin analogues; mitonafide; mitotoxin fibroblast growth factor-saporin; mitoxantrone; mofarotene; molgramostim; monoclonal antibody, human chorionic gonadotrophin; monophosphoryl
- 20 lipid A+myobacterium cell wall sk; mopidamol; multiple drug resistance gene inhibitor; multiple tumor suppressor 1-based therapy; mustard anticancer agent; mycaperoxide B; mycobacterial cell wall extract; myriaporone; N-acetyldinaline; N-substituted benzamides; nafarelin; nagrestip; naloxone+pentazocine; napavin; naphterpin; nartograstim; nedaplatin; nemorubicin; neridronic acid; neutral endopeptidase;
- 25 nilutamide; nisamycin; nitric oxide modulators; nitroxide antioxidant; nitrullyn; O6-benzylguanine; octreotide; okicenone; oligonucleotides; onapristone; ondansetron; ondansetron; oracin; oral cytokine inducer; ormaplatin; osaterone; oxaliplatin; oxaunomycin; paclitaxel; paclitaxel analogues; paclitaxel derivatives; palauamine; palmitoylrhizoxin; pamidronic acid; panaxytriol; panomifene; parabactin; pazelliptine;
- 30 pegaspargase; peldesine; pentosan polysulfate sodium; pentostatin; pentozole; perflubron; perfosfamide; perillyl alcohol; phenazinomycin; phenylacetate; phosphatase inhibitors; picibanil; pilocarpine hydrochloride; pirarubicin; piritrexim; placetin A; placetin B; plasminogen activator inhibitor; platinum complex; platinum compounds; platinum-triamine complex; porfimer sodium; porfironomycin; prednisone; propyl bis-

- acridone; prostaglandin J2; proteasome inhibitors; protein A-based immune modulator; protein kinase C inhibitor; protein kinase C inhibitors, microalgal; protein tyrosine phosphatase inhibitors; purine nucleoside phosphorylase inhibitors; purpurins; pyrazoloacridine; pyridoxylated hemoglobin polyoxyethylene conjugate; raf antagonists; raltitrexed; ramosetron; ras farnesyl protein transferase inhibitors; ras inhibitors; ras-GAP inhibitor; retelliptine demethylated; rhenium Re 186 etidronate; rhizoxin; ribozymes; RII retinamide; rogletimide; rohitukine; romurtide; roquinimex; rubiginone B1; ruboxyl; safingol; saintopin; SarCNU; sarcophytol A; sargramostim; Sdi 1 mimetics; semustine; senescence derived inhibitor 1; sense oligonucleotides; signal transduction inhibitors; signal transduction modulators; single chain antigen binding protein; sizofuran; sobuzoxane; sodium borocaptate; sodium phenylacetate; solverol; somatomedin binding protein; sonermin; sparfosic acid; spicamycin D; spiromustine; splenopentin; spongistatin I; squalamine; stem cell inhibitor; stem-cell division inhibitors; stipiamide; stromelysin inhibitors; sulfinosine; superactive vasoactive intestinal peptide antagonist; suradista; suramin; swainsonine; synthetic glycosaminoglycans; tallimustine; tamoxifen methiodide; tauromustine; tazarotene; tecogalan sodium; tegafur; tellurapyrylium; telomerase inhibitors; temoporfin; temozolomide; teniposide; tetrachlorodecaoxide; tetrazomine; thaliblastine; thiocoraline; thrombopoietin; thrombopoietin mimetic; thymalfasin; thymopoietin receptor agonist; thymotrigan; thyroid stimulating hormone; tin ethyl etiopurpurin; tirapazamine; titanocene bichloride; topsentin; toremifene; totipotent stem cell factor; translation inhibitors; tretinoin; triacetyluridine; triciribine; trimetrexate; triptorelin; tropisetron; turosteride; tyrosine kinase inhibitors; tyrphostins; UBC inhibitors; ubenimex; urogenital sinus-derived growth inhibitory factor; urokinase receptor antagonists; vapreotide; variolin B; vector system, erythrocyte gene therapy; velaresol; veramine; verdins; verteporfin; vinorelbine; vinxaltine; vitaxin; vorozole; zanoterone; zeniplatin; zilascorb; and zinostatin stimalamer. In one embodiment, the anti-cancer drug is 5-fluorouracil, taxol, or leucovorin.

Dosage and Formulation (Pharmaceutical compositions)

- The present invention envisions treating a disease, for example, cancer and the like, in a mammal by the administration of therapeutic agent.

Administration of the therapeutic agent in accordance with the present invention may be continuous or intermittent, depending, for example, upon the recipient's physiological condition, whether the purpose of the administration is therapeutic or

prophylactic, and other factors known to skilled practitioners. The administration of the agents of the invention may be essentially continuous over a preselected period of time or may be in a series of spaced doses. Both local and systemic administration is contemplated. The amount administered will vary depending on various factors including, but not limited to, the composition chosen, the particular disease, the weight, the physical condition, and the age of the mammal, and whether prevention or treatment is to be achieved. Such factors can be readily determined by the clinician employing animal models or other test systems which are well known to the art

One or more suitable unit dosage forms having the therapeutic agent(s) of the invention, which, as discussed below, may optionally be formulated for sustained release (for example using microencapsulation, see WO 94/07529, and U.S. Pat. No. 4,962,091 the disclosures of which are incorporated by reference herein), can be administered by a variety of routes including parenteral, including by intravenous and intramuscular routes, as well as by direct injection into the diseased tissue. For example, the therapeutic agent or modified cell may be directly injected into the tumor. The formulations may, where appropriate, be conveniently presented in discrete unit dosage forms and may be prepared by any of the methods well known to pharmacy. Such methods may include the step of bringing into association the therapeutic agent with liquid carriers, solid matrices, semi-solid carriers, finely divided solid carriers or combinations thereof, and then, if necessary, introducing or shaping the product into the desired delivery system.

When the therapeutic agents of the invention are prepared for administration, they are preferably combined with a pharmaceutically acceptable carrier, diluent or excipient to form a pharmaceutical formulation, or unit dosage form. The total active ingredients in such formulations include from 0.1 to 99.9% by weight of the formulation. A "pharmaceutically acceptable" is a carrier, diluent, excipient, and/or salt that is compatible with the other ingredients of the formulation, and not deleterious to the recipient thereof. The active ingredient for administration may be present as a powder or as granules; as a solution, a suspension or an emulsion.

Pharmaceutical formulations containing the therapeutic agents of the invention can be prepared by procedures known in the art using well known and readily available ingredients. The therapeutic agents of the invention can also be formulated as solutions appropriate for parenteral administration, for instance by intramuscular, subcutaneous or intravenous routes.

The pharmaceutical formulations of the therapeutic agents of the invention can also take the form of an aqueous or anhydrous solution or dispersion, or alternatively the form of an emulsion or suspension.

Thus, the therapeutic agent may be formulated for parenteral
5 administration (e.g., by injection, for example, bolus injection or continuous infusion) and may be presented in unit dose form in ampules, pre-filled syringes, small volume infusion containers or in multi-dose containers with an added preservative. The active ingredients may take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing
10 agents. Alternatively, the active ingredients may be in powder form, obtained by aseptic isolation of sterile solid or by lyophilization from solution, for constitution with a suitable vehicle, e.g., sterile, pyrogen-free water, before use.

It will be appreciated that the unit content of active ingredient or ingredients contained in an individual aerosol dose of each dosage form need not in itself
15 constitute an effective amount for treating the particular indication or disease since the necessary effective amount can be reached by administration of a plurality of dosage units. Moreover, the effective amount may be achieved using less than the dose in the dosage form, either individually, or in a series of administrations.

The pharmaceutical formulations of the present invention may include, as
20 optional ingredients, pharmaceutically acceptable carriers, diluents, solubilizing or emulsifying agents, and salts of the type that are well-known in the art. Specific non-limiting examples of the carriers and/or diluents that are useful in the pharmaceutical formulations of the present invention include water and physiologically acceptable buffered saline solutions, such as phosphate buffered saline solutions pH 7.0-8.0.

25 The agents of this invention can be formulated and administered to treat a variety of disease states by any means that produces contact of the active ingredient with the agent's site of action in the body of the organism. They can be administered by any conventional means available for use in conjunction with pharmaceuticals, either as individual therapeutic active ingredients or in a combination of therapeutic active
30 ingredients. They can be administered alone, but are generally administered with a pharmaceutical carrier selected on the basis of the chosen route of administration and standard pharmaceutical practice.

In general, water, suitable oil, saline, aqueous dextrose (glucose), and related sugar solutions and glycols such as propylene glycol or polyethylene glycols are

suitable carriers for parenteral solutions. Solutions for parenteral administration contain the active ingredient, suitable stabilizing agents and, if necessary, buffer substances. Antioxidizing agents such as sodium bisulfate, sodium sulfite or ascorbic acid, either alone or combined, are suitable stabilizing agents. Also used are citric acid and its salts and sodium Ethylenediaminetetraacetic acid (EDTA). In addition, parenteral solutions can contain preservatives such as benzalkonium chloride, methyl- or propyl-paraben and chlorobutanol. Suitable pharmaceutical carriers are described in Remington's Pharmaceutical Sciences, a standard reference text in this field.

The active ingredients of the invention may be formulated to be suspended in a pharmaceutically acceptable composition suitable for use in mammals and in particular, in humans. Such formulations include the use of adjuvants such as muranyl dipeptide derivatives (MDP) or analogs that are described in U.S. Patent Nos. 4,082,735; 4,082,736; 4,101,536; 4,185,089; 4,235,771; and 4,406,890. Other adjuvants, which are useful, include alum (Pierce Chemical Co.), lipid A, trehalose dimycolate and dimethyldioctadecylammonium bromide (DDA), Freund's adjuvant, and IL-12. Other components may include a polyoxypropylene-polyoxyethylene block polymer (Pluronic®), a non-ionic surfactant, and a metabolizable oil such as squalene (U.S. Patent No. 4,606,918).

Additionally, standard pharmaceutical methods can be employed to control the duration of action. These are well known in the art and include control release preparations and can include appropriate macromolecules, for example polymers, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methyl cellulose, carboxymethyl cellulose or protamine sulfate. The concentration of macromolecules as well as the methods of incorporation can be adjusted in order to control release. Additionally, the agent can be incorporated into particles of polymeric materials such as polyesters, polyamino acids, hydrogels, poly (lactic acid) or ethylenevinylacetate copolymers. In addition to being incorporated, these agents can also be used to trap the compound in microcapsules.

Accordingly, the pharmaceutical composition of the present invention may be delivered via various routes and to various sites in a mammal body to achieve a particular effect (see, e.g., Rosenfeld et al., 1991; Rosenfeld et al., 1991a; Jaffe et al., supra; Berkner, supra). One skilled in the art will recognize that although more than one route can be used for administration, a particular route can provide a more immediate and more effective reaction than another route. Local or systemic delivery can be

accomplished by administration comprising application or instillation of the formulation into body cavities, inhalation or insufflation of an aerosol, or by parenteral introduction, comprising intramuscular, intravenous, peritoneal, subcutaneous, intradermal, as well as topical administration.

5 The active ingredients of the present invention can be provided in unit dosage form wherein each dosage unit, e.g., a teaspoonful, tablet, solution, or suppository, contains a predetermined amount of the composition, alone or in appropriate combination with other active agents. The term “unit dosage form” as used herein refers to physically discrete units suitable as unitary dosages for human and mammal subjects, each unit
10 containing a predetermined quantity of the compositions of the present invention, alone or in combination with other active agents, calculated in an amount sufficient to produce the desired effect, in association with a pharmaceutically acceptable diluent, carrier, or vehicle, where appropriate. The specifications for the unit dosage forms of the present invention depend on the particular effect to be achieved and the particular
15 pharmacodynamics associated with the pharmaceutical composition in the particular host.

Detecting and Diagnosing Cancer

In some embodiments, the invention provides methods of detecting or diagnosing cancer, the method comprising identifying the presence of an epicenter in a
20 cancer in a patient. In some embodiments, the presence of an epicenter is determined using a method such as ATAC-seq in biological samples collected from an individual. In some embodiments, the method relates to evaluating expression levels or activity of one or more transcription factors relative to a comparator control. In some embodiments, the one or more transcription factors is determined using transcriptome analysis of a DNA
25 sample collected from a biological sample determined to have one or more epicenters or super-enhancers or transposase accessible regions. In some embodiments, the DNA sample is collected, isolated, and purified using methods known to one skilled in the art. In some embodiments, the transcriptome analysis includes using ATAC-seq data identifying regions of DNA, and using HOMER software (Heinz et al., 2010) in addition
30 to the MolSigDB database (Mootha et al., 2003) to search for and identify genes with binding motifs of transcription factors determined to be enriched in epicenters or super-enhancers differentially regulated in a cancer sample relative to a comparator control.

The invention contemplates the identification of differentially accessible epicenters, transcription factors binding thereto, and genes regulated thereby, in order to

identify markers differentially expressed between normal, and cancer subjects. The invention further contemplates using methods known to those skilled in the art to detect and to measure the level of differentially expressed marker of expression products, such as RNA and protein, to measure the level of one or more differentially expressed marker
5 expression products, for example transcription factor or targeted gene expression level or expression products.

Methods of detecting or measuring gene expression may utilize methods that focus on cellular components (cellular examination), or methods that focus on examining extracellular components (fluid examination). Because gene expression
10 involves the ordered production of a number of different molecules, a cellular or fluid examination may be used to detect or measure a variety of molecules including RNA, protein, and a number of molecules that may be modified as a result of the protein's function. Typical diagnostic methods focusing on nucleic acids include amplification techniques such as PCR and RT-PCR (including quantitative variants), and hybridization
15 techniques such as in situ hybridization, microarrays, blots, and others. Typical diagnostic methods focusing on proteins include binding techniques such as ELISA, immunohistochemistry, microarray and functional techniques such as enzymatic assays.

The genes identified as being differentially expressed may be assessed in a variety of nucleic acid detection assays to detect or quantify the expression level of a gene
20 or multiple genes in a given sample. For example, traditional Northern blotting, nuclease protection, RT-PCR, microarray, and differential display methods may be used for detecting gene expression levels. Methods for assaying for mRNA include Northern blots, slot blots, dot blots, and hybridization to an ordered array of oligonucleotides. Any method for specifically and quantitatively measuring a specific protein or mRNA or DNA
25 product can be used. However, methods and assays are most efficiently designed with array or chip hybridization-based methods for detecting the expression of a large number of genes. Any hybridization assay format may be used, including solution-based and solid support-based assay formats.

The protein products of the genes identified herein can also be assayed to
30 determine the amount of expression. Methods for assaying for a protein include Western blot, immunoprecipitation, and radioimmunoassay. The proteins analyzed may be localized intracellularly (most commonly an application of immunohistochemistry) or extracellularly (most commonly an application of immunoassays such as ELISA).

Biological samples may be of any biological tissue or fluid. Frequently the sample will be a “clinical sample” which is a sample derived from a patient.

Controls groups may either be normal or samples from known stages of cancer. As described below, comparison of the expression patterns of the sample to be tested with those of the controls can be used to diagnose between normal and cancer subjects. In some instances, the control groups are only for the purposes of establishing initial cutoffs for the assays of the invention. Therefore, in some instances, the systems and methods of the invention can diagnose between normal and cancer subjects without the need to compare with a control group.

10 The present invention relates to the identification of biomarkers associated with cancer, and/or cancer related conditions. Accordingly, the present invention features methods for identifying subjects who are at risk of developing cancer and/or cancer related conditions, including those subjects who are asymptomatic or only exhibit non-specific indicators of cancer and/or cancer related conditions by detection of the biomarkers disclosed herein. These biomarkers are also useful for monitoring subjects undergoing treatments and therapies for cancer and/or cancer conditions, and for selecting or modifying therapies and treatments that would be efficacious in subjects having cancer and/or cancer conditions, wherein selection and use of such treatments and therapies slow the progression of cancer and/or cancer conditions, or prevent their onset.

20 The invention provides improved diagnosis and prognosis of cancer and/or cancer conditions. The risk of developing cancer and/or cancer conditions can be assessed by measuring one or more of the biomarkers described herein, and comparing the measured values to reference or index values. Such a comparison can be undertaken with mathematical algorithms or formula in order to combine information from results of multiple individual biomarkers and other parameters into a single measurement or index. Subjects identified as having an increased risk of cancer and/or cancer conditions can optionally be selected to receive treatment regimens, such as administration of prophylactic or therapeutic compounds or combination with other anti-cancer agents, including cytotoxic/antineoplastic agents and anti-angiogenic agents to prevent, treat or delay the onset of cancer and/or cancer conditions.

Identifying a subject before they develop cancer and/or cancer related conditions enables the selection and initiation of various therapeutic interventions or treatment regimens in order to delay, reduce or prevent that subject's conversion to a disease state. Monitoring the levels of at least one biomarker also allows for the course of

treatment of cancer and/or cancer conditions to be monitored. For example, a sample can be provided from a subject undergoing treatment regimens or therapeutic interventions, e.g., drug treatments, for cancer and/or cancer conditions. Such treatment regimens or therapeutic interventions can include dietary modification, dietary supplementation, surgical intervention, administration of pharmaceuticals, anti-cancer agents, including cytotoxic/antineoplastic agents and anti-angiogenic agents, and treatment with therapeutics or prophylactics used in subjects diagnosed or identified with cancer and/or cancer related conditions. Samples can be obtained from the subject at various time points before, during, or after treatment.

The biomarkers of the present invention can thus be used to generate a biomarker profile or signature of subjects: (i) who do not have and are not expected to develop cancer and/or cancer related conditions and/or (ii) who have or expected to develop cancer and/or cancer related conditions. The biomarker profile of a subject can be compared to a predetermined or reference biomarker profile to diagnose or identify subjects at risk for developing cancer and/or cancer related conditions, to monitor the progression of disease, as well as the rate of progression of disease, and to monitor the effectiveness of cancer and/or cancer related condition treatments. Data concerning the biomarkers of the present invention can also be combined or correlated with other data or test results, such as, without limitation, measurements of clinical parameters or other algorithms for cancer and/or cancer related conditions. The machine-readable media can also comprise subject information such as medical history and any relevant family history.

The present invention also provides methods for identifying agents for treating cancer and/or cancer related conditions that are appropriate or otherwise customized for a specific subject. In this regard, a test sample from a subject, exposed to a therapeutic agent or a drug, can be taken and the level of one or more biomarkers can be determined. The level of one or more biomarkers can be compared to a sample derived from the subject before and after treatment, or can be compared to samples derived from one or more subjects who have shown improvements in risk factors as a result of such treatment or exposure.

In one embodiment, the invention is a method of diagnosing cancer. In one embodiment, the method includes distinguishing between normal and cancer subjects.

In various embodiments, methods are disclosed herein that may be of use to determine whether a subject has a cancer. In some embodiments, these methods may

utilize a biological sample (such as urine, saliva, blood, serum, amniotic fluid, or tears), for the detection of one or more markers of the invention in the sample.

In one embodiment, the invention provides a biomarker for the detection of cancer from non-cancer. In one embodiment, the biomarker for the detection of cancer
5 from non-cancer includes but is not limited to SOX9, KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3.

In one embodiment, the method comprises detecting one or more markers in a biological sample of the subject. In various embodiments, the level of one or more of markers of the invention in the biological sample of the subject is compared with the level
10 of a corresponding biomarker in a comparator. Non-limiting examples of comparators include, but are not limited to, a negative control, a positive control, an expected normal background value of the subject, a historical normal background value of the subject, an expected normal background value of a population that the subject is a member of, or a historical normal background value of a population that the subject is a member of.

15 In another embodiment, the invention is a method of monitoring the progression of cancer in a subject by assessing the level of one or more of the markers of the invention in a biological sample of the subject.

In various embodiments, the subject is a human subject, and may be of any race, sex and age.

20 Information obtained from the methods of the invention described herein can be used alone, or in combination with other information (e.g., disease status, disease history, vital signs, blood chemistry, etc.) from the subject or from the biological sample obtained from the subject.

In other various embodiments of the methods of the invention, the level of
25 one or more markers of the invention is determined to be increased when the level of one or more of the markers of the invention is increased by at least 10%, by at least 20%, by at least 30%, by at least 40%, by at least 50%, by at least 60%, by at least 70%, by at least 80%, by at least 90%, or by at least 100%, when compared to with a comparator control.

In the methods of the invention, a biological sample from a subject is
30 assessed for the level of one or more of the markers of the invention in the biological sample obtained from the patient. The level of one or more of the markers of the invention in the biological sample can be determined by assessing the amount of polypeptide of one or more of the biomarkers of the invention in the biological sample, the amount of mRNA of one or more of the biomarkers of the invention in the biological

sample, the amount of enzymatic activity of one or more of the biomarkers of the invention in the biological sample, or a combination thereof.

Detecting a biomarker

5 In one embodiment, the invention includes detecting an extracellular mRNA in a biological sample, wherein the extracellular mRNA is detected in a cell-free fluid phase portion of the biological sample.

 Biomarkers generally can be measured and detected through a variety of assays, methods and detection systems known to one of skill in the art. Various methods
10 include but are not limited to refractive index spectroscopy (RI), ultra-violet spectroscopy (UV), fluorescence analysis, electrochemical analysis, radiochemical analysis, near-infrared spectroscopy (near-IR), infrared (IR) spectroscopy, nuclear magnetic resonance spectroscopy (NMR), light scattering analysis (LS), mass spectrometry, pyrolysis mass spectrometry, nephelometry, dispersive Raman spectroscopy, gas chromatography, liquid
15 chromatography, gas chromatography combined with mass spectrometry, liquid chromatography combined with mass spectrometry, matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) combined with mass spectrometry, ion spray spectroscopy combined with mass spectrometry, capillary electrophoresis, colorimetry and surface plasmon resonance (such as according to systems provided by Biacore Life
20 Sciences). See also PCT Publications WO/2004/056456 and WO/2004/088309. In this regard, biomarkers can be measured using the above-mentioned detection methods, or other methods known to the skilled artisan. Other biomarkers can be similarly detected using reagents that are specifically designed or tailored to detect them.

 Different types of biomarkers and their measurements can be combined in
25 the compositions and methods of the present invention. In various embodiments, the protein form of the biomarkers is measured. In various embodiments, the nucleic acid form of the biomarkers is measured. In exemplary embodiments, the nucleic acid form is mRNA. In various embodiments, measurements of protein biomarkers are used in conjunction with measurements of nucleic acid biomarkers.

30 Methods for detecting mRNA, such as RT-PCR, real time PCR, branch DNA, NASBA and others, are well known in the art. Using sequence information provided by the database entries for the biomarker sequences, expression of the biomarker sequences can be detected (if present) and measured using techniques well known to one of ordinary skill in the art. For example, sequences in sequence database

entries or sequences disclosed herein can be used to construct probes for detecting biomarker RNA sequences in, e.g., Northern blot hybridization analyses or methods which specifically, and, preferably, quantitatively amplify specific nucleic acid sequences. As another example, the sequences can be used to construct primers for specifically amplifying the biomarker sequences in, e.g., amplification-based detection methods such as reverse-transcription based polymerase chain reaction (RT-PCR). When alterations in gene expression are associated with gene amplification, deletion, polymorphisms and mutations, sequence comparisons in test and reference populations can be made by comparing relative amounts of the examined DNA sequences in the test and reference cell populations. In addition to Northern blot and RT-PCR, RNA can also be measured using, for example, other target amplification methods (e.g., TMA, SDA, NASBA), signal amplification methods (e.g., bDNA), nuclease protection assays, in situ hybridization and the like.

The concentration of the biomarker in a sample may be determined by any suitable assay. A suitable assay may include one or more of the following methods, an enzyme assay, an immunoassay, mass spectrometry, chromatography, electrophoresis or an antibody microarray, or any combination thereof. Thus, as would be understood by one skilled in the art, the system and methods of the invention may include any method known in the art to detect a biomarker in a sample.

The invention described herein also relates to methods for a multiplex analysis platform. In one embodiment, the method comprises an analytical method for multiplexing analytical measurements of markers.

Wound Healing

In one embodiment, the present invention is a method for treating a wound. In one embodiment, method comprising administering to a subject a composition as described herein, that when administered to said subject improves wound healing. In one embodiment, the method comprises administering to a subject a modulator of epicenter activity, wherein the modulator activates transcription factor activity wherein the transcription factor activity improves wound healing. In one embodiment, the method comprises administering to subject a modulator of epicenter activity, wherein the modulator inhibits transcription factor activity, wherein the inhibition of transcription factor activity improves wound healing. In one embodiment, the modulated transcription

factors include but are not limited to SOX9, KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3.

These methods described herein are by no means all-inclusive, and further methods to suit the specific application will be apparent to the ordinary skilled artisan.

- 5 Moreover, the effective amount of the compositions can be further approximated through analogy to compounds known to exert the desired effect.

EXPERIMENTAL EXAMPLES

- 10 The invention is further described in detail by reference to the following experimental examples. These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified. Thus, the invention should in no way be construed as being limited to the following examples, but rather, should be construed to encompass any and all variations which become evident as a result of the teaching provided herein.

- 15 Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the compounds of the present invention and practice the claimed methods. The following working examples therefore, specifically point out the preferred embodiments of the present invention, and are not to be construed as limiting in any way the remainder
20 of the disclosure.

Example 1: Stem Cell Lineage Infidelity at the Crossroads of Wound-Repair and Cancer

- Tissue stem cells govern tissue regeneration and wound-repair. Tumors often hijack these normal cellular programs and exploit them for malignancy. Here, such
25 a phenomenon was identified in skin, where stem cells of the epidermis and hair follicle remain faithfully restricted to fueling their own tissue during homeostasis. They lose lineage fidelity during tumorigenesis. Moreover, breakdown of stem cell lineage confinement – granting privileges associated with both fates – is not only a hallmark, but also obligatory for malignancy. Intriguingly, it was found that lineage plasticity is also
30 critical in wound-repair, where it functions transiently to redirect fates. Probing mechanism, it was shown that irrespective of cellular origin, lineage infidelity occurs in wounding when stress-responsive enhancers are activated and override the homeostatic enhancers that govern lineage-specificity. In cancer, stress-responsive transcription factor

levels rise, causing lineage commanders to reach excess. When lineage and stress factors collaborate, they activate new oncogenic enhancers that distinguish cancers from wounds.

The methods used to perform the experiments are now described

5

Fluorescence Activated Cell Sorting (FACS) and Analysis

Purification of adult populations was performed using P60 second telogen WT mice. Purification of tumor populations was performed using transplants from TGFbRII-deficient; HRasG12V background (SCCs). Purification of wound population was performed using SOX9CreER; R26YFP mice, treated with Tamoxifen at P53, 10 wounded at P60 (partial thickness wound using a Dremel head gently scraping against backskin), and collected at P67. For adult cell isolation, the backskin was first scraped from dermal side to remove fat; for tumor isolation, tumor was first chopped and the placed in collagenase (Sigma, 0.25% in HBSS) for 1 hour at 37°C and spin down 300 g 15 4°C; for wound isolation, the wound center was dissected and EDTA (50 mM) treated for 30min at 37°C after which the epidermis is peeled off. The remaining epidermal side (adult skin) or cell mixture (tumor) or epidermis (wound) was then transferred to trypsin (Gibco, 0.25% in PBS) at 37°C for 10 minutes. Single-cell suspensions were obtained by scraping the skin from epidermal side (adult skin) or pipetting (tumor or wound) gently. 20 The cells were then filtered with 70 µM followed by 40 µM strainers, and pelleted at 300 g 4°C. For FACS analysis, EdU was administrated through intraperitoneal injection at 100 µl (5 mg/ml) per 20 g mouse weight for 1 hour before euthanasia. Cells were first stained with Live/Dead Blue (Life Tech, 1:100), and then fixed, permeabilized, stained with anti-GFP (1: 1,000), followed by EdU Click-iT reactions and Alexa Fluor (IntA6) staining. 25 For FACS sorting, cell suspensions were incubated with the appropriate antibodies for 30 min on ice and washed. Signal cell suspensions were first gated against DAPI to exclude dead cells, and then with forward and side scatters to gate for singlets. Lineage-negative gating was as follows: CD31 for endothelial cells, CD45 for immune cells, CD117 for melanocytes, and CD140a for fibroblasts. Adult EpdSC were further gated 30 IntA6+Sca1+CD34-, HF-SC gated as IntA6+Sca1-CD34+. Tumor grafts formed from previously characterized aggressive SCCs (Yang et al, GFP tagged) were gated as GFP+IntA6+IntB4+ for SCCSCs. Whether cultured, or injected directly into recipient mice, these basal SCC cells have been previously shown to have tumor-initiating properties characteristic of cancer stem cells (Schober et al, Yang et al). The findings

presented herein corroborate this behavior as shown by the tumor-initiating assays in figures throughout the manuscript. Wounded SCs were gated as YFP+IntA6+ for repairing cells that are HFSCs originated. The following antibodies were used: CD34_eFluor660 (1:100, eBioscience 50-0341-80), IntA6_PE (1:100, BD Bioscience 551129), IntB4_Alexa647 (1:1,000, eBioscience 14-0291-81), Scal_PerCP-Cy5.5 (1:1,000, eBioscience 45-5981-80), CD140a_PE-Cy7(1:100, eBioscience 14-1401-81), CD31_PE-Cy7 (1:1,000, eBioscience 25-0311-81), CD117_PE-Cy7 (1:1,000, eBioscience 25-1171- 82), CD45_APC-eFluor450(1:1,000, eBioscience 48-0451-80). DAPI (0.2 µg/ml) was used to exclude dead cells. Sorting was performed on a BD FACS Aria II equipped with Diva software (BD Biosciences). Analyses were performed on LSR II FACS analyzers and FlowJo.

