(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 January 2017 (26.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/015637 A1

(51) International Patent Classification:
A61K 48/00 (2006.01) *C12N 15/63* (2006.01)
C07K 19/00 (2006.01) *C12N 15/65* (2006.01)
C12N 5/22 (2006.01) *C12N 15/867* (2006.01)

(21) International Application Number:
PCT/US2016/043756

(22) International Filing Date:
22 July 2016 (22.07.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/195,680 22 July 2015 (22.07.2015) US
62/293,313 9 February 2016 (09.02.2016) US

(71) Applicant: **DUKE UNIVERSITY** [US/US]; 2812 Erwin Road, Suite 306, Durham, NC 27705 (US).

(72) Inventors: **GERSBACH, Charles, A.**; 3220 Wilderness Road, Durham, NC 27712 (US). **CRAWFORD, Gregory, E.**; 208 Maywood Way, Chapel Hill, NC 27516 (US). **REDDY, Timothy, E.**; 242 Sweet Bay Place, Carrboro, NC 27510 (US). **KLANN, Tyler, S.**; 3759 SW Durham Drive, Apartment 205, Durham, NC 27707 (US).

(74) Agent: **YEH, Sansun**; Michael Best & Friedrich LLP, 100 East Wisconsin Avenue, Suite 3300, Milwaukee, WI 53202-4108 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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))*

(54) Title: HIGH-THROUGHPUT SCREENING OF REGULATORY ELEMENT FUNCTION WITH EPIGENOME EDITING TECHNOLOGIES

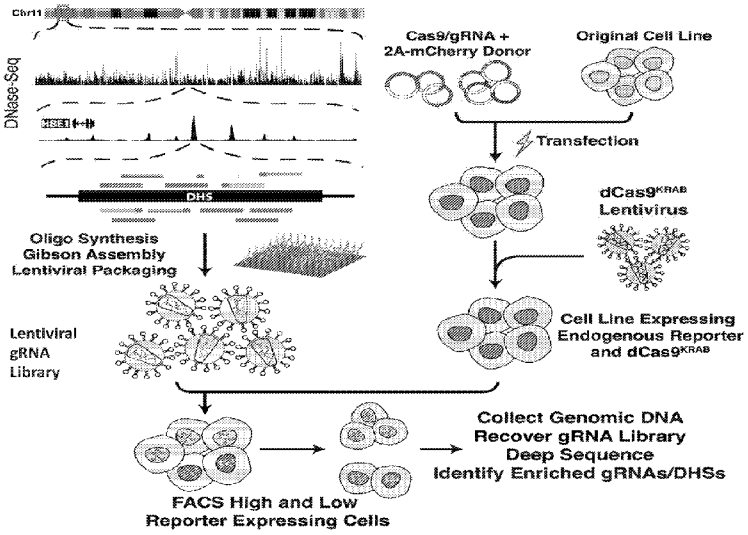


FIG. 1

(57) Abstract: Disclosed herein are methods of using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas) 9-based epigenomic editing systems for high-throughput screening of regulatory element function.

WO 2017/015637 A1

HIGH-THROUGHPUT SCREENING OF REGULATORY ELEMENT FUNCTION WITH EPIGENOME EDITING TECHNOLOGIES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 62/293,313, filed February 9, 2016, and U.S. Provisional Application No. 62/195,680, filed July 22, 2015, which are incorporated herein by reference in their entirety.

STATEMENT OF GOVERNMENT INTEREST

[0002] This invention was made with Government support under Federal Grant Nos. R01DA036865 and 1DP2OD008586 awarded by the National Institutes of Health. The Government has certain rights to this invention.

TECHNICAL FIELD

[0003] The present disclosure is directed to methods of using CRISPR/Cas9-based epigenomic editing systems in high-throughput screening of regulatory element function.

BACKGROUND

[0004] The Human Genome Project was funded and pursued based on the premise that the sequencing of the human genome would reveal the genetic basis for complex diseases that have a strong inheritable component, including cardiovascular disease, neurodegenerative conditions, and metabolic diseases such as diabetes. It was believed that this information would lead to new drug targets for these widespread diseases. However, thousands of genome-wide association studies (GWAS) have shown that the genetic variation associated with these complex diseases does not occur within genes, but rather in intergenic regulatory regions that control the levels of particular genes. Similarly, approximately 20% of Mendelian disorders do not have a detectable coding mutation, suggesting that the causal mutation is in a gene regulatory element. It is very difficult to assign functional roles to these regulatory elements as they often are located in distant locations from their target genes.

[0005] The human genome encodes approximately 50,000 genes. Understanding how those genes are regulated and how this correlates to complex cell phenotypes is a focus. Many genes

and regulatory elements fall into each positive hit of each GWAS study, and the actual target gene(s) that causes disease may fall outside of the regions identified by GWAS studies. Follow-up projects to the Human Genome Project, such as the NIH-funded Encyclopedia of DNA Elements (ENCODE) and the Roadmap Epigenomics Project, have identified millions of putative regulatory elements across the human genome for many human cell types and tissues. These regulatory elements determine the gene expression patterns responsible for complex cell phenotypes including cell differentiation, tissue specificity, oncogenesis, immunomodulation, and disease. However, the function of these regulatory elements and their relationships to these phenotypes are largely unknown. Additionally, conventional screening tools for perturbing cellular processes, such as small molecules and RNA interference, cannot directly target genomic regulatory elements.

[0006] Conventional screening technologies include small molecule screens that inhibit protein function and RNA interference screens that block protein translation. Although successful in many cases, these screening technologies have also been plagued by confounding off-target effects. Furthermore, as described above, it has been discovered that gene regulatory elements play a critical role in determining cell phenotype, susceptibility to various diseases and disorders, and response to drug treatment. Conventional screening technologies are unable to directly probe the function of gene regulatory elements. There remains a need for the ability to target direct manipulation of epigenetic properties.

SUMMARY

[0007] The present invention is directed a method of high-throughput screening for one or more putative gene regulatory elements in a genome that modulate a phenotype. The method includes a) contacting a plurality of modified target cells with a library of single guide RNAs (sgRNAs) that target a plurality of gene regulatory elements within the genome, thereby generating a plurality of test cells, b) selecting a population of test cells or an organism having a modulated phenotype; c) quantitating the frequency of the sgRNAs within the population of selected cells or the organism, wherein the sgRNAs that target gene regulatory elements that modulate the phenotype are overrepresented or underrepresented in the selected cells; and d) identifying and characterizing the sgRNAs within the population of selected test cells or the organism thereby identifying the gene regulatory elements that modulate the phenotype. The

modified target cell or organism includes a fusion protein, the fusion protein comprising a first polypeptide domain comprising a nuclease-deficient Cas9 and a second polypeptide domain having an activity selected from the group consisting of transcription activation activity, transcription repression activity, nuclease activity, transcription release factor activity, histone modification activity, nucleic acid association activity, DNA methylase activity, and direct or indirect DNA demethylase activity.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIG. 1 shows a schematic of CRISPR/Cas9-based Epigenetic Regulatory Element Screening (CERES).

[0009] FIGS. 2A-2C show generation of a reporter cell line. FIG. 2A shows the location of the control gRNAs for validating the reporter cell line. FIG. 2B shows the 2A-mCherry donor construct. FIG. 2C show the results of the reporter cell line validation.

[0010] FIGS. 3A-3C show the validation of a human K562 erythroid cell line engineered to express red fluorescence protein (RFP) from the HBE1 (epsilon globin) endogenous gene locus, as well as a CRISPR/dCas9^{KRAB} repressor. FIG. 3A shows that following lentiviral transduction of a library of 10,739 gRNAs targeting all DNase I hypersensitive sites in a 5 megabase window around HBE1, the fraction of RFP-negative cells increased from 6% to 20%. These RFP-negative cells were isolated by fluorescence-activated cell sorting, and the gRNAs within in these cells were recovered by PCR from the lentiviral vector. FIGS. 3B and 3C show that high-throughput sequencing of the sgRNA library showed enrichment of sequences targeting known regulatory elements at the HBE1 promoter (FIG. 2B) and the HS1 and HS2 enhancers (FIG. 2C), demonstrating successful identification of specific known regulator elements from this complex library.

[0011] FIG. 4 shows a Manhattan plot showing the results of a high-throughput screen for regulatory elements in the 4.5 Mb surrounding the globin locus using the dCas9^{KRAB} repressor. Every DHS in this region was targeted by up to 50 gRNAs each. Enriched gRNAs following selection for decreased HBE expression were found only in the well-known promoter and enhancers (HS2-4) of this gene.

[0012] FIGS. 5A-5D show HBE1 dCas9^{KRAB} screening. FIG. 5A shows the enrichment and location of the pooled gRNA library targeting DHSs (indicated by DNase I susceptibility) in

K562 cells, both the HBE1 full library and β-globin locus. FIG. 5B shows representative flow cytometry data of dCas9^{KRAB} repressing a mCherry reporter knocked into the endogenous HBE1 locus in K562 cells. FIG. 5C shows a Manhattan plot showing the results of a high-throughput screen for regulatory elements in the 4.5 Mb surrounding the globin locus (HBE1 gene) of 10,739 gRNAs targeting 281 DHSs (limit 50 gRNAs per DHS) using the dCas9 ^{KRAB} repressor. FIG. 5D shows that enriched DHSs following selection for decreased HBE expression were found only in the well-known HBE promoter and enhancers (HS1-4).

[0013] FIGS. 6A-6C show HBE1 screen gRNA enrichment. FIG. 6A shows a Manhattan plot showing enrichment of individual gRNAs across the entire library in the 4.5 Mb region surrounding HBE1. FIG. 6B shows a closer view of the globin locus control region (HS1-5). FIG. 6C shows enriched gRNAs within the HS2 enhancer.

[0014] FIGS. 7A-7D shows the screening of HER2 enhancers with HER2 dCas9^{KRAB}. FIG. 7A shows flow cytometry data of dCas9^{KRAB} repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 7B shows a Manhattan plot showing the results of a screen for regulatory elements in the 4 Mb surrounding the HER2 gene using the dCas9^{KRAB} repressor. FIG. 7C shows a zoomed view of the HER2 region. FIG. 7D shows a Manhattan plot showing the results of a high-throughput screen for regulatory elements in the 4 Mb surrounding the globin locus (HBE1 gene) of 12,189 gRNAs targeting 433 DHSs (limit 30 gRNAs per DHS) using the dCas9 ^{KRAB} repressor.

[0015] FIGS. 8A-8C show HER2 screen gRNA enrichment. FIG. 8A shows a Manhattan plot showing enrichment of individual gRNAs across the entire library surrounding the 4 Mb region surrounding HER2. FIG. 8B shows enrichment of gRNAs near HER2. FIG. 8C shows enriched gRNAs by the HER2 promoter.

[0016] FIGS. 9A-9D show screening HER2 enhancers with HER2 dCas9^{p300}. FIG. 9A shows flow cytometry data of dCas9^{p300} activating HER2 in 293Tcells measured via cell surface staining with a monoclonal antibody. FIG. 9B shows a Manhattan plot showing results of a screen for regulatory elements in the 4 Mb surrounding the HER2 gene using the dCas9^{p300} activator. FIG. 9C shows a zoomed view of the HER2 region. When comparing high and low expressing HER2 cell populations, gRNAs are enriched in the promoter and three intronic DHSs of HER2 as well as several nearby DHSs. FIG. 9D shows a closer view of the beginning of the

HER2 gene. Interestingly, the intronic DHS enriched in the dCas9^{KRAB} screen did not appear enriched in the p300 screen.

[0017] FIGS. 10A-10B show HER2 screen validation results. FIG. 10A shows HER2 mRNA fold change versus log2 fold change in the screen for dCas9^{KRAB} hits. FIG. 10B shows HER2 mean fluorescent intensity measured via antibody staining versus log2 fold change in the screen for dCas9^{KRAB} hits.

[0018] FIG. 10C shows HER2 mRNA fold change versus log2 fold change in the screen for dCas9^{p300} hits.

[0019] FIG. 10D shows HER2 mean fluorescent intensity measured via antibody staining us log2 fold change in the screen for dCas9^{p300} hits.

[0020] FIGS. 11A and 11B show the fold change in mean fluorescent intensity (FIG. 11A) and expression of HER2 mRNA (FIG. 11B) for individual hits (gRNAs 1549.1, 1553.1, 1553.2, 1543.1, 1544.1, 1561.1, 1769.1, 1592.1, 1856.1, 1562.1, 1826.1, 1559.2, 1560.1, 1782.1, 1671.1, 1556.1 ("1556.3"), 1562.2, 1612.1, and 1535.1) that were identified using the CRISPR/dCas9^{KRAB} repression system and combinations of gRNAs (i.e., the combinations of 1549.1 and 1553.1, 1549.1 and 1553.2, 1543.1 and 1544.1, 1553.1 and 1562.1, 1553.2 and 1562.1, 1562.1 and 1562.1, 1553.2, 1543.1, 1544.1, 1561.1, and 1562.1, 1559.2 and 1553.1).

[0021] FIG. 11C shows flow cytometry data of dCas9^{KRAB} repressing HER2 in control A431 cells measured via cell surface staining with a monoclonal antibody (Monoclonal Mouse IgG2B Clone # 191924, R&D Systems).

[0022] FIGS. 12A-12C show enrichment of gRNA 1549.1 ("1549"). FIG. 12A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1549. FIG. 12B shows flow cytometry data of dCas9^{KRAB} repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 12C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0023] FIGS. 13A-13F show enrichment of gRNA 1553.1 and 1553.2. FIGS. 13A and 13D show the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1553.1 (FIG. 13A) and 1553.2 (FIG. 13D). FIGS. 13B and 13E show flow cytometry data of dCas9^{KRAB} repressing

HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIGS. 13C and 13F show normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0024] FIGS. 14A-14C show enrichment of gRNA 1543.1 ("1543"). FIG. 14A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1543. FIG. 14B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 14C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0025] FIGS. 15A-15C show enrichment of gRNA 1544.1 ("1544"). FIG. 15A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1544. FIG. 15B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 15C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0026] FIGS. 16A-16C show enrichment of gRNA 1561.1 ("1561"). FIG. 16A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1561. FIG. 16B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 16C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0027] FIGS. 17A-17C show enrichment of gRNA 1769.1 ("1769"). FIG. 17A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1769. FIG. 17B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 17C shows normalized counts for all gRNAs per DHS (left panel)

and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0028] FIGS. 18A-18C show enrichment of gRNA 1592.1 ("1592"). FIG. 18A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1592. FIG. 18B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 18C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0029] FIGS. 19A-19C show enrichment of gRNA 1856.1 ("1856"). FIG. 19A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1856. FIG. 19B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 19C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0030] FIGS. 20A-20C show enrichment of gRNA 1562.1 ("1562"). FIG. 20A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1562. FIG. 20B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 20C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0031] FIGS. 21A-21C show enrichment of gRNA 1826.1 ("1826"). FIG. 21A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1826. FIG. 21B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 21C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0032] FIGS. 22A-22C show enrichment of gRNA 1559.1 ("1559"). FIG. 22A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1559. FIG. 22B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 22C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0033] FIGS. 23A-23C show enrichment of gRNA 1560.1 ("1560"). FIG. 23A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1560. FIG. 23B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 23C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0034] FIGS. 24A-24C show enrichment of gRNA 1782.1 ("1782"). FIG. 24A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1782. FIG. 24B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 24C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0035] FIGS. 25A-25C show enrichment of gRNA 1671.1 ("1671"). FIG. 25A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1671. FIG. 25B shows flow cytometry data of $dCas9^{KRAB}$ repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 25C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0036] FIGS. 26A-26C show enrichment of gRNA 1556.1 ("1556"). FIG. 26A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1556. FIG. 26B shows flow cytometry

data of dCas9^{KRAB} repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 26C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0037] FIGS. 27A-27C show enrichment of gRNA 1612.1 ("1612"). FIG. 27A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1612. FIG. 27B shows flow cytometry data of dCas9^{KRAB} repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 27C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0038] FIGS. 28A-28C show enrichment of gRNA 1535.1 ("1535"). FIG. 8A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of gRNA 1535. FIG. 28B shows flow cytometry data of dCas9^{KRAB} repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 28C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0039] FIG. 29 shows the fold change in expression of HER2 mRNA individual hits (gRNAs 1549, 1550, 1551, 1552, 1548, 1544, 1542, 1530, 1531, 1561, and 1563) that were identified using the CRISPR/dCas9$^{p300\ Core}$ activation system compared with Control and GFP control.