ATAC-Seq

ATAC-seq libraries were made from freshly FACS-sorted SCs, with two biologically independent replicates for cell population. Library preparation and analysis was performed as described (Buenrostro et al., 2013). Briefly, Freshly FACS-sorted cells (20K-100K) are subject to tagmentation reaction with 2-10 µl TnD transposase, cleaned up and PCR amplified with 10-15 cycles. Library concentration and quality was confirmed with D1000 Tape Station prior to sequencing.

RNA-Seq and Quantitative PCR

Normal HF-, Epd- and SCC-SCs were isolated as previously described (Ge et al., 2016). For Wound-SCs (Wd-SCs), adult second telogen (P60) SOX9CreER; R26YFP female mice were subjected to full-thickness 7 mm punch wound, and 7 days post-wounding, single cell resuspensions were obtained by FACS as Lineage-YFP+Integrin α 6+ (HF-derived) and Lineage-YFP-Integrin α 6+ (Epd-derived). Cells were lysed with TrizolLS (Invitrogen) and total RNA was isolated with the Direct-zol RNA MiniPrep kit (Zymo Research) and submitted to the Genomics Resources Core Facility of the Weill Cornell Medical College for quality control (determined using Agilent 2100 Bioanalyzer, with all samples passing the quality threshold of RNA integrity numbers (RIN 8). Library construction was performed using IlluminaTruSeq Stranded mRNA Sample Prep Kit, and sequencing was performed on an Illumina HiSeq2000 sequencing machine.

For RNA quantitative PCR, complementary DNAs were generated from 1 µg of total RNA using the SuperScript Vilo cDNA synthesis kit (Life Tech), diluted and used as templates for real-time PCR performed with the 7900HT Fast Real-Time PCR System (Applied Biosystems) and gene-specific primers listed in Table 1. GAPDH was used as a loading reference.

Table 1. qPCR primers

KLF5-F	GGCTCTCCCGAGTTCAC TA (SEQ ID NO:1)
KLF5-R	ATTACTGCCGTCTGGTTTGTC (SEQ ID NO:2)
SOX9-F	CCACGGAACAGACTCACATC (SEQ ID NO:3)
SOX9-R	CCCTCTCGCTTCAGATCAAC (SEQ ID NO:4)

10 *Immunofluorescence and Microscopy*

For immunofluorescence, skin samples (sections) or embryos (whole mount) were pre-fixed in 4% PFA in PBS for 1 hour at 4°C and washed extensively in PBS. For sections, samples incubated with 30% sucrose in PBS overnight, embedded in OCT (Tissue Tek), cut at a thickness of 12 µm. For whole-mount, embryos were washed with PBS overnight and dissected under dissection scope. Tissue samples were then blocked at room temperature in Gelatin Block (2% fish gelatin, 5% normal donkey serum, 1% BSA, 0.3% Triton, in PBS). When immuno-labeling with mouse antibodies, sections were incubated with the M.O.M. blocking kit according to manufacturer's instructions (Vector Laboratories). Samples are then incubated with primary antibody overnight at 4°C, washed in PBS, secondary antibody 1 hour at room temperature (sections) or overnight at 4°C, washed in PBS, and mounted in ProLong Gold with DAPI (Life Tech). The following primary antibodies and dilutions were used: KLF-5 (goat, 1:50, R&D), SOX9 (rabbit, 1:300, E.Fuchs), P-Cadherin (goat, 1:400, R&D), LHX2 (rabbit, 1:2000, E. Fuchs), GFP (chicken, 1:2000, Abcam), JUN (rabbit, 1:1000, Cell Signaling), FOS (rabbit, 1:100, Abcam), pSTAT3 (rabbit, 1:1000, Cell Signaling), pETS2 (rabbit, 1:200, Thermo Fisher Sci), K5 (rabbit, 1:500, E. Fuchs), K10 (rabbit, 1:1,000, Covance), CD104 (rat, eBioscience, 1:1,000). Secondary antibodies were conjugated to Alexa488, 546, 647 (1:1,000, Life Technologies A-11006). Images were captured on a Zeiss Axioplan2 using a Plan- Apochromat 20x/0.8 air objective. For whole mount imaging, z stacks of 20–40

planes (0.25 mm) were acquired. Images were processed using ImageJ and Adobe Photoshop CS5.

Human SCC Xenografts and Metastases

- 5 Human HNSCC cell line A431 cells were LV-transduced with EC reporter in culture, FACS sorted with internal control color H2BRFP, and grafted (50K) onto nude backskin and harvest in 3 weeks for primary tumors, or tail vein injected (1M) into nude and harvest in 2 months for lung mets.

10 *CRISPR/CAS Knockouts in Vivo*

- Initial screening of efficient CRISPR guide RNAs (sgRNAs) was carried out by transiently transfecting pLentiGuide-Puro carrying sgRNAs against Klf5 or SOX9 (Doench et al., 2016; Sanjana et al., 2014) into cultured keratinocytes, selecting with puromycin (3µg/mL) for 2 days, isolating RNA, and measuring gene expression by
- 15 qPCR. The most efficient guides (as judged by qPCR) were then subcloned into LV-Cre or LV-CreER LV vectors and packaged into high-titer virus, followed by in utero LV injection into E9.5 R26-LSL-Cas9-P2A-GFP embryos. In vivo ablation was validated by immunostaining of E18.5 (sections) or E16.5 (whole mount) embryos. For wounding assays, LV-CreER driven CAS9 activation (hence gene ablation) was induced by
- 20 administering a single dose of tamoxifen (200 µl at 20 mg/mL per 40 g body weight) by oral gavage at E18.5, as intraperitoneal injection of tamoxifen at doses sufficient to induce recombination frequently led to aborted litters.

CRISPR/CAS Knockouts in SCC-SC Cells

- 25 Guides against KLF5 or SOX9 were selected as described above and the most efficient guides were then subcloned into plentiCRISPRv2 vectors (Sanjana et al., 2014). sgRNAs against Scrambled sequences were used as controls and done side-by-side with sgRNAs against target sites throughout the experiments to rule out phenotypic changes due to nonspecific targeting. plentiCRISPRv2 vector carrying targeting sgRNA
- 30 was then packaged into LV and used to transduce SCC-SCs cells, which were then selected with puromycin (3 µg/ml) for 2 days, and then transplanted into the backs of immunosuppressed Nude mice. Quantitative PCR was performed prior to transplantation to validate the effectiveness of the gene knockout. At least two small guides for each gene

were tested and results were consistent. The small guide sequences used are found in Table 2.

Table 2. Small guide sequences used for CRISPR

KLF5-guide 1	GATCCGGTGTATTCAACAGC (SEQ ID NO:5)
KLF5-guide 2	AGCTCACCTGAGGACTCATA (SEQ ID NO:6)
SOX9-guide 1	CAGCGTAGTCGTATTGCCGAG (SEQ ID NO:7)
SOX9-guide 2	GAGACCCTTCGTGGAGGAGG (SEQ ID NO:8)

5

Tumor Growth Curves

SCC-SCs were transduced with pLentiV2 Crispr guides against Scramble, KLF5 or SOX9, respectively, briefly selected with puromycin for 2 days, and then
 10 transplanted onto immunocompromised nude back skin via intradermal injection as previously described (Ge et al., 2016). Animals were monitored every 3 days for a month. Tumor size was measured using a digital caliper, and tumor volume was calculated using the formula $(\pi(\text{length} \times \text{width})^2)/6$. GraphPad Prism software was used to generate the tumor growth curves and to calculate the P value by two-way ANOVA with repeated
 15 measurement test.

Partial- and Full-thickness Wounding

For full-thickness wound, adult mice at the second telogen were anesthetized and 6 mm punch wounds were created on the shaved back skin. For partial-
 20 thickness wound, the wounding method is modified from (Levy et al., 2007). Briefly, second-telogen mice were anesthetized. Skin was shaved and remaining hair cleared with hair removal cream. Skin was gently stretched between two fingers, and a Dremel tool with polishing wheel was used to generate abrasions by polishing the skin laterally for 4-5 times (100 series rotary tool and 520 polishing wheel from Dremel Inc.

25

Split- and Full-thickness Engraftments

Split- and Full-thickness grafts were performed as previously described (Nowak et al., 2008; Rhee et al., 2006). Briefly, for full-thickness graft, head skins from early postnatal (P3) mice were dissected from the scramble guide (LV-CreER-sgScr) and

guide against gene of interest (LV-CreER-sgGene) transduced R26-LSL-Cas9-P2A-GFP animals and placed onto the backs of anesthetized female nu/nu (Nude) recipient mice. Grafts were secured by sterile gauze and cloth bandages, which were removed after healing (14 days). For split-thickness grafts, head skin from P3 mice was placed dermis-side up in 50 mM EDTA in PBS for 1 hour at 37°C. The epidermis was removed as a single sheet from hair follicle/dermis. Dermis was then grafted in the same manner as full thickness grafts. In all cases, paired sgScr and sgGene partial-thickness grafts of comparable size were grafted side-by-side, with full-thickness grafts as control on the same nude recipient.

Cell Culture and Boyden Chamber Assay

The mouse SCC-SC cell lines isolated from malignant SCCs were generated previously in the Fuchs' laboratory (Yang et al., 2015). Mouse keratinocytes were isolated and cultured as previously described (Blanpain et al., 2004; Guasch et al., 2007). SCC-SC cells were cultured in E intermediate calcium medium (contains 300 μ M calcium); mouse keratinocytes were cultured in E low calcium medium (contains 50 μ M calcium). For Boyden chamber assay, keratinocytes are cultured in regular culture media until close to confluent, and starved in serum-free P media containing 0.3 mM calcium overnight. The second day, cells are treated with mitomycin C (10 μ g/ml) for 2 hours at 37°C to arrest cell proliferation. Coat insert (24 mm Corning Transwell polycarbonate membrane cell culture inserts with 8.0 μ m pore) with fibronectin on bottom side of, and matrigel on top side of the membrane. Place 2.5 mL 3T3-fibroblast-feeder conditioned E low calcium media (24 hours on feeder) into bottom well, seed cells at 2 million cells per insert in 1.5 mL P media onto the top of membrane to reach around 70%. Incubate 24 hours, trypsinize cells from both sides of membrane, add media slowly and remove top or bottom cells carefully, resuspend in collection tubes and count cells from both sides.

CRISPR in SCC-SC Cells

Guides against the KLF5 or SOX9 were selected as described above and the most efficient guide was then subcloned into plentiCRISPRv2 vector (Sanjana et al., 2014). sgRNAs against Scrambled sequences were used as controls and done side-by-side with sgRNAs against target sites throughout the experiments to rule out phenotypic changes due to nonspecific targeting. plentiCRISPRv2 vector carrying targeting sgRNA was packaged into lentivirus and used to transduce SCC-SCs cells, which were then

selected with puromycin (3 µg/ml) for 2 days, and transplanted onto immunosuppressant mice back. Real-time PCR were performed prior to transplanting to validate gene knockout by effective sgRNA.

5 *Sample Selection and Blinding*

Sample sizes were selected empirically based on previous experience of performing in utero lentiviral transductions (Beronja et al., 2010) and transgenic mouse analysis (Vassar et al., 1991). Mice were only excluded from further analysis if they did not carry the expected lentiviral transgene or if they were not at the correct developmental
10 stage. All mice were treated with identical experimental conditions. Randomization and experimenter blinding were unnecessary and not performed.

Human HNSCC grafting and epicenter reporter test

Cells from human head and neck squamous cell carcinoma (HNSCC) cell
15 lines (SCC9, SCC15, SCC25) were infected in culture with lentiviruses (LVs) harboring the desired epicenter reporter and a ubiquitously active control PGK-H2BRFP gene. RFP+ transduced cells were purified by FACS, expanded, and then grafted (50K) onto Nude mouse backskin. After 3 weeks, primary tumors were harvested, pre-fixed in 4% paraformaldehyde (PFA) in phosphate buffered saline (PBS) for 1 hour at 4°C, washed
20 extensively in PBS, incubated with 30% sucrose in PBS overnight, embedded in Optimal Cutting Temperature (OCT) compound (Tissue Tek), and sectioned at a thickness of 12 µm. Tumor samples were then blocked at room temperature in Gelatin Block (2% fish gelatin, 5% normal donkey serum, 1% bovine serum albumin, 0.3% Triton, in PBS), incubated with primary antibody overnight at 4°C, washed in PBS, and then exposed to
25 secondary antibody for 1hr at room temperature. Following these steps, samples were then washed in PBS, and mounted in ProLong Gold with DAPI (Life Tech) prior to imaging.

Quantification and Statistical Analysis

30 *RNA-Seq Alignment and Differential Expression Analysis*

Raw sequenced reads were aligned to the Mouse reference genome (Version mm10 from UCSC) using STAR (Version 2.4.2) aligner. Aligned reads were quantified against the reference annotation (mm10 from UCSC) to obtain FPKM (Fragments per Kilobase per million) and raw counts using CuffLinks (v 2.2.1) and

HTSeq, respectively. Differential expression to compare expression profiles of respective groups was performed on normalized raw counts using the limma package in R. Genes with absolute log₂ fold change > 2 and FDR < 0.05 were considered to be significantly differentially expressed. Hierarchical Clustering and Principle Component Analysis in
 5 order to classify the samples based on gene expression profiles, unsupervised clustering, namely Hierarchical data clustering and Principal Component analysis (PCA) were performed. Both methods were performed on the log₂ transformed FPKM expression values in R statistical software.

10 *Gene Set Enrichment Analysis*

Pathway analysis using GSEA (Gene Set Enrichment Analysis Software) Software from Broad institute was used to identify functions of differentially expressed genes. Genes were ranked by the t-statistic value obtained from comparisons and the pre-ranked version of the tool was used to identify significantly enriched biological pathways.
 15 Pathways enriched with FDR < 0.25 were considered to be significant.

ATAC-Seq Alignment and Peak Calling

50-bp paired-end reads were aligned to 10 mm using bowtie with the parameters -X 2000 and -m 1. Duplicates were removed using Picard. Peaks were called
 20 using MACS2 (2.1.1.20160309) (Zhang et al., 2008) with the parameter -keep-dup all.

TSS and CTCF plots

Average ATAC signal for TSS (+/- 3000 bp) and CTCF binding sites (+/- 1000 bp) were plotted in each sample. CTCF sites were derived from CTCF Peaks called
 25 in E14.5 C57BL/6 Limb embryo from ENCODE (Accession: ENCFF001YAK) and converted to mm10 coordinates using liftOver.

Principal Component Analysis of ATAC signals

PCA analysis was performed on the union set of peak calls for Epd-SC, HFSC, SCC-SC (Figure 1), and Wd-SC (Figure 5), combined and merged using bedtools.
 30 Read counts were summed for each genomic region, normalized for sequencing depth, and log transformed.

Z-Score Transformation and Hierarchical Clustering of ATAC Signal and Cumulative Distribution Plot

Z-score transformed wiggle files were created using the average ATAC signal per base-pair over the genome, excluding chromM and chromY for each sample. Scores were then averaged over 100-bp non-overlapping windows. To filter out regions of background signal it required that the windows average greater than 1 in replicates. Significance was assayed between groups using a t-test (p-value < 0.05). Filtered data was then clustered using Cluster 3.0 (Eisen et al., 1998) and visualized using Java Treeview (Saldanha, 2004). Significantly altered windows were then associated with genes and subjected to pathway enrichment using GREAT (McLean et al., 2010).

Cumulative Distribution Plot

ATAC-seq peaks for tumor and epidermis were associated with genes using GREAT and quantified as peaks per gene. Log₂FC (Tumor/Normal) of gene expression was presented as a cumulative distribution for a) genes demonstrating an increase in number of ATAC peak associations in tumor relative epidermis (gain of at least 6 peaks, green), loss of ATAC peaks (loss of at least 1 peak, red), and all genes (black).

Motif Analysis of ATAC Peaks

Intersect or unique ATAC peaks were generated via BEDTools v2.25.0 (Quinlan and Hall, 2010), and HOMER (Heinz et al., 2010) de novo motif search was used to generate the list of enriched sequence motifs matching known transcription factor binding sites. For putative SOX9 and KLF5 targets, HOMER derived target gene list was then subjected to MolSigDB (Broad Institute) and scored for significantly enriched pathways associated with each gene.

The results of the experiments are now described.

Tumor SCs Undergo Global Changes in Chromatin Accessibility Compared to their normal counterparts

Open chromatin regions bound by TFs typically exhibit higher accessibility than otherwise closed regions. Since ATAC-seq can be applied on small numbers of SCs, this approach was used to delineate open chromatin states across

homeostatic, wound-induced and tumorigenic SC populations in vivo. Importantly, by complementing these data with transcriptional profiling, TF binding motifs could be predicted clustered that functionally dictate each cellular state (Figure 8).

The HFSC and SCC-SC open chromatin domains determined by genome-wide ATAC signals showed strong parallels to those identified by prior H3K27Ac chromatin immunoprecipitation and deep sequencing (ChIPseq) on these two populations (Yang et al., 2015). Charting new territory, EpdSC ATAC-seq revealed greater similarities in open chromatin states between HFSCs and EpdSCs than SCC-SCs (Figure 1A), indicating that oncogenic stress had elicited large-scale chromatin remodeling not seen in either of the homeostatic skin SC lineages.

To identify genomic regions where chromatin accessibility was significantly altered by tumorigenesis, ATAC signals were z-score-transformed and averaged them across 100-bp genomic windows (see Method Details). The hierarchical clustering heatmap revealed numerous accessible regions specific to SCC-SCs (Figure 1B, yellow: gain of signal; blue: loss of signal). In addition, when combined with transcriptome profiling (Adam et al., 2015; Lapouge et al., 2012; Schober and Fuchs, 2011; Yang et al., 2015), it was found that genes associated with gain or loss of ATAC peaks constituted a subset which displayed stronger than average transcriptional changes between their normal and malignant states (Figure 1C).

These analyses indicated that the datasets had captured key regulatory elements in chromatin that dictate transcriptional activity during these SC remodeling events. As illustrated in Figure 1D, ATAC-seq identified not only pan SC marker genes that were invariably expressed regardless of malignancy states (e.g., miR-205, Krt14), but also many genes that were either induced (e.g. miR-21, Gsta4, K17) or suppressed (e.g. Isml, Ackr4) specifically in tumor SCs.

Tumors Co-express Epidermal Lineage Marker KLF5 and HF Lineage marker SOX9

To track these distinct changes during malignant transformation, HOMER de novo motif search (Heinz et al., 2010) was used, and the TF motifs that are uniquely enriched within the open chromatin regions of each SC population were identified. It was then analyzed how this differs according to SC identity and state. As expected, motifs for the known HFSC identity TFs (Blanpain et al., 2004; Tumbar et al., 2004) were uniquely found within the HFSC profile when compared to EpdSCs (Figure 2A). Of significance, HFSC-specific ATAC peaks precisely captured previously annotated regulatory regions

from ChIP-seq where the full suite of HFSC identity TFs is bound and shouldered by active histone marker H3K27Ac (Figure 8C) (Folgueras et al., 2013; Kadaja et al., 2014; Keyes et al., 2013; Lien et al., 2014), so called ‘epicenters’ (ECs). Numerous of such ECs sit within H3K27Ac marked, broad open chromatin domains known as ‘super-enhancers’
 5 (Whyte et al., 2013), notable as regulators of stem cell key identity genes in the HFSCs (Adam et al., 2015).

Based upon these comparisons, it was posited that when combined with transcriptome analyses, ATAC datasets across different SC lineages and microenvironments would serve as valuable predictive tools to identify ECs, key TF-
 10 controlled bona fide enhancer elements that orchestrate gene expression specific to the corresponding SC identity and state. It was first tested on the unique ATAC peaks of EpdSCs compared to HFSCs, and found enriched motifs for a cohort of TFs including KLF5, GATA, GRHL (Cangkrama et al., 2016) and AP2 (Wang et al., 2008) (Figure 2A). As the top enriched EpdSC factor, KLF5 was particularly interesting in that during
 15 embryogenesis, it started as ubiquitously expressed in epidermal progenitors but then was turned off in the newly emerged hair follicles (Figure 2B). Immunofluorescence further documented that KLF5 and SOX9 demarcated Epd and HF lineages under adult homeostasis (Figure 2C). Indeed, just as the SOX9 locus was accessible uniquely to HFSCs, the KLF5 locus was accessible largely to EpdSCs (Figure 2D).

It was then asked what is unique to the SCC-SC ATAC peaks compared to those of each lineage-specific SCs in their homeostatic state. Curiously, the SCC-SC-unique ATAC peaks contained TF motifs for both EpdSCs (KLF5, GRHL, AP2) and HFSCs (SOX9, TCF, NFI) (Figure 2E). Moreover, in the tumor SCs, the regulatory regions for both KLF5 and SOX9 displayed a considerably broadened open-chromatin
 25 state not seen in the opposite SC lineage (Figure 2C). By immunofluorescence, it was documented that SOX9 and KLF5 are co-localized to the expanded layers of invasive SCC progenitors near the tumor-stromal interface (Figure 2F). This dual lineage expression pattern emerged as early as skin hyperplasia, and further escalated during the progression to benign and finally malignant tumors (Figure 9). This condition was
 30 referred to as “lineage infidelity”.

Lineage Infidelity is Functionally Required for Tumor Progression

Having documented the existence of lineage infidelity in SCC-SCs, its physiological relevance was addressed. To this end, CRISPR/CAS was used to inactivate

the KLF5 and SOX9 genes in SCs that were FACS-purified from established SCCs in vivo. qPCR confirmed that relative to cells transduced with a scrambled small guide (Scr-sg) RNA, cells transduced with the KLF5 or SOX9 sgRNAs displayed a marked transcriptional suppression of these genes (Figure 3A). Moreover, in subsequent

5 tumorigenesis assays where Scr-sgRNA- and TF-sgRNA-transduced cells were injected onto the respective left and right sides of host recipient mice, tumor growth from SCC-SCs was markedly suppressed when the SCs lacked either KLF5 or SOX9 (Figure 3B). Together, these results underscored the physiological importance of each of these lineage-specific TFs in malignancy.

10 To better understand the functional requirement of lineage infidelity, ATAC-seq data was used to investigate the genes targeted by KLF5 and SOX9 in SCC-SCs. Using HOMER software (Heinz et al., 2010) and the MolSigDB database (Mootha et al., 2003; Subramanian et al., 2005), genes that contain SOX9 or KLF5 motifs enriched in their associated ATAC peaks in SCC-SCs within previously mapped super-enhancers

15 (Yang et al., 2015) were searched for, and asked what molecular pathways are associated (see Method Details). Under these criteria, both SOX9 and KLF5 scored as putative regulators of pathways including cellular response to stimuli, immune regulation, extracellular matrix (matrisome) and adhesion (Figure 3C). By contrast, the invasive carcinoma genes showed selective enhanced regulation by SOX9, while the regulation of

20 cell proliferation genes showed enhanced regulation by KLF5.

These analyses were particularly intriguing given that SOX9 levels rose late during invasive carcinoma progression, while KLF5 expression was induced early in tumorigenesis. To determine the physiological significance of these correlations, in utero lentiviral delivery method was used and generated mice whose epidermis was selectively

25 transduced with inducible SOX9 or KLF5 transgenes on the rtTA background (Figure 3D). Three days after doxycycline administration to adult mice, a two-hour pulse with EdU (5-ethynyl-2'-deoxyuridine) was added prior to analyses.

Skin epidermis over-expressing KLF5 exhibited heightened proliferation in vivo (Figure 3D), in agreement with prior studies showing KLF5-associated

30 hyperkeratosis and epidermal erosions (Sur et al., 2006). Consistent with KLF5's role in sustaining progenitor proliferation, when KLF5 was depleted from tumors, it was observed that reductions in K5+ basal cells and expansion of K10+ differentiated cells as compared to SCCs, which are dominated by K5+ basal SCs (Figure 3E; Figure S3). By contrast, EpdSCs that ectopically expressed SOX9 showed only a modest increase in

proliferation relative to control skin (Figure 3D), but instead displayed a robust ability to penetrate matrigel and migrate to the bottom well in a Boyden chamber assay (Figure 3F). These findings functionally corroborate in silico analysis of chromatin and transcriptional landscapes of SCC-SCs, and support a more prominent role for KLF5 in SC proliferation and an enhanced role for SOX9 in invasion.

Epithelial Wounding Transiently Inflicts Lineage Infidelity and Relies Upon It For Healing

To acquire further molecular insights into how and why tumor lineage infidelity occurs, it was asked whether there might be a phenotypically analogous, but otherwise physiologically normal process where lineage infidelity arises. To this end, the SC response following injury was turned to, where cell proliferation, migration and invasion transpire naturally during the course of wound-repair. It could be reasoned that if tumors were to hijack and rewire normal processes to assist in malignant transformation, wound healing would be one such process to explore.

To test this possibility, partial-thickness wounding was employed, a controlled process where the skin epidermis and uppermost dermis can be selectively removed, challenging the bulge HFSCs remaining in the dermis to re-epithelialize and restore the skin barrier (see Method Details). Notably, by day three or five post denuding of the epidermis, mobilized HFSCs in and above the bulge niche induced KLF5, resulting in its co-expression with SOX9 (Figure 4A). As the healing process proceeded and the wounded tissue was gradually repaired, SOX9 and KLF5 returned to their homeostatic expression patterns, demarcating the HF and epidermis, respectively (Figure 4A, 2 weeks).

It was next asked whether this transient lineage infidelity is functionally relevant for wound repair. To address this, CRISPR/CAS was used to specifically ablate SOX9 and KLF5 individually in the skin epithelium in vivo. This was accomplished by engineering lentiviral (LV) vectors harboring CRISPR small guide (sg) RNAs and PGK-CreER. LVs were then delivered in utero into the amniotic sacs of living E9.5 embryos harboring a Lox-Stop-Lox-Cas9-P2A-eGFP cassette knocked into the ubiquitously active Rosa26 locus (Platt et al., 2014) (Figure 4B). With this method, the LV is transduced within a day exclusively into the single layer of unspecified epidermal progenitors, and thereafter is stably propagated to the developing epidermis and HF progeny (Beronja et al., 2010).

At E18.5, the pregnant females harboring the transduced embryos were treated with tamoxifen to induce Cre recombinase and permanently activate CAS9 and eGFP expression in transduced clones. Four days later, after allowing the sgRNAs and CAS9 to ablate KLF5 and SOX9 in the skin epithelium of neonatal pups, split-thickness grafting was conducted. This method is analogous to partial-thickness wounding in that it challenges targeted HFSCs to re-epithelialize the skin denuded of its epidermis (Figure 4B).

Prior to enzymatic separation of epidermis and engraftment of denuded dermis, both epidermal and HF compartments showed high rate of transduction, with efficient ablation of the targeted TF gene selectively in the GFP-marked, CAS9-activated cells (Figures 4B P3 pre-split and 11). Two weeks post engraftment, it was clear that while in the sgScr control group, GFP+, CAS9-activated cells were efficiently recruited and contributed to re-epithelialization, in either sgKLF5 or sgSOX9 group, only WT (GFP-, CAS9-inactive) cells had contributed to reforming the epidermis and its barrier (Figures 4B 2wks post-dermal graft).

The importance of KLF5 to HFSC-derived re-epithelialization was intriguing, since KLF5 was only transiently induced in HFSCs upon injury (see Figure 4A), and it had not been previously implicated in wound-healing. Additionally, despite SOX9 being turned off in HFSCs that had re-populated the damaged epidermis, SOX9 was also essential during the repair process. Of note, and in contrast to Scr-sgRNA controls, neither KLF5- nor SOX9-deficient HFs survived the graft, suggesting that analogous to what was seen in the tumors, the expression of both KLF5 and SOX9 during wounding – lineage infidelity – is functionally obligatory for wound-repair.

25 *Similarities in Transcriptome and Genome-wide Chromatin Accessibility of Wound and Tumor*

To search for common stress signals that might be instrumental in triggering the breakdown of SOX9-KLF5 lineage confinement in wound and tumorigenic states SCs from wounded skin, were purified and transcriptionally profiled (Wd-SCs, see 30 Method Details). Unsupervised clustering analyses revealed that the transcriptomes of wounded and tumorigenic SCs clustered together, and were markedly distinct from normal homeostatic SCs (Figures 5A and 5B).

The relation between tumor and wound responses was even more striking when the transcriptome data were subjected to Gene Set Enrichment Analysis (GSEA)

(Subramanian et al., 2005), which revealed statistically significant, concordant similarities (>98%) between the two biological states (Figures 5C and 12A). By contrast, neither tumor nor wound mirrored the transcriptome of short-lived committed proliferating progenies of HFSCs (Figure 12B). These data indicate that the stress response mounted by wound-mobilized and tumorigenic SCs goes beyond mere proliferation.

Paralleling the relations revealed by transcriptome analyses, accessible chromatin regions marked by ATAC tags in Wd-SCs were more akin to SCC-SCs than to their homeostatic normal counterparts (Figure 5D, $R^2 = 0.71$, see Figure 1A, 12C and 12D). A hierarchical clustering heatmap of gain (yellow) or loss (blue) of ATAC signals highlighted these concordant largescale chromatin deviations between stress-experienced and normal homeostatic states (Figure 5E).

Enriched motifs of AP-1, ETS2 and often STAT3 stood out to distinguish the accessible regions of wound and tumor SCs from homeostatic SCs (Figure 5F). This was true even for genes e.g. KLF5, active in homeostatic EpdSCs, but acquired new AP1 and ETS2-associated ATAC peaks in tumor and wound. Other tumor super-enhancer (SE) regulated genes, e.g. miR-21, were not expressed in EpdSCs or HFSCs, but acquired ATAC peaks shouldered by the H3K27Ac signals where activated histones bound.

It was posited that ATAC-marked-ECs (regulatory elements within super-enhancers that contain ATAC peaks with densely clustered TF motifs) specifically initiated in wounds and maintained in tumors might be reflective of unique enhancers that drive the activity of these genes in a stress-induced microenvironment. To test this hypothesis, they were cloned into lentiviral reporters (with internal H2BRFP control) and examined their ability to drive eGFP expression in vivo via in utero delivery.

In the homeostatic state of adult animals, skin displayed only nuclear H2BRFP expression, indicating efficient transduction but no EC activity (Figure 5G). However, when wounded, the skin showed striking activation of eGFP, indicating commissioning of the stress-responsive EC. Moreover, invasive SCC tumors generated by SCC-SC transduction and engraftment showed marked signs of eGFP expression as well (Figure 5G). These findings suggested that wound ECs activate genes in response to wound and tumor stress.

Stress-induced TFs Drive Lineage Infidelity Regardless of SC Origin

Having established the strong resemblance between transcriptome and chromatin landscapes between wound and tumor, and documented the efficacy of wound ECs in driving gene activation in wound and tumor, it was next wondered whether activation of stress-induced TFs might be causal to lineage infidelity, and if so, whether infidelity occurs irrespective of SC origin. First to address how HFSCs activate EpdSC genes, wound- and tumor-induced ATAC tags were examined within genes that are normally only expressed in EpdSCs and not HFSCs (Adam et al., 2015). Like KLF5, the chromatin encompassing regulatory regions of such epidermal genes displayed a plethora of new ATAC peaks with ETS, AP1 and STAT motifs in both wound and tumor induced states (Figure 13A Hes1 and Scal, also see 5F). This finding implied HFSCs turn on EpdSC genes downstream of stress-responsive TFs.

Immunofluorescence microscopy confirmed that activated (phosphorylated) ETS2, AP-1 (JUN/FOS) and (phosphorylated) STAT3 were all potently induced in both tumor and wound states (Figures 6A and 13B). ETS2 was particularly intriguing given that it is phosphorylated at T72 and activated by ERK1/2, which is transiently elevated during wound-repair and constitutively activated in HRASG12V induced tumorigenesis (Yang et al., 1996). Moreover, when engineered to be constitutively active (T72D mutation) in skin epidermis, ETS2 functions as an oncogenic driver in SCCs (Yang et al., 2015).

The mutant, ETS2 was therefore used in vivo to address whether stress-responsive TFs cause EpdSC genes activation in HFSCs. Indeed, in addition to inducing the cohort of stress responsive TFs (Figure 13C), KLF5 was activated in HFSCs (Figure 6B), similar to the ectopic KLF5 induction that had been seen in wound-experienced HFSCs (see Figure 4A).

Next, it was wondered whether analogously, the HFSC identity gene SOX9 would be induced in EpdSCs under stress. Notably, the SOX9 locus also acquired stress-specific ATAC peaks unique to the wound and tumor (Figure 13D). Indeed, SOX9 was also ectopically induced in EpdSCs upon ETS2 (T72D) expression (Figure 6B). Moreover, in response to full-thickness wounding, a condition that forces re-epithelialization from surrounding mobilized EpdSCs where ETS2/MAPK was naturally activated, SOX9 was again ectopically induced (Figure 6C). Together, these findings provide compelling evidence that stress-responsive of TFs, in particular pETS2, triggers lineage infidelity regardless of lineage SC origin.

Human HNSCC Grafting and Epicenter Reporting

While all lines exhibited comparable LV transduction reflected by internal RFP signal, only human HNSCCs transduced with LV-epicenter reporters (miR-21EC and Klf5EC) but not control-EC, conveyed activated GFP expression in developed tumor xenografts, indicating HNSCC specific reporter activation.

Resolution of Lineage Infidelity After Wound Repair

In contrast to tumorigenesis, lineage infidelity occurs during wound repair transiently and is resolved following re-epithelialization of the epidermis. For instance, when HFSCs participate in wound-repair, they switch their identity from HFSC to EpdSC at the conclusion of the healing process. To understand how this happens, it was first considered that SOX9 ablation within HFSCs causes their epidermal conversion (Kadaja et al., 2014) and ectopic SOX9 in EpdSCs results in HFSC gene activation (Adam et al., 2015; Kadaja et al., 2014). Thus, to achieve the HFSC→EpdSC fate switch, SOX9 must be silenced.

One possibility is that KLF5 antagonizes SOX9, as posited previously based upon expression patterns and the presence of KLF5 motifs in the SOX9 promoter (Bell et al., 2013; McConnell et al., 2011; Nandan et al., 2014). Consistent with this hypothesis, when K14rtTA embryos were transduced with a doxycycline-inducible KLF5 LV, and then induced KLF5 in adult HFs, SOX9 expression waned (Figure 6D), along with HFSC marker TENASCIN C, while EpdSC markers K10 and LORICRIN were activated (Figure 14A). These results are consistent with a model whereby as ETS2 is activated during wound-repair driving KLF5 expression in HFSCs, KLF5 suppresses the HFSC-homeostatic regulation of SOX9 and completes the reprogramming from the HF to Epd fate at the conclusion of the repair process.

New Epicenters Involving SOX9 and Stress-Responsive TFs Are Activated in Malignancy to Lock Tumor SCs into Sustained Lineage Infidelity

If under stress, KLF5 is indeed sufficiently potent to suppress SOX9 and reprogram the HF lineage into epidermis, why then is lineage infidelity constitutively activated upon malignant transformation? To understand how SOX9 escapes KLF5 suppression during tumor progression, the ATAC tags within the SOX9 locus of SCC-SCs were examined and compared them to those of normal and wounded SCs.

Interestingly in SCC-SCs, additional ATAC peaks appeared and these opened chromatin segments were silenced altogether in HFSCs and were barely accessible in the wound state (Figure 6E, right red boxes, 'tumor-EC'). In SCC-SCs, wound-ECs remained open (red shaded ECs exhibiting peaks in both wound and tumor).

- 5 However, the new, tumor-specific ECs that were activated in SCCs now included not only KLF, ETS and AP1 motifs but also SOX motifs.

- Since the ATAC-seq data indicated that these newly commissioned elements were silent in the wound state, an intuitive possibility is that the increasing levels of oncogenic RAS/MAPK known to occur during SCC progression were responsible for activating these composite enhancer elements composed of motifs for both lineage TFs and stress-responsive TFs. This is consistent with findings indicating that in stress-induced conditions, SOX9 and KLF5 became dependent upon ETS (Figure 5F and 13D), and that forced ETS2 activation caused elevation of SOX9 and KLF5 (Figure 6B), thereby favoring activation of compound stress-responsive TF and lineage TF enhancers. Additionally, as expected for the elevation of KLF5, the homeostatic HFSC-ECs dampened in a wound-response, were now completely decommissioned in the tumor state (Figure 6E, green boxes at left, HFSC-EC).
- 10
15

- Observations so far suggested an appealing hypothesis that during tumor-progression, skin SCs first silence homeostatic regulatory enhancer elements and activate low-stress wound-induced regulatory elements (wound-EC). At the point of tumor bifurcation from wound, oncogenic high-stress elements (tumor-EC) are now activated, driven by a composite of lineage- and stress-associated TFs. If this model is correct, then in contrast to the wound ECs, which were active in both wound and SCC (see Figure 5G), the tumor-ECs should only be active in SCCs, whereas HFSC-EC is expected silenced all together once exiting homeostasis.
- 20
25

- Indeed, when analyzed in vivo, this is precisely what was observed (Figure 6F). Finally, since SOX9 wanes upon wound healing and sustains in SCCs, the consequences of sustaining SOX9 during wounding were tested. To this end, TRE-SOX9; LV-rtTA-H2BGFP mice were engineered, doxycycline was applied, mice were wounded and then the analyzed was two weeks later. Strikingly, in contrast to control skin, which had returned to normal homeostasis, sustained SOX9 resulted in aberrant epidermal thickening reminiscent of a neoplastic state (Figures 6G, 14B and C). Of importance, KLF5 and pETS2 were both induced (Figure 6G), suggesting that the antagonistic
- 30

circuitry between KLF5 and SOX9, seen in both homeostatic and wounded HFSCs, had now been overcome.

Lineage Constraints on SCs are Transiently Unleashed in Wound-repair and

5 *Permanently Unhinged in Cancer*

In tackling the age-old debate as to whether a tumor is a wound that never heals, we've unearthed a striking lineage infidelity phenotype that occurs transiently in a wound response and constitutively in malignancy. In this process, epidermal and HF lineages, which are distinct and spatially confined during normal homeostasis, are
10 transiently blurred during wound-repair. By contrast, in malignancy, all vestiges of lineage confinement break down, as lineage infidelity becomes a permanent feature of the cancer. At the root of these lineage decisions are two master regulators of fate determination: KLF5 (EpdSCs) and SOX9 (HFSCs). The combined use of inducible epithelial CRISPR/CAS targeting, and in utero lentiviral CreER and guide RNA delivery
15 enabled us to expose hitherto unrecognized roles for KLF5 and SOX9 in wound and SCC as elaborated upon below.

KLF5 is an Epidermal Lineage Factor Required for Wound-repair and SCC

Compared to SOX9, which as is known, governs HFSC identity (Adam et al., 2015; Kadaja et al., 2014; Nowak et al., 2008), relative little was known about
20 KLF5's skin functions (Ohnishi et al., 2000; Shindo et al., 2002; Sur et al., 2002). It was discovered that KLF5 is not only spatially associated with but also functionally essential for EpdSC fate. KLF5 is induced early in HFSCs upon epidermal injury, and by virtue of its ability to suppress SOX9 and reprogram HFSCs to EpdSCs, KLF5 plays an integral
25 role in the fate switch, enabling HFSCs to participate in epidermal re-epithelialization during wound-repair. A similar antagonistic relation between KLF5 and SOX9 has also been reported in the intestine and colon (Bell et al., 2013; McConnell et al., 2011; Nandan et al., 2014), suggesting a broader importance for these two factors in cell fate decision.

Studies further showed that by activating a group of targets that are
30 preferentially associated with cell proliferative activities, KLF5 reinforces the proliferative features of the malignant state, an attribute of KLF5 that was previously recognized (Sur et al., 2006). Importantly, however, the loss of function studies shows that once KLF5 is lost, malignant progression is stalled and SCs terminally differentiate, resulting in SCC collapse.

SOX9 is Essential for HFSCs to Repair Wounds and Progress to SCC.

On the flip side of KLF5's role, results described herein also unveil a hitherto unrecognized importance for SOX9 in adult HFSC function during wound-repair and also in malignant SCC progression. These results were particularly intriguing in that SOX9 was recently found to be essential for basal cell carcinoma, which in contrast to SCC, is rooted in super-activation of the Sonic Hedgehog pathway (Larsimont et al., 2015; Vidal et al., 2005), rather than the RAS/MAPK pathway (Balmain and Yuspa, 2014). Prior to these results, SOX2 had been thought to be the major SOX factor in SCC tumorigenesis (Boumahdi et al., 2014; Siegle et al., 2014). Given that both SOX factors are expressed in SCCs and display phenotypic consequences of their loss, their possible non-redundant roles merit future investigation.

Wound and Tumor Converge on Lineage Infidelity Irrespective of SC Origin.

Regardless of SC origin, lineage infidelity surfaced in HFSCs and in EpdSCs in response to injury. In full-thickness wounds, mobilized EpdSCs at the wound edge adopted certain features of HFSCs before resolving back to EpdSC. In partial-thickness wounds, mobilized HFSCs adopted features of EpdSCs before losing their own identity as they re-epithelialized denuded epidermis and restored the skin barrier. Similarly, even though both HFSCs and EpdSCs can be the cell of origin for SCCs, the co-expression of SOX9 and KLF5 and the obligatory requirement of this dual lineage state is a hallmark of SCCs.

Tumor and Wound Diverge as Transient Lineage Plasticity Becomes Sustained.

How then does a wound enter the lineage infidelity state but consequently resolve it, while a cancer is permanently locked in? The collective results suggest the model illustrated in Figure 7. In normal homeostatic EpdSCs, KLF5 keeps SOX9 in check, while in HFSCs, SOX9 reigns, driven by HFSC-specific promoters and enhancer elements governed by TCF3/4, LHX2, NFIB, NFATC1 and SOX9 (Adam et al., 2015). In response to injury or oncogenic hyperplasia, HRas/MAPK signaling phosphorylates and activates ETS2. This spawns the activation of stress-associated regulatory elements driven by ETS and AP1 TFs within both the KLF5 and SOX9 super-enhancers, thus overriding the normal homeostatic regulatory elements, and marking the onset of lineage infidelity.

As MAPK and pETS2 levels wane near the conclusion of wound-repair, the reliance upon such stress-associated enhancers diminishes. In the epidermis, KLF5 expression returns to being controlled by EpdSC homeostatic enhancers. At this time, however, SOX9 becomes silenced in both EpdSCs and in HFSC-derived progeny that have reached the epidermis. In this regard, KLF5 drives the silencing of SOX9 via its homeostatic regulatory elements, ensuring fate switch being completed and epidermal homeostasis restored.

During malignant progression, however, the “low-stress activated” epicenters (shared by both wound and tumor) are now joined by newly emerging “high-stress” epicenters (tumor specific), signifying the divergence from wound to tumor. This expanded enhancer landscape is rooted in regulatory elements that contain not only ETS2 and AP1 motifs but also KLF5 and now SOX9 motifs. The commissioning of these combined lineage-infidelity and stress-induced regulatory elements is likely a direct consequence of one, the concomitant rise in HRAS/MAPK activity known to occur during the progression of normal SCs→hyperproliferation→benign papilloma→SCC (Rodriguez-Puebla ML et al., 1999; Guasch et al., 2007; Lapouge et al., 2012; Oft et al., 2002; Oshimori et al., 2015; Quintanilla et al., 1986; Schober and Fuchs, 2011; Wong et al., 2013; Yang et al., 2015); and two, the subsequently elevated levels of SOX9 and KLF5, driven by sustained ETS2 (T72D), an established target of MAPK activity (Foulds et al., 2004; Yang et al., 1996).

Taken together, while KLF5 antagonizes SOX9, causing the closing of its HFSC-specific epicenters in SCC-SCs, the acquisition of tumor-specific epicenters renders SOX9’s expression refractory to such normal homeostatic regulations. The outcome is sustained lineage infidelity. Indeed, as it was shown that at the crossroad between wound and tumor, forcing SOX9 expression during wound-repair is sufficient to convert the normal wound into a neoplastic state.

A sensitive method, “assay for transposase accessible chromatin with high throughput sequencing” (ATAC-seq) (Buenrostro et al., 2013) was exploited a, to systemically interrogate the open chromatin landscape of purified tumorigenic SCs and compare them to those found in normal EpdSCs and HFSCs first during homeostatic and then under wound conditions. Such comparative analysis led to a surprising phenotype where two otherwise confined skin lineages – EpdSCs and HFSCs – are breached and manifested simultaneously in the tumor SCs. Intriguingly, it was found that this feature

also occurs during a wound-response, but it is transient and resolves itself upon wound-closure.

To evaluate the functional importance of this lineage infidelity, CRISPR/CAS was used to ablate genes encoding two distinguishing lineage TFs, SOX9 (HFSCs) and KLF5 (EpdSCs). It was demonstrated that in tumorigenesis, co-expression of SOX9 and KLF5 is not only a hallmark of SCC, but also a requirement for malignant transformation. Lineage infidelity also appears to be crucial for wound-repair, as even though co-expression of SOX9 and KLF5 is transient, without either one, wounded SCs fail to contribute effectively to the healing process.

To investigate how lineage infidelity occurs, it was unveiled that at both the transcriptome- and chromatin-level a striking convergence between the wounded SCs and their tumorigenic counterparts. It was found and documented genetically that many EpdSC and HFSC genes whose expressions are otherwise restricted to their respective lineages under homeostasis, are now expressed in a lineage independent fashion in wound-activated and oncogenic SCs. Teasing apart the mechanisms involved, it was unraveled how homeostatic enhancers are repressed as stress-induced enhancers become activated during wound-healing, and subsequently how this state becomes locked in as oncogenic HRAS/MAPK levels rise and are sustained in malignancy. Finally, as the roots that distinguish a cancer from a wound are unearthed, these differences are exploited to generate tumor-specific enhancer drivers, which now pave the way for future advancements in therapeutic targeting.

Example 2: Epicenter activity in multiple squamous cell carcinoma models

Experiments were designed to evaluate epicenter activity in at least skin and head and neck squamous cell carcinoma models.

Figure 15 is an image demonstrating that epicenter activity in cultured mouse and human SCC lines. Lentiviral epicenter reporters were stably transduced into SCC line (mouse skin SCC, human skin SCC, and human head and neck SCCs). Candidate test ECs (Klf5EC and miR21EC) are compared against negative control EC (Ctrl EC) in each experiment. In each cases, significant EC reporter activity induction was observed in SCCs (** $p < 0.01$, *** $p < 0.001$, paired student's test). Five biologically independent experiments were performed.

Figure 16 is an image demonstrating epicenter activity in human head and neck squamous cell carcinoma models. Human head and neck squamous carcinoma

(SCC9) cells were transduced with lentiviral reporters for either stress-specific epicenter Klf5-EC or control EC (Ctrl-EC), xenografted onto immunosuppressant mice. Tumors were dissected 6 weeks post graft, and examined for GFP expression. RFP expression served as internal controls for lentiviral reporter transduction. Specific GFP expression was only observed in Klf5EC but not in CtrlEC. Three independent results were conducted.

The disclosures of each and every patent, patent application, and publication cited herein are hereby incorporated herein by reference in their entirety.

While this invention has been disclosed with reference to specific embodiments, it is apparent that other embodiments and variations of this invention may be devised by others skilled in the art without departing from the true spirit and scope of the invention. The appended claims are intended to be construed to include all such embodiments and equivalent variations.

CLAIMS

What is claimed is:

1. A composition for treating cancer, wherein the composition comprises a modulator of a transcription epicenter, wherein the epicenter regulates cancer.
2. The composition of claim 1, wherein the modulator comprises an inhibitor of the activity of the transcription epicenter.
3. The composition of claim 1, wherein the modulator comprises an activator of the activity of the transcription epicenter.
4. The composition of claim 1, wherein the modulator modulates the activation of an epicenter regulator, wherein the regulator is selected from the group consisting of polymerases, acetyltransferases, histone deacetylases, methylases, histone demethylases, transcription factors, coactivators, corepressors, and enhancers.
5. The composition of claim 1, wherein the modulator modulates the activation of one or more transcription factors selected from the group consisting of SOX9, KLF5, GATA, AP1 (JUN/FOS), AP2, TCF, LHX2, NFI, ETS2, and STAT3.
6. The composition of claim 2, wherein the inhibitor is one or more molecules selected from the group consisting of a small interfering RNA (siRNA), a small guide RNA (gRNA), a microRNA, an antisense nucleic acid, a ribozyme, an expression vector encoding a transdominant negative mutant, an antibody, a peptide, a chemical compound and a small molecule.
7. The composition of claim 3, wherein the activator is one or more compound selected from the group consisting of a chemical compound, a protein, a peptidomimetic, an antibody, a nucleic acid molecule.
8. The composition of claim 1, wherein the modulator binds a region of DNA, wherein the region of DNA comprising at least one binding transcription factor binding motif.

9. The composition of claim 8, wherein the transcription factor binding motif is selected from the group consisting of a SOX9 binding motif, a KLF5 binding motif, and a combination thereof.
10. The composition of claim 1, wherein the cancer is squamous cell carcinoma.
11. The composition of claim 10, wherein the squamous cell carcinoma is head and neck squamous cell carcinoma.
12. A method for treating cancer in a subject, the method comprising administering to subject in need thereof a modulator of an epicenter, wherein the epicenter is accessible, and wherein the epicenter regulates cancer.
13. The method of claim 12, wherein the modulator comprises an inhibitor of the activity of the transcription epicenter.
14. The method of claim 12, wherein the modulator comprises an activator of the activity of the transcription epicenter.
15. The method of claim 12, wherein the modulator modulates the activation of an epicenter regulator, wherein the regulator is selected from the group consisting of polymerases, acetyltransferases, histone deacetylases, methylases, histone demethylases, transcription factors, coactivators, corepressors, and enhancers.
16. The method of claim 12, wherein the modulator modulates the activation of one or more transcription factors selected from the group consisting of SOX9, KLF5, ETS, AP-1, GATA, GRHL, AP2, TCF, and NFI.
17. The method of claim 13, wherein the inhibitor is one or more molecules selected from the group consisting of a small interfering RNA (siRNA), a small guide RNA (gRNA), a microRNA, an antisense nucleic acid, a ribozyme, an expression vector encoding a transdominant negative mutant, an antibody, a peptide, a chemical compound

and a small molecule.

18. The method of claim 14, wherein the activator is one or more compound selected from the group consisting of a chemical compound, a protein, a peptidomimetic, an antibody, a nucleic acid molecule.

19. The method of claim 12, wherein the modulator binds a region of DNA, wherein the region of DNA comprising at least one binding transcription factor binding motif.

20. The method of claim 19, wherein the transcription factor binding motif is selected from the group consisting of a SOX9 binding motif, a KLF5 binding motif, and a combination thereof.

21. The method of claim 12, wherein the cancer is squamous cell carcinoma.

22. The method of claim 12, wherein the cancer is head and neck squamous cell carcinoma.

23. A method of diagnosing cancer in a subject, the method comprising the step of: (a) obtaining a sample from the subject, (b) detecting epicenter activity relative to a comparator control in the sample, and (c) thereby diagnosing the subject with cancer.

24. The method of claim 23, wherein the sample is a human sample.

25. The method of claim 23, wherein the sample is skin tissue.

26. The method of claim 23, wherein the cancer is squamous cell carcinoma.

27. The method of claim 26, wherein the squamous cell carcinoma is head and neck squamous cell carcinoma.

28. The method of claim 23, wherein the comparator control is at least one selected from the group consisting of: a positive control, a negative control, a historical control, a

historical norm, or the level of a reference molecule in the biological sample.

29. A method of identifying a cancer marker in a sample, the method comprising the steps of (a) classifying the transcriptional state of the sample, (b) performing transcriptome analysis to identify modulated transcription factors, and (c) identifying a cancer marker.

30. The method of claim 29, wherein the sample is a population of cells.

31. The method of claim 29, wherein the cells are tumor stem cells.

32. The method of claim 29, wherein the samples are sorted by transcriptional state wherein transcriptional state is characterized as significantly altered open chromatin state relative to a comparator control.

33. The method of claim 32, wherein transcriptionally sorted samples are statistically evaluated.

34. The method of claim 32, wherein the sorted samples are mapped to genes.

35. The method of claim 34, wherein the genes are classified as either induced or suppressed relative to a comparator control.

36. The method of claim 35, wherein the comparator control is at least one selected from the group consisting of: a positive control, a negative control, a historical control, a historical norm, or the level of a reference molecule in the biological sample.

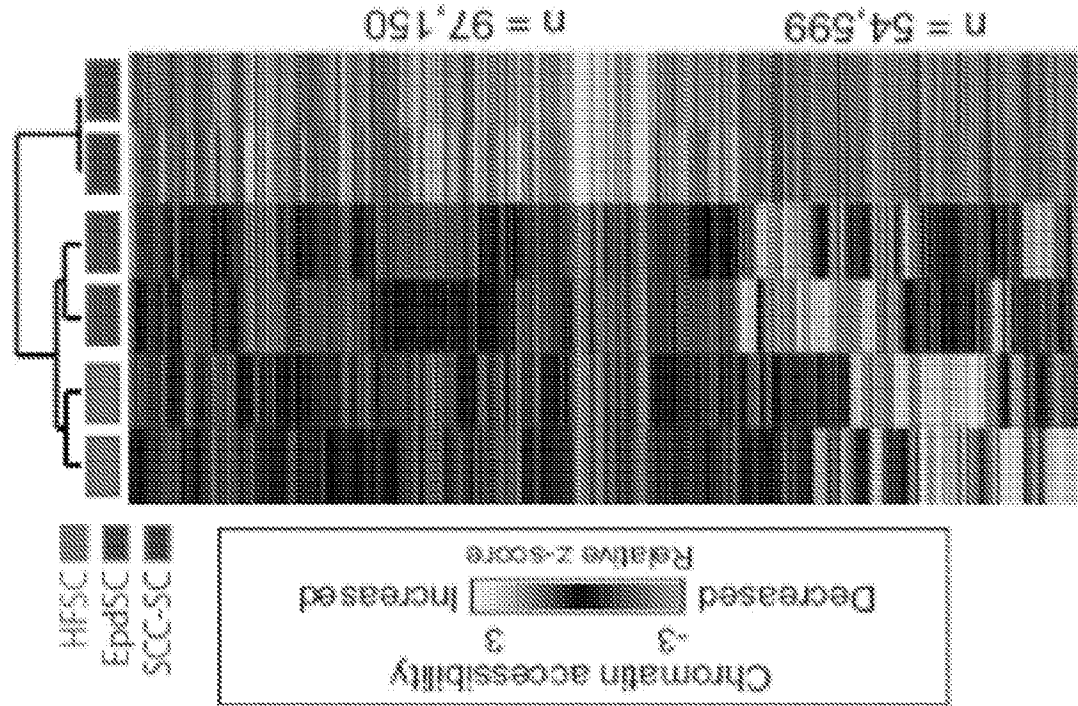
37. The method of claim 35 wherein the genes are subjected to pathway enrichment analysis.

38. The method of claim 35, wherein the genes are screened for key regulatory elements.

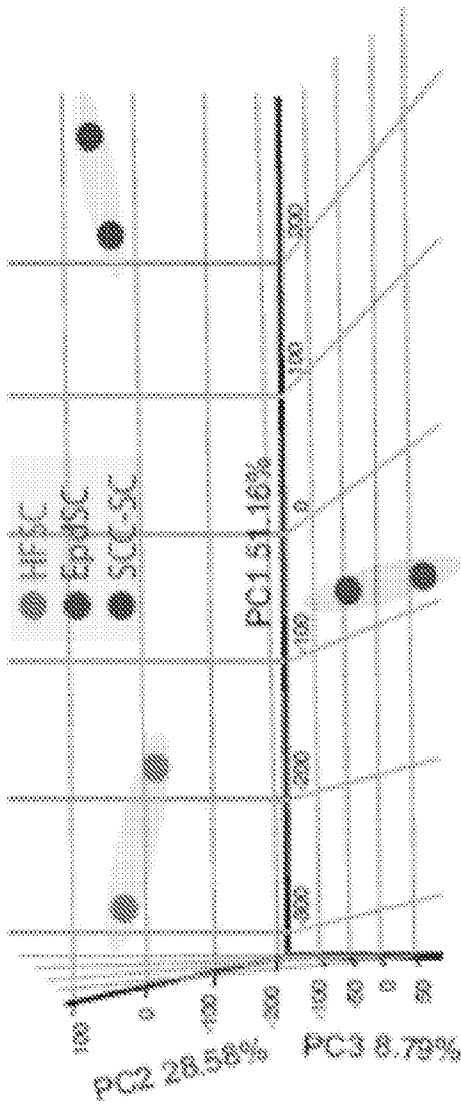
39. The method of claim 38, wherein the regulatory elements are mapped to transcription epicenters.
40. The method of claim 39, wherein the epicenters are screened for genes corresponding to the regions containing the epicenters.
41. The method of claim 40, wherein the epicenters are linked to specific transcription factors.
42. The methods of claim 41, wherein the transcription factors are screened in databases.
43. A composition for treating wounds comprising an inhibitor of SOX9.
44. A method for treating a subject having a wound, the method comprising administering an inhibitor to a subject in need thereof and inhibitor of SOX9.
45. A composition for detecting transcription epicenters comprising a plurality of probes and a plurality of promoters.
46. The composition of claim 45, wherein the plurality of probes comprises two probes.
47. The composition of claim 46, wherein the plurality of probes comprises a fluorescent protein.
48. The composition of claim 47, wherein the fluorescent protein is selected from the group consisting of RFP, iRFP and GFP.
49. The composition of claim 45, wherein the plurality of promoters comprises two promoters.
50. The composition of claim 49, wherein the two promoters are selected from a group consisting of a H2B, SOX9, KLF5, PGK and miR21.