[0040] FIGS. 30A-30C show the enrichment of gRNA 1549. FIG. 30A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1549. FIG. 30B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 30C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0041] FIGS. 31A-31C show the enrichment of gRNA 1550. FIG. 31A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the

vertical highlighted band indicates the location of gRNA 1550. FIG. 31B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 31C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0042] FIGS. 32A-32C show the enrichment of gRNA 1551. FIG. 32A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1551. FIG. 32B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 32C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0043] FIGS. 33A-33C show the enrichment of gRNA 1552. FIG. 33A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1552. FIG. 33B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 33C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0044] FIGS. 34A-34C show the enrichment of gRNA 1548. FIG. 34A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1548. FIG. 34B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 34C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0045] FIGS. 35A-35C show the enrichment of gRNA 1544. FIG. 35A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 154. FIG. 35B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 35C shows normalized counts for all gRNAs per

DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0046] FIGS. 36A-36C show the enrichment of gRNA 1542. FIG. 36A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1542. FIG. 36B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 36C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0047] FIGS. 37A-37C show the enrichment of gRNA 1530. FIG. 37A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1530. FIG. 37B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 37C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0048] FIGS. 38A-38C show the enrichment of gRNA 1531. FIG. 38A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1531. FIG. 38B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 38C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0049] FIGS. 39A-39C show the enrichment of gRNA 1561. FIG. 39A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1516. FIG. 39B shows the flow cytometry data of dCas9^{p300} activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 39C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0050] FIGS. 40A-40C show the enrichment of gRNA 1563. FIG. 40A shows the frequency and location of the pooled gRNA library targeting DHSs in HEK293T cells, where the vertical highlighted band indicates the location of gRNA 1563. FIG. 40B shows the flow cytometry data of $dCas9^{p300}$ activating HER2 in HEK293T cells measured via cell surface staining with a monoclonal antibody. FIG. 40C shows normalized counts for all gRNAs per DHS (left panel) and individual gRNA used (right panel) where "0" is input (no Cas9 Effector), "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0051] FIG. 41A shows HBE1 mRNA fold change versus log2 fold change in the screen for $dCas9^{KRAB}$ hits. FIG. 41B shows HBE1 mean fluorescent intensity (MFI) versus log2 fold change in the screen for $dCas9^{KRAB}$ hits.

[0052] FIGS. 42A and 42B show mean fluorescent intensity fold change (FIG. 42A) and mRNA fold change (FIG. 42B) for sgRNAs targeting HBG1/2 promoter, HBE1 promoter, HS1, HS2, HS3, and HS4.

[0053] FIGS. 43A-43C show enrichment of gRNA targeting HBG1/2 promoter (gRNA 783_784). FIG. 43A shows the frequency and location of the pooled gRNA library targeting DHSs in K562 cells, where the vertical highlighted band indicates the location of gRNA 783_784. FIG. 43B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in K562 cells. FIG. 43C shows normalized counts for all conditions: HBG1 promoter, HBG1 gRNA, HBG2 promoter, and HBG2 gRNA, where "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0054] FIGS. 44A-44C show enrichment of gRNA targeting HBE1 promoter (gRNA 787). FIG. 44A shows the frequency and location of the pooled gRNA library targeting DHSs in K562 cells, where the vertical highlighted band indicates the location of gRNA 787. FIG. 44B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in K562 cells. FIG. 44C shows normalized counts for all conditions: HBE1 promoter and HBE1 gRNA, where "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0055] FIGS. 45A-45C show enrichment of gRNA targeting HS1 (gRNA 788). FIG. 45A shows the frequency and location of the pooled gRNA library targeting DHSs in K562 cells, where the vertical highlighted band indicates the location of gRNA 788. FIG. 45B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in K562 cells. FIG. 45C shows normalized

counts for all conditions: HS1 and HS1 gRNA, where "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0056] FIGS. 46A-46C show enrichment of gRNA targeting HS2 (gRNA 790). FIG. 46A shows the frequency and location of the pooled gRNA library targeting DHSs in K562 cells, where the vertical highlighted band indicates the location of gRNA 790. FIG. 46B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in K562 cells. FIG. 46C shows normalized counts for all conditions: HS2 and HS2 gRNA,, where "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0057] FIGS. 47A-47C show enrichment of gRNA targeting HS3 (gRNA 791). FIG. 47A shows the frequency and location of the pooled gRNA library targeting DHSs in K562 cells, where the vertical highlighted band indicates the location of gRNA 791. FIG. 47B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in K562 cells. FIG. 47C shows normalized counts for all conditions: HS3 and HS3 gRNA,, where "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0058] FIGS. 48A-48C show enrichment of gRNA targeting HS4 (gRNA 792). FIG. 48A shows the frequency and location of the pooled gRNA library targeting DHSs in K562 cells, where the vertical highlighted band indicates the location of gRNA 792. FIG. 48B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in K562 cells. FIG. 48C shows normalized counts for all conditions: HS4 and HS4 gRNA,, where "1" is unsorted bulk population, "2" is lowest 10% sorted, and "3" is highest 10% sorted.

[0059] FIGS. 49A-49B show enrichment of combinations of 1543.1 and 1544.1 gRNAs identified in HER2 screen. FIG. 49A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 49B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1543.1 and 1544.1 gRNAs alone and in combination.

[0060] FIGS. 50A-50B show enrichment of combinations of 1549.1 and 1553.1 gRNAs identified in HER2 screen. FIG. 50A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 50B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1549.1 and 1553.1 gRNAs alone and in combination.

[0061] FIGS. 51A-51B show enrichment of combinations of 1549.1 and 1553.2 gRNAs identified in HER2 screen. FIG. 51A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 51B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1549.1 and 1553.2 gRNAs alone and in combination.

[0062] FIGS. 52A-52B show enrichment of combinations of 1553.1 and 1561.1 gRNAs identified in HER2 screen. FIG. 52A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 52B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1553.1 and 1561.1 gRNAs alone and in combination.

[0063] FIGS. 53A-53C show enrichment of combinations of 1553.2 and 1562.1 gRNAs identified in HER2 screen. FIG. 53A and FIG. 53B show the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 53C shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1553.2 and 1562.1 gRNAs alone and in combination.

[0064] FIGS. 54A-54B show enrichment of combinations of gRNAs identified in HER2 screen. FIG. 54A shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1553.2, 1543.1, and 1544.1 gRNAs alone. FIG. 54B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1561.1 and 1562.1 gRNAs alone and the combination of 1553.2, 1543.1, 1544.1, 1561.1, and 1562.1.

[0065] FIGS. 55A-55B show enrichment of combinations of 1556.1 and 1559.2 gRNAs identified in HER2 screen. FIG. 55A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 55B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1556.1 and 1559.2 gRNAs alone and in combination.

[0066] FIGS. 56A-56B show enrichment of combinations of 1559.2 and 1560.1 gRNAs identified in HER2 screen. FIG. 56A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 56B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1559.2 and 1560.1 gRNAs alone and in combination.

[0067] FIGS. 57A-57B show enrichment of combinations of 1561.1 and 1562.1 gRNAs identified in HER2 screen. FIG. 57A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 57B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1561.1 and 1562.1 gRNAs alone and in combination.

[0068] FIGS. 58A-58B show enrichment of combinations of 1782.1 and 1671.1 gRNAs identified in HER2 screen. FIG. 58A shows the frequency and location of the pooled gRNA library targeting DHSs in A431 cells, where the vertical highlighted band indicates the location of the gRNAs. FIG. 58B shows flow cytometry data of $dCas9^{KRAB}$ repressing HBE1 in A431 cells for 1782.1 and 1671.1 gRNAs alone and in combination.

[0069] FIG. 59 shows percent of cells having 0, 1, or more than 1 copy of sgRNA.

DETAILED DESCRIPTION

[0070] Disclosed herein are methods of high throughput screening for regulatory element function using CRISPR/Cas9-based epigenomic editing systems. The CRISPR/Cas9-based epigenomic editing systems directly activate or repress genomic regulatory elements in their natural chromosomal location. In particular, the CRISPR/Cas9-based epigenomic editing system includes a Cas9 fusion protein of a Cas9 protein that does not have nuclease activity and a protein having histone acetyltransferase activity, such as the catalytic histone acetyltransferase (HAT) core domain of the human E1A-associated protein p300, or repression activity, such as KRAB. Recruitment of the acetyltransferase function or repression function by dCas9 and a gRNA to the genomic target site allow direct modulation of epigenetic structure, and thus provide an effective means of epigenomic editing. The methods described herein use these tools for high-throughput identification of gene regulatory element activity. These methods allow the mapping and characterization of regulatory elements across the genome.

[0071] The disclosed methods can be used for high-throughput identification of the regulatory elements responsible for the expression level of any gene in any organism, such as any animal. For example, the disclosed methods can be used to identify how genetic variation in the human population is responsible for inherited risk for any genetically linked disease, including most cardiovascular, neurodegenerative, and metabolic disorders, or other traits such as intelligence, size, and regenerative potential. As a result, the technology can be used to

identify drug targets for a vast array of disorders and conditions. Such screening of gene regulatory elements in their natural chromosomal location is not currently possible with any other technology.

1. Definitions

[0072] The terms "comprise(s)," "include(s)," "having," "has," "can," "contain(s)," and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. The singular forms "a," "and" and "the" include plural references unless the context clearly dictates otherwise. The present disclosure also contemplates other embodiments "comprising," "consisting of" and "consisting essentially of," the embodiments or elements presented herein, whether explicitly set forth or not.

[0073] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[0074] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present invention. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

[0075] "Cell phenotype" as used herein refers to the observable characteristics or traits of a cell, such as its morphology, development, biochemical or physiological properties, phenology, or behavior. A phenotype results from the expression of the cell's genes as well as the influence of environmental factors and the interactions between the two.

[0076] "Chromatin" as used herein refers to an organized complex of chromosomal DNA associated with histones.

[0077] "Cis-regulatory elements" or "CREs" as used interchangeably herein refers to regions of non-coding DNA which regulate the transcription of nearby genes. CREs are found in

the vicinity of the gene, or genes, they regulate. CREs typically regulate gene transcription by functioning as binding sites for transcription factors. Examples of CREs include promoters, enhancers, super-enhancers, silencers, insulators, and locus control regions.

[0078] "Clustered Regularly Interspaced Short Palindromic Repeats" and "CRISPRs", as used interchangeably herein refers to loci containing multiple short direct repeats that are found in the genomes of approximately 40% of sequenced bacteria and 90% of sequenced archaea.

[0079] "Coding sequence" or "encoding nucleic acid" as used herein means the nucleic acids (RNA or DNA molecule) that comprise a nucleotide sequence which encodes a protein. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is administered. The coding sequence may be codon optimize.

[0080] "Complement" or "complementary" as used herein means a nucleic acid can mean Watson-Crick (*e.g.*, A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. "Complementarity" refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

[0081] "Demethylases" as used herein refers to an enzyme that removes methy (CH3-) groups from nucleic acids, proteins (in particular histones), and other molecules. Demethylase enzymes are important in epigenetic modification mechanisms. The demethylase proteins alter transcriptional regulation of the genome by controlling the methylation levels that occur on DNA and histones and, in turn, regulate the chromatin state at specific gene loci within organisms. "Histone demethylase" refers to a methylase that removes methy groups from histones. There are several families of histone demethylases, which act on different substrates and play different roles in cellular function. The Fe(II)-dependent lysine demethylases may be a JMJC demethylase. A JMJC demethylase is a histone demethylase containing a JumonjiC (JmjC) domain. The JMJC demethylase may be a member of the KDM3, KDM4, KDM5, or KDM6 family of histone demethylases.

[0082] "DNase I hypersensitive sites" or "DHS" as used interchangeably herein refers to docking sites for the transcription factors and chromatin modifiers, including p300 that coordinate distal target gene expression.

[0083] "Endogenous gene" as used herein refers to a gene that originates from within an organism, tissue, or cell. An endogenous gene is native to a cell, which is in its normal genomic and chromatin context, and which is not heterologous to the cell. Such cellular genes include, e.g., animal genes, plant genes, bacterial genes, protozoal genes, fungal genes, mitochondrial genes, and chloroplastic genes. An "endogenous target gene" as used herein refers to an endogenous gene that is targeted by gRNA and CRISPR/Cas9-based epigenomic editing system.

[0084] "Enhancer" as used herein refers to non-coding DNA sequences containing multiple activator and repressor binding sites. Enhancers range from 50 bp to 1500 bp in length and may be either proximal, 5' upstream to the promoter, within any intron of the regulated gene, or distal, in introns of neighboring genes, or intergenic regions far away from the locus, or on regions on different chromosomes. More than one enhancer may interact with a promoter. Similarly, enhancers may regulate more than one gene without linkage restriction and may "skip" neighboring genes to regulate more distant ones. Transcriptional regulation may involve elements located in a chromosome different to one where the promoter resides. Proximal enhancers or promoters of neighboring genes may serve as platforms to recruit more distal elements.

[0085] "Fusion protein" as used herein refers to a chimeric protein created through the joining of two or more genes that originally coded for separate proteins. The translation of the fusion gene results in a single polypeptide with functional properties derived from each of the original proteins.

[0086] "Genetic construct" as used herein refers to the DNA or RNA molecules that comprise a nucleotide sequence that encodes a protein. The coding sequence includes initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of the individual to whom the nucleic acid molecule is administered. As used herein, the term "expressible form" refers to gene constructs that contain the necessary regulatory elements operable linked to a coding sequence that encodes a protein such that when present in the cell of the individual, the coding sequence will be expressed.

[0087] "Genome" as used herein refers to the complete set of genes or genetic material present in a cell or organism. The genome includes DNA or RNA in RNA viruses. The genome

includes both the genes, (the coding regions), the noncoding DNA and the genomes of the mitochondria and chloroplasts.

[0088] "Histone acetyltransferases" or "HATs" are used interchangeably herein refers to enzymes that acetylate conserved lysine amino acids on histone proteins by transferring an acetyl group from acetyl CoA to form ε-N-acetyllysine. DNA is wrapped around histones, and, by transferring an acetyl group to the histones, genes can be turned on and off. In general, histone acetylation increases gene expression as it is linked to transcriptional activation and associated with euchromatin. Histone acetyltransferases can also acetylate non-histone proteins, such as nuclear receptors and other transcription factors to facilitate gene expression.

[0089] "Histone deacetylases" or "HDACs" as used interchangeably herein refers to a class of enzymes that remove acetyl groups ($O=C-CH_3$) from an ε-N-acetyl lysine amino acid on a histone, allowing the histones to wrap the DNA more tightly. HDACs are also called lysine deacetylases (KDAC), to describe their function rather than their target, which also includes non-histone proteins.

[0090] "Histone methyltransferase" or "HMTs" as used interchangeably herein refers to histone-modifying enzymes (e.g., histone-lysine N-methyltransferases and histone-arginine N-methyltransferases), that catalyze the transfer of one, two, or three methyl groups tolysine and arginine residues of histone proteins. The attachment of methyl groups occurs predominantly at specific lysine or arginine residues on histones H3 and H4.

[0091] "Identical" or "identity" as used herein in the context of two or more nucleic acids or polypeptide sequences means that the sequences have a specified percentage of residues that are the same over a specified region. The percentage may be calculated by optimally aligning the two sequences, comparing the two sequences over the specified region, determining the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the specified region, and multiplying the result by 100 to yield the percentage of sequence identity. In cases where the two sequences are of different lengths or the alignment produces one or more staggered ends and the specified region of comparison includes only a single sequence, the residues of single sequence are included in the denominator but not the numerator of the calculation. When comparing DNA and RNA, thymine (T) and uracil (U) may be considered

equivalent. Identity may be performed manually or by using a computer sequence algorithm such as BLAST or BLAST 2.0.

[0092] "Insulators" as used herein refers to a genetic boundary element that blocks the interaction between enhancers and promoters. By residing between the enhancer and promoter, the insulator may inhibit their subsequent interactions. Insulators can determine the set of genes an enhancer can influence. Insulators are needed where two adjacent genes on a chromosome have very different transcription patterns and the inducing or repressing mechanisms of one does not interfere with the neighboring gene. Insulators have also been found to cluster at the boundaries of topological association domains (TADs) and may have a role in partitioning the genome into "chromosome neighborhoods" - genomic regions within which regulation occurs. Insulator activity is thought to occur primarily through the 3D structure of DNA mediated by proteins including CTCF. Insulators are likely to function through multiple mechanisms. Many enhancers form DNA loops that put them in close physical proximity to promoter regions during transcriptional activation. Insulators may promote the formation of DNA loops that prevent the promoter-enhancer loops from forming. Barrier insulators may prevent the spread of heterochromatin from a silenced gene to an actively transcribed gene.

[0093] "Locus control regions" as used herein refers to a long-range cis-regulatory element that enhances expression of linked genes at distal chromatin sites. It functions in a copy number-dependent manner and is tissue-specific, as seen in the selective expression of β-globin genes in erythroid cells. Expression levels of genes can be modified by the LCR and gene-proximal elements, such as promoters, enhancers, and silencers. The LCR functions by recruiting chromatin-modifying, coactivator, and transcription complexes. Its sequence is conserved in many vertebrates, and conservation of specific sites may suggest importance in function.

[0094] "Modulate" as used herein may mean any altering of activity, such as regulate, down regulate, upregulate, reduce, inhibit, increase, decrease, deactivate, or activate.

[0095] "Multiplicity of infection" or "MOI" as used interchangeably herein refers to a ratio of agents (e.g., phage or more generally virus, bacteria) to infection targets (e.g., cell), such as when referring to a group of cells inoculated with virus particles, the multiplicity of infection or MOI is the ratio of the number of virus particles, which can include the sgRNA, to the number of target cells present in a defined space.

[0096] "Nucleic acid" or "oligonucleotide" or "polynucleotide" as used herein means at least two nucleotides covalently linked together. The depiction of a single strand also defines the sequence of the complementary strand. Thus, a nucleic acid also encompasses the complementary strand of a depicted single strand. Many variants of a nucleic acid may be used for the same purpose as a given nucleic acid. Thus, a nucleic acid also encompasses substantially identical nucleic acids and complements thereof. A single strand provides a probe that may hybridize to a target sequence under stringent hybridization conditions. Thus, a nucleic acid also encompasses a probe that hybridizes under stringent hybridization conditions.

[0097] Nucleic acids may be single stranded or double stranded, or may contain portions of both double stranded and single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine and isoguanine. Nucleic acids may be obtained by chemical synthesis methods or by recombinant methods.

[0098] "Operably linked" as used herein means that expression of a gene is under the control of a promoter with which it is spatially connected. A promoter may be positioned 5' (upstream) or 3' (downstream) of a gene under its control. The distance between the promoter and a gene may be approximately the same as the distance between that promoter and the gene it controls in the gene from which the promoter is derived. As is known in the art, variation in this distance may be accommodated without loss of promoter function.