51. A method of screening for a composition that modulates a transcription epicenter, wherein the epicenter regulates cancer, the method comprises: (a) administering an epicenter reporter construct into cell, (b) detecting epicenter activity relative to a comparator control in the cell, and (c) thereby identifying a composition that modulate the epicenter.

B Global changes of SCC-SCs in chromatin accessibility



A Normal and SCC-SCs exhibit distinct chromatin accessibility patterns



C Genes that gain or lose ATAC peaks during tumorigenesis undergo larger transcriptional changes

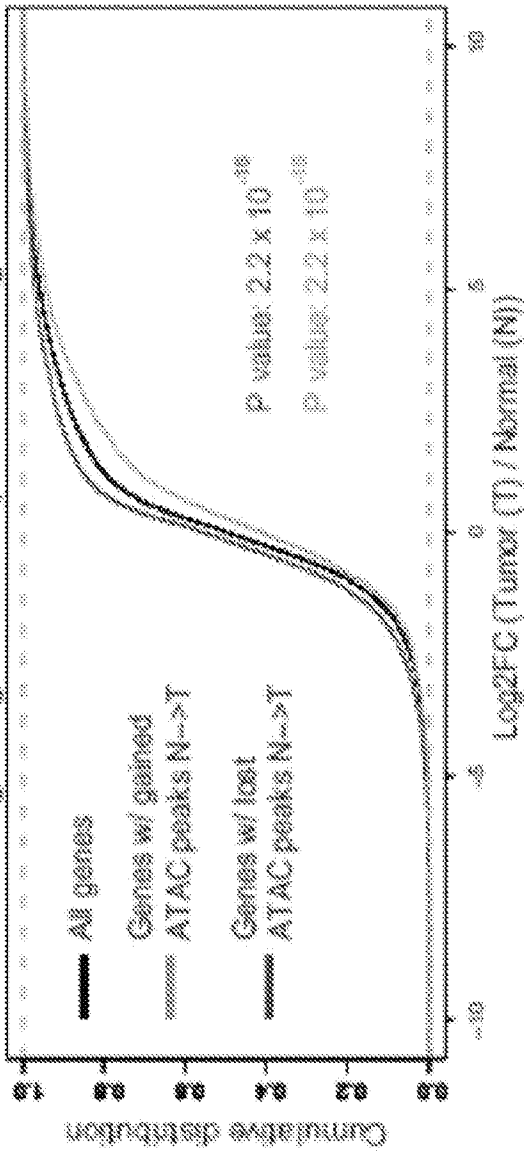
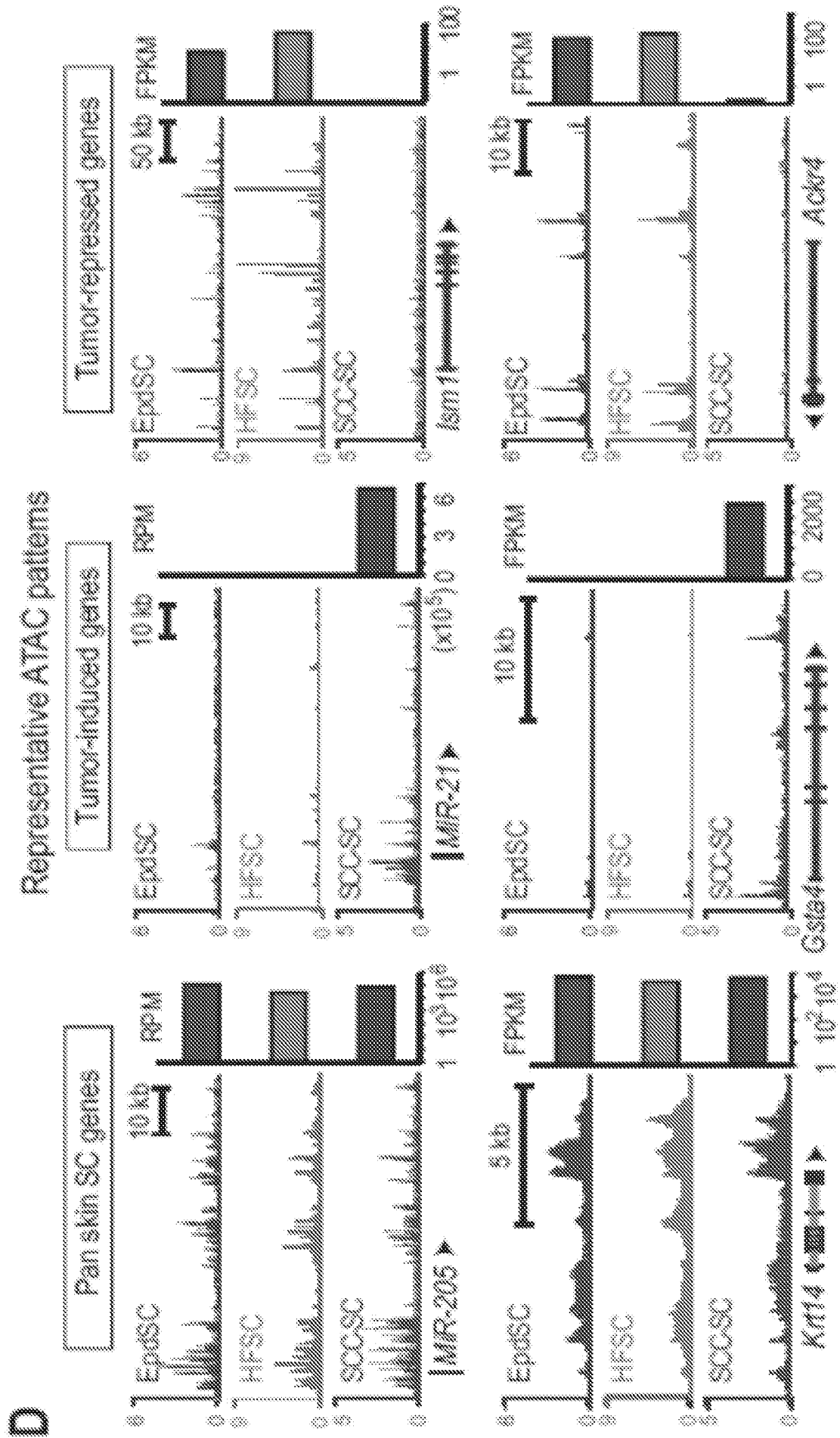


Figure 1A-1C



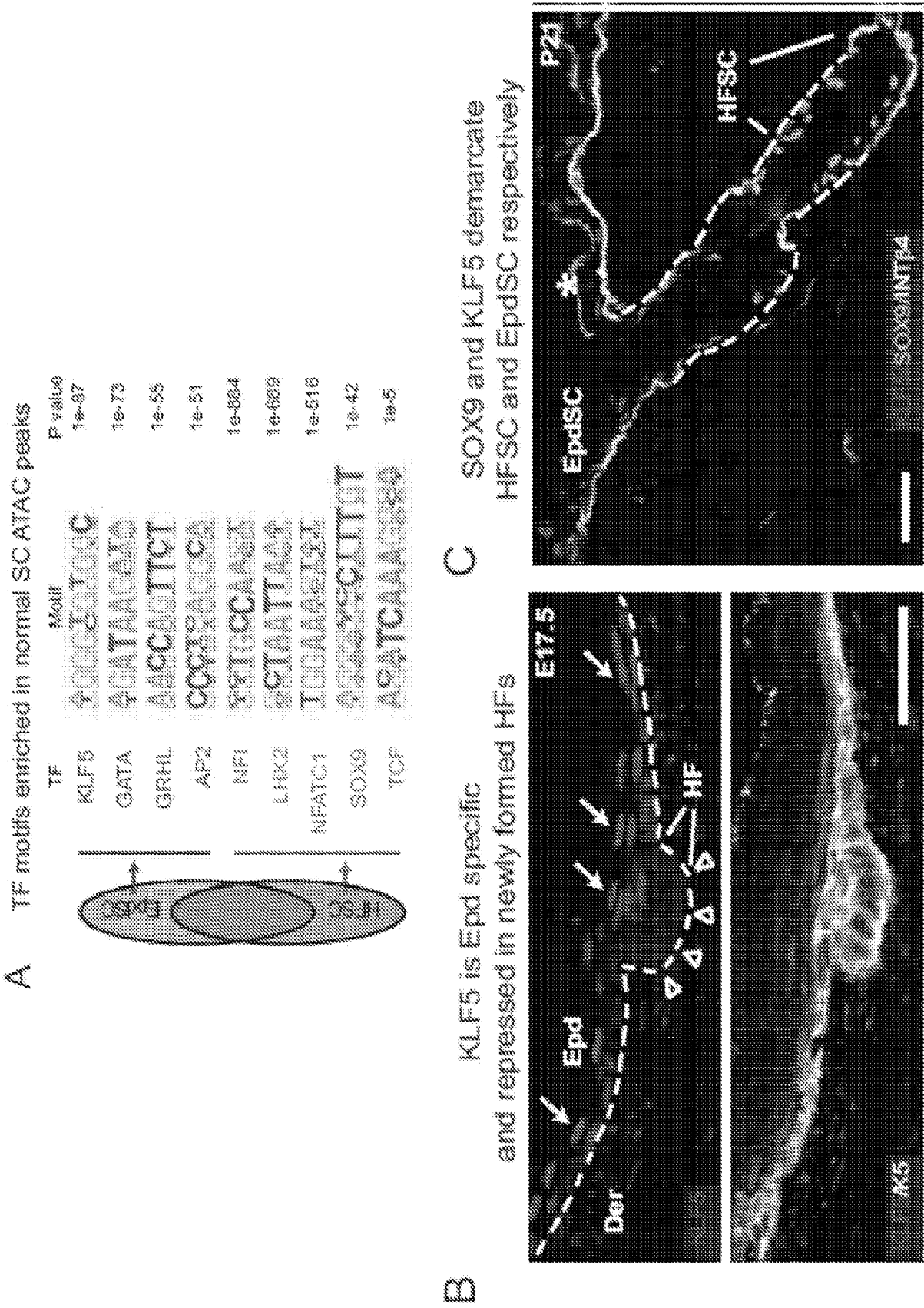
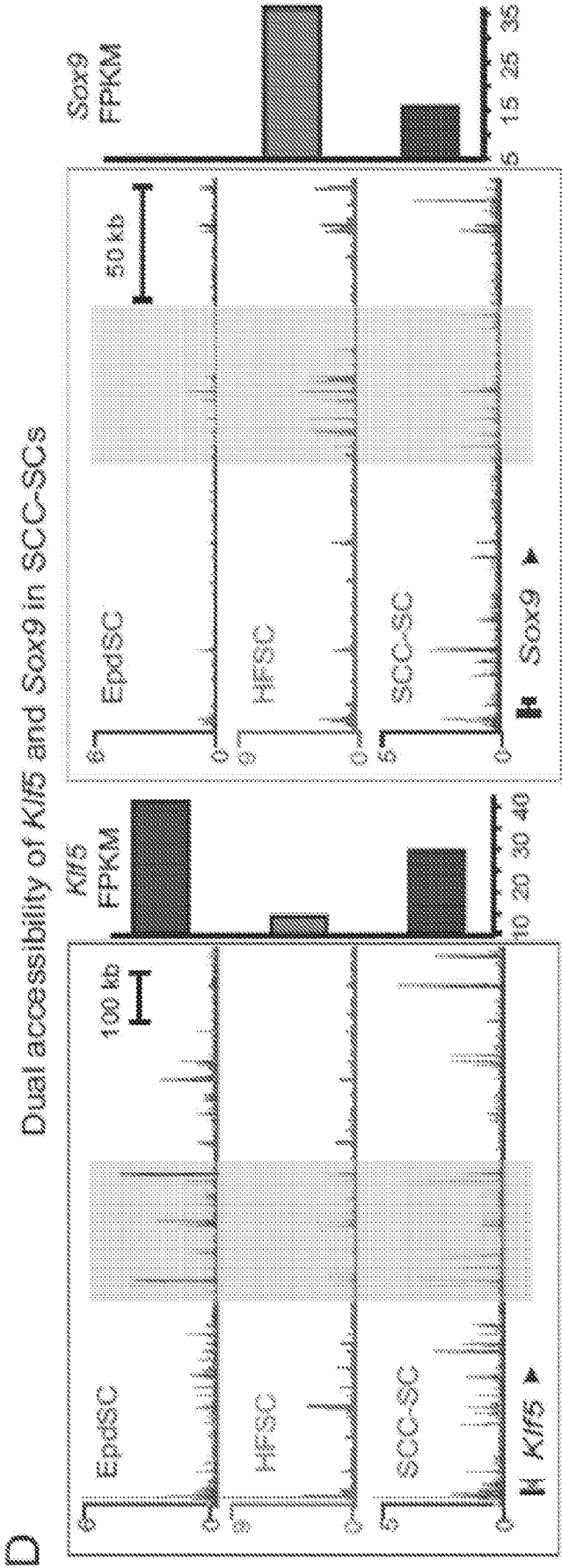


Figure 2A – 2C



E Both lineage TF motifs are enriched in SCC-SCs: sign of lineage infidelity

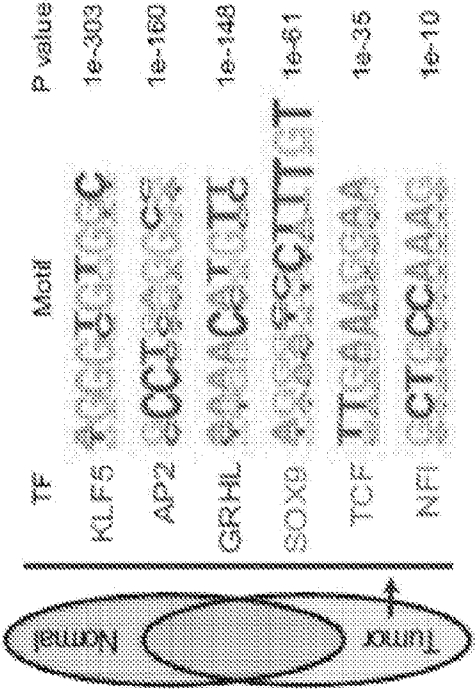


Figure 2D – 2E

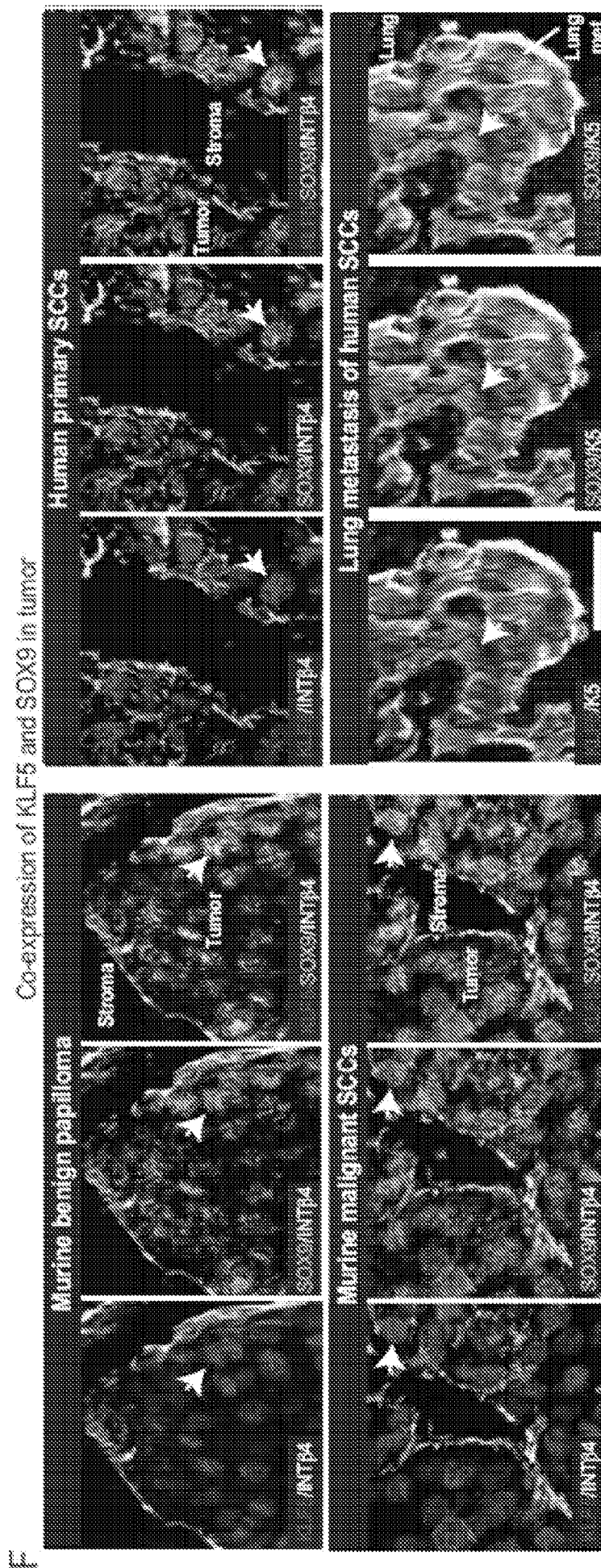


Figure 2F

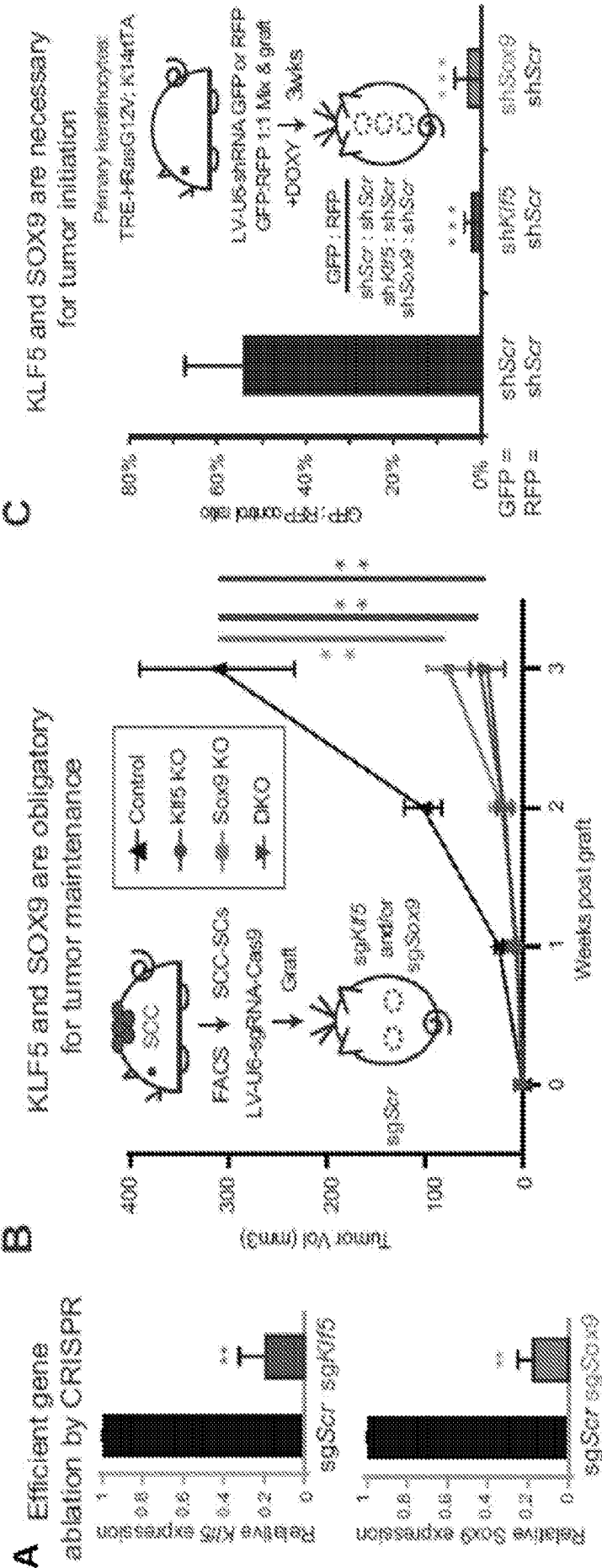


Figure 3A – 3C

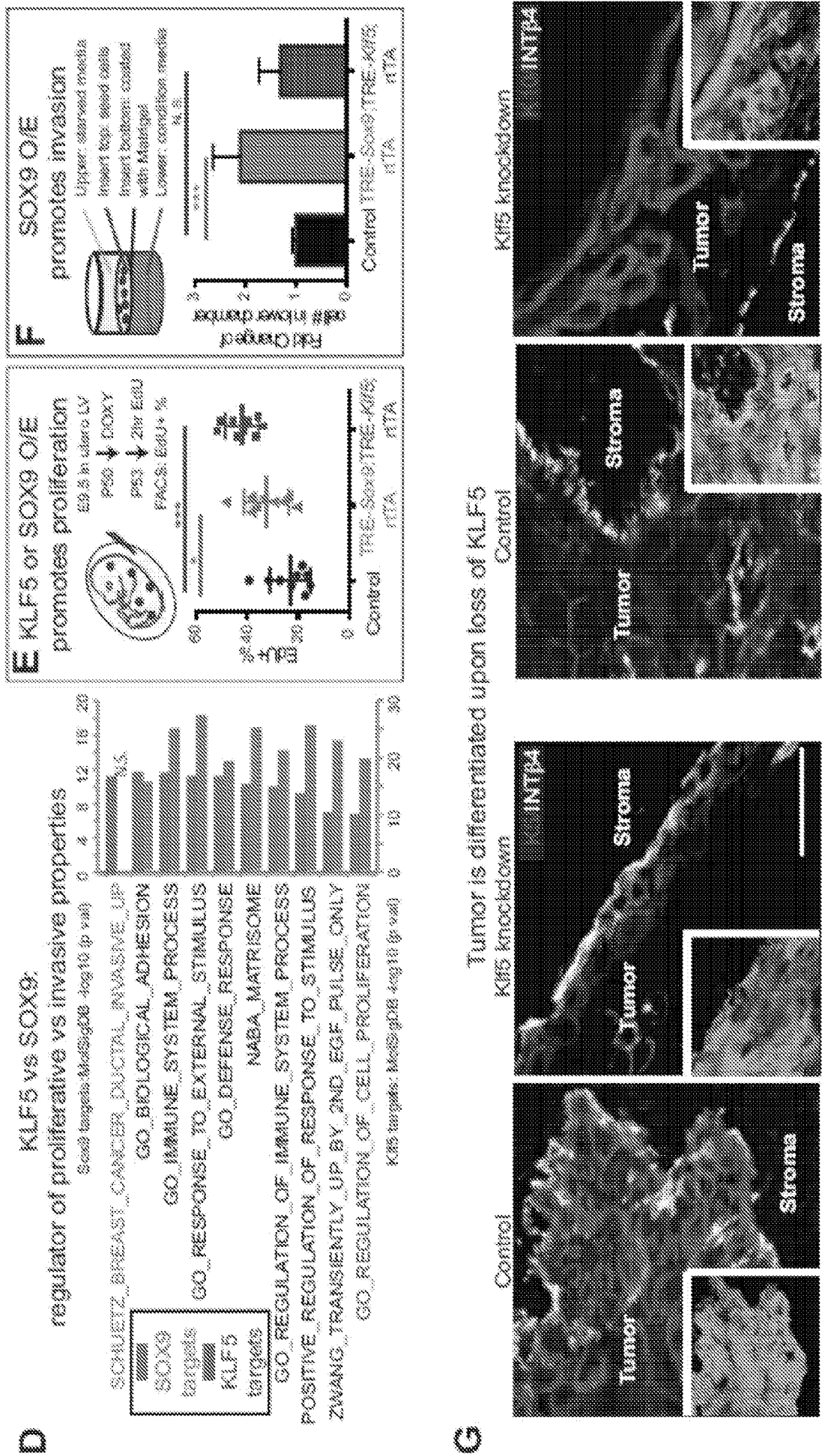


Figure 3D – 3G

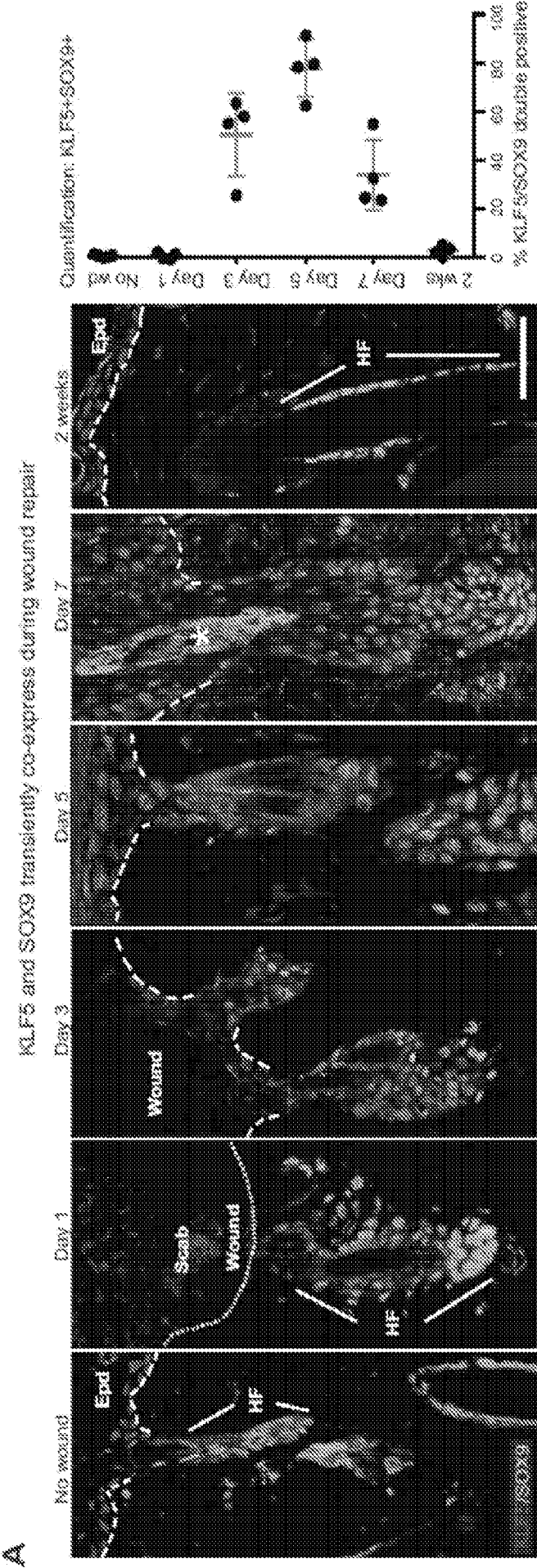


Figure 4A

B

KLF5 and SOX9 are both required for wound repair

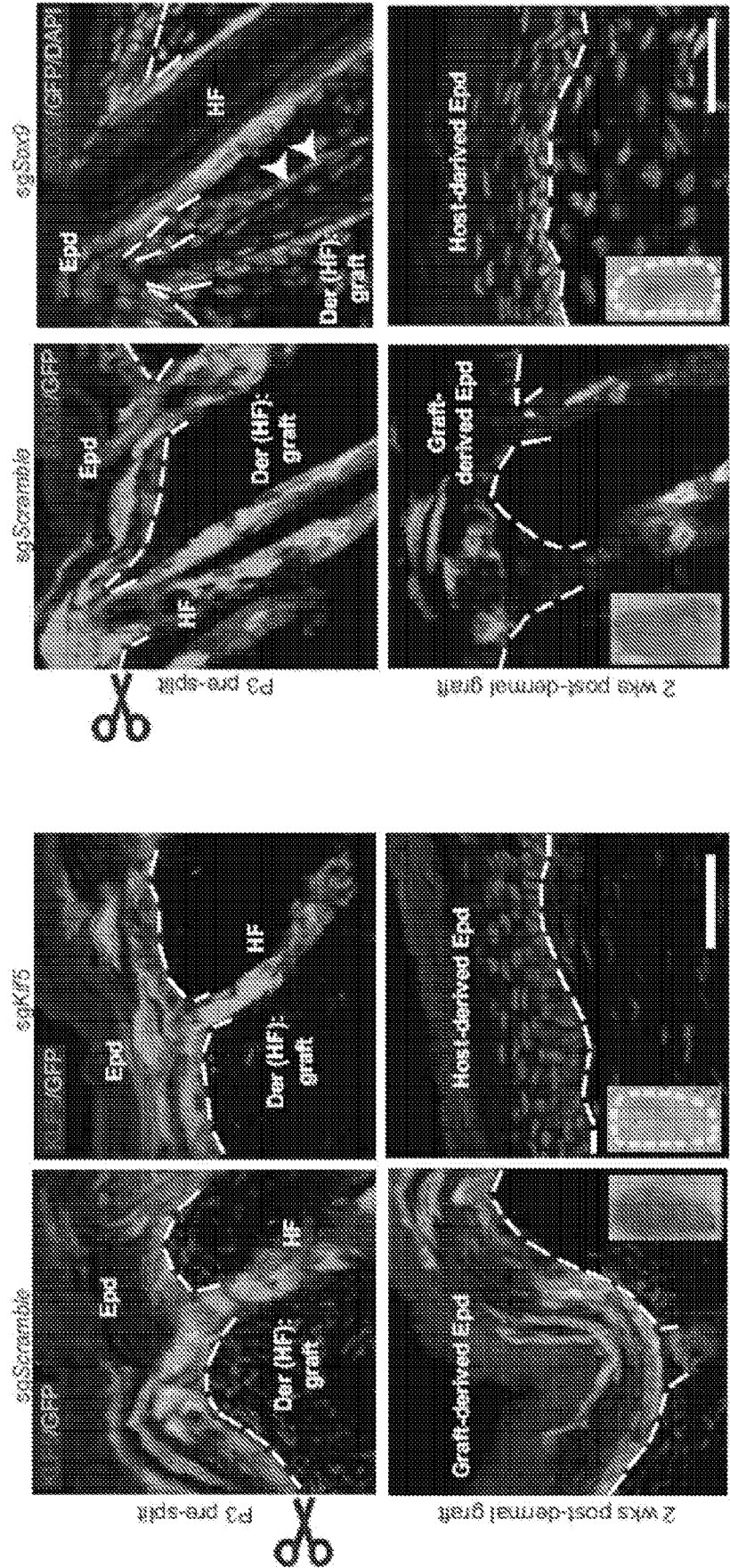
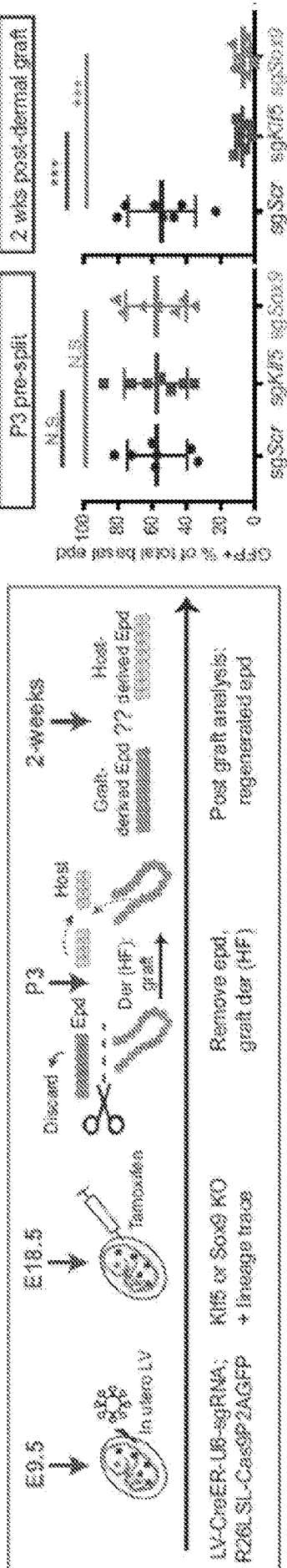