[0099] "Organismal phenotype" as used herein refers to refers to the observable characteristics or traits of an organism, such as its morphology, development, biochemical or physiological properties, phenology, or behavior. A phenotype results from the expression of the organism's genes as well as the influence of environmental factors and the interactions between the two.

[0100] "p300 protein," "EP300," or "E1A binding protein p300" as used interchangeably herein refers to the adenovirus E1A-associated cellular p300 transcriptional co-activator protein encoded by the *EP300* gene. p300 is a highly conserved acetyltransferase involved in a wide range of cellular processes. p300 functions as a histone acetyltransferase that regulates transcription via chromatin remodeling and is involved with the processes of cell proliferation and cell differentiation.

[0101] "Primary cell" as used herein refers to cells taken directly from living tissue (e.g. biopsy material). Primary cells can be established for growth in vitro. These cells have undergone very few population doublings and are therefore more representative of the main functional component of the tissue from which they are derived in comparison to continuous (tumor or artificially immortalized) cell lines thus representing a more representative model to the in vivo state. Primary cells may be taken from different species, such as mouse or humans.

[0102] "Promoter" as used herein means a synthetic or naturally-derived molecule which is capable of conferring, activating or enhancing expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs, or anywhere in the genome, from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, hormones, toxins, drugs, pathogens, metal ions, or inducing agents. Representative examples of promoters include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter and the CMV IE promoter.

[0103] "Protospacer adjacent motif" or "PAM" as used herein refers to a DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. PAM is a component of the invading virus or plasmid, but is not a component of the bacterial CRISPR locus. Cas9 will not successfully bind to or cleave the target DNA sequence if it is not followed by the PAM sequence. PAM is an essential targeting component (not found in bacterial genome) which distinguishes bacterial self from non-self DNA, thereby preventing the CRISPR locus from being targeted and destroyed by nuclease.

[0104] "Silencers" or "repressors" as used interchangeably herein refer to a DNA sequence capable of binding transcription regulation factors and preventing genes from being expressed as

proteins. A silencer is a sequence-specific element that induces a negative effect on the transcription of its particular gene. There are many positions in which a silencer element can be located in DNA. The most common position is found upstream of the target gene where it can help repress the transcription of the gene. This distance can vary greatly between approximately -20 bp to -2000 bp upstream of a gene. Certain silencers can be found downstream of a promoter located within the intron or exon of the gene itself. Silencers have also been found within the 3 prime untranslated region (3' UTR) of mRNA. There are two main types of silencers in DNA, which are the classical silencer element and the non-classical negative regulatory element (NRE). In classical silencers, the gene is actively repressed by the silencer element, mostly by interfering with general transcription factor (GTF) assembly. NREs passively repress the gene, usually by inhibiting other elements that are upstream of the gene.

[0105] "Super enhancer" as used herein refers to a region of the mammalian genome comprising multiple enhancers that is collectively bound by an array of transcription factor proteins to drive transcriptionof genes involved in cell identity. Super-enhancers are frequently identified near genes important for controlling and defining cell identity and can be used to quickly identify key nodes regulating cell identity. Enhancers have several quantifiable traits that have a range of values, and these traits are generally elevated at super-enhancers. Super-enhancers are bound by higher levels of transcription-regulating proteins and are associated with genes that are more highly expressed. Expression of genes associated with super-enhancers is particularly sensitive to perturbations, which may facilitate cell state transitions or explain sensitivity of super-enhancer—associated genes to small molecules that target transcription.

[0106] "Target enhancer" as used herein refers to enhancer that is targeted by a gRNA and CRISPR/Cas9-based epigenomic editing system. The target enhancer may be within the target region.

[0107] "Target gene" as used herein refers to any nucleotide sequence encoding a known or putative gene product. The target gene includes the regulatory regions, such as the promoter and enhancer regions, the transcribed regions, which include the coding regions, and other functional sequence regions.

[0108] "Target region" as used herein refers to a cis-regulatory region or a trans-regulatory region of a target gene to which the guide RNA is designed to recruit the CRISPR/Cas9-based

epigenomic editing system to modulate the epigenetic structure and allow the activation or repression of gene expression of the target gene.

[0109] "Target regulatory element" as used herein refers to a regulatory element that is targeted by a gRNA and CRISPR/Cas9-based epigenomic editing system. The target regulatory element may be within the target region.

[0110] "Transcribed region" as used herein refers to the region of DNA that is transcribed into single-stranded RNA molecule, known as messenger RNA, resulting in the transfer of genetic information from the DNA molecule to the messenger RNA. During transcription, RNA polymerase reads the template strand in the 3' to 5' direction and synthesizes the RNA from 5' to 3'. The mRNA sequence is complementary to the DNA strand.

[0111] "Transcriptional Start Site" or "TSS" as used interchangeably herein refers to the first nucleotide of a transcribed DNA sequence where RNA polymerase begins synthesizing the RNA transcript.

[0112] "Transgene" as used herein refers to a gene or genetic material containing a gene sequence that has been isolated from one organism and is introduced into a different organism. This non-native segment of DNA may retain the ability to produce RNA or protein in the transgenic organism, or it may alter the normal function of the transgenic organism's genetic code. The introduction of a transgene has the potential to change the phenotype of an organism.

[0113] "Trans-regulatory elements" as used herein refers to regions of non-coding DNA which regulate the transcription of genes distant from the gene from which they were transcribed. Trans-regulatory elements may be on the same or different chromosome from the target gene. Examples of trans-regulatory elements include enhancers, super-enhancers, silencers, insulators, and locus control regions.

[0114] "Variant" used herein with respect to a nucleic acid means (i) a portion or fragment of a referenced nucleotide sequence (including nucleotide sequences that have insertions or deletions as compared to the referenced nucleotide sequences); (ii) the complement of a referenced nucleotide sequence or portion thereof; (iii) a nucleic acid that is substantially identical to a referenced nucleic acid or the complement thereof; or (iv) a nucleic acid that hybridizes under stringent conditions to the referenced nucleic acid, complement thereof, or a sequences substantially identical thereto.

[0115] "Variant" with respect to a peptide or polypeptide that differs in amino acid sequence by the insertion, deletion, or conservative substitution of amino acids, but retain at least one biological activity. Variant may also mean a protein with an amino acid sequence that is substantially identical to a referenced protein with an amino acid sequence that retains at least one biological activity. A conservative substitution of an amino acid, *i.e.*, replacing an amino acid with a different amino acid of similar properties (*e.g.*, hydrophilicity, degree and distribution of charged regions) is recognized in the art as typically involving a minor change. These minor changes may be identified, in part, by considering the hydropathic index of amino acids, as understood in the art. Kyte *et al.*, *J. Mol. Biol.* 157:105-132 (1982). The hydropathic index of an amino acid is based on a consideration of its hydrophobicity and charge. It is known in the art that amino acids of similar hydropathic indexes may be substituted and still retain protein function. In one aspect, amino acids having hydropathic indexes of ±2 are substituted. The hydrophilicity of amino acids may also be used to reveal substitutions that would result in proteins retaining biological function. A consideration of the hydrophilicity of amino acids in the context of a peptide permits calculation of the greatest local average hydrophilicity of that peptide. Substitutions may be performed with amino acids having hydrophilicity values within ±2 of each other. Both the hydrophobicity index and the hydrophilicity value of amino acids are influenced by the particular side chain of that amino acid. Consistent with that observation, amino acid substitutions that are compatible with biological function are understood to depend on the relative similarity of the amino acids, and particularly the side chains of those amino acids, as revealed by the hydrophobicity, hydrophilicity, charge, size, and other properties.

[0116] "Vector" as used herein means a nucleic acid sequence containing an origin of replication. A vector may be a viral vector, bacteriophage, bacterial artificial chromosome or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be a self-replicating extrachromosomal vector, and preferably, is a DNA plasmid. For example, the vector may encode a CRISPR/Cas9-based epigenomic editing system having an amino acid sequence of SEQ ID NO: 2 or 3.

2. Methods of High Throughput Screening of Regulatory Element Function

[0117] The present disclosure is directed towards methods for high-throughput screening of regulatory element activity that involves the design and synthesis of libraries of gRNAs, using custom array synthesis, targeted to all active gene regulatory elements in a particular genomic

region (targeted screen), or the entire genome (genome-wide screen), and screening for gain or loss of function following co-delivery with an epigenomic editing system that is a targeted modifier of gene expression or epigenetic state, such as a CRISPR/dCas9-based epigenomic editing system,. The screening is based on cell phenotype, organismal phenotype, or expression of a specific target gene. Sequencing the gRNAs in cells enriched with a gain or loss of gene expression can then be used to reveal active regulatory elements. The library may be screened in vitro, e.g., in cells, or in vivo, such as mouse heart. By selecting the cells with modulated phenotype or gene expression, the regulatory elements are identified that are responsible for these characteristics directly in the cell type of interest.

[0118] The present disclosure is directed to a method of high-throughput screening for one or more putative gene regulatory elements in a genome that modulate a phenotype. The method comprises: a) contacting a plurality of modified target cells with a library of single guide RNAs (sgRNAs, also referred to herein as gRNA) that target a plurality of gene regulatory elements within the genome, thereby generating a plurality of test cells; b) selecting a single cell or population of test cells having a modulated phenotype; c) quantitating the frequency of the sgRNAs within the population of selected cells, wherein the sgRNAs that target gene regulatory elements that modulate the phenotype are overrepresented or underrepresented in the selected cells; and d) identifying and characterizing the sgRNAs within the population of selected test cells thereby identifying the gene regulatory elements that modulate the phenotype. The modified target cell comprises a fusion protein. The fusion protein comprises a first polypeptide domain comprising a nuclease-deficient Cas9 and a second polypeptide domain. The second polypeptide domain has an activity such as transcription activation activity, transcription repression activity, nuclease activity, transcription release factor activity, histone modification activity, nucleic acid association activity, methylase activity, and direct or indirect demethylase activity. Examples of fusion proteins are described in U.S. Patent Publication No. 20160201089 and International Patent Publication NO. 2014/197748, which are incorporated by reference herein in their entirely.

[0119] The disclosed methods can include screening for differential endogenous gene expression using targeted qRT-PCR or genome-wide gene expression by RNA-seq. Hits from these screens can be further validated and characterized by RNA-seq, deletion of the regulatory element, and/or introduction of variants or other modifications using CRISPR/Cas9, as described

below. In contrast to the development of CRISPR-based screens for genetic knockout or targeting of gene promoter regions, the disclosed method is specifically focused on targeting regulatory elements and is uniquely enabled by the development of tools that alter the activity of these elements.

[0120] The disclosed method involves the design and synthesis of libraries of gRNAs targeted to all candidate gene regulatory elements in a genomic region or whole or entire genome, such as those defined by DNase-seq, ATAC-seq, FAIRE-seq, or ChIP-seq. Lentiviral vectors encoding the gRNA library can be used to deliver the gRNA library to cell lines expressing a CRISPR/Cas9-based epigenomic editing system, such as $dCas9^{p300}$ for activation or $dCas9^{KRAB}$ for repression. The modified target cells can be screened for gain or loss of gene expression. Sequencing the gRNAs in the selected cell subpopulations then reveals active regulatory elements

3. CRISPR/Cas9-based Epigenomic Editing System

[0121] Provided herein are CRISPR/Cas9-based epigenomic editing systems for use in activating or repressing gene expression of a target gene. The CRISPR/Cas9-based epigenomic editing system includes a fusion protein of a Cas9 protein that does not have nuclease activity, such as dCas9, and a second domain. The second domain may include a transcription activation domain, such as a VP64 domain or p300 domain, transcription repression domain, such as KRAB domain, nuclease domain, transcription release factor domain, histone modification domain, nucleic acid association domain, acetylase domain, deacetylase domain, methylase domain, such as a DNA methylase domain, demethylase domain, phosphorylation domain, ubiquitylation domain, or sumoylation domain. The second domain may be a modifier of DNA methylation or chromatin looping.

[0122] In some embodiments, the fusion protein can include a dCas9 domain and a transcriptional activator. For example, the fusion protein can include the amino acid sequence of SEQ ID NO: 2. In other embodiments, the fusion protein can include a dCas9 domain and a transcriptional repressor. For example, the fusion protein comprises the amino acid sequence of SEQ ID NO:3. In further aspects, the fusion protein can include a dCas9 domain and a site-specific nuclease.

a) CRISPR system

[0123] The CRISPR system is a microbial nuclease system involved in defense against invading phages and plasmids that provides a form of acquired immunity. The CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage. Short segments of foreign DNA, called spacers, are incorporated into the genome between CRISPR repeats, and serve as a 'memory' of past exposures. Cas9 forms a complex with the 3' end of the single guide RNA ("sgRNA"), and the protein-RNA pair recognizes its genomic target by complementary base pairing between the 5' end of the sgRNA sequence and a predefined 20 bp DNA sequence, known as the protospacer. This complex is directed to homologous loci of pathogen DNA via regions encoded within the CRISPR RNA ("crRNA"), i.e., the protospacers, and protospacer-adjacent motifs (PAMs) within the pathogen genome. The non-coding CRISPR array is transcribed and cleaved within direct repeats into short crRNAs containing individual spacer sequences, which direct Cas nucleases to the target site (protospacer). By simply exchanging the 20 bp recognition sequence of the expressed chimeric sgRNA, the Cas9 nuclease can be directed to new genomic targets. CRISPR spacers are used to recognize and silence exogenous genetic elements in a manner analogous to RNAi in eukaryotic organisms.

[0124] Three classes of CRISPR systems (Types I, II and III effector systems) are known. The Type II effector system carries out targeted DNA double-strand break in four sequential steps, using a single effector enzyme, Cas9, to cleave dsDNA. Compared to the Type I and Type III effector systems, which require multiple distinct effectors acting as a complex, the Type II effector system may function in alternative contexts such as eukaryotic cells. The Type II effector system consists of a long pre-crRNA, which is transcribed from the spacer-containing CRISPR locus, the Cas9 protein, and a tracrRNA, which is involved in pre-crRNA processing. The tracrRNAs hybridize to the repeat regions separating the spacers of the pre-crRNA, thus initiating dsRNA cleavage by endogenous RNase III. This cleavage is followed by a second cleavage event within each spacer by Cas9, producing mature crRNAs that remain associated with the tracrRNA and Cas9, forming a Cas9:crRNA-tracrRNA complex.

[0125] An engineered form of the Type II effector system of *Streptococcus pyogenes* was shown to function in human cells for genome engineering. In this system, the Cas9 protein was

directed to genomic target sites by a synthetically reconstituted "guide RNA" ("gRNA", also used interchangeably herein as a chimeric sgRNA, which is a crRNA-tracrRNA fusion that obviates the need for RNase III and crRNA processing in general.

[0126] The Cas9:crRNA-tracrRNA complex unwinds the DNA duplex and searches for sequences matching the crRNA to cleave. Target recognition occurs upon detection of complementarity between a "protospacer" sequence in the target DNA and the remaining spacer sequence in the crRNA. Cas9 mediates cleavage of target DNA if a correct protospacer-adjacent motif (PAM) is also present at the 3' end of the protospacer. For protospacer targeting, the sequence must be immediately followed by the protospacer-adjacent motif (PAM), a short sequence recognized by the Cas9 nuclease that is required for DNA cleavage. Different Type II systems have differing PAM requirements. The *S. pyogenes* CRISPR system may have the PAM sequence for this Cas9 (SpCas9) as 5'-NRG-3', where R is either A or G, and characterized the specificity of this system in human cells. A unique capability of the CRISPR/Cas9 system is the straightforward ability to simultaneously target multiple distinct genomic loci by co-expressing a single Cas9 protein with two or more sgRNAs. For example, the *Streptococcus pyogenes* Type II system naturally prefers to use an "NGG" sequence, where "N" can be any nucleotide, but also accepts other PAM sequences, such as "NAG" in engineered systems (Hsu et al., *Nature Biotechnology* (2013) doi:10.1038/nbt.2647). Similarly, the Cas9 derived from *Neisseria meningitidis* (NmCas9) normally has a native PAM of NNNNGATT, but has activity across a variety of PAMs, including a highly degenerate NNNNGNNN PAM (Esvelt et al. *Nature Methods* (2013) doi:10.1038/nmeth.2681).

b) Cas9

[0127] The CRISPR/Cas9-based epigenomic editing system may include a Cas9 protein or a Cas9 fusion protein. Cas9 protein is an endonuclease that cleaves nucleic acid and is encoded by the CRISPR loci and is involved in the Type II CRISPR system. The Cas9 protein may be from any bacterial or archaea species, such as *Streptococcus pyogenes, Streptococcus thermophiles*, or *Neisseria meningitides*. The Cas9 protein may be mutated so that the nuclease activity is inactivated. In some embodiments, an inactivated Cas9 protein from *Streptococcus pyogenes* (iCas9, also referred to as "dCas9"; SEQ ID NO: 1) may be used. As used herein, "iCas9" and "dCas9" both refer to a Cas9 protein that has the amino acid substitutions D10A and H840A and

has its nuclease activity inactivated. In some embodiments, an inactivated Cas9 protein from *Neisseria meningitides*, such as NmCas9, may be used.

c) CRISPR/Cas9-Based Gene Activation System

[0128] The CRISPR/Cas9-based epigenomic editing system can be a CRISPR/Cas9-based gene activation system that can activate regulatory element function with exceptional specificity of epigenome editing. The CRISPR/Cas9-based gene activation system can be used to screen for enhancers, insulators, silencers, and locus control regions that can be targeted to increase or decrease target gene expression. This technology can be used to assign function to putative regulatory elements identified through genomic studies such as the ENCODE and the Roadmap Epigenomics projects.

[0129] The CRISPR/Cas9-based gene activation system may activate gene expression by modifying DNA methylation, chromatin looping or catalyzing acetylation of histone H3 lysine 27 at its target sites, leading to robust transcriptional activation of target genes from promoters and proximal and distal enhancers. The CRISPR/Cas9-based gene activation system is highly specific and may be guided to the target gene using as few as one guide RNA. The CRISPR/Cas9-based gene activation system may activate the expression of one gene or a family of genes by targeting enhancers at distant locations in the genome.

i) Histone acetyltransferase (HAT) protein

[0130] The CRISPR/Cas9-based gene activation system may include a histone acetyltransferase protein, such as a p300 protein, CREB binding protein (CBP; an analog of p300), GCN5, or PCAF, or fragment thereof. Acetylating histones in regulatory elements using a programmable CRISPR/Cas9-based fusion protein is an effective strategy to increase the expression of target genes. A CRISPR/Cas9-based histone acetyltransferase that can be targeted to any site in the genome is uniquely capable of activating distal regulatory elements. The histone acetyltransferase protein may include a human p300 protein or a fragment thereof. The histone acetyltransferase protein may include a wild-type human p300 protein or a mutant human p300 protein, or fragments thereof. The histone acetyltransferase protein may include the core lysine-acetyltransferase domain of the human p300 protein, i.e., the p300 HAT Core (also known as “p300 Core”).

ii) CRISPR/dCas9$^{p300\ Core}$ Activation System

[0131] The p300 protein regulates the activity of many genes in tissues throughout the body. The p300 protein plays a role in regulating cell growth and division, prompting cells to mature and assume specialized functions (differentiate) and preventing the growth of cancerous tumors. The p300 protein may activate transcription by connecting transcription factors with a complex of proteins that carry out transcription in the cell's nucleus. The p300 protein also functions as a histone acetyltransferase that regulates transcription via chromatin remodeling.

[0132] The dCas9$^{p300\ Core}$ fusion protein is a potent and easily programmable tool to synthetically manipulate acetylation at targeted endogenous loci, leading to regulation of proximal and distal enhancer-regulated genes. The p300 Core acetylates lysine 27 on histone H3 (H3K27ac) and may provide H3K27ac enrichment. The fusion of the catalytic core domain of p300 to dCas9 may result in substantially higher transactivation of downstream genes than the direct fusion of full-length p300 protein despite robust protein expression. The dCas9$^{p300\ Core}$ fusion protein may also exhibit an increased transactivation capacity relative to dCas9^{VP64}, including in the context of the *Nm*-dCas9 scaffold, especially at distal enhancer regions, at which dCas9^{VP64} displayed little, if any, measurable downstream transcriptional activity. Additionally, the dCas9$^{p300\ Core}$ displays precise and robust genome-wide transcriptional specificity. dCas9$^{p300\ Core}$ may be capable of potent transcriptional activation and co-enrichment of acetylation at promoters targeted by the epigenetically modified enhancer.

[0133] The dCas9$^{p300\ Core}$ may activate gene expression through a single gRNA that target and bind a promoter and/or a characterized enhancer. This technology also affords the ability to synthetically transactivate distal genes from putative and known regulatory regions and simplifies transactivation via the application of a single programmable effector and single target site. These capabilities allow multiplexing to target several promoters and/or enhancers simultaneously. The mammalian origin of p300 may provide advantages over virally-derived effector domains for *in vivo* applications by minimizing potential immunogenicity.

[0134] Gene activation by dCas9$^{p300\text{-}Core}$ is highly specific for the target gene. In some embodiments, the p300 Core includes amino acids 1048-1664 of SEQ ID NO: 2 (i.e., SEQ ID NO: 4). In some embodiments, the CRISPR/Cas9-based gene activation system includes a dCas9$^{p300\ Core}$ fusion protein of SEQ ID NO: 2 or an *Nm*-dCas9$^{p300\ Core}$ fusion protein of SEQ ID NO: 5.

d) CRISPR/Cas9-Based Gene Repression System

[0135] The CRISPR/Cas9-based epigenomic editing system can be a CRISPR/Cas9-based gene repression system which can inhibit regulatory element function with exceptional specificity of epigenome editing. In some embodiments, the CRISPR/Cas9-based gene repression system, such as one that include dCas9^{KRAB}, can interfere with distal enhancer activity by highly specific remodeling of the epigenetic state of targeted genetic loci.

i) CRISPR/dCas9^{KRAB} Gene Repression System

[0136] The dCas9^{KRAB} repressor is a highly specific epigenome editing tool that can be used in loss-of-function screens to study gene function and discover targets for drug development. The dCas9^{KRAB} has exceptional specificity to target a particular enhancer, silence only the target genes of that enhancer, and create a repressive heterochromatin environment at that site. dCas9-^{KRAB} can be used to screen for novel regulatory elements within the endogenous genomic context by silencing proximal or distal regulatory elements and corresponding gene targets. The specificity of dCas9-KRAB repressors allows it to be used for transcriptome-wide specificity for silencing endogenous genes. Epigenetic mechanisms for disruption at targeted locus such as histone methylation.

[0137] The KRAB domain, a common heterochromatin-forming motif in naturally occurring zinc finger transcription factors, has been genetically linked to dCas9 to create an RNA-guided synthetic repressor, dCas9^{KRAB}. The Kruppel-associated box ("KRAB") recruits heterochromatin-forming factors: Kap1, HP1, SETDB1, NuRD. It induces H3K0 tri-methylation, histone deacetylation. KRAB-based synthetic repressors can effectively silence the expression of single genes and have been employed to repress oncogenes, inhibit viral replication, and treat dominant negative diseases.

4. sgRNA Library

[0138] As disclosed herein, libraries of gRNAs are targeted to the set of regulatory elements defined for each screen are used in the disclosed methods. In some embodiments, the libraries of gRNAs may be used for a genome wide screen for regulatory elements associated with a particular modulated phenotype. In some embodiments, the libraries of gRNAs may be used to for a genome wide screen of regulatory elements of at least one target gene known or unknown to be associated with a particular modulated phenotype. The libraries of gRNAs may be

designed using a custom algorithm that identifies gRNA target sites in user-defined lists of regulatory elements that have a high degree of targeting specificity in the human genome (Genomic Target Scan (GT-Scan); see O'Brien et al., Bioinformatics (2014) 30:2673-2675). In some embodiments, 100,000 gRNAs can be generated on a single array, providing more than enough gRNAs to comprehensively screen regulatory elements in a 5-10 Mb window around a particular gene of interest. This approach can also be scaled to genome-wide screens by the synthesis of multiple gRNA libraries in parallel. The exact number of gRNAs in the library can depend on whether the screen 1) targets genes or regulatory elements, 2) targets the complete genome, or 3) the number of regulatory elements determined by open chromatin studies such as DNase-seq, ChIP-seq, FAIRE-seq, ATAC-seq, MNase-seq, or related approaches. In some embodiments, between 1 and 100 gRNAs per regulatory element can be generated, depending on the scope of the screen, to compensate for variation in individual gRNA activity. For example, between about 1 and about 100, between about 5 and about 100, between about 10 and about 100, between about 50 and about 100, between about 1 and about 50, between about 5 and about 50, between about 10 and about 50 gRNAs are generated per regulatory element.

[0139] In other embodiments, gRNAs can be generated to target every PAM sequence overlapping a regulatory element. In other embodiments, the gRNAs may be designed to target individual alleles of a regulatory element whose DNA sequence varies between individuals. In other embodiments, gRNAs may be designed to target every or a subset of the PAM sequence within 10 million base pairs of one gene's transcription start site, regardless of the existence of naturally occurring regulatory elements in that region. In other embodiments, gRNAs may be designed to target every or a subset of the PAM sequence within 10 million base pairs of several genes' transcription start sites, regardless of the existence of naturally occurring regulatory elements in those regions. In some embodiments, the gRNA libraries can be synthesized by electrochemical means on arrays (e.g., CustomArray, Twist, Gen9), DNA printing method (e.g., Agilent), or generating individual oligos (e.g., by IDT). In some embodiments, gRNAs that do not target putative regulatory elements can be included as negative controls. The gRNAs can be amplified by PCR and cloned into a lentiviral expression vector that also expresses the CRISPR/Cas9-based epigenomic editing system, such as $dCas9^{p300\text{-}Core}$ or $dCas9^{KRAB}$.

[0140] As disclosed herein, the library of sgRNAs can target a plurality of target sites. The target sites can include promoters, DNAse I hypersensitivity sites, Transposase-Accessible

Chromatin sites, DNA methylation sites, transcription factor binding sites, epigenetic marks, expression quantitative trait loci, and/or regions associated with human traits or phenotypes in genetic association studies. The target sites can be determined by DNase-sequencing (DNase-seq), Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq), ChIP-sequencing, self-transcribing active regulatory region sequencing (STARR-Seq), single molecule real time sequencing (SMRT), Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE–seq), micrococcal nuclease sequencing (MNase-seq), reduced representation bisulfite sequencing (RRBS-seq), whole genome bisulfite sequencing, methyl-binding DNA immunoprecipitation (MEDIP-seq), or genetic association studies.

[0141] In some embodiments, the library of sgRNAs targets one or more gene regulatory elements in a genomic region of a target gene. In some embodiments, the library of sgRNAs targets one or more gene regulatory elements in the entire genome. The target gene can include an endogenous target gene. In some embodiments, between 5 and 50 gRNAs are generated per gene regulatory element. In some embodiments, the genomic region is between about 0 bp to about 10 Mb upstream and/or downstream of the transcription start site of at least one target gene. In some embodiments, the genomic region is greater than 10 Mb or on a different chromosome of at least one target gene.

[0142] In some embodiments, the library of sgRNAs includes between at least about 1,000 to about 150,000 or between at least about 1,000 to about 100,000,000 structurally distinct sgRNAs. For example, the library may include at least about 1,000 to about 200,000,000, at least about 1,000 to about 150,000,000, at least about 1,000 to about 100,000,000, at least about 1,000 to about 50,000,000, at least about 1,000 to about 25,000,000, at least about 1,000 to about 10,000,000, at least about 1,000 to about 1,000,000, at least about 1,000 to about 500,000, at least about 1,000 to about 250,000, at least about 1,000 to about 200,000, at least about 1,000 to about 150,000, at least about 1,000 to about 100,000, at least about 1,000 to about 75,000, at least about 1,000 to about 50,000, at least about 1,000 to about 25,000, at least about 1,000 to about 10,000, at least about 1,000 to about 5,000, at least about 5,000 to about 200,000,000, at least about 5,000 to about 150,000,000, at least about 5,000 to about 100,000,000, at least about 5,000 to about 50,000,000, at least about 5,000 to about 25,000,000, at least about 5,000 to about 10,000,000, at least about 5,000 to about 1,000,000, at least about 5,000 to about 500,000, at least about 5,000 to about 250,000, at least about 5,000 to about 200,000, at least about 5,000 to

about 150,000, at least about 5,000 to about 100,000, at least about 5,000 to about 75,000, at least about 5,000 to about 50,000, at least about 5,000 to about 25,000, at least about 5,000 to about 10,000, at least about 10,000 to about 200,000,000, at least about 10,000 to about 150,000,000, at least about 10,000 to about 100,000,000, at least about 10,000 to about 50,000,000, at least about 10,000 to about 25,000,000, at least about 10,000 to about 10,000,000, at least about 10,000 to about 1,000,000, at least about 10,000 to about 500,000, at least about 10,000 to about 250,000, at least about 10,000 to about 200,000, at least about 10,000 to about 150,000, at least about 10,000 to about 100,000, at least about 10,000 to about 75,000, at least about 10,000 to about 50,000, at least about 10,000 to about 25,000, at least about 25,000 to about 200,000,000, at least about 25,000 to about 150,000,000, at least about 25,000 to about 100,000,000, at least about 25,000 to about 50,000,000, at least about 25,000 to about 25,000,000, at least about 25,000 to about 10,000,000, at least about 25,000 to about 1,000,000, at least about 25,000 to about 500,000, at least about 25,000 to about 250,000, at least about 25,000 to about 200,000, at least about 25,000 to about 150,000, at least about 25,000 to about 100,000, at least about 25,000 to about 75,000, at least about 25,000 to about 50,000, at least about 50,000 to about 200,000,000, at least about 50,000 to about 150,000,000, at least about 50,000 to about 100,000,000, at least about 50,000 to about 50,000,000, at least about 50,000 to about 25,000,000, at least about 50,000 to about 10,000,000, at least about 50,000 to about 1,000,000, at least about 50,000 to about 500,000, at least about 50,000 to about 250,000, at least about 50,000 to about 200,000, at least about 50,000 to about 150,000, at least about 50,000 to about 100,000, at least about 50,000 to about 75,000, at least about 100,000 to about 200,000,000, at least about 100,000 to about 150,000,000, at least about 100,000 to about 100,000,000, at least about 100,000 to about 50,000,000, at least about 100,000 to about 25,000,000, at least about 100,000 to about 10,000,000, at least about 100,000 to about 1,000,000, at least about 100,000 to about 500,000, at least about 100,000 to about 250,000, at least about 100,000 to about 200,000, or at least about 100,000 to about 150,000 sgRNAs. In some embodiments, the library of sgRNAs can tile across the entire genome, and not be limited to known cis-regulatory elements.

[0143] The CRISPR/Cas9-based epigenomic editing system may include at least one gRNA that targets a nucleic acid sequence. The gRNA provides the targeting of the CRISPR/Cas9-based epigenomic editing system. The gRNA is a fusion of two noncoding RNAs: a crRNA and

a tracrRNA. The sgRNA may target any desired DNA sequence by exchanging the sequence encoding a 20 bp protospacer which confers targeting specificity through complementary base pairing with the desired DNA target. gRNA mimics the naturally occurring crRNA:tracrRNA duplex involved in the Type II Effector system. This duplex, which may include, for example, a 42-nucleotide crRNA and a 75-nucleotide tracrRNA, acts as a guide for the Cas9.

[0144] The gRNA may target and bind a target region of a target gene. The target region may be a cis-regulatory region or trans-regulatory region of a target gene. In some embodiments, the target region is a distal or proximal cis-regulatory region of the target gene. The gRNA may target and bind a cis-regulatory region or trans-regulatory region of a target gene. In some embodiments, the gRNA may target and bind an enhancer region, a promoter region, an insulator region, a silencer region, a locus control region, or a transcribed region of a target gene. For example, the gRNA may target and bind the target region is at least one of HS2 enhancer of the human β-globin locus, distal regulatory region (DRR) of the *MYOD* gene, core enhancer (CE) of the *MYOD* gene, proximal (PE) enhancer region of the OCT4 gene, or distal (DE) enhancer region of the *OCT4* gene. In some embodiments, the target region may be a viral promoter, such as an HIV promoter.

[0145] There are ~100,000 candidate regulatory elements in a typical cell type, and >2 million regulatory elements have been identified across all human cell types. While individually synthesizing gRNAs to target each enhancer is trivial, screening ~100,000 enhancers in that manner is infeasible. Hundreds of thousands of targeting gRNAs can be generated quickly and inexpensively by high-throughput DNA synthesis on arrays.

[0146] The target region may include a target enhancer or a target regulatory element. In some embodiments, the target enhancer or target regulatory element controls the gene expression of several target genes. In some embodiments, the target enhancer or target regulatory element controls a cell phenotype that involves the gene expression of one or more target genes. In some embodiments, the identity of one or more of the target genes is known. In some embodiments, the identity of one or more of the target genes is unknown. The CRISPR/Cas9-based epigenomic editing system allows the determination of the identity of these unknown genes that are involved in a cell phenotype. Examples of cell phenotypes include, but not limited to, T-cell phenotype, cell differentiation, such as T-cell, dendritic cell, or hematopoietic stem cell differentiation,

oncogenesis, immunomodulation, cell response to stimuli, cell death, cell growth, cell motility, cell metabolism, cell immunogenicity, drug resistance, or drug sensitivity.

[0147] In some embodiments, at least one gRNA may target and bind a target enhancer or target regulatory element, whereby the expression of one or more genes is activated or repressed. For example, between 1 gene and 20 genes, between 1 gene and 15 genes, between 1 gene and 10 genes, between 1 gene and 5 genes, between 2 genes and 20 genes, between 2 genes and 15 genes, between 2 genes and 10 genes, between 2 genes and 5 genes, between 5 genes and 20 genes, between 5 genes and 15 genes, or between 5 genes and 10 genes are activated or repressed by at least one gRNA. In some embodiments, at least 1 gene, at least 2 genes, at least 3 genes, at least 4 genes, at least 5 gene, at least 6 genes, at least 7 genes, at least 8 genes, at least 9 gene, at least 10 genes, at least 11 genes, at least 12 genes, at least 13 gene, at least 14 genes, at least 15 genes, or at least 20 genes are activated or repressed by at least one gRNA.