Figure 4B

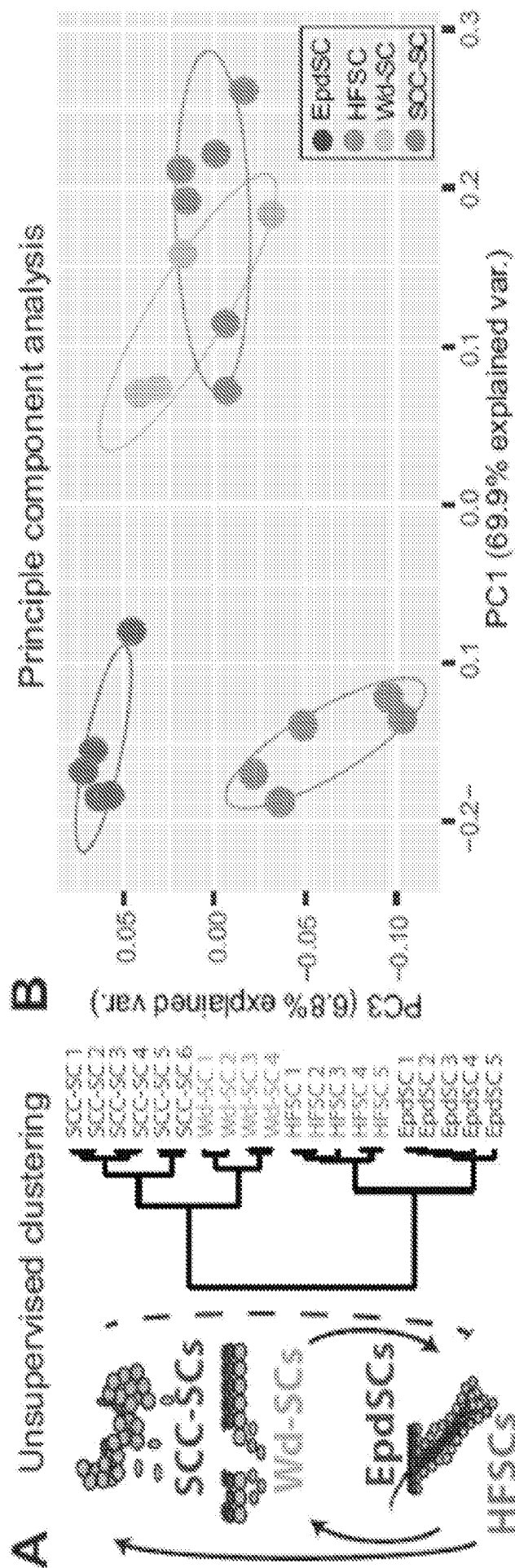


Figure 5A – 5B

C

GSEA: shared changes in tumor and wound

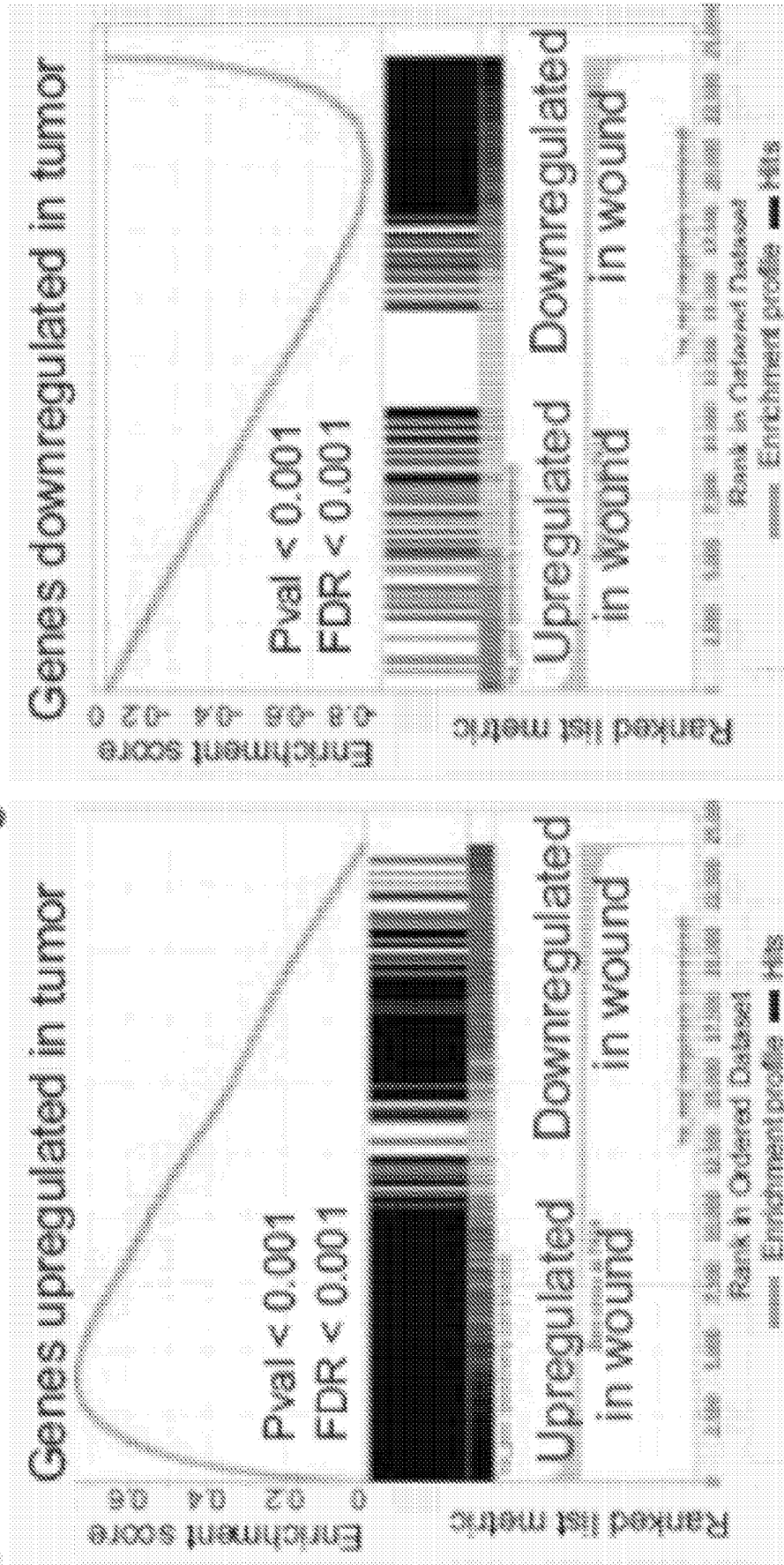


Figure 5C

D Wound and tumor exhibit similar open chromatin landscape

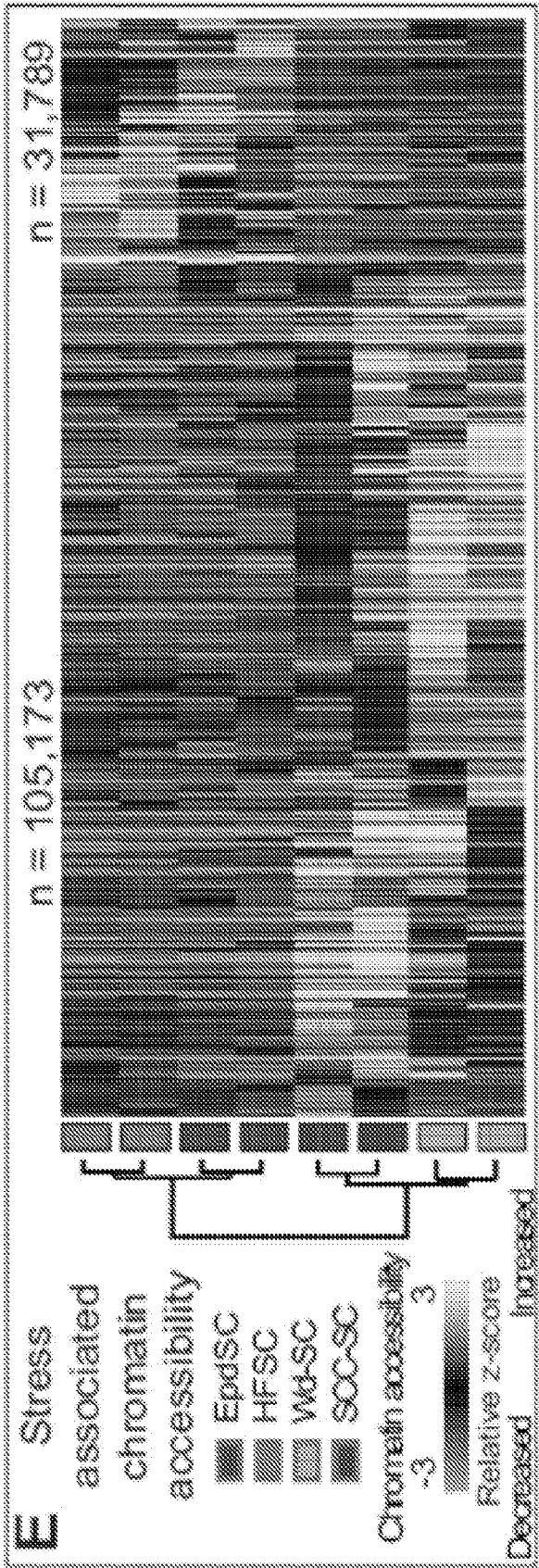
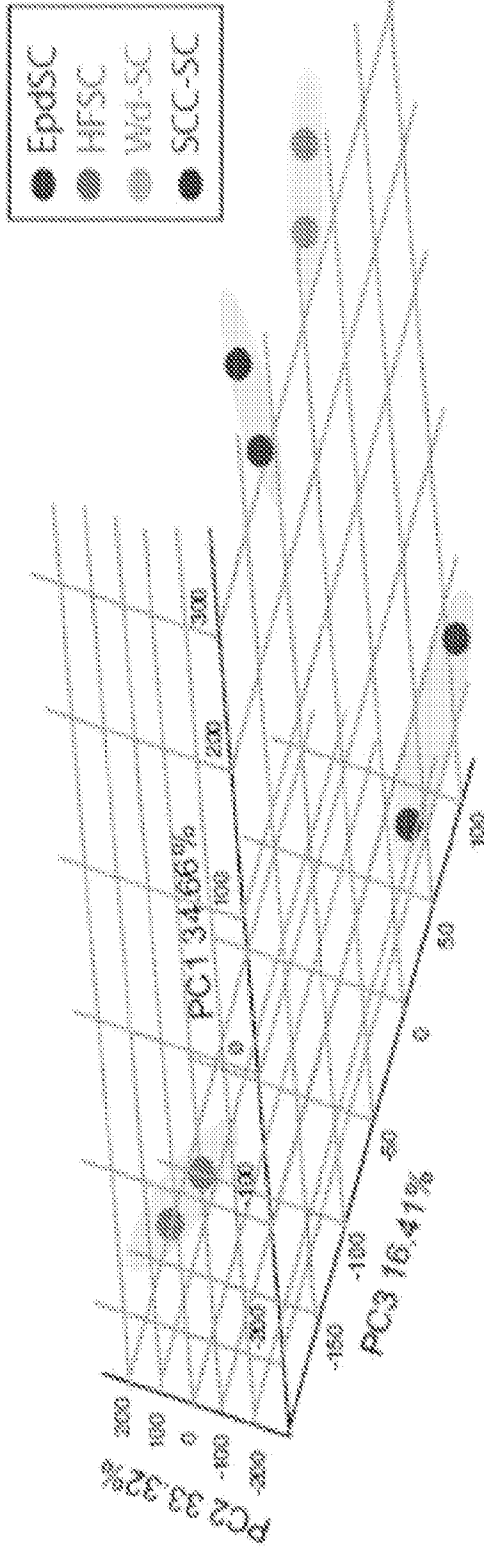


Figure 5D-5E

A Wound epi-centers (ECs) mark stress-experiencing open chromatin states in tumor and wound

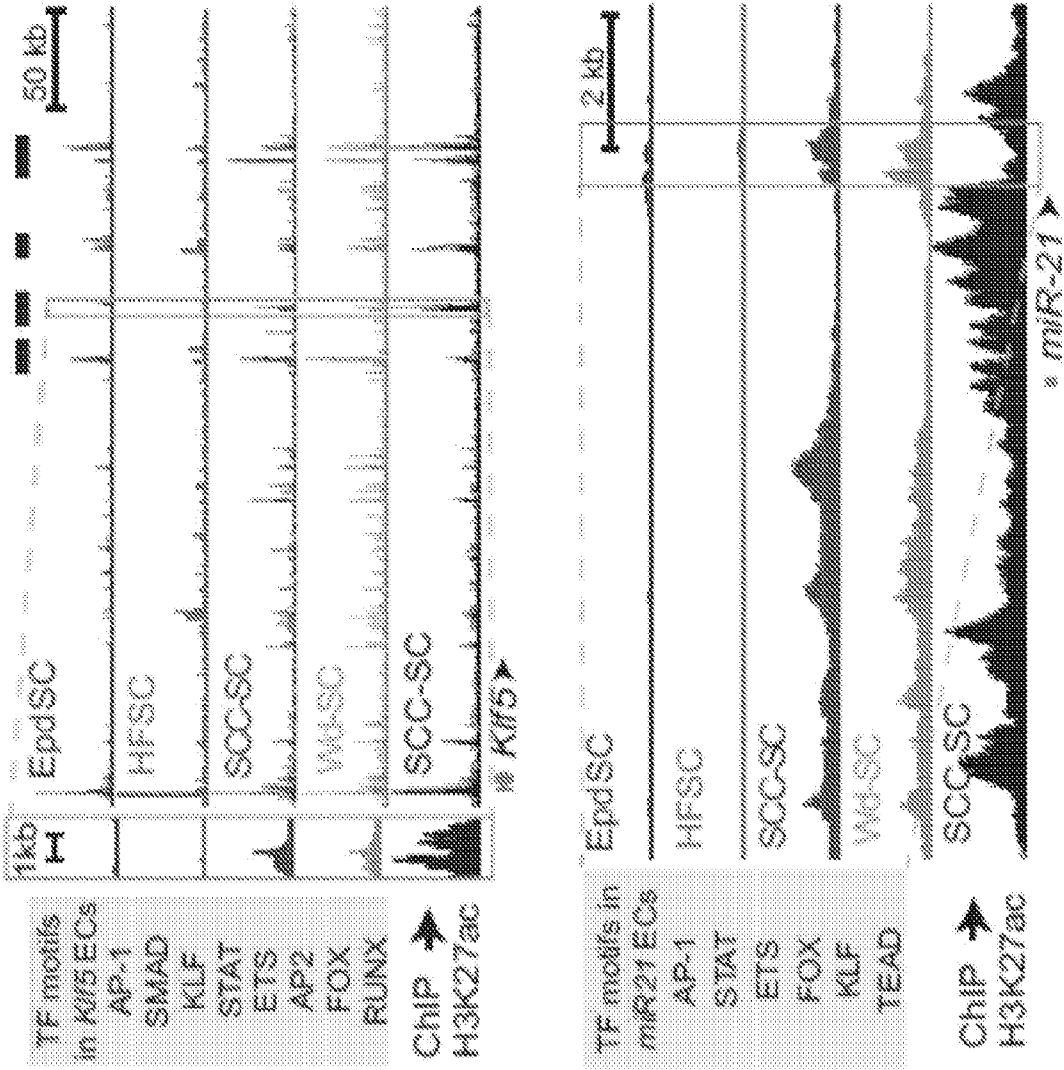


Figure 6A

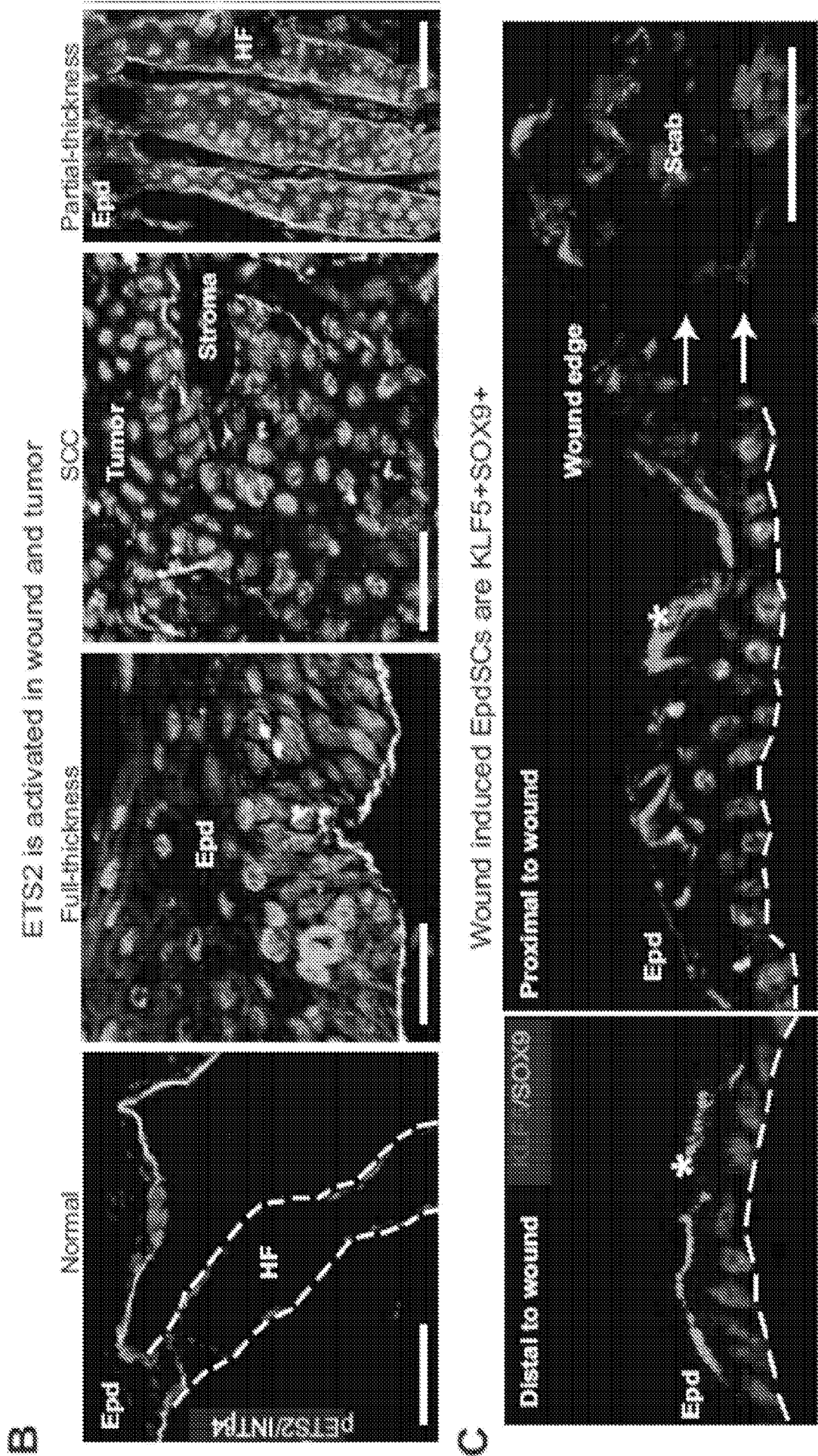
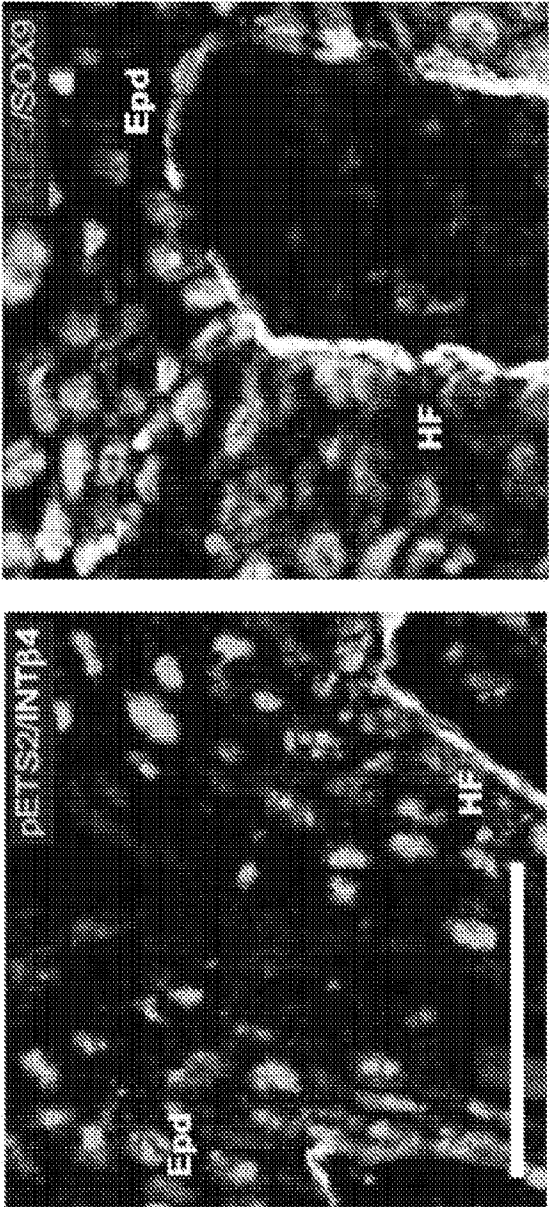