[0148] The CRISPR/Cas9-based epigenomic editing system may activate genes at both proximal and distal locations relative the transcriptional start site (TSS). The CRISPR/Cas9-based epigenomic editing system may target a region anywhere in the genome to determine and identify regulatory elements. The CRISPR/Cas9-based epigenomic editing system may target a region that is at least about 1 base pair to about 10,000,000 base pairs, at least about 100 base pairs to about 10,000,000 base pairs, at least about 250 base pairs to about 10,000,000 base pairs, at least about 500 base pairs to about 10,000,000 base pairs, at least about 1,000 base pairs to about 10,000,000 base pairs, at least about 2,000 base pairs to about 10,000,000 base pairs, at least about 5,000 base pairs to about 10,000,000 base pairs, at least about 10,000 base pairs to about 10,000,000 base pairs, at least about 20,000 base pairs to about 10,000,000 base pairs, at least about 50,000 base pairs to about 10,000,000 base pairs, at least about 75,000 base pairs to about 10,000,000 base pairs, at least about 1 base pair to about 1,000,000 base pairs, at least about 100 base pairs to about 1,000,000 base pairs, at least about 250 base pairs to about 1,000,000 base pairs, at least about 500 base pairs to about 1,000,000 base pairs, at least about 1,000 base pairs to about 1,000,000 base pairs, at least about 2,000 base pairs to about 1,000,000 base pairs, at least about 5,000 base pairs to about 1,000,000 base pairs, at least about 10,000 base pairs to about 1,000,000 base pairs, at least about 20,000 base pairs to about 1,000,000 base pairs, at least about 50,000 base pairs to about 1,000,000 base pairs, at least about 75,000 base pairs to about 1,000,000 base pairs, at least about 1 base pair to about 500,000 base pairs, at least

about 100 base pairs to about 500,000 base pairs, at least about 250 base pairs to about 500,000 base pairs, at least about 500 base pairs to about 500,000 base pairs, at least about 1,000 base pairs to about 500,000 base pairs, at least about 2,000 base pairs to about 500,000 base pairs, at least about 5,000 base pairs to about 500,000 base pairs, at least about 10,000 base pairs to about 500,000 base pairs, at least about 20,000 base pairs to about 500,000 base pairs, at least about 50,000 base pairs to about 500,000 base pairs, at least about 75,000 base pairs to about 500,000 base pairs, at least about 1 base pair to about 200,000 base pairs, at least about 100 base pairs to about 200,000 base pairs, at least about 250 base pairs to about 200,000 base pairs, at least about 500 base pairs to about 200,000 base pairs, at least about 1,000 base pairs to about 200,000 base pairs, at least about 2,000 base pairs to about 200,000 base pairs, at least about 5,000 base pairs to about 200,000 base pairs, at least about 10,000 base pairs to about 200,000 base pairs, at least about 20,000 base pairs to about 200,000 base pairs, at least about 50,000 base pairs to about 200,000 base pairs, at least about 75,000 base pairs to about 200,000 base pairs, at least about 1 base pair to about 150,000 base pairs, at least about 100 base pairs to about 150,000 base pairs, at least about 250 base pairs to about 150,000 base pairs, at least about 500 base pairs to about 150,000 base pairs, at least about 1,000 base pairs to about 150,000 base pairs, at least about 2,000 base pairs to about 150,000 base pairs, at least about 5,000 base pairs to about 150,000 base pairs, at least about 10,000 base pairs to about 150,000 base pairs, at least about 20,000 base pairs to about 150,000 base pairs, at least about 50,000 base pairs to about 150,000 base pairs, at least about 75,000 base pairs to about 150,000 base pairs, at least about 1 base pair to about 75,000 base pairs, at least about 100 base pairs to about 75,000 base pairs, at least about 250 base pairs to about 75,000 base pairs, at least about 500 base pairs to about 75,000 base pairs, at least about 1,000 base pairs to about 75,000 base pairs, at least about 2,000 base pairs to about 75,000 base pairs, at least about 5,000 base pairs to about 75,000 base pairs, at least about 10,000 base pairs to about 75,000 base pairs, at least about 20,000 base pairs to about 75,000 base pairs, at least about 50,000 base pairs to about 75,000 base pairs, at least about 1 base pair to about 50,000 base pairs, at least about 100 base pairs to about 50,000 base pairs, at least about 250 base pairs to about 50,000 base pairs, at least about 500 base pairs to about 50,000 base pairs, at least about 1,000 base pairs to about 50,000 base pairs, at least about 2,000 base pairs to about 50,000 base pairs, at least about 5,000 base pairs to about 50,000 base pairs, at least about 10,000 base pairs to about 50,000 base pairs, at least about 20,000 base pairs to about 50,000 base pairs, at

least about 1 base pair to about 25,000 base pairs, at least about 100 base pairs to about 25,000 base pairs, at least about 250 base pairs to about 25,000 base pairs, at least about 500 base pairs to about 25,000 base pairs, at least about 1,000 base pairs to about 25,000 base pairs, at least about 2,000 base pairs to about 25,000 base pairs, at least about 5,000 base pairs to about 25,000 base pairs, at least about 10,000 base pairs to about 25,000 base pairs, at least about 20,000 base pairs to about 25,000 base pairs, at least about 1 base pair to about 10,000 base pairs, at least about 100 base pairs to about 10,000 base pairs, at least about 250 base pairs to about 10,000 base pairs, at least about 500 base pairs to about 10,000 base pairs, at least about 1,000 base pairs to about 10,000 base pairs, at least about 2,000 base pairs to about 10,000 base pairs, at least about 5,000 base pairs to about 10,000 base pairs, at least about 1 base pair to about 5,000 base pairs, at least about 100 base pairs to about 5,000 base pairs, at least about 250 base pairs to about 5,000 base pairs, at least about 500 base pairs to about 5,000 base pairs, at least about 1,000 base pairs to about 5,000 base pairs, or at least about 2,000 base pairs to about 5,000 base pairs upstream from the TSS of a target gene. The CRISPR/Cas9-based epigenomic editing system may target a region that is at least about 1 base pair, at least about 100 base pairs, at least about 500 base pairs, at least about 1,000 base pairs, at least about 1,250 base pairs, at least about 2,000 base pairs, at least about 2,250 base pairs, at least about 2,500base pairs, at least about 5,000 base pairs, at least about 10,000 base pairs, at least about 11,000 base pairs, at least about 20,000 base pairs, at least about 30,000 base pairs, at least about 46,000 base pairs, at least about 50,000 base pairs, at least about 54,000 base pairs, at least about 75,000 base pairs, at least about 150,000 base pairs, at least about 200,000 base pairs, at least about 250,000 base pairs, at least about 300,000 base pairs, at least about 500,000 base pairs, at least about 750,000 base pairs, at least about 1,000,000 base pairs, at least about 2,000,000 base pairs, at least about 4,000,000 base pairs, at least about 5,000,000 base pairs, at least about 6,000,000 base pairs, at least about 7,000,000 base pairs, at least about 8,000,000 base pairs, at least about 9,000,000 base pairs, or at least 10,000,000 base pairs upstream or downstream from the TSS.

[0149] The CRISPR/Cas9-based epigenomic editing system may target a region that is at least about 1 base pair to at least about 1500 base pairs, at least about 1 base pair to at least about 1000 base pairs, at least about 1 base pair to at least about 500 base pairs, at least about 1 base pair to at least about 250 base pairs, at least about 1 base pair to at least about 200 base pairs, at least about 1 base pair to at least about 100 base pairs, is at least about 50 base pair to at least

about 1500 base pairs, at least about 50 base pair to at least about 1000 base pairs, at least about 50 base pairs to at least about 500 base pairs, at least about 50 base pairs to at least about 250 base pairs at least about 50 base pairs to at least about 200 base pairs, at least about 50 base pairs to at least about 100 base pairs, is at least about 100 base pair to at least about 1500 base pairs, at least about 100 base pair to at least about 1000 base pairs, at least about 100 base pairs to at least about 500 base pairs, at least about 100 base pairs to at least about 250 base pairs, or at least about 100 base pairs to at least about 200 base pairs upstream downstream from the TSS. The CRISPR/Cas9-based epigenomic editing system may target a region that is at least about 1 base pair, at least about 2 base pairs, at least about 3 base pairs, at least about 4 base pairs, at least about 5 base pairs, at least about 10 base pairs, at least about 15 base pairs, at least about 20 base pairs, at least about 25 base pairs, at least about 30 base pairs, at least about 40 base pairs, at least about 50 base pairs, at least about 60 base pairs, at least about 70 base pairs, at least about 80 base pairs, at least about 90 base pairs, at least about 100 base pairs, at least about 110 base pairs, at least about 120, at least about 130, at least about 140 base pairs, at least about 150 base pairs, at least about 160 base pairs, at least about 170 base pairs, at least about 180 base pairs, at least about 190 base pairs, at least about 200 base pairs, at least about 210 base pairs, at least about 220, at least about 230, at least about 240 base pairs, at least about 250 base pairs, at least about 500 base pairs, at least about 750 base pairs, at least about 1000 base pairs, or at least about 1500 base pairs upstream or downstream from the TSS.

[0150] In some embodiments, the CRISPR/Cas9-based epigenomic editing system may target and bind a target region that is on the same chromosome as the target gene but more than 150,000 base pairs upstream or downstream from the TSS. In some embodiments, the CRISPR/Cas9-based epigenomic editing system may target and bind a target region that is on a different chromosome from the target gene.

[0151] The CRISPR/Cas9-based epigenomic editing system may use gRNA of varying sequences and lengths. The gRNA may comprise a complementary polynucleotide sequence of the target DNA sequence followed by NGG. The gRNA may comprise a "G" at the 5' end of the complementary polynucleotide sequence. The gRNA may comprise at least a 10 base pair, at least a 11 base pair, at least a 12 base pair, at least a 13 base pair, at least a 14 base pair, at least a 15 base pair, at least a 16 base pair, at least a 17 base pair, at least a 18 base pair, at least a 19 base pair, at least a 20 base pair, at least a 21 base pair, at least a 22 base pair, at least a 23 base

pair, at least a 24 base pair, at least a 25 base pair, at least a 30 base pair, or at least a 35 base pair complementary polynucleotide sequence of the target DNA sequence followed by NGG. The gRNA may target at least one of a promoter region, an enhancer region, a repressor region, an insulator region, a silencer region, a region involved in DNA looping with the promoter region, a gene splicing region, or the transcribed region of the target gene.

[0152] The CRISPR/Cas9-based epigenomic editing system may include at least 10,000 gRNA, at least 20,000 different gRNAs, at least 30,000 different gRNAs at least 40,000 different gRNAs, at least 50,000 different gRNAs, at least 60,000 different gRNAs, at least 70,000 different gRNAs, at least 80,000 different gRNAs, at least 90,000 different gRNAs, at least 100,000 different gRNAs, at least 150, 000 different gRNAs, at least 200,000 different gRNAs, at least 300,000 different gRNAs at least 400,000 different gRNAs, at least 500,000 different gRNAs, at least 600,000 different gRNAs, at least 700,000 different gRNAs, at least 800,000 different gRNAs, at least 900,000 different gRNAs, at least 1,000,000 different gRNAs, at least 5,000,000 different gRNAs, at least 10,000,000 different gRNAs, at least 50,000,000 different gRNAs, or at least 100,000,000 different gRNAs. The CRISPR/Cas9-based epigenomic editing system may include between at least 10,000 gRNA to at least 1,000,000 different gRNAs, at least 10,000 gRNA to at least 800,000 different gRNAs, at least 10,000 gRNA to at least 400,000 different gRNAs, at least 20,000 gRNA to at least 150,000 different gRNAs, at least 20,000 gRNA to at least 800,000 different gRNAs, at least 20,000 different gRNAs to at least 400,000 different gRNAs, at least 40,000 gRNA to at least 100,000 different gRNAs, at least 40,000 different gRNAs to at least 150,000 different gRNAs.

a) Target Gene

[0153] As disclosed herein, the CRISPR/Cas9-based epigenomic editing system may be designed to target and activate the expression of any target gene. The target gene may be an endogenous gene, a transgene, or a viral gene in a cell line. In some embodiments, the target gene may be a known gene that is known to be associated with a particular phenotype. In some embodiments, the target gene is an unknown gene that is associated with a particular phenotype, such as a known gene that is not known to be associated with a particular phenotype or a unknown gene that has not been characterized. In some embodiments, the target region is located on a different chromosome as the target gene. In some embodiments, the CRISPR/Cas9-based epigenomic editing system may activate or repress between about one target gene to about

ten target genes, about one target genes to about five target genes, about one target genes to about four target genes, about one target genes to about three target genes, about one target genes to about two target genes, about two target gene to about ten target genes, about two target genes to about five target genes, about two target genes to about four target genes, about two target genes to about three target genes, about three target genes to about ten target genes, about three target genes to about five target genes, or about three target genes to about four target genes. In some embodiments, the CRISPR/Cas9-based epigenomic editing system may activate or repress at least one target gene, at least two target genes, at least three target genes, at least four target genes, at least five target genes, or at least ten target genes. For example, the may target the hypersensitive site 2 (HS2) enhancer region of the human β-globin locus and activate downstream genes (HBE, HBG, HBD and HBB).

[0154] In some embodiments, the CRISPR/Cas9-based epigenomic editing system induces or represses the gene expression of a target gene by at least about 1 fold, at least about 2 fold, at least about 3 fold, at least about 4 fold, at least about 5 fold, at least about 6 fold, at least about 7 fold, at least about 8 fold, at least about 9 fold, at least about 10 fold, at least 15 fold, at least 20 fold, at least 30 fold, at least 40 fold, at least 50 fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, at least 100 fold, at least about 110 fold, at least 120 fold, at least 130 fold, at least 140 fold, at least 150 fold, at least 160 fold, at least 170 fold, at least 180 fold, at least 190 fold, at least 200 fold, at least about 300 fold, at least 400 fold, at least 500 fold, at least 600 fold, at least 700 fold, at least 800 fold, at least 900 fold, at least 1000 fold, at least 1500 fold, at least 2000 fold, at least 2500 fold, at least 3000 fold, at least 3500 fold, at least 4000 fold, at least 4500 fold, at least 5000 fold, at least 600 fold, at least 7000 fold, at least 8000 fold, at least 9000 fold, at least 10000 fold, at least 100000 fold compared to a control level of gene expression. A control level of gene expression of the target gene may be the level of gene expression of the target gene in a cell that is not treated with any CRISPR/Cas9-based epigenomic editing system

[0155] In some aspects, the target gene is a disease-relevant gene. In some embodiments, the unmodified target cell is a mammalian cell. In some embodiments, the genome includes a human genome. In some embodiments, the target gene may be a mammalian gene. For example, the CRISPR/Cas9-based epigenomic editing system may target a mammalian gene, such as *IL1RN*, *MYOD1*, *OCT4*, *HBE*, *HBG*, *HBD*, *HBB*, *MYOCD* (Myocardin), *PAX7* (Paired box protein Pax-7), *FGF1* (fibroblast growth factor-1) genes, such as *FGF1A*, *FGF1B*, and

FGF1C. Other target genes include, but not limited to, Atf3, Axud1, Btg2, c-Fos, c-Jun, Cxcl1, Cxcl2, Edn1, Ereg, Fos, Gadd45b, Ier2, Ier3, Ifrd1, Il1b, Il6, Irf1, Junb, Lif, Nfkbia, Nfkbiz, Ptgs2, Slc25a25, Sqstm1, Tieg, Tnf, Tnfaip3, Zfp36, Birc2, Ccl2, Ccl20, Ccl7, Cebpd, Ch25h, CSF1, Cx3cl1, Cxcl10, Cxcl5, Gch, Icam1, Ifi47, Ifngr2, Mmp10, Nfkbie, Npal1, p21, Relb, Ripk2, Rnd1, S1pr3, Stx11, Tgtp, Tlr2, Tmem140, Tnfaip2, Tnfrsf6, Vcam1, 1110004C05Rik (GenBank accession number BC010291), Abca1, AI561871 (GenBank accession number BI143915), AI882074 (GenBank accession number BB730912), Arts1, AW049765 (GenBank accession number BC026642.1), C3, Casp4, Ccl5, Ccl9, Cdsn, Enpp2, Gbp2, H2-D1, H2-K, H2-L, Ifit1, Ii, Il13ra1, Il1rl1, Lcn2, Lhfpl2, LOC677168 (GenBank accession number AK019325), Mmp13, Mmp3, Mt2, Naf1, Ppicap, Prnd, Psmb10, Saa3, Serpina3g, Serpinf1, Sod3, Stat1, Tapbp, U90926 (GenBank accession number NM_020562), Ubd, A2AR (Adenosine A2A receptor), B7-H3 (also called CD276), B7-H4 (also called VTCN1), BTLA (B and T Lymphocyte Attenuator; also called CD272), CTLA-4 (Cytotoxic T-Lymphocyte-Associated protein 4; also called CD152), IDO (Indoleamine 2,3-dioxygenase) KIR (Killer-cell Immunoglobulin-like Receptor), LAG3 (Lymphocyte Activation Gene-3), PD-1 (Programmed Death 1 (PD-1) receptor), TIM-3 (T-cell Immunoglobulin domain and Mucin domain 3), and VISTA (V-domain Ig suppressor of T cell activation).

[0156] In some embodiments, the genomic region can be the β-globin locus. In some embodiments, the modified target cell comprises an endogenous gene encoding hemoglobin subunit epsilon (HBE1) chain tagged with a reporter gene. In some embodiments, the reporter gene is red fluorescent protein. In some embodiments, the at least one target gene is HER2, KRAS, or Proprotein convertase subtilisin/kexin type 9 (PCSK9) gene.

i) HBE1

[0157] The HS2 enhancer is a potent activator of the developmentally regulated globin genes, which are involved in sickle cell anemia and beta thalassemia. $dCas9^{KRAB}$ targeted to HS2 by a single guide RNA (gRNA) can disrupt the expression of multiple downstream globin genes. These findings expand the capability capabilities of synthetic repressors by showing that $dCas9^{KRAB}$ targeted to a distal regulatory element by a gRNA can silence expression of multiple genes located 10 to 50 kb away

ii) HER2

[0158] 20% of breast cancers are driven by HER2. Herceptin (anti-HERS) extends life by 20-25 months. Less than 35% of patients respond while 70% of patients gain resistance. Resistance is often from protein regulation.

5. Modified Target Cells

[0159] As disclosed herein, the modified target cell can include an intracellular or cell-surface marker. In some embodiments, the intracellular or cell-surface marker is labelled by immunofluorescence staining. The modified target cell includes a target gene includes a reporter gene inserted within the coding region of the target gene. In some embodiments, the reporter gene encodes a fluorescent protein. In some embodiments, the cell-surface marker is labelled by a fluorescently-labelled specific binding protein for the cell-surface marker. For example, the fluorescent marker can be a green fluorescent protein (GFP), enhanced green fluorescent protein (EGFP), a yellow fluorescent protein (YFP), such as mBanana, a red fluorescent protein (RFP), such as mCherry, DsRed, dTomato, tdTomato, mHoneydew, or mStrawberry, TagRFP, far-red fluorescent pamidronate (FRFP), such as mGrape1 or mGrape2, a cyan fluorescent protein (CFP), a blue fluorescent protein (BFP), enhanced cyan fluorescent protein (ECFP), ultramarine fluorescent protein (UMFP), orange fluorescent protein (OFP), such as mOrange or mTangerine, red (orange) fluorescent protein (mROFP), TagCFP, or a tetracystein fluorescent motif.

[0160] In some embodiments, the modified target cell is a reporter cell line that has GFP gene inserted into the endogenous coding region of a target gene, such as a disease-relevant gene. In some embodiments, the modified target cell is a reporter cell line that produces a cell surface protein that is recognized by a cell surface antibody for staining. In some embodiments, the reporter cell line can be engineered to express a modifier of gene regulatory elements, such as the $dCas9^{p300\ Core}$ acetyltransferase or the $dCas9^{KRAB}$ repressor, and is then transduced with the gRNA libraries.

[0161] The CRISPR/Cas9-based epigenomic editing system may be used with any type of cell. In some embodiments, the cell is a bacterial cell, a fungal cell, an archaea cell, a plant cell or an animal cell. In some embodiments, this may be an organ or an animal organism. In some embodiments, the cell may be any cell type or cell line, including but not limited to, 293-T cells, 3T3 cells, 721 cells, 9L cells, A2780 cells, A2780ADR cells, A2780cis cells, A172 cells, A20

cells, A253 cells, A431 cells, A-549 cells, ALC cells, B16 cells, B35 cells, BCP-1 cells, BEAS-2B cells, bEnd.3 cells, BHK-21 cells, BR 293 cells, BxPC3 cells, C2C12 cells, C3H-10T1/2 cells, C6/36 cells, Cal-27 cells, CHO cells, COR-L23 cells, COR-L23/CPR cells, COR-L23/5010 cells, COR-L23/R23 cells, COS-7 cells, COV-434 cells, CML T1 cells, CMT cells, CT26 cells, D17 cells, DH82 cells, DU145 cells, DuCaP cells, EL4 cells, EM2 cells, EM3 cells, EMT6/AR1 cells, EMT6/AR10.0 cells, FM3 cells, H1299 cells, H69 cells, HB54 cells, HB55 cells, HCA2 cells, HEK-293 cells, HeLa cells, Hepa1c1c7 cells, HL-60 cells, HMEC cells, HT-29 cells, Jurkat cells, J558L cells, JY cells, K562 cells, Ku812 cells, KCL22 cells, KG1 cells, KYO1 cells, LNCap cells, Ma-MeI 1, 2, 3 . . . 48 cells, MC-38 cells, MCF-7 cells, MCF-10A cells, MDA-MB-231 cells, MDA-MB-468 cells, MDA-MB-435 cells, MDCK II cells, MDCK II cells, MG63 cells, MOR/0.2R cells, MONO-MAC 6 cells, MRC5 cells, MTD-1A cells, MyEnd cells, NCI-H69/CPR cells, NCI-H69/LX10 cells, NCI-H69/LX20 cells, NCI-H69/LX4 cells, NIH-3T3 cells, NALM-1 cells, NW-145 cells, OPCN/OPCT cells, Peer cells, PNT-1A/PNT 2 cells, Raji cells, RBL cells, RenCa cells, RIN-5F cells, RMA/RMAS cells, Saos-2 cells, Sf-9 cells, SiHa cells, SkBr3 cells, T2 cells, T-47D cells, T84 cells, THP1 cells, U373 cells, U87 cells, U937 cells, VCaP cells, Vero cells, WM39 cells, WT-49 cells, X63 cells, YAC-1 cells, YAR cells, GM12878, K562, H1 human embryonic stem cells, HeLa-S3, HepG2, HUVEC, SK-N-SH, IMR90, A549, MCF7, HMEC or LHCM, CD14+, CD20+, primary heart or liver cells, differentiated H1 cells, 8988T, Adult_CD4_naive, Adult_CD4_Th0, Adult_CD4_Th1, AG04449, AG04450, AG09309, AG09319, AG10803, AoAF, AoSMC, BC_Adipose_UHN00001, BC_Adrenal_Gland_H12803N, BC_Bladder_01-11002, BC_Brain_H11058N, BC_Breast_02-03015, BC_Colon_01-11002, BC_Colon_H12817N, BC_Esophagus_01-11002, BC_Esophagus_H12817N, BC_Jejunum_H12817N, BC_Kidney_01-11002, BC_Kidney_H12817N, BC_Left_Ventricle_N41, BC_Leukocyte_UHN00204, BC_Liver_01-11002, BC_Lung_01-11002, BC_Lung_H12817N, BC_Pancreas_H12817N, BC_Penis_H12817N, BC_Pericardium_H12529N, BC_Placenta_UHN00189, BC_Prostate_Gland_H12817N, BC_Rectum_N29, BC_Skeletal_Muscle_01-11002, BC_Skeletal_Muscle_H12817N, BC_Skin_01-11002, BC_Small_Intestine_01-11002, BC_Spleen_H12817N, BC_Stomach_01-11002, BC_Stomach_H12817N, BC_Testis_N30, BC_Uterus_BN0765, BE2_C, BG02ES, BG02ES-EBD, BJ, bone_marrow_HS27a, bone_marrow_HS5, bone_marrow_MSC, Breast_OC, Caco-2, CD20+_RO01778,

CD20+_RO01794, CD34+_Mobilized, CD4+_Naive_Wb11970640, CD4+_Naive_Wb78495824, Cerebellum_OC, Cerebrum_frontal_OC, Chorion, CLL, CMK, Colo829, Colon_BC, Colon_OC, Cord_CD4_naive, Cord_CD4_Th0, Cord_CD4_Th1, Decidua, Dnd41, ECC-1, Endometrium_OC, Esophagus_BC, Fibrobl, Fibrobl_GM03348, FibroP, FibroP_AG08395, FibroP_AG08396, FibroP_AG20443, Frontal_cortex_OC, GC_B_cell, Gliobla, GM04503, GM04504, GM06990, GM08714, GM10248, GM10266, GM10847, GM12801, GM12812, GM12813, GM12864, GM12865, GM12866, GM12867, GM12868, GM12869, GM12870, GM12871, GM12872, GM12873, GM12874, GM12875, GM12878-XiMat, GM12891, GM12892, GM13976, GM13977, GM15510, GM18505, GM18507, GM18526, GM18951, GM19099, GM19193, GM19238, GM19239, GM19240, GM20000, H0287, H1-neurons, H7-hESC, H9ES, H9ES-AFP-, H9ES-AFP+, H9ES-CM, H9ES-E, H9ES-EB, H9ES-EBD, HAc, HAEpiC, HA-h, HAL, HAoAF, HAoAF_6090101.11, HAoAF_6111301.9, HAoEC, HAoEC_7071706.1, HAoEC_8061102.1, HA-sp, HBMEC, HBVP, HBVSMC, HCF, HCFaa, HCH, HCH_0011308.2P, HCH_8100808.2, HCM, HConF, HCPEpiC, HCT-116, Heart_OC, Heart_STL003, HEEpiC, HEK293, HEK293T, HEK293-T-REx, Hepatocytes, HFDPC, HFDPC_0100503.2, HFDPC_0102703.3, HFF,HFF-Myc, HFL11W, HFL24W, HGF, HHSEC, HIPEpiC, HL-60, HMEpC, HMEpC_6022801.3, HMF, hMNC-CB, hMNC-CB_8072802.6, hMNC-CB_9111701.6, hMNC-PB, hMNC-PB_0022330.9, hMNC-PB_0082430.9, hMSC-AT, hMSC-AT_0102604.12, hMSC-AT_9061601.12, hMSC-BM, hMSC-BM_0050602.11, hMSC-BM_0051105.11, hMSC-UC, hMSC-UC_0052501.7, hMSC-UC_0081101.7, HMVEC-dAd, HMVEC-dBl-Ad, HMVEC-dBl-Neo, HMVEC-dLy-Ad, HMVEC-dLy-Neo, HMVEC-dNeo, HMVEC-LBl, HMVEC-LLy, HNPCEpiC, HOB, HOB_0090202.1, HOB_0091301, HPAEC, HPAEpiC, HPAF, HPC-PL, HPC-PL_0032601.13, HPC-PL_0101504.13, HPDE6-E6E7, HPdLF, HPF, HPIEpC, HPIEpC_9012801.2, HPIEpC_9041503.2, HRCEpiC, HRE, HRGEC, HRPEpiC, HSaVEC, HSaVEC_0022202.16, HSaVEC_9100101.15, HSMM, HSMM_emb, HSMM_FSHD, HSMMtube, HSMMtube_emb, HSMMtube_FSHD, HT-1080, HTR8svn, Huh-7, Huh-7.5, HVMF, HVMF_6091203.3, HVMF_6100401.3, HWP, HWP_0092205, HWP_8120201.5, iPS, iPS_CWRU1, iPS_hFib2_iPS4, iPS_hFib2_iPS5, iPS_NIHi11, iPS_NIHi7, Ishikawa, Jurkat, Kidney_BC, Kidney_OC, LHCN-M2, LHSR, Liver_OC, Liver_STL004, Liver_STL011, LNCaP, Loucy, Lung_BC, Lung_OC, Lymphoblastoid_cell_line, M059J, MCF10A-Er-Src, MCF-7, MDA-MB-

231, Medullo, Medullo_D341, Mel_2183, Melano, Monocytes-CD14+, Monocytes-CD14+_RO01746, Monocytes-CD14+_RO01826, MRT_A204, MRT_G401, MRT_TTC549, Myometr, Naive_B_cell, NB4, NH-A, NHBE, NHBE_RA, NHDF, NHDF_0060801.3, NHDF_7071701.2, NHDF-Ad, NHDF-neo, NHEK, NHEM.f_M2, NHEM.f_M2_5071302.2, NHEM.f_M2_6022001, NHEM_M2, NHEM_M2_7011001.2, NHEM_M2_7012303, NHLF, NT2-D1, Olf_neurosphere, Osteobl, ovcar-3, PANC-1, Pancreas_OC, PanIsletD, PanIslets, PBDE, PBDEFetal, PBMC, PFSK-1, pHTE, Pons_OC, PrEC, ProgFib, Prostate, Prostate_OC, Psoas_muscle_OC, Raji, RCC_7860, RPMI-7951, RPTEC, RWPE1, SAEC, SH-SY5Y, Skeletal_Muscle_BC, SkMC, SKMC, SkMC_8121902.17, SkMC_9011302, SK-N-MC, SK-N-SH_RA, Small_intestine_OC, Spleen_OC, Stellate, Stomach_BC, T_cells_CD4+, T-47D, T98G, TBEC, Th1, Th1_Wb33676984, Th1_Wb54553204, Th17, Th2, Th2_Wb33676984, Th2_Wb54553204, Treg_Wb78495824, Treg_Wb83319432, U2OS, U87, UCH-1, Urothelia, WERI-Rb-1, and WI-38. In some embodiments, the unmodified target cell can be any cell, such as a primary cell, a HEK293 cell, 293Ts cell, SKBR3 cell, A431 cell, K562 cell, HCT116 cell, HepG2 cell, or K-Ras-dependent and K-Ras-independent cell groups.

6. Modulated Phenotypes

[0162] As described herein, the modified target cells can be transduced with a sgRNA library, such as a gRNA library lentiviral pool, at low multiplicity of infection (MOI) such that >90% of cells express a single gRNA (see FIG. 59). In some embodiments, the optimal balance between percent of cells transduced and percent of cells with one gRNA can be obtained. For example, the optimal balance may be obtained with at least about 10% transduction efficiency or about MOI 0.1, at least about 20% transduction efficiency or about MOI 0.2, at least about 30% transduction efficiency or about MOI 0.3, at least about 40% transduction efficiency or about MOI 0.4, at least about 50% transduction efficiency or about MOI 0.5, at least about 60% transduction efficiency or about MOI 0.6, at least about 70% transduction efficiency or about MOI 0.7, at least about 80% transduction efficiency or about MOI 0.8, at least about 90% transduction efficiency or about MOI 0.9, at least about 99% transduction efficiency or about MOI 0.99, or at least about 100% transduction efficiency or about MOI 1.0. In some embodiments, the cells can be transduced with at least about 100-fold to about 2000-fold coverage of the library at a particular MOI, such as 0.2. For example, the cells can be transduced with at least about 100-fold to about 2000-fold, at least about 100-fold to about 1500-fold, at

least about 100-fold to about 1000-fold, at least about 100-fold to about 500-fold, at least about 200-fold to about 2000-fold, at least about 200-fold to about 1500-fold, at least about 200-fold to about 1000-fold, at least about 200-fold to about 500-fold, at least about 500-fold to about 2000-fold, at least about 500-fold to about 1500-fold, or at least about 500-fold to about 1000-fold, or with at least about 100-fold, at least about 200-fold, at least about 500-fold,at least about 1000-fold, at least about 1500-fold, or at least about 2000-fold coverage of the library. In some embodiments, the modified target cells are transduced with a high multiplicity of infection so that target cells express multiple gRNAs. The number of cells that are transfected

[0163] After an appropriate period of time, cells with modulated gene expression or cell phenotype will be isolated using techniques described below, for example by fluorescence-activated cell sorting (FACS) or by magnetic-activated cell sorting. The modulated phenotype may be recognized by intracellular or cell-surface markers. In some embodiments, the intracellular or cell-surface markers can be labelled by immunofluorescence staining. In some embodiments, an endogenous target gene can be tagged with fluorescent reporters, such as by genome editing. Other possible modulated phenotypic screens include isolating unique cell populations based on a change in response to stimuli, cell death, cell growth, cell proliferation, cell survival, drug resistance, or drug sensitivity. In some embodiments, this phenotypic change in response can be detected in an intact animal organism.

[0164] As disclosed herein, the modulated phenotype can be a change in gene expression of at least one target gene or a change in cell or organismal phenotype. The gene regulatory element can modulate the gene expression or the cell phenotype. In some embodiments, the cell phenotype can be a cell response to stimuli, cell death, cell growth, drug resistance, drug sensitivity, or combinations thereof. The stimuli can be a physical signal, an environmental signal, a hormone, a growth factor, an inflammatory cytokine, an anti-inflammatory cytokine, a transcription factor, or combinations thereof. In some embodiments, the cell phenotype is T-cell phenotype or hematopoietic cell differentiation. In some embodiments, the modulated phenotype can be a modulation in T-cell phenotype or hematopoietic cell differentiation.

7. Identification of regulatory elements that modulate target gene expression or cell phenotype.

[0165] As disclosed herein, the disclosed methods may further include enriching cells expressing differential levels of the reporter gene. In some embodiments, enriching cells may

include flow cytometry, for example by fluorescence-activated cell sorting (FACS) or immunopurification with a cell surface marker. In some embodiments, enriching cells may include magnetic-activated cell sorting. In some embodiments, the regulatory elements responsible for the altered target gene expression or cell phenotype is identified based on an alignment of the significantly enriched gRNAs to the genome.

[0166] As disclosed herein, the sgRNAs are selected to inhibit transcription of the target gene, activate transcription of the target gene, or knockout at least one target gene. In some embodiments, the identifying and characterizing the sgRNAs within the population of selected test cells can include high-throughput sequencing of the sgRNA-expressing DNA, RT-PCR, qRT-PCR, RNA-seq, emulsion PCR, affinity hybridization of the sgRNAs or of DNA expressing the sgRNA to microarrays, affinity hybridization of the sgRNA or of the DNA expressing the sgRNA to colored probes followed by image-based quantitation potentially with a Nanostring nCounter instrument, or direct RNA sequencing. In some embodiments, identifying and characterizing the sgRNAs within the population of selected test cells includes deep sequencing of the genomic DNA of the test cell. Statistical methods may be used to confirm the identity of the sgRNAs, including such methods known to one of skill, such a linear regression, generalized linear regression, hierarchical regression, penalized regression, mixed effects regression models, Bayesian graphical models, support vector machines, random forests, principle component analysis, or neural networks.

[0167] As disclosed herein, the methods include further comparing the gene expression of at least one target gene in the test cells with the gene expression of the target gene in a plurality of control cells. In some embodiments, the change in gene expression of the target gene is an increase or decrease in gene expression compared to a control plurality of cells. The change in gene expression can be determined by a change in protein expression, RNA expression, or protein activity. In some embodiments, the control cells are cells that do not express the CRISPR/Cas9-based epigenomic editing system. In some embodiments, the control cells are cells that have not been exposed to the stimuli or drug.