Figure 6B – 6C

D

Ectopic ETS2-T27D elicits lineage infidelity



E

ETS2 is required for wound repair

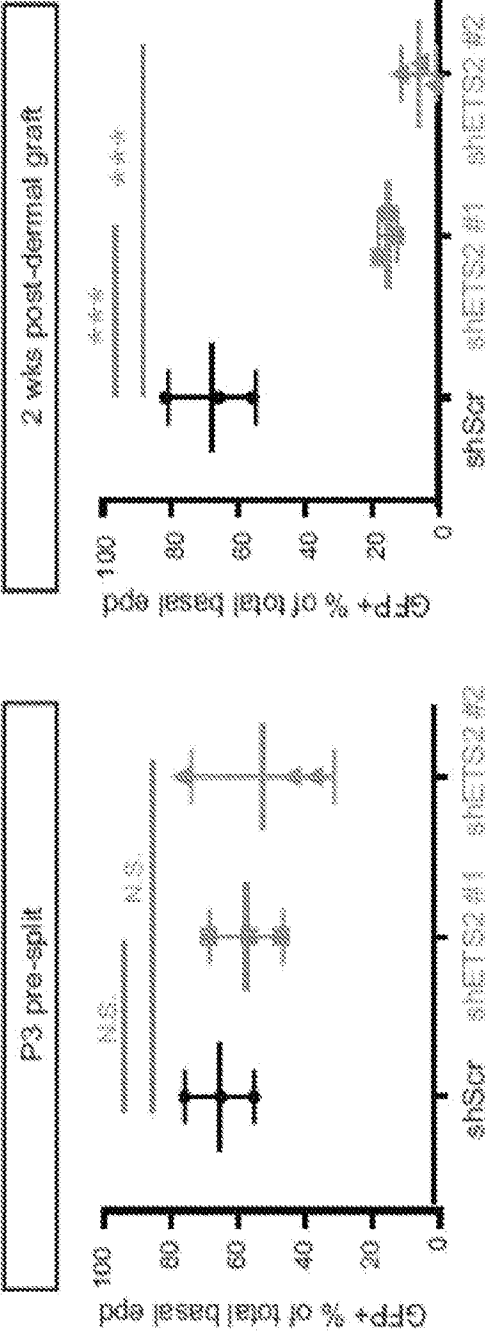


Figure 6D – 6E

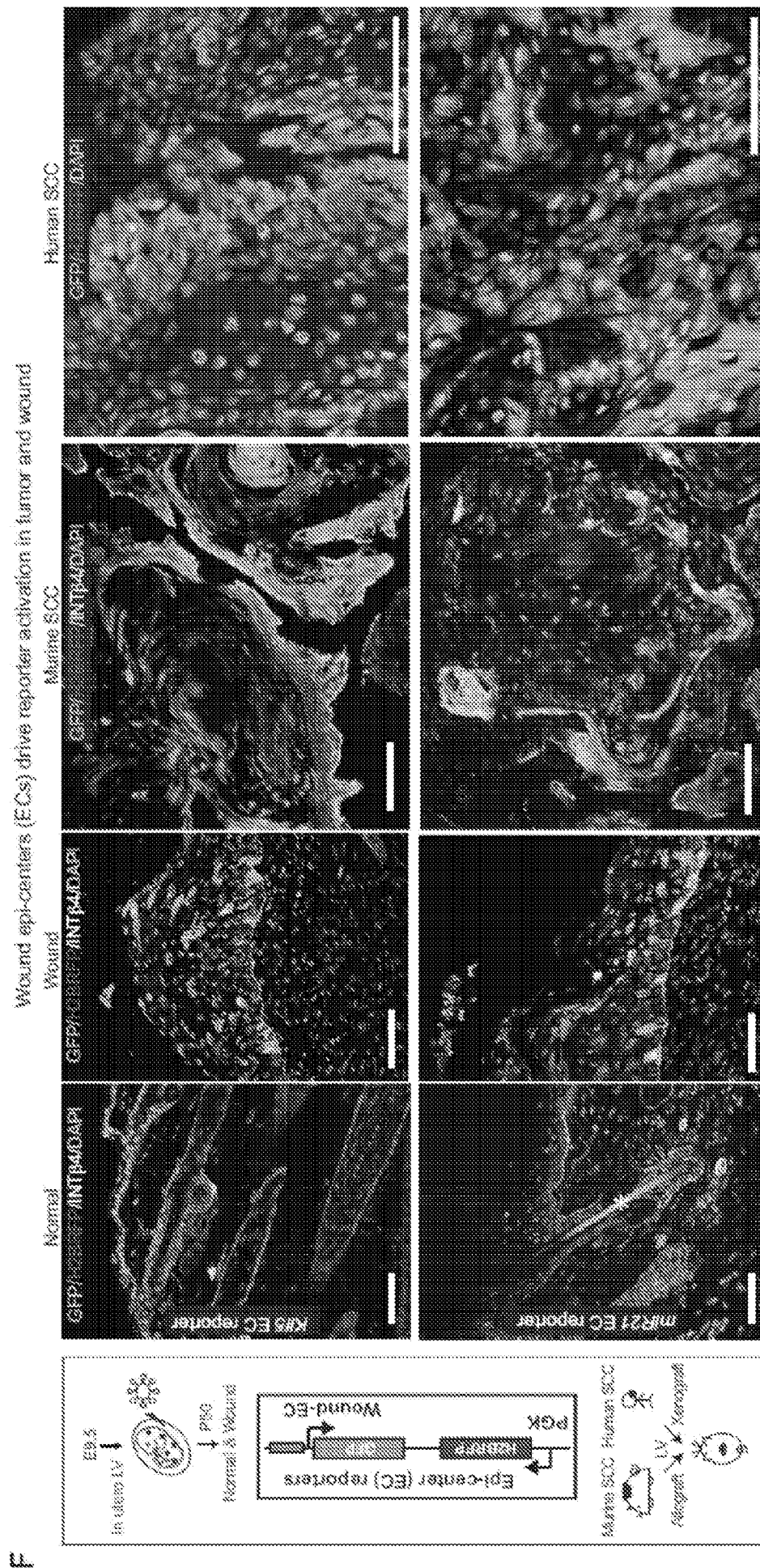


Figure 6F

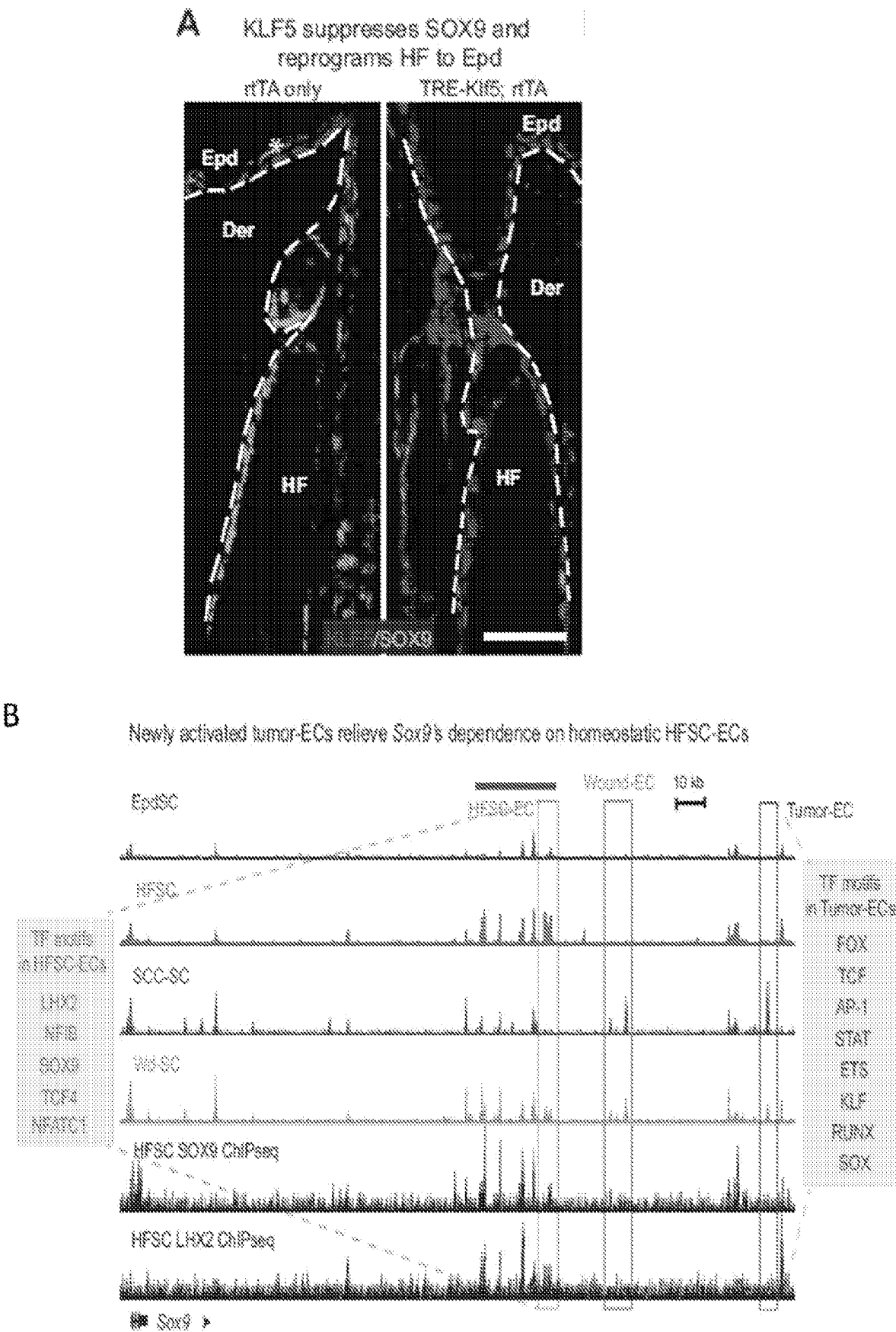


Figure 7A - 7B

Homeostatic HFSC ECs are turned off in wound and tumor,
while high-stress tumor ECs are activated in tumor

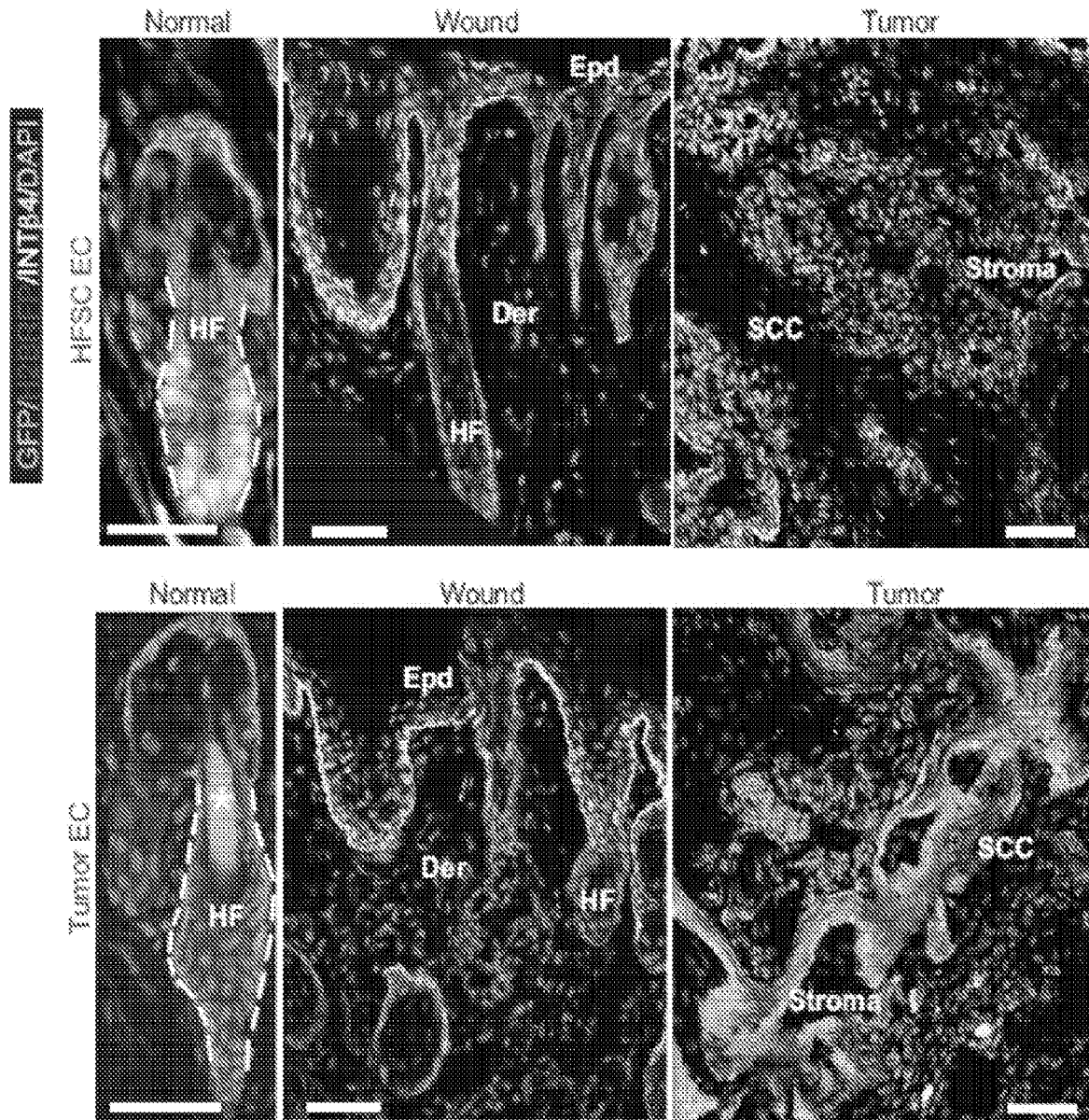
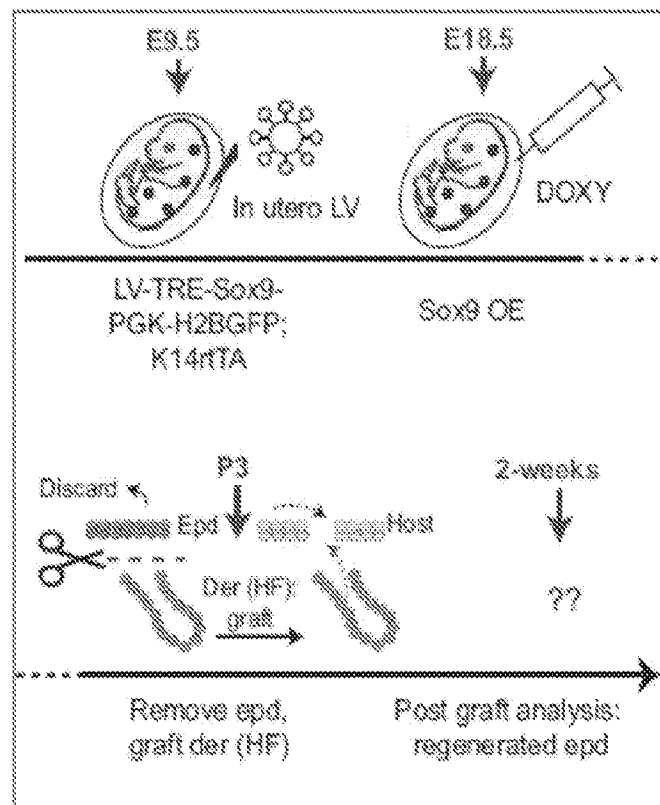


Figure 7C



Wounds do not heal when SOX9 is sustained

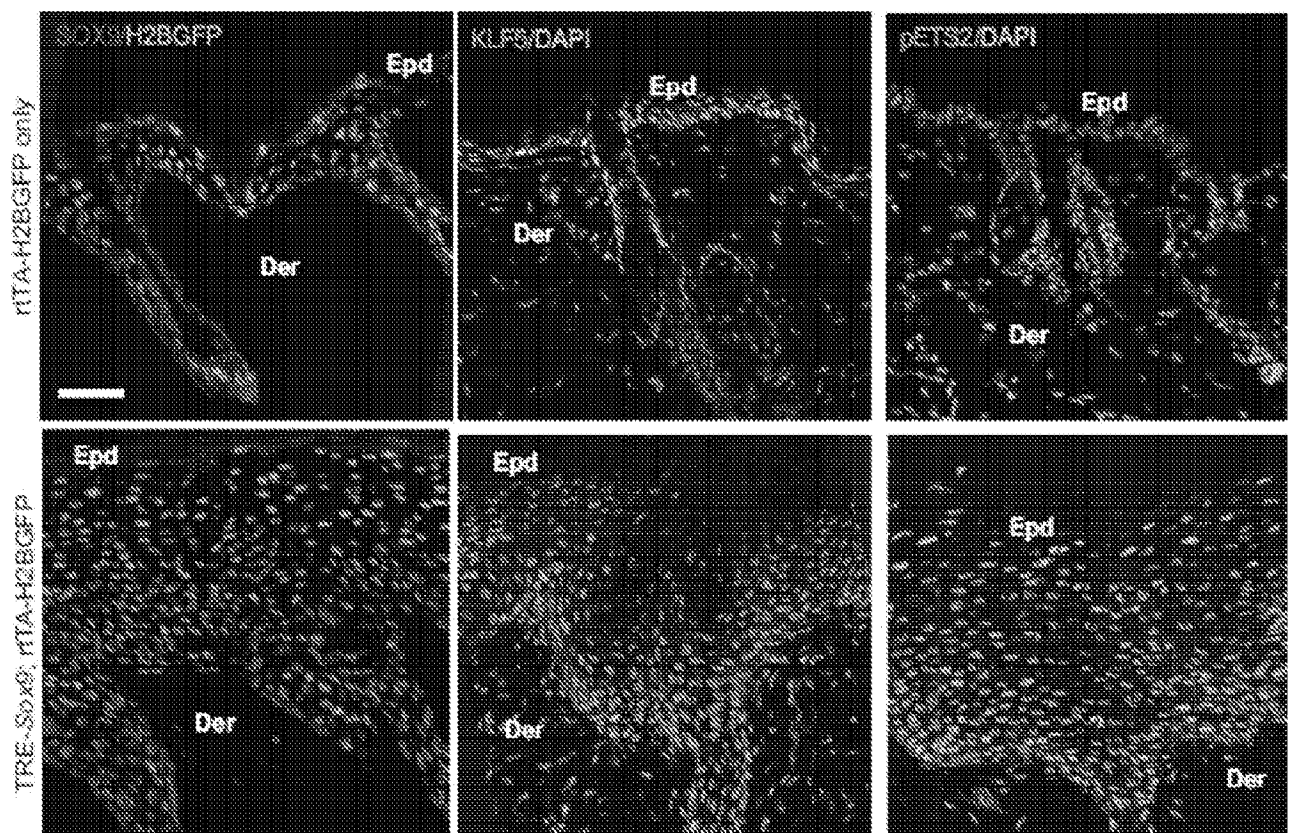


Figure 7D

A ATAC signals are found at expected TSS and CTCF sites throughout the Epd-, HF- and SCC-SC genome

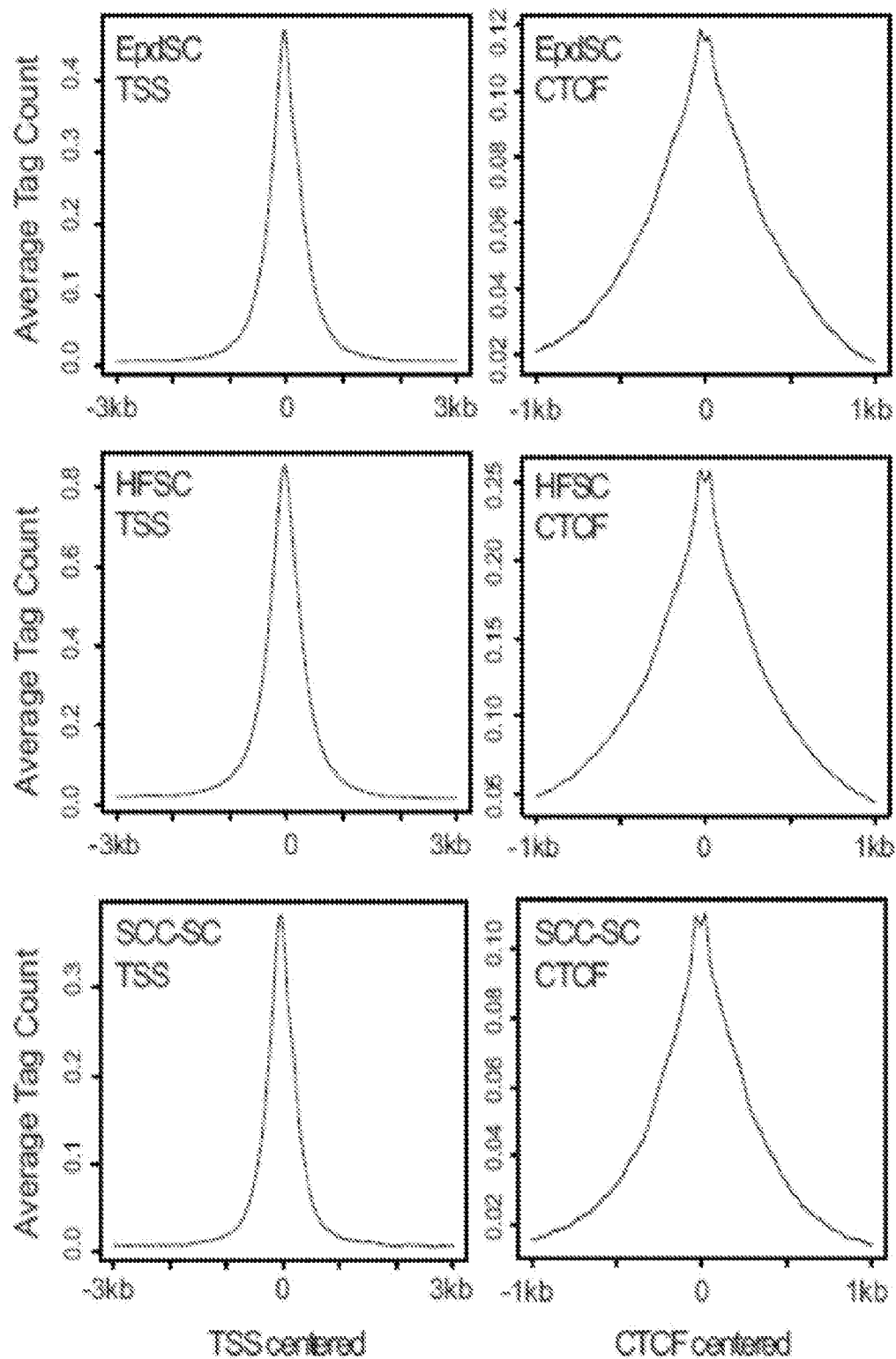
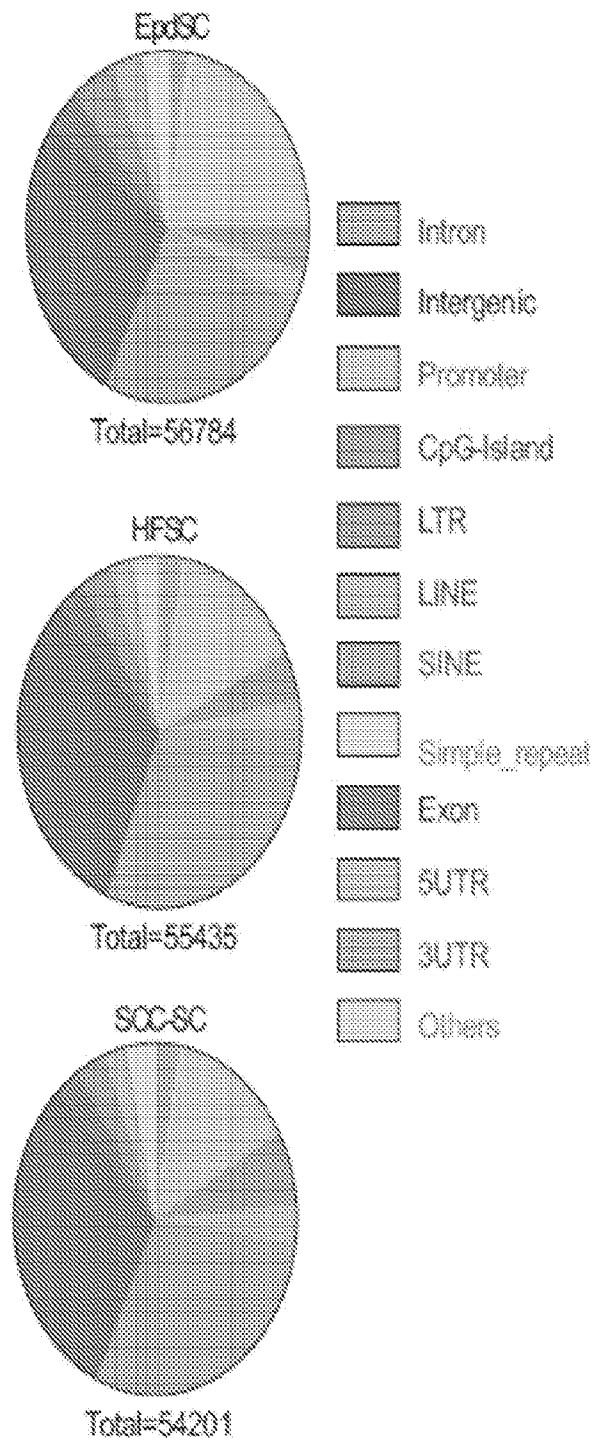


Figure 8A

B Genomic distribution of ATAC peaks



C HFSC specific ATAC peak marks homeostatic epi-centers (ECs) in HFSCs

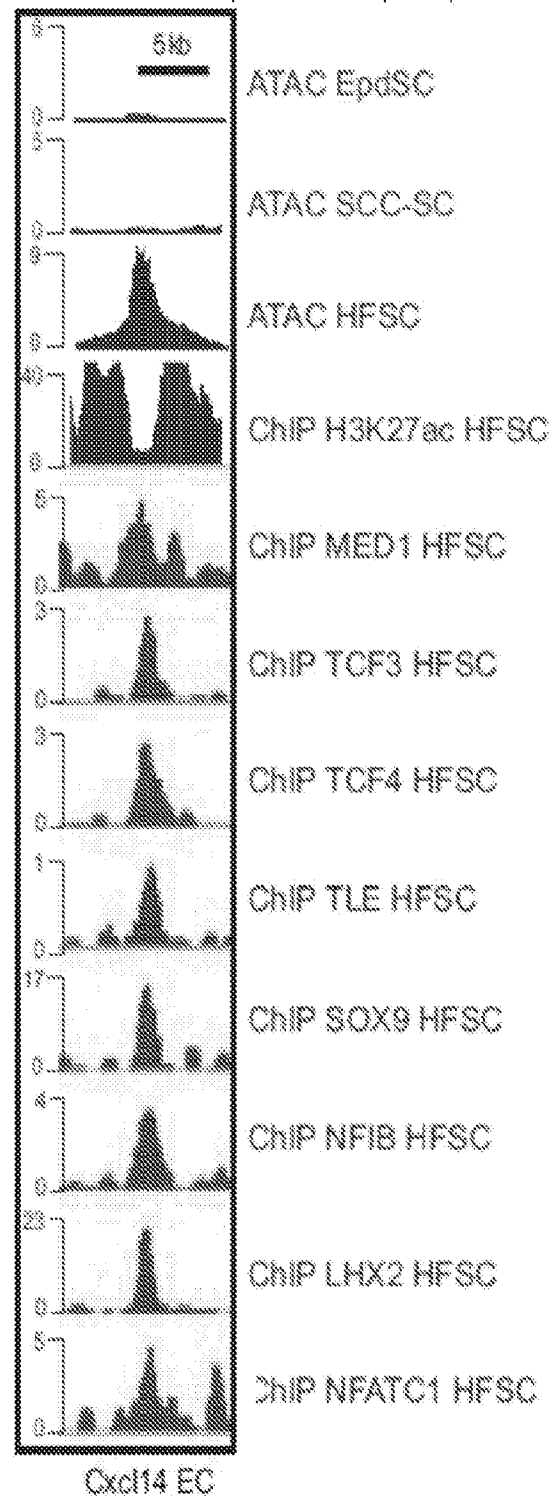
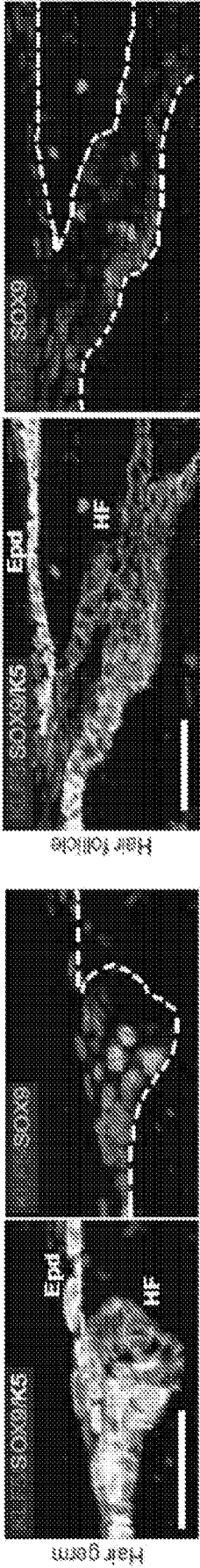


Figure 8B – 8C

KLF5 wanes and SOX9 emerges in early hair precursors

A



B KLF5 and SOX9 are co-expressed in skin epithelial hyperplasia

C

Both EpdSC and HFSC lineage genes are pervasively co-expressed in SCC-SCs

D

Both HF and Epd lineage genes are expressed in SCCs

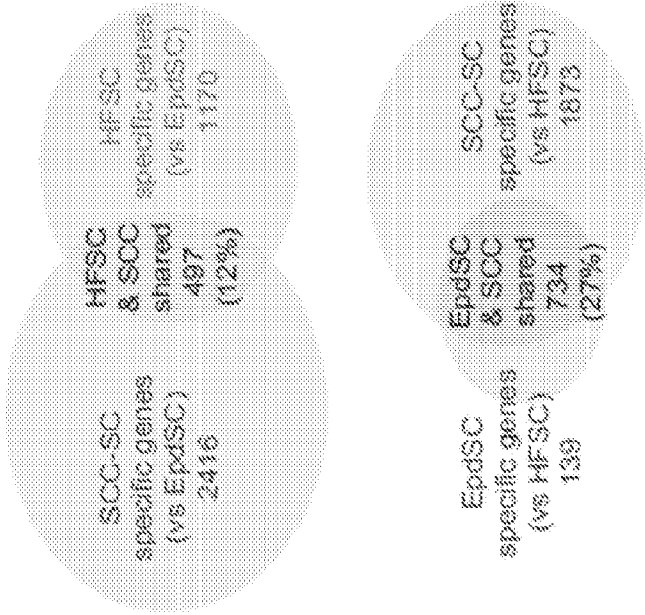
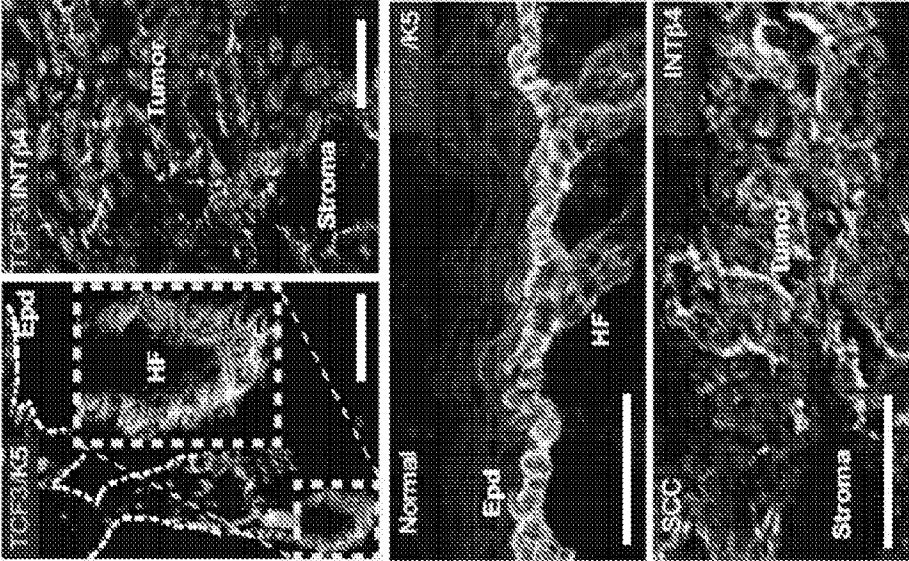
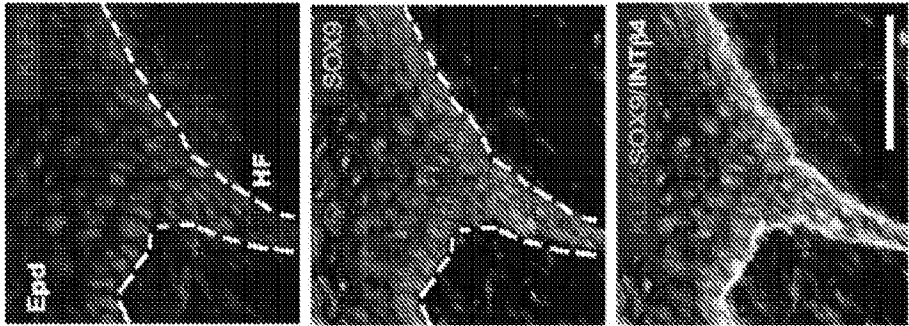