[0168] In some aspects, selecting the test cells for a change in gene expression of the target gene further includes culturing the cells and selecting the cells on the basis of cellular proliferation or survival. In some embodiments, the culturing the cells is performed in the presence of a selection agent. The selection agent can be a chemotherapeutic, a DNA damaging

agent, a cytotoxic agent, a growth factor, a transcription factor, a kinase, a drug, an exogenous gene under the control of a heterologous promoter, or a hormone. In some embodiments, the plurality of control cells is cultured in the same conditions as the plurality of test cell but without the presence of the selection agent. In some embodiments, selection can occur at the level of entire animal organism.

[0169] For example, the gRNAs in the sorted cell pool can be recovered by PCR of the integrated lentiviral vector and sequenced on Illumina MiSeq. In some embodiments, the screening of sgRNA library can be repeated in one or more independent cell populations to adequately power the identification of gRNAs that modulate gene expression. In some embodiments, between about 1 and about 100, between about 1 and about 50, between about 1 and about 10, between about 1 and about 9, between about 1 and about 8, between about 1 and about 7, between about 1 and about 6, between about 1 and about 5, between about 2 and about 20, between about 2 and about 9, between about 2 and about 8, between about 2 and about 7, between about 2 and about 6, between about 2 and about 5, between about 3 and about 30, between about 3 and about 9, between about 3 and about 8, between about 3 and about 7, between about 3 and about 6, between about 3 and about 5, or at least one, at least two, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 independent cell population can be used.

[0170] In some embodiments, at least 1 or more replicates are used to detect the regulatory elements. For example, between about 1 and about 100, between about 1 and about 50, between about 1 and about 10, between about 1 and about 9, between about 1 and about 8, between about 1 and about 7, between about 1 and about 6, between about 1 and about 5, between about 2 and about 20, between about 2 and about 9, between about 2 and about 8, between about 2 and about 7, between about 2 and about 6, between about 2 and about 5, between about 3 and about 30, between about 3 and about 9, between about 3 and about 8, between about 3 and about 7, between about 3 and about 6, between about 3 and about 5, or at least one, at least two, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100 replicates are used to detect the regulatory elements.

8. Validation of regulatory element activity

[0171] As disclosed herein, the hits from the CRISPR/Cas9-based epigenomic editing system, such as $dCas9^{p300\text{-}Core}$ and $dCas9^{KRAB}$, library screen can be validated by delivering each strongly enriched gRNA individually to independent cell cultures and measuring changes in gene expression by qRT-PCR and immunofluorescence cell surface staining. In some embodiments, ChIP PCR, ChIP-qPCR, or ChIP-seq can be used to validate the gRNA target site. In some embodiments, the specificity of gene regulation can be determined by RNA-seq. The activity of any particular enhancer can be further validated by excising the element from the genome via genome editing, by editing the sequences of putative transcription factor binding sites that fall within these elements, or by confirming the effect of regulatory element function on various aspects of cell phenotype or differentiation state.

9. Constructs and Plasmids

[0172] The compositions, as described above, may comprise genetic constructs that encodes the CRISPR/Cas9-based epigenomic editing system, as disclosed herein. The genetic construct, such as a plasmid or expression vector, may comprise a nucleic acid that encodes the CRISPR/Cas9-based epigenomic editing system, such as the CRISPR/Cas9-based gene activation or gene repression systems, and/or at least one of the gRNAs. The compositions, as described above, may comprise genetic constructs that encodes the modified lentiviral vector and a nucleic acid sequence that encodes the CRISPR/Cas9-based epigenomic editing system, as disclosed herein. The genetic construct, such as a plasmid, may comprise a nucleic acid that encodes the CRISPR/Cas9-based epigenomic editing system. The compositions, as described above, may comprise genetic constructs that encodes a modified lentiviral vector. The genetic construct, such as a plasmid, may comprise a nucleic acid that encodes the CRISPR/Cas9-based epigenomic editing system, such as the CRISPR/Cas9-based gene activation or gene repression systems, and at least one sgRNA. The genetic construct may be present in the cell as a functioning extrachromosomal molecule. The genetic construct may be a linear minichromosome including centromere, telomeres or plasmids or cosmids.

[0173] The genetic construct may also be part of a genome of a recombinant viral vector, including recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus. The genetic construct may be part of the genetic material in attenuated live microorganisms or recombinant microbial vectors which live in cells. The genetic constructs

may comprise regulatory elements for gene expression of the coding sequences of the nucleic acid. The regulatory elements may be a promoter, an enhancer, an initiation codon, a stop codon, or a polyadenylation signal.

[0174] The nucleic acid sequences may make up a genetic construct that may be a vector. The vector may be capable of expressing the fusion protein, such as the CRISPR/Cas9-based epigenomic editing system, in the cell of a mammal. The vector may be recombinant. The vector may comprise heterologous nucleic acid encoding the fusion protein, such as the CRISPR/Cas9-based epigenomic editing system. The vector may be a plasmid. The vector may be useful for transfecting cells with nucleic acid encoding the CRISPR/Cas9-based epigenomic editing system, which the transformed host cell is cultured and maintained under conditions wherein expression of the CRISPR/Cas9-based epigenomic editing system takes place.

[0175] Coding sequences may be optimized for stability and high levels of expression. In some instances, codons are selected to reduce secondary structure formation of the RNA such as that formed due to intramolecular bonding.

[0176] The vector may comprise heterologous nucleic acid encoding the CRISPR/Cas9-based epigenomic editing system and may further comprise an initiation codon, which may be upstream of the CRISPR/Cas9-based epigenomic editing system coding sequence, and a stop codon, which may be downstream of the CRISPR/Cas9-based epigenomic editing system coding sequence. The initiation and termination codon may be in frame with the CRISPR/Cas9-based epigenomic editing system coding sequence. The vector may also comprise a promoter that is operably linked to the CRISPR/Cas9-based epigenomic editing system coding sequence. The CRISPR/Cas9-based epigenomic editing system may be under the light-inducible or chemically inducible control to enable the dynamic control of epigenomic editing in space and time. The promoter operably linked to the CRISPR/Cas9-based epigenomic editing system coding sequence may be a promoter from simian virus 40 (SV40), a mouse mammary tumor virus (MMTV) promoter, a human immunodeficiency virus (HIV) promoter such as the bovine immunodeficiency virus (BIV) long terminal repeat (LTR) promoter, a Moloney virus promoter, an avian leukosis virus (ALV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter, Epstein Barr virus (EBV) promoter, or a Rous sarcoma virus (RSV) promoter. The promoter may also be a promoter from a human gene such as human ubiquitin C (hUbC), human actin, human myosin, human hemoglobin, human muscle creatine, or human

metalothionein. The promoter may also be a tissue specific promoter, such as a muscle or skin specific promoter, natural or synthetic. Examples of such promoters are described in US Patent Application Publication No. US20040175727, the contents of which are incorporated herein in its entirety.

[0177] The vector may also comprise a polyadenylation signal, which may be downstream of the CRISPR/Cas9-based epigenomic editing system. The polyadenylation signal may be a SV40 polyadenylation signal, LTR polyadenylation signal, bovine growth hormone (bGH) polyadenylation signal, human growth hormone (hGH) polyadenylation signal, or human β-globin polyadenylation signal. The SV40 polyadenylation signal may be a polyadenylation signal from a pCEP4 vector (Invitrogen, San Diego, CA).

[0178] The vector may also comprise an enhancer upstream of the CRISPR/Cas9-based epigenomic editing system, *i.e.*, the CRISPR/Cas9-based gene activation or gene repression systems coding sequence, and/or sgRNAs. The enhancer may be necessary for DNA expression. The enhancer may be human actin, human myosin, human hemoglobin, human muscle creatine or a viral enhancer such as one from CMV, HA, RSV or EBV. Polynucleotide function enhancers are described in U.S. Patent Nos. 5,593,972, 5,962,428, and WO94/016737, the contents of each are fully incorporated by reference. The vector may also comprise a mammalian origin of replication in order to maintain the vector extrachromosomally and produce multiple copies of the vector in a cell. The vector may also comprise a regulatory sequence, which may be well suited for gene expression in a mammalian or human cell into which the vector is administered. The vector may also comprise a reporter gene, such as green fluorescent protein ("GFP") and/or a selectable marker, such as hygromycin ("Hygro").

[0179] The vector may be expression vectors or systems to produce protein by routine techniques and readily available starting materials including Sambrook *et al.*, Molecular Cloning and Laboratory Manual, Second Ed., Cold Spring Harbor (1989), which is incorporated fully by reference. In some embodiments the vector may comprise the nucleic acid sequence encoding the CRISPR/Cas9-based epigenomic editing system, including the nucleic acid sequence encoding the CRISPR/Cas9-based gene activation or gene repression systems, and the nucleic acid sequence encoding the at least one gRNA.

[0180] In some embodiments, the sgRNA is encoded by a polynucleotide sequence and packaged into a lentiviral vector, thereby generating a gRNA library lentiviral pool. In some

embodiments, the lentiviral vector includes an expression cassette. The expression cassette can includes a promoter operably linked to the polynucleotide sequence encoding the sgRNA. In some embodiments, the promoter operably linked to the polynucleotide encoding the sgRNA is inducible. In some embodiments, the target cells are transduced with the gRNA library lentiviral pool at a low multiplicity of infection such that >90% of the target cells express a single gRNA. In some embodiments, the target cells are transduced with the gRNA library lentiviral pool at a high multiplicity of infection so that target cells express multiple gRNAs.

10. Methods of Delivery

[0181] Provided herein is a method for delivering the CRISPR/Cas9-based epigenomic editing system for providing genetic constructs and/or proteins of the CRISPR/Cas9-based epigenomic editing system. The delivery of the CRISPR/Cas9-based epigenomic editing system may be the transfection or electroporation of the CRISPR/Cas9-based epigenomic editing system as one or more nucleic acid molecules that is expressed in the cell and delivered to the surface of the cell. The CRISPR/Cas9-based epigenomic editing system protein may be delivered to the cell. The nucleic acid molecules may be electroporated using BioRad Gene Pulser Xcell or Amaxa Nucleofector IIb devices or other electroporation device. Several different buffers may be used, including BioRad electroporation solution, Sigma phosphate-buffered saline product #D8537 (PBS), Invitrogen OptiMEM I (OM), or Amaxa Nucleofector solution V (N.V.). Transfections may include a transfection reagent, such as Lipofectamine 2000.

[0182] The vector encoding a CRISPR/Cas9-based epigenomic editing system protein may be delivered to the modified target cell in a tissue or subject by DNA injection (also referred to as DNA vaccination) with and without in vivo electroporation, liposome mediated, nanoparticle facilitated, and/or recombinant vectors. The recombinant vector may be delivered by any viral mode. The viral mode may be recombinant lentivirus, recombinant adenovirus, and/or recombinant adeno-associated virus.

[0183] The nucleotide encoding a CRISPR/Cas9-based epigenomic editing system protein may be introduced into a cell to induce gene expression of the target gene. For example, one or more nucleotide sequences encoding the CRISPR/Cas9-based epigenomic editing system directed towards a target gene may be introduced into a mammalian cell. Upon delivery of the CRISPR/Cas9-based epigenomic editing system to the cell, and thereupon the vector into the cells of the mammal, the transfected cells will express the CRISPR/Cas9-based epigenomic

editing system. The CRISPR/Cas9-based epigenomic editing system may be administered to a mammal to induce or modulate gene expression of the target gene in a mammal. The mammal may be human, non-human primate, cow, pig, sheep, goat, antelope, bison, water buffalo, bovids, deer, hedgehogs, elephants, llama, alpaca, mice, rats, or chicken, and preferably human, cow, pig, or chicken.

[0184] Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a cell. Suitable methods include, include e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery, and the like. In some embodiments, the composition may be delivered by mRNA delivery and ribonucleoprotein (RNP) complex delivery.

11. Kits

[0185] Provided herein is a kit, which may be used in high throughput screening of regulatory elements of a target gene. The kit comprises a composition for epigenomic editing, as described above, and instructions for using said composition. Instructions included in kits may be affixed to packaging material or may be included as a package insert. While the instructions are typically written or printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this disclosure. Such media include, but are not limited to, electronic storage media (*e.g.*, magnetic discs, tapes, cartridges, chips), optical media (*e.g.*, CD ROM), and the like. As used herein, the term "instructions" may include the address of an internet site that provides the instructions.

[0186] The composition for epigenomic editing may include a modified lentiviral vector and a nucleotide sequence encoding a CRISPR/Cas9-based epigenomic editing system, as described above. The CRISPR/Cas9-based epigenomic editing system may include CRISPR/Cas9-based activation or repression system, as described above, that specifically binds and targets a cis-regulatory region or trans-regulatory region of a target gene and modulates gene expression of a target gene or target a target enhancer or target regulatory element. The

CRISPR/Cas9-based epigenomic editing system, as described above, may be included in the kit to specifically bind and target a particular regulatory region of the target gene.

12. Examples

[0187] The foregoing may be better understood by reference to the following examples, which are presented for purposes of illustration and are not intended to limit the scope of the invention.

EXAMPLE 1

gRNA Library Generation

[0188] gRNA Library. Two libraries of gRNAs targeting DNase I hypersensitivity sites (DHSs), which are strong predictors of gene regulatory element activity, were developed within ~5 megabase regions surrounding target genes of interest. The gRNA libraries were generated using DNase-seq data generated by us as part of the ENCODE Consortium, which were derived from the cell lines in which we will do the screening. The DHSs were used as input to search for gRNA sequences using a modified version of GT-Scan software. Identified gRNAs were then aligned to the genome using Bowtie to identify possible off-target binding locations. For each DHS, the top 50 gRNAs ranked based on predicted off-targets were selected, which resulted in 10,739 gRNAs in a 4.5 megabase region surrounding the β-globin locus (*HBE1* locus) and 10,137 gRNAs in a 5 megabase region surrounding the HER2 gene. Another ~2,000 gRNAs targeting non-DHS regions were included as controls. These sequences were synthesized in parallel on arrays, cloned as a pool into a lentiviral transfer plasmid, and packaged into lentivirus for pooled delivery to the cell type of interest.

EXAMPLE 2

gRNA Library Cell Lines and Screening

[0189] To screen these libraries, two cell lines were used. For the globin-targeting library, the K562 chronic myelogenous leukemia cell line was used because it expresses all the genes in the globin locus. Also, the globin locus has well-characterized regulatory elements that serve as a positive control for this technology. This line was modified by tagging hemoglobin subunit epsilon (HBE1) with red fluorescent protein. This allows monitoring of gene expression through fluorescence intensity using flow cytometry.

[0190] For the HER2 targeting library, the A431 cells (which are used in studies of the cell cycle and cancer-associated cell signaling pathways since they express abnormally high levels of the Epidermal growth factor receptor (EGFR)) were used. The SKBR3 adenocarcinoma HER2 over-expressing line was also used. This line has multiple copies of HER2 duplicated in tandem in its genome, providing an ideal target for identifying regions of the genome in this area that modulate its expression. Modulation of HER2 levels by gRNA library members was evaluated by cell surface immunostaining.

[0191] Both cell types were also transduced with a lentivirus to stably express a suppressor dCas9-KRAB fusion protein. For both libraries, cells with differential target gene expression were isolated by fluorescence-activated cell sorting and the responsible gRNAs recovered from the lentiviral vector in the genomic DNA by PCR. These recovered libraries were then subjected to high-throughput next generation DNA sequencing to identify the genomic regulatory elements associated with altering the expression of each gene.

EXAMPLE 3

Individual gRNA validation methods

[0192] The same modified cell lines used in the screen are used for the individual gRNA validations. For the HBE1 dCas9^{KRAB} screen, K562s with an HBE1-2A-mCherry reporter and constitutively expressing dCas9^{KRAB} were used. For the HER2 dCas9^{KRAB} screen, A431s with constitutively expressing dCas9^{KRAB} were used. For the HER2 dCas9^{p300} screen, HEK293Ts with constitutively expressing dCas9^{p300} were used. The cells were transduced with individual gRNAs and after 2 days were selected with Blasticidin S (ThermoFisher, A1113903). Cells were selected for 12 days for the HBE1 dCas9^{KRAB} screen hits, 9 days for HER2 dCas9^{KRAB} screen hits, and 7 days for the HER2 dCas9^{p300} screen hits.

[0193] For all screen validations, mRNA expression was done in triplicate. Total mRNA was harvested from cells using the Qiagen RNeasy Plus Mini kit (Qiagen, 74136). cDNA was generated using the SuperScript VILO cDNA Synthesis Kit (ThermoFisher, 11754250). qRT-PCR was performed using the Perfecta SYBR Green FastMix (Quanta Biosciences, 95072-012) with the FX96 Real-Time PCR Detection System (Bio-Rad) with the primers listed in Table 1. The results are expressed as fold-increase mRNA expression of the gene of interest normalized to GAPDH expression by the $\Delta\Delta$Ct method.

Table 1

Primer Name	Sequence	SEQ ID NO:	Annealing Temp (°C)
hGAPDH fwd	CAATGACCCCTTCATTGACC	6	53
hGAPDH rev	TTGATTTTGGAGGGATCTCG	7	53
hERBB2 fwd	AGACCATGTCCGGGAAAACC	8	57.5
hERBB2 Rev	GCCAGCCCGAAGTCTGTAAT	9	57.5
hHBG1 fwd	GCTGAGTGAACTGCACTGTGA	10	62
hHBG1 rev	GAATTCTTTGCCGAAATGGA	11	62
hHBE1 fwd	TCACTAGCAAGCTCTCAGGC	12	58
hHBE1 rev	AACAACGAGGAGTCTGCCC	13	58

[0194] For flow cytometry analysis of the HBE1 dCas9^{KRAB} screen validations, cells were harvested, washed once in PBS, and resuspended in PBS. For the HER2 dCas9^{KRAB} and dCas9^{p300} screen validations, cells were harvested, washed once with PBS, then resuspended in 5% goat serum in PBS and blocked for 30 minutes at 4°C. HER2 primary antibody (Monoclonal Mouse IgG2B Clone 191924, R&D Systems) was then added and allowed to incubated for 30 minutes at 4°C. Cells were then washed once in 5% goat serum in PBS. Secondary antibody (Goat anti-Mouse IgG2b Secondary Antibody, Alexa Fluor® 488 conjugate, ThermoFisher A-21141) was then added and cells were allowed to incubate at 4°C for 30 minutes. Cells were then washed once in PBS. All cells were analyzed using the MACSQuant VYB flow cytometer (Miltenyi Biotec).