Figure 9A - 9D

E

Both HF and Epd lineage genes exhibit open chromatin states in SCCs

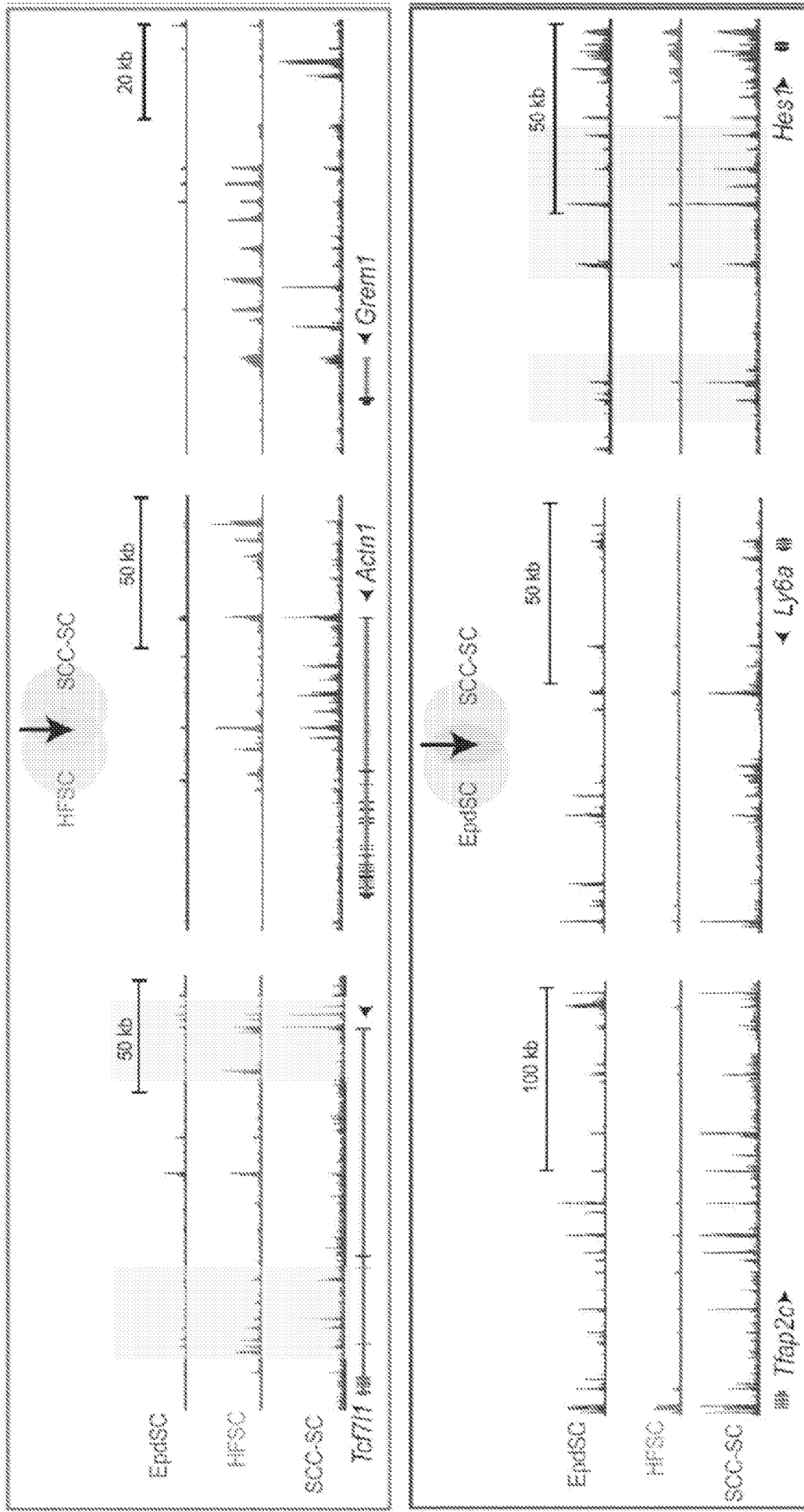


Figure 9E

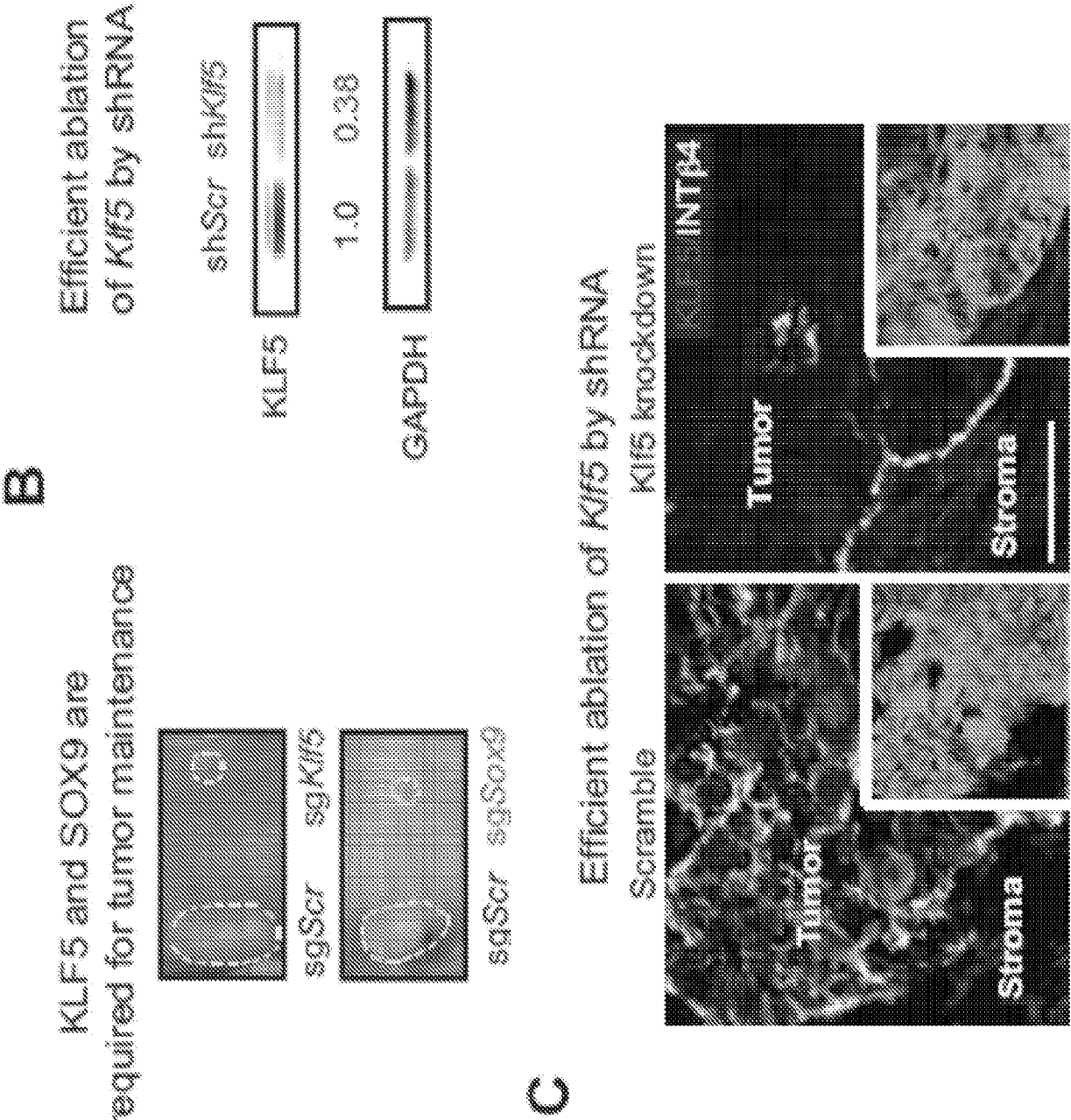


Figure 10A - 10C

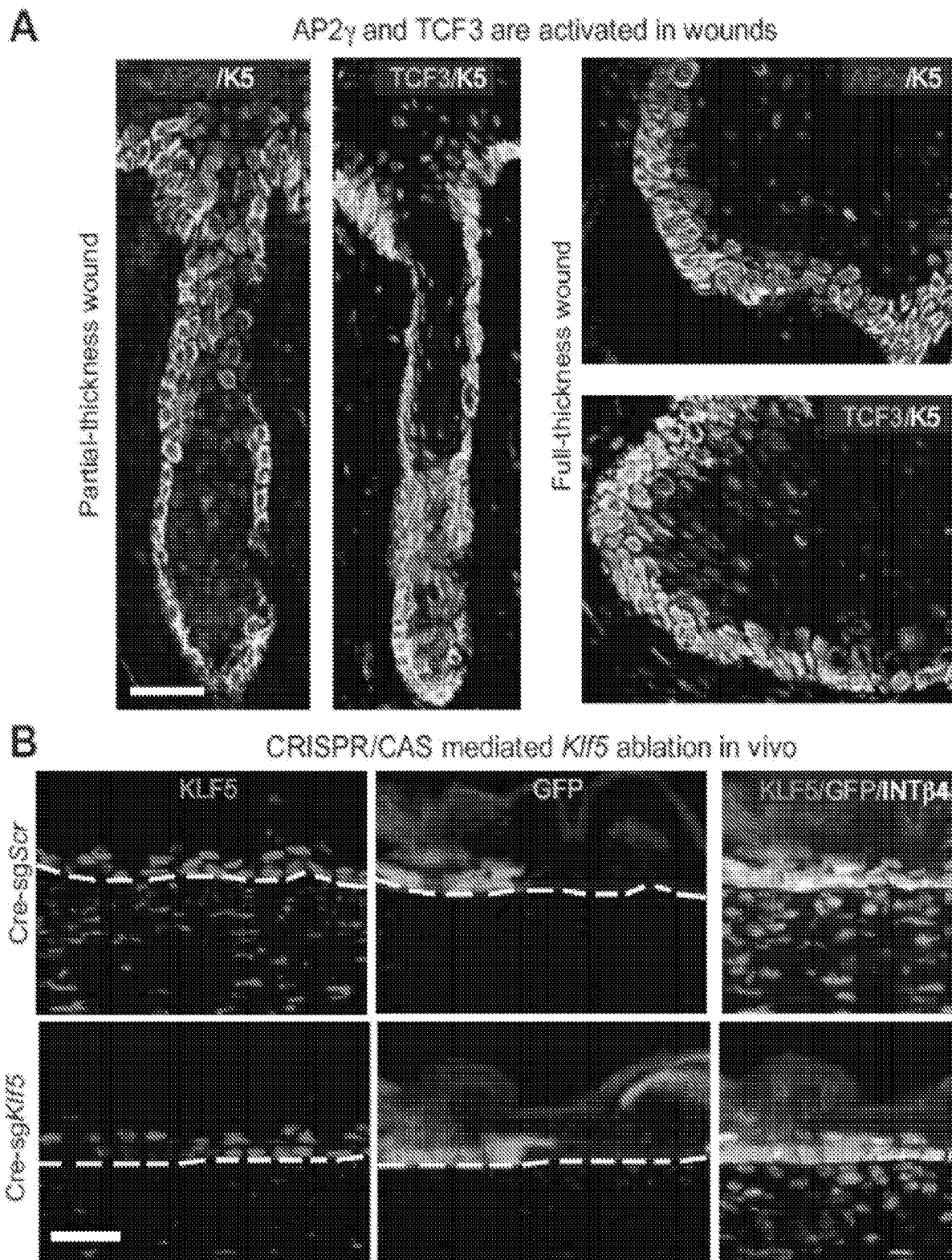


Figure 11A - 11B

C

CRISPR/CAS mediated Sox9 ablation in vivo

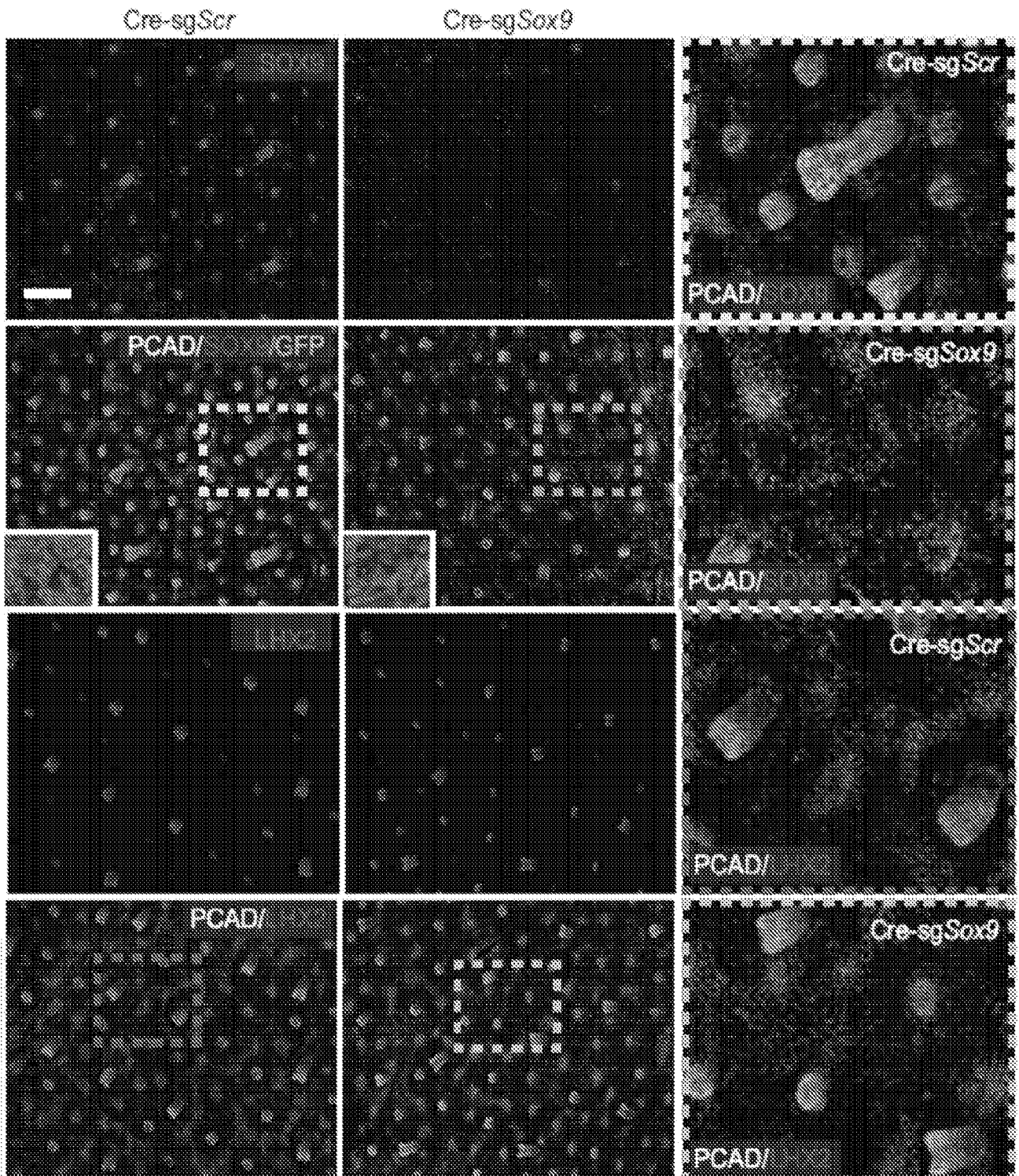


Figure 11C

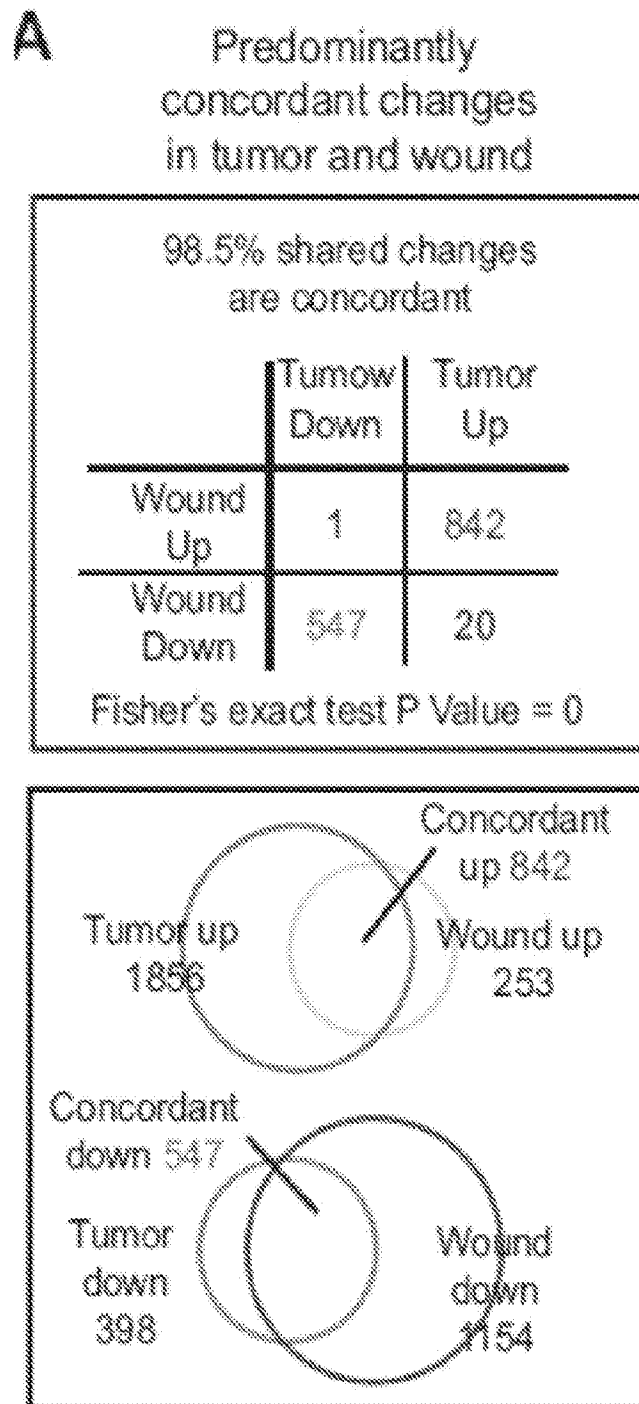
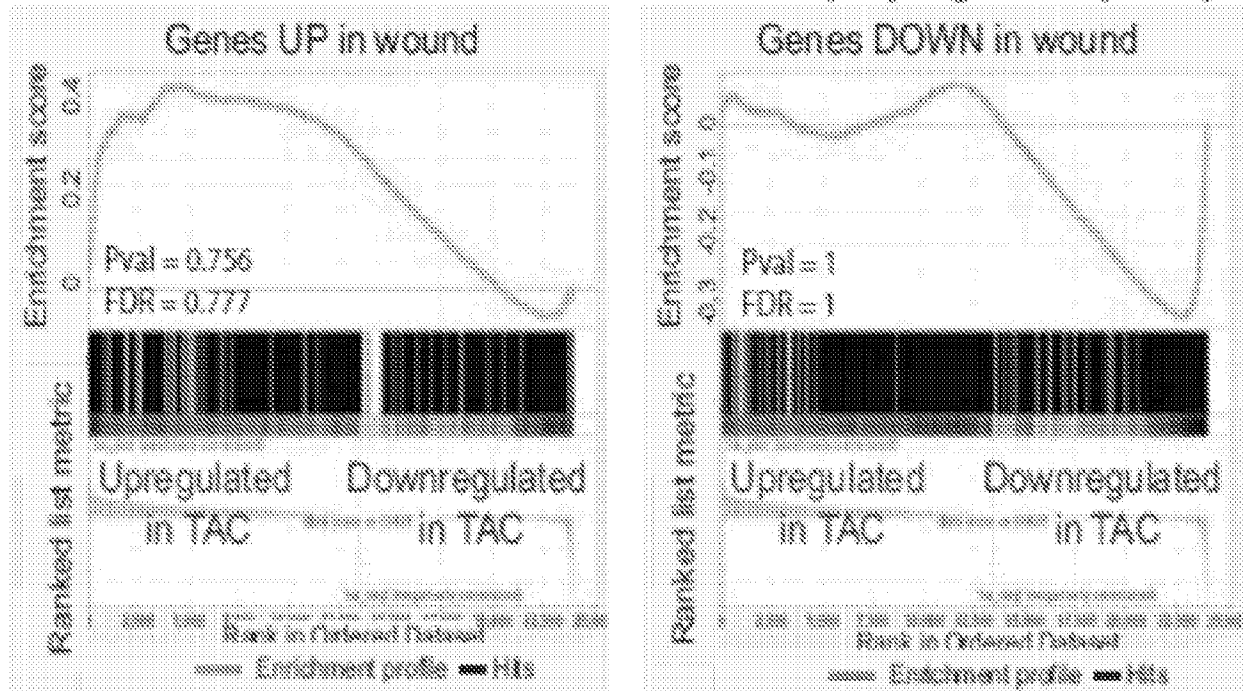


Figure 12A

B GSEA: changes in wound are distinct from those in committed transient amplifying cells (TAC)



GSEA: changes in tumor are distinct from those in committed transient amplifying cell (TAC)

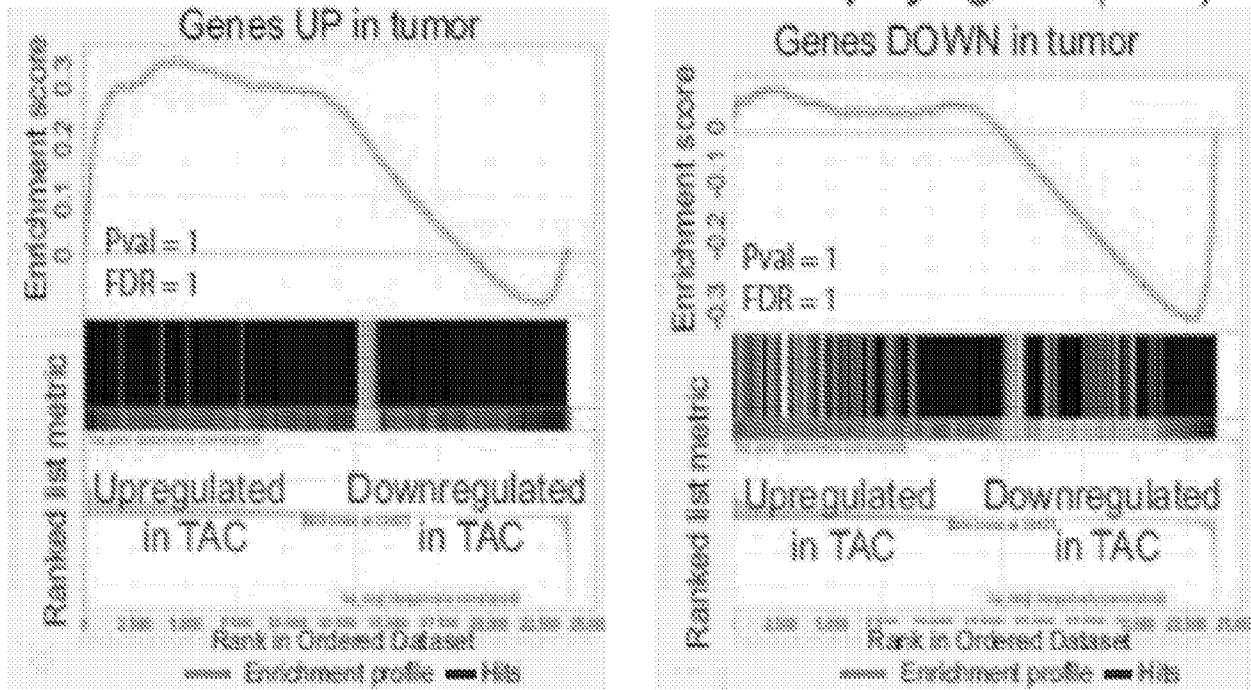
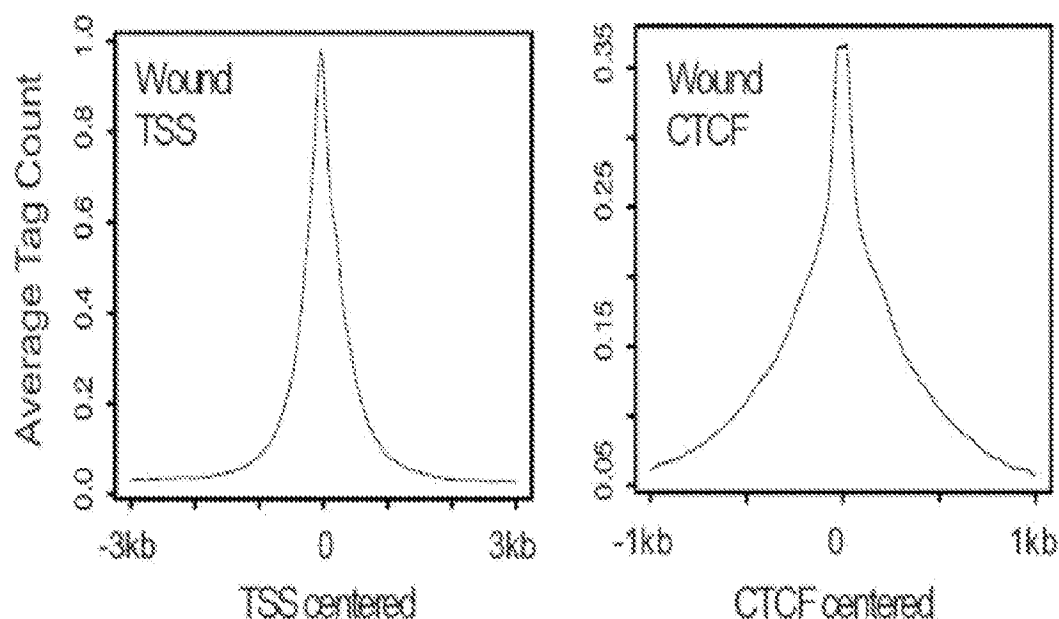


Figure 12B

C

ATAC signals are found at expected
TSS and CTCF sites throughout the Wd-SC genome

**D**

Genomic distribution of ATAC peaks in Wd-SCs

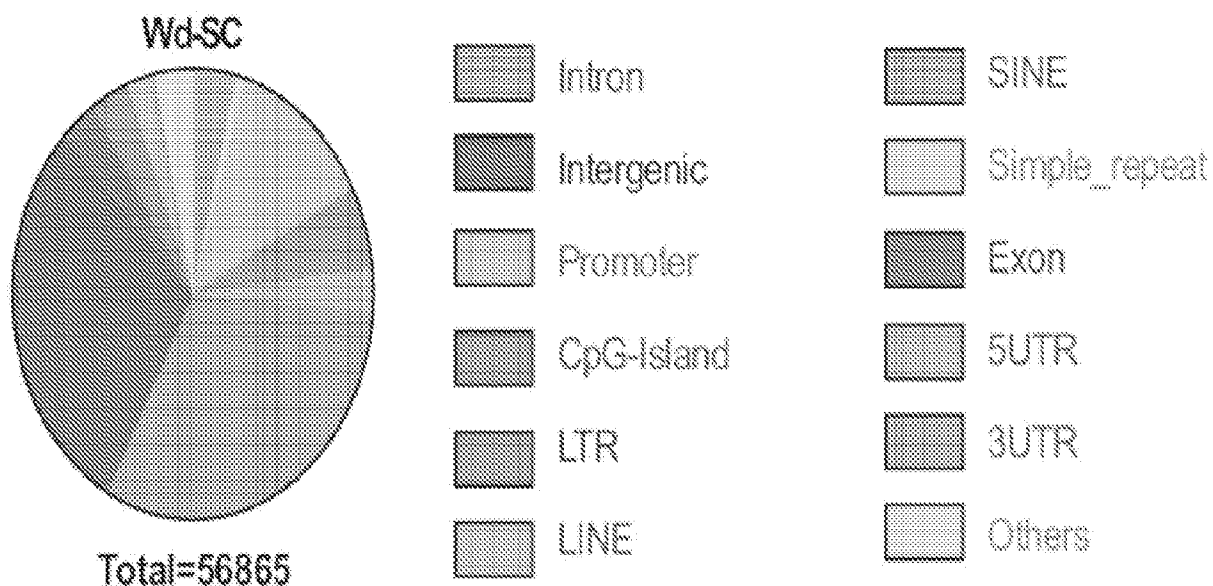


Figure 12C - 12D

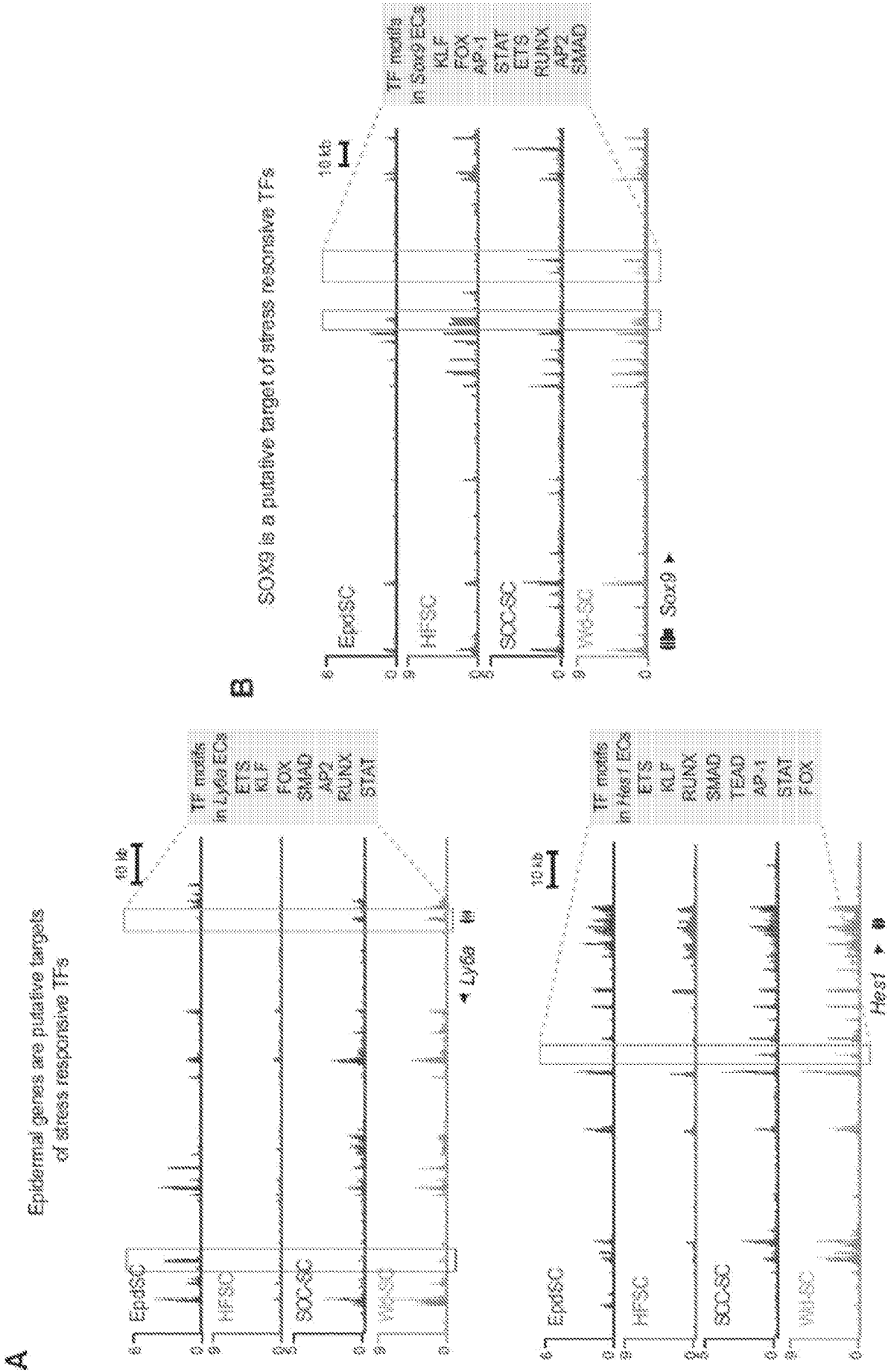


Figure 13A - 13B

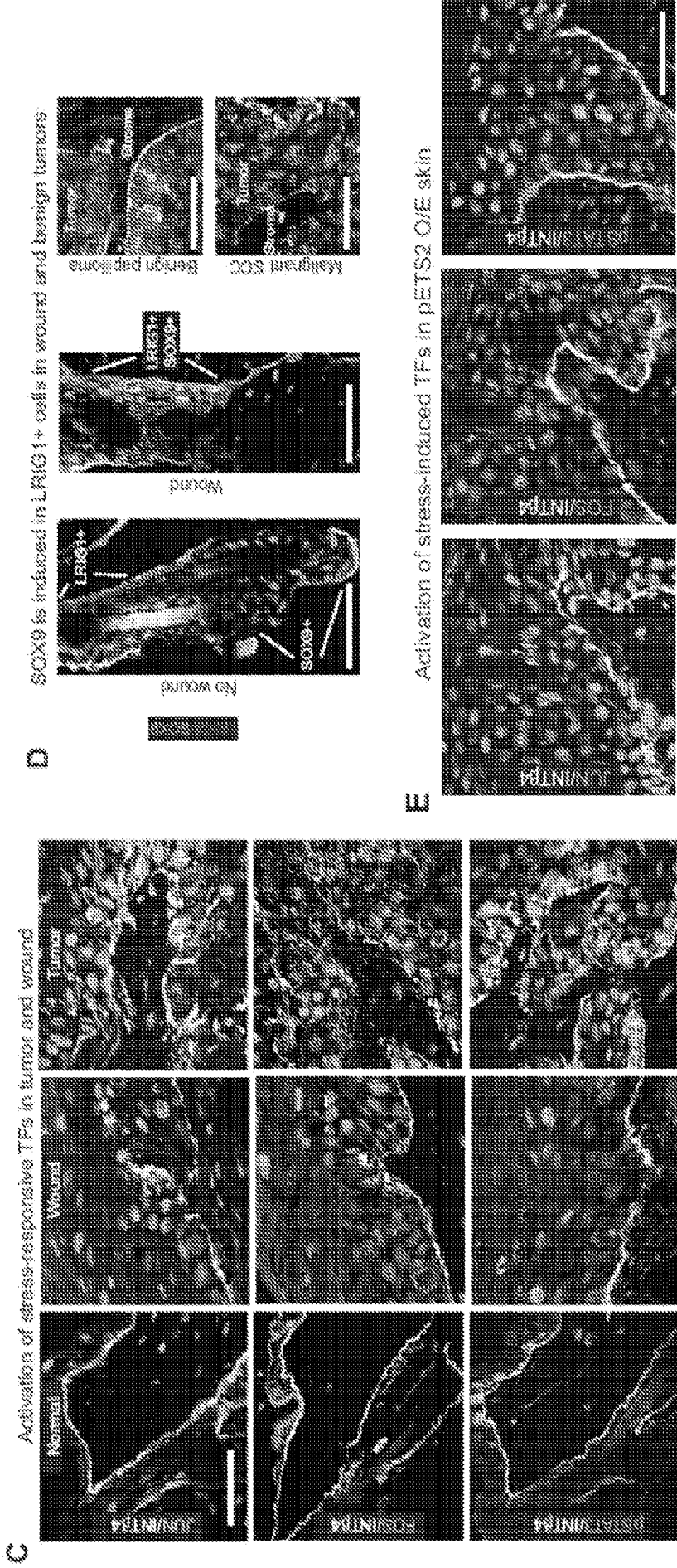


Figure 13C -13E

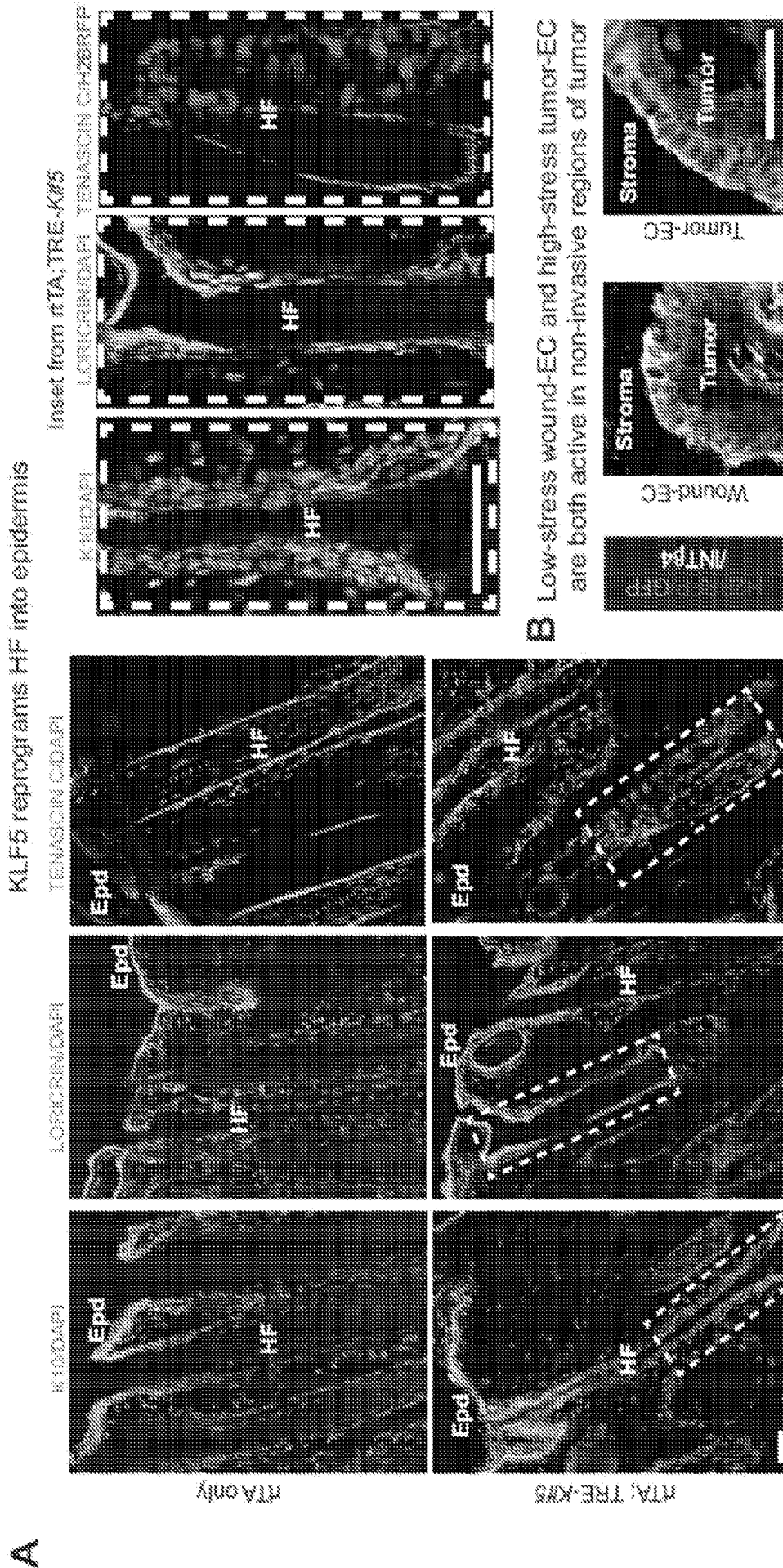
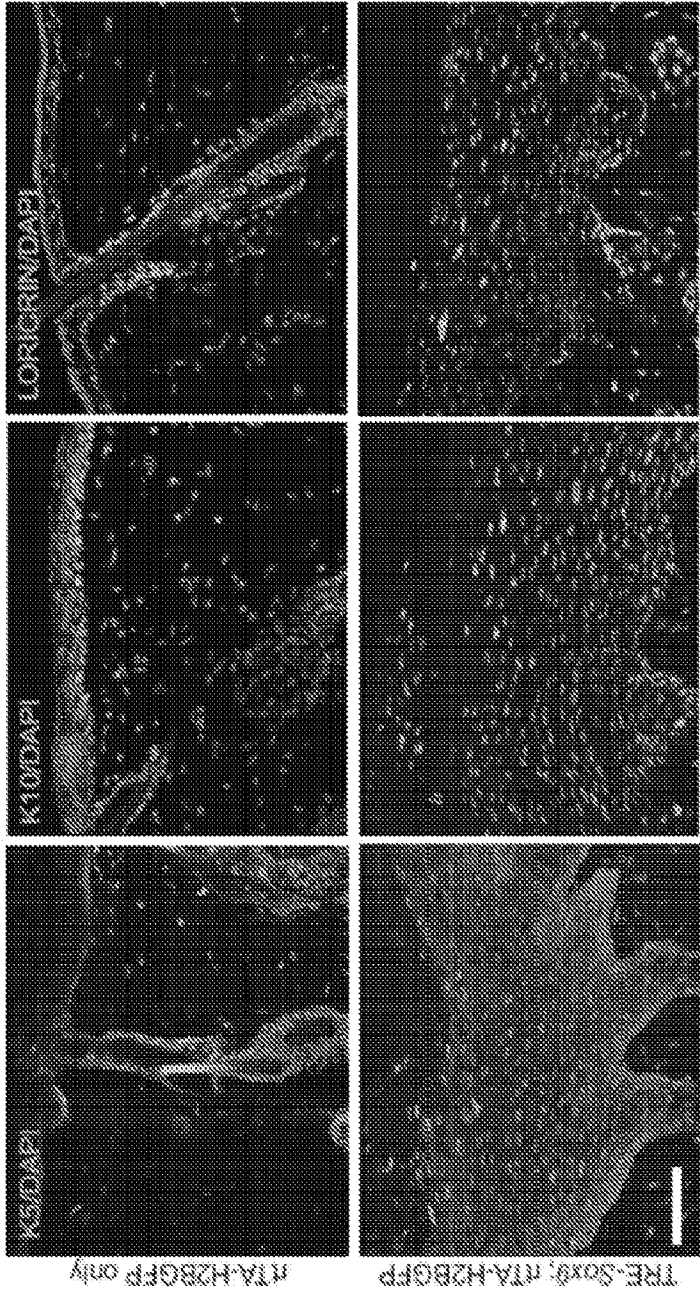


Figure 14A and Figure 14B

C

Wounds do not heal when SOX9 is sustained



D

Sustained SOX9 in wounding causes hyper-proliferation

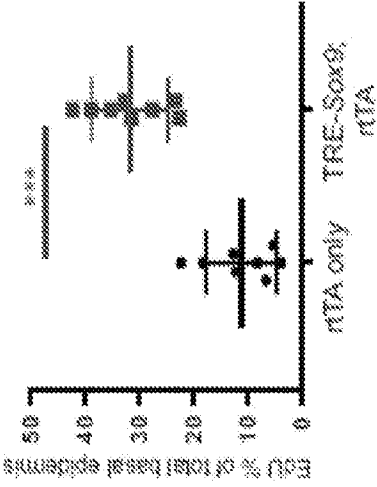


Figure 14C – 14D

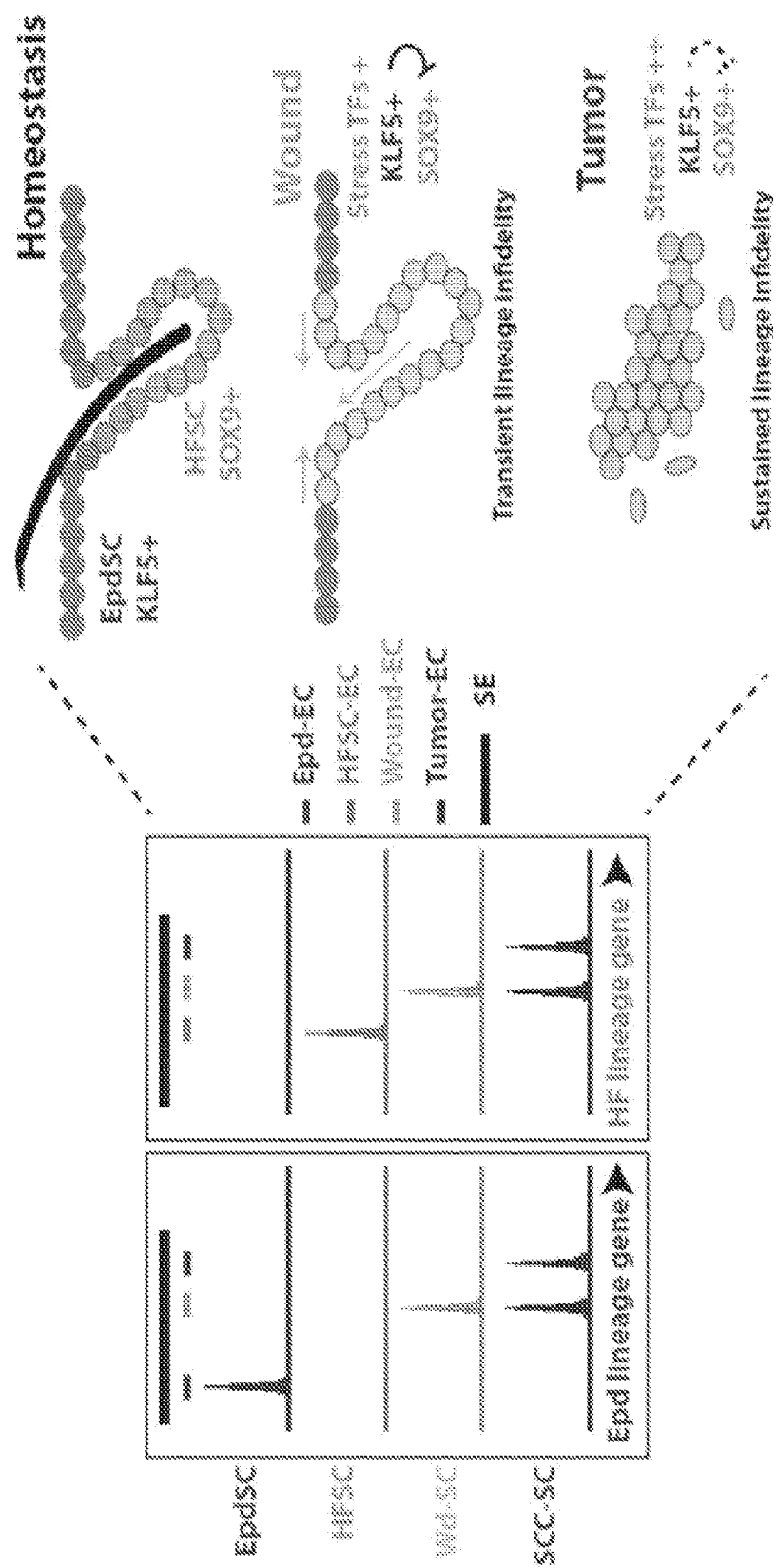


Figure 14E

E

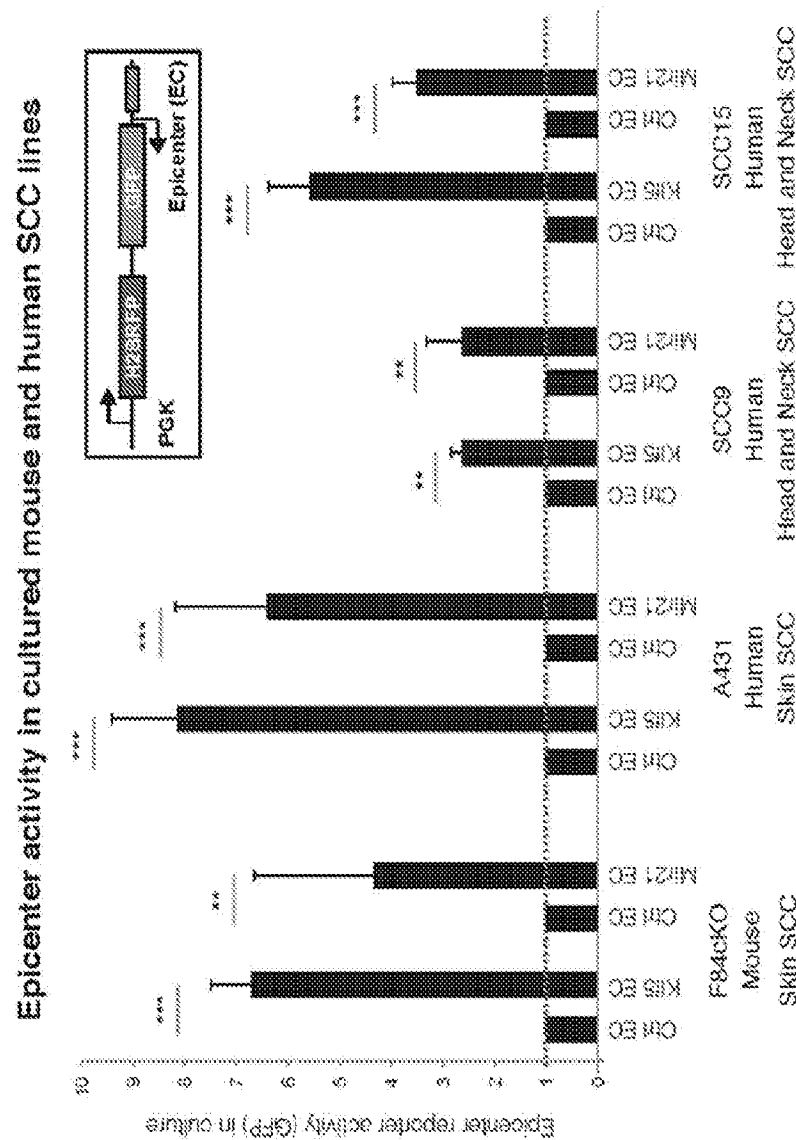


Figure 15

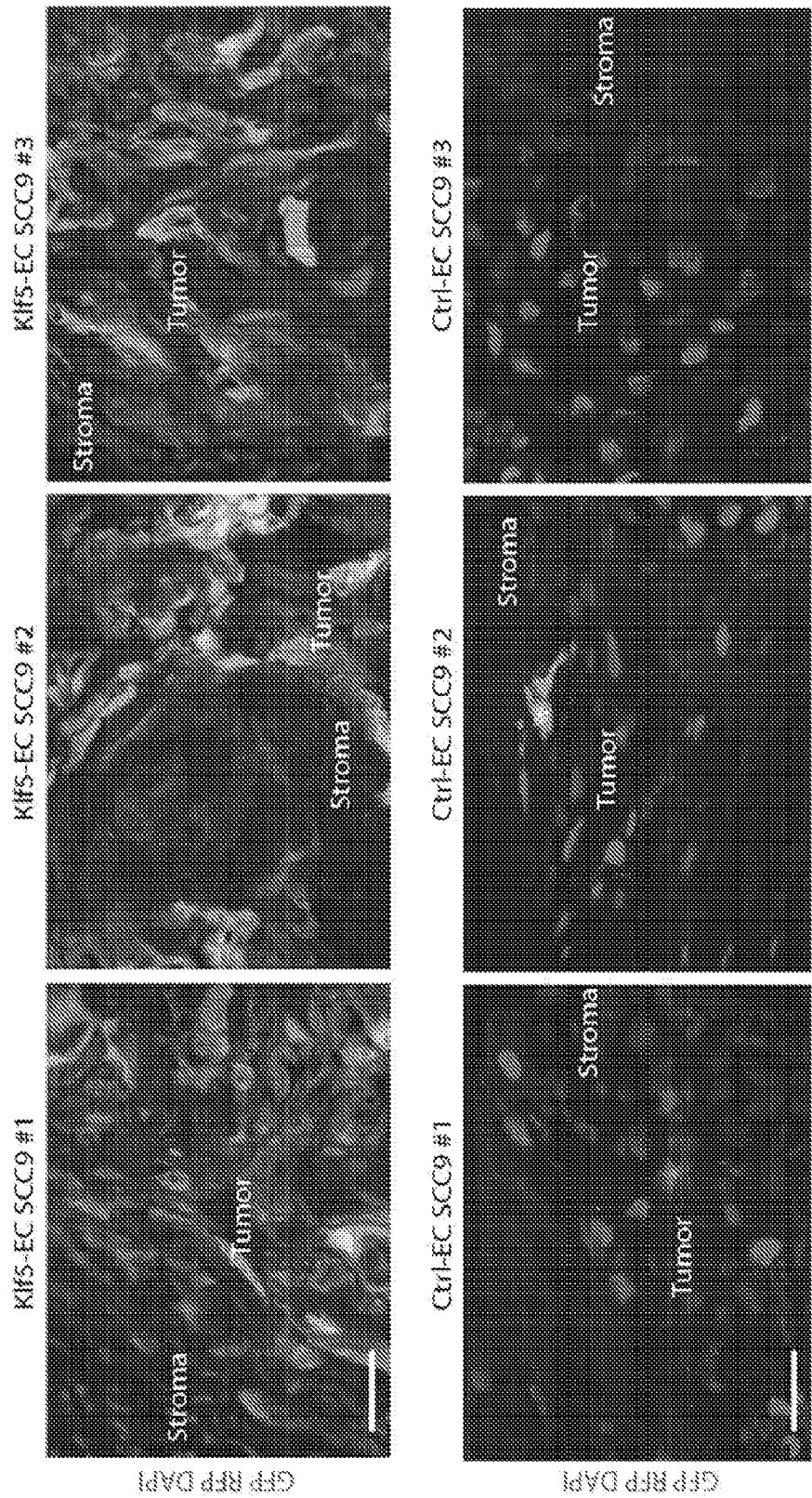


Figure 16

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/23683

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68 (2018.01)

CPC - C12Q 1/6886, C12Q 1/6837, C12Q 2600/106, C12Q 2600/158

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y ✓	ADAM et al., Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature, 21 May 2015, Vol. 521, pp 366-370. Abstract; p2, col 2, para 2; p5, col 1, para 3-para 5	1-22, 43, 44
Y ✓	LOVEN et al., Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell. 11 April 2013, Vol. 153, No. 2, pp 320-334. Abstract; p325, col 2, para 3; p331, col 2, para 1	1-22
Y	US 2011/0269691 A1 (HEARING et al.) 3 November 2011 (3.11.2011) Abstract; para [0031]	10, 11, 21, 22
Y	US 2010/0028364 A1 (BROWN et al.) 4 February 2010 (4.02.2010) Abstract; Claim 19; para [0011]	43, 44



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

5 June 2018

Date of mailing of the international search report

01 AUG 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/23683

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-22, 43, 44, drawn to a composition/method for treating a disorder.

Group II, claims 23-42, 45-51, drawn to a method/composition of diagnosing cancers, identifying cancer biomarkers, detecting transcription epicenters, and screening modulators.

----Please see continuation in first extra sheet -----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-22, 43, 44

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Continuation of Box No. III. Observations where unity of invention is lacking.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features

Group I includes the special technical feature of administering to subject in need thereof a modulator of an epicenter, not required by Group II.

Group II includes the special technical feature of obtaining a sample from the subject, detecting epicenter activity and identifying a cancer marker, not required by Group I.

Common Technical Features

The inventions of Groups I and II share the technical feature of a modulator of a transcription epicenter, wherein the epicenter regulates cancer.

However, these shared technical features do not represent a contribution over prior art in view of the article entitled 'Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice' by Adam et al., (hereinafter 'Adam') (Nature. 21 May 2015, Vol 521, No 7552, pp 366-370) and the article entitled 'Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers' by Loven et al., (hereinafter 'Loven') (Cell. 11 April 2013, Vol 153, No 2, pp 320-334).

Adam teaches (instant claim 1) a composition comprising a modulator of a transcription epicenter (Abstract - 'Here we show that super-enhancers underlie the identity, lineage commitment and plasticity of adult stem cells in vivo....We show that super-enhancers and their dense clusters ("epicentres") of transcription factor binding sites undergo remodelling upon lineage progression...Finally, we identify SOX9 as a crucial chromatin rheostat of hair follicle stem cell super-enhancers, and provide functional evidence that super-enhancers are dynamic, dense transcription-factor binding platforms which are acutely sensitive to pioneer master regulators whose levels define not only spatial and temporal features of lineage-status but also stemness, plasticity in transitional states and differentiation.'). Adam does not specifically teach a composition comprising a modulator of a transcription epicenter, wherein the epicenter regulates cancer. Loven teaches a composition for treating cancer, wherein the composition comprises a modulator of super-enhancers (Abstract - 'Here, we investigate how inhibition of the widely expressed transcriptional coactivator BRD4 leads to selective inhibition of the MYC oncogene in multiple myeloma (MM). BRD4 and Mediator were found to co-occupy thousands of enhancers associated with active genes. They also co-occupied a small set of exceptionally large super-enhancers associated with genes that feature prominently in MM biology, including the MYC oncogene. Treatment of MM tumor cells with the BET-bromodomain inhibitor JQ1 led to preferential loss of BRD4 at super-enhancers and consequent transcription elongation defects that preferentially impacted genes with super-enhancers, including MYC. Super enhancers were found at key oncogenic drivers in many other tumor cells. These observations have implications for the discovery of cancer therapeutics directed at components of super-enhancers in diverse tumor types.'). It would have been obvious to one of ordinary skill in the art to have recognized that the super-enhancers of Loven could be located in the dense clusters, epicentres, as taught by Adam. Given that modulator of super-enhancers can be used to treat cancer, as taught by Loven, it would have been obvious to one of ordinary skill in the art to have determined whether a modulator of transcription epicenter can be used to treat cancer, and thus to have arrived at the claimed limitation.

As said technical features were known in the art at the time of the invention, these cannot be considered special technical features that would otherwise unify the groups.

Groups I and II therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.