EXAMPLE 4

HBE1 Regulatory Element Screening with CRISPR/dCas9-^{KRAB} repressor

[0195] As described above, the human K562 erythroid cell line was engineered to express red fluorescence protein (RFP) from the HBE1 (epsilon globin) endogenous gene locus, as well as a CRISPR/dCas9-KRAB repressor. *See* FIGS. 2A-2C. FIGS. 3A-3C show the validation of this reporter line. After lentiviral transduction of a library of 10,739 gRNAs targeting all DNase I hypersensitive sites in a 5 megabase window around HBE1 into K562 cells, cells were sorted with diminished *HBE1* expression according to a fluorescent reporter that was introduced into the gene by genome editing. These RFP-negative cells were isolated by fluorescence-activated cell sorting, and the gRNAs within in these cells were recovered by PCR from the lentiviral vector. The fraction of RFP-negative cells increased from 6% to 20%. The gRNAs were then

recovered from the *HBE1*-depleted cells and sequenced to determine their identity. High-throughput sequencing of the sgRNA library showed enrichment of sequences targeting known regulatory elements at the HBE1 promoter and the HS1 and HS2 enhancers, demonstrating successful identification of specific known regulator elements in their natural chromosomal location from this complex library (**FIGS. 4 and 5C-5C**).

[0196] The enrichment of individual gRNAs across the entire library surrounding the 4.5 Mb region surrounding HBE1 is shown in further detail in FIG. 6A. Enriched gRNAs were determined using differential expression analysis instead of grouping gRNAs per DHS and performing linear regression. FIG. 6B shows a closer view of the globin-locus control region (HS1-5). Significant enrichment of gRNAs was found in the promoter region of HBE1 as well as HS1-4. Additionally, gRNAs targeting the promoter of nearby HBG1/2 genes were enriched in the population of cells expressing high levels of HBE1. FIG. 6C shows enriched gRNAs within the HS2 enhancer. Table 2 shows sgRNAs that were identified, wherein the bolded nucleotides indicate the protospacer. dCas9-^{KRAB} directed to the HS2 enhancer with a single gRNA specifically silenced multiple globin genes located 10-150 kb away. FIGS. 41-48 show the enrichment of gRNAs.

Table 2: K562-dCas9^{KRAB} HBE1 Screen Validation sgRNAs

Name	Target	SEQ ID NO:	sgRNA Sequence
783_784	HBG1/2 promoter	14	**GGCAAGGCTGGCCAACCCAT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
787	HBE1 promoter	15	**CGGACCTGACTCCACCCCTGG**TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
788	HS1	16	**ACTATGCTGAGCTGTGATGA**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
790	HS2	17	**TGCCCTGTAAGCATCCTGCT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
791	HS3	18	**CCATGAGTAGAGGGCAGACA**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
792	HS4	19	**TACTAGGCTGACTCACTCCA**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

EXAMPLE 5

HER2 Regulatory Element Screening with CRISPR/dCas9-^{KRAB} repressor

[0197] The A431 cells were used to screen for HER2 regulatory elements. Modulation of HER2 levels by gRNA library members was evaluated by cell surface immunostaining using a cell surface antibody. FIG. 7A shows flow cytometry data of dCas9^{KRAB} repressing HER2 in A431 cells measured via cell surface staining with a monoclonal antibody. FIG. 7B shows a Manhattan plot showing the results of a screen for regulatory elements in the 4 Mb surrounding the HER2 gene using the dCas9^{KRAB} repressor. FIG. 7C shows a zoomed view of the HER2 region. When comparing high and low expressing HER2 cell populations, gRNAs are enriched in the promoter and an intronic DHS of HER2, as welling DHSs including the promoter of GRB7, an adapter protein that associates with tyrosine kinases. FIG. 7D shows a Manhattan plot showing the results of a high-throughput screen for regulatory elements in the 4 Mb surrounding the globin locus (HBE1 gene) of 12,189 gRNAs targeting 433 DHSs (limit 30 gRNAs per DHS) using the dCas9 ^{KRAB} repressor.

[0198] The enrichment of individual gRNAs across the entire library surrounding the 4 Mb region surrounding HER2 is shown in FIG. 8A. FIG. 8B shows enrichment of gRNAs near HER2. FIG. 8C shows enriched gRNAs by the HER2 promoter.

[0199] Modulation of HER2 levels by gRNA library members was evaluated by cell surface immunostaining using a cell surface antibody. See Table 3, where the protospacer sequence is bolded. FIGS. 49-58 show enrichment data using combinations of gRNAs (i.e., the combinations of 1549.1 and 1553.1, 1549.1 and 1553.2, 1543.1 and 1544.1, 1553.1 and 1562.1, 1553.2 and 1562.1, 1562.1 and 1562.1, 1553.2, 1543.1, 1544.1, 1561.1, and 1562.1, 1559.2 and 1553.1).

Table 3: A431-dCas9^{KRAB} HER2 Screen Validation sgRNAs

Name	SEQ ID NO:	sgRNA Sequence
1549.1	20	**GGTGCGTCCCTCCTAGCGCC**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1553.1	21	**TACCCCGGCGCCCCTCACCA**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

1553.2	22	**CTTTGGCGCATGCTTCACCC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1543.1	23	**AAGGCCTCAAACATTCCCCT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1544.1	24	**GCGTTAGGTGGTGTGGTCTA**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1561.1	25	**GTGCCTCCCCCATTCCGCTC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1769.1	26	**TTCCCACACCCCGTTCCTGG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1592.1	27	**AACCAGGTGCCCACCCGGGC**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1856.1	28	**TCGCCTGCAGCCTTACGGGC**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1562.1	29	**TCCTTGGCTAACTCCAGGCT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1826.1	30	**CGCGGCGCCGCTCTCCGCTG**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1559.2	31	**AGCCGGGTGTCCTGACGCTC**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1560.1	32	**GTTTAATTCGGGAAGAATGC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1782.1	33	**AGCGCTCGATAAATACTTAC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1671.1	34	**TGTTGGGCATTAAGAGGGAG**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1556.1	35	**CATCGCGGCCGGCTCCGCTC**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1562.2	36	**TGACAGCAATAGTGGCCTAC**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1612.1	37	**CAGGTTAGACTTACAAGGTG**GTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1535.1	38	**CTACTCCTTCCTCTTGAGAC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

[0200] The fold change in expression of HER2 mRNA individual hits (gRNAs 1549, 1553, 1543, 1544, 1561, 1769, 1592, 1856, 1562, 1826, 1559, 1560, 1782, 1671, 1556, 1612, and 1535) that were identified using the CRISPR/dCas9^{KRAB} repression system are shown in FIG. 11A. FIG. 11B shows flow cytometry data of dCas9^{KRAB} repressing HER2 in control SKBR3 cells measured via cell surface staining with a monoclonal antibody.

[0201] The enrichment of gRNA 1549 is shown in FIGS. 12A-12C. The enrichment of gRNA 1553 is shown in FIGS. 13A-13C. The enrichment of gRNA 1543 is shown in FIGS. 14A-14C. The enrichment of gRNA 1544 is shown in FIGS. 15A-15C. The enrichment of gRNA 1561 is shown in FIGS. 16A-16C. The enrichment of gRNA 1769 is shown in FIGS.

17A-17C. The enrichment of gRNA 1592 is shown in FIGS. 18A-18C. The enrichment of gRNA 1856 is shown in FIGS. 19A-19C. The enrichment of gRNA 1562 is shown in FIGS. 20A-20C. The enrichment of gRNA 1826 is shown in FIGS. 21A-21C. The enrichment of gRNA 1559 is shown in FIGS. 22A-22C. The enrichment of gRNA 1560 is shown in FIGS. 23A-23C. The enrichment of gRNA 1782 is shown in FIGS. 24A-24C. The enrichment of gRNA 1671 is shown in FIGS. 25A-25C. The enrichment of gRNA 1556 is shown in FIGS. 26A-26C. The enrichment of gRNA 1612 is shown in FIGS. 27A-27C. The enrichment of gRNA 1535 is shown in FIGS. 28A-28C.

EXAMPLE 6

HER2 Regulatory Element Screening with CRISPR/dCas9^{p300} Screen

[0202] The HER2 gRNA library was also screened with HER2 dCas9^{p300} for enhancers. FIG. 9A shows flow cytometry data of dCas9^{p300} activating HER2 in 293Tcells measured via cell surface staining with a monoclonal antibody. FIG. 9B shows a Manhattan plot showing results of a screen for regulatory elements in the 4 Mb surrounding the HER2 gene using the dCas9^{p300} activator. FIG. 9C shows a zoomed view of the HER2 region. When comparing high and low expressing HER2 cell populations, gRNAs are enriched in the promoter and three intronic DHSs of HER2 as well as several nearby DHSs. FIG. 9D shows a closer view of the beginning of the HER2 gene. Interestingly, the intronic DHS enriched in the dCas9^{KRAB} screen did not appear enriched in the p300 screen.

[0203] The fold change in expression of HER2 mRNA individual hits (gRNAs 1549, 1550, 1551, 1552, 1548, 1544, 1542, 1530, 1531, 1561, and 1563) that were identified using the CRISPR/dCas9$^{p300\ Core}$ activation system compared with Control and GFP control is shown in FIG. 29. See Table 4, where the protospacer sequence is bolded.

Table 4: 293T-dCas9^{p300} HER2 Screen Validation sgRNAs

Name	SEQ ID NO:	sgRNA Sequence
1549	39	**GGGTGCGTCCCTCCTAGCGC**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1550	40	**GGATGTACTCCCTGGAAGAG**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1551	41	**GTAGAAAGTGGAGCTGAGCT**GTTTTAGAGCTAGAAATAGCAAGTTAAA

		ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1552	42	**TCAGGCCTGACATCAGACCA**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1548	43	**GCTGGAGGTATAGACCACTG**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1544	44	**GACTGCAGGACTGAGCTAAG**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1542	45	**AGGGACCAAGCTGCTGGGAT**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1530	46	**GACTGAGGCTACCCCTGCTG**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1531	47	**TTCCTGATAGCCCTTGACCG**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1561	48	**TGCTAGCCAGAGCGGAATGG**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
1563	49	**CGATGGCGCGTCCCCTCGTG**GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

[0204] The enrichment of gRNA 1549 is shown in FIGS. 30A-30C. The enrichment of gRNA 1550 is shown in FIGS. 31A-31C. The enrichment of gRNA 1551 is shown in FIGS. 32A-32C. The enrichment of gRNA 1552 is shown in FIGS. 33A-33C. The enrichment of gRNA 1548 is shown in FIGS. 34A-34C. The enrichment of gRNA 1544 is shown in FIGS. 35A-35C. The enrichment of gRNA 1542 is shown in FIGS. 36A-36C. The enrichment of gRNA 1530 is shown in FIGS. 37A-37C. The enrichment of gRNA 1531 is shown in FIGS. 38A-38C. The enrichment of gRNA 1561 is shown in FIGS. 39A-39C. The enrichment of gRNA 1563 is shown in FIGS. 40A-40C.

EXAMPLE 7

HER2 Screen Validations

[0205] To validate enriched hits from both the $dCas9^{KRAB}$ and $dCas9^{p300}$ screens, several of the most enriched gRNAs were individually delivered via lentivirus to A431 or 293T cells expressing $dCas9^{KRAB}$ or $dCas9^{p300}$, respectfully, and HER2 mRNA and protein expression were measured. FIG. 10A shows $dCas9^{KRAB}$ hits: a HER2 mRNA fold change versus log2 fold change in the screen. FIG. 10B shows $dCas9^{KRAB}$ hits: HER2 mean fluorescent intensity measured via antibody staining versus log2 fold change in the screen. FIG. 10C shows $dCas9^{p300}$ hits: a HER2 mRNA fold change versus log2 fold change in the screen. FIG. 10D shows $dCas9^{p300}$ hits: HER2 mean fluorescent intensity measured via antibody staining versus log2 fold change in the screen.

EXAMPLE 8

Genome-wide screen experimental design

[0206] All identified DNase I Hypersensitivity sites (DHSs) from K562 cells from the ENCODE consortium data are used as genomic coordinates for designing a gRNA library. Each DHS is limited to 10 gRNAs for a total of 112,025 gRNAs targeting 1,120,250 DHSs. The library will be synthesized by electrochemical means on arrays and assembled into a lentiviral expression plasmid and packaged into lentivirus in HEK293T cells. To achieve 1,000-fold coverage of the library at an MOI of 0.2, ~5.6 billion K562 cells expressing $dCas9^{KRAB}$ or $dCas9^{p300}$ will be transduced in a 6-liter spinner flask. After day 2, cells will be selected with antibiotic for at least 7 days to isolate transduced cells expressing gRNAs. To detect regulatory elements essential for survival in normal conditions, cells will be grown for at least 4 weeks. For identifying regulatory elements involved in drug resistance, cells will be grown until a non-transduced control is completely killed in the presence of the drug. Cells will be harvested for genomic DNA using Phenol-Chloroform extraction and the library will be recovered using PCR Enrichment or depletion of each gRNA will be determined through next generation sequencing of the library and compared to cells transduced with the same gRNA library not expressing $dCas9^{KRAB}$ or $dCas9^{p300}$ or to cells expressing $dCas9^{KRAB}$ or $dCas9^{p300}$ at earlier time points.

[0207] It is understood that the foregoing detailed description and accompanying examples are merely illustrative and are not to be taken as limitations upon the scope of the invention, which is defined solely by the appended claims and their equivalents.

[0208] Various changes and modifications to the disclosed embodiments will be apparent to those skilled in the art. Such changes and modifications, including without limitation those relating to the chemical structures, substituents, derivatives, intermediates, syntheses, compositions, formulations, or methods of use of the invention, may be made without departing from the spirit and scope thereof.

[0209] For reasons of completeness, various aspects of the invention are set out in the following numbered clauses:

[0210] Clause 1. A method of high-throughput screening for one or more putative gene regulatory elements in a genome that modulate a phenotype, the method comprising: a) contacting a plurality of modified target cells with a library of single guide RNAs (sgRNAs) that target a plurality of gene regulatory elements within the genome, thereby generating a plurality

of test cells, b) selecting a population of test cells or an organism having a modulated phenotype; c) quantitating the frequency of the sgRNAs within the population of selected cells or the organism, wherein the sgRNAs that target gene regulatory elements that modulate the phenotype are overrepresented or underrepresented in the selected cells; and d) identifying and characterizing the sgRNAs within the population of selected test cells or the organism thereby identifying the gene regulatory elements that modulate the phenotype, wherein the modified target cell or organism comprises a fusion protein, the fusion protein comprising a first polypeptide domain comprising a nuclease-deficient Cas9 and a second polypeptide domain having an activity selected from the group consisting of transcription activation activity, transcription repression activity, nuclease activity, transcription release factor activity, histone modification activity, nucleic acid association activity, DNA methylase activity, and direct or indirect DNA demethylase activity.

[0211] Clause 2. The method of clause 1, wherein the modulated phenotype is a change in gene expression of at least one target gene or a change in cell or organismal phenotype and the gene regulatory element modulates the gene expression or the cell or organismal phenotype.

[0212] Clause 3. The method of clause 2, wherein the cell phenotype is at least one of cell response to stimuli, cell death, cell growth, drug resistance, or drug sensitivity.

[0213] Clause 4. The method of clause 3, wherein the stimuli comprises at least one of a physical signal, an environmental signal, a hormone, a growth factor, an inflammatory cytokine, an anti-inflammatory cytokine, a small molecule drug, or a transcription factor.

[0214] Clause 5. The method of clause 2, wherein the cell phenotype is T-cell phenotype, cell differentiation, oncogenesis, immunomodulation, cell response to stimuli, cell death, cell growth, cell motility, cell metabolism, cell immunogenicity, drug resistance, or drug sensitivity.

[0215] Clause 6. The method of any one of clauses 1-5, wherein the library of sgRNAs targets a plurality of target sites in the genome.

[0216] Clause 7. The method of clause 6, wherein the comprising promoters, DNAse I hypersensitivity sites, Transposase-Accessible Chromatin sites, DNA methylation, and/or epigenetic marks.

[0217] Clause 8. The method of clause 7, wherein the target sites are determined by DNase-sequencing (DNase-seq), Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq), ChIP-sequencing, self-transcribing active regulatory region

sequencing, single molecule real time sequencing (SMRT), or Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE–seq), or micrococcal nuclease sequencing (MNase-seq).

[0218] Clause 9. The method of any one of clauses 1-8, wherein the library of sgRNAs targets one or more gene regulatory elements in a genomic region of a target gene.

[0219] Clause 10. The method of clause 9, wherein the target gene comprises an endogenous target gene.

[0220] Clause 11. The method of any one of clauses 1-10, wherein between 5 and 50 gRNAs are generated per gene regulatory element.

[0221] Clause 12. The method of any one of clauses 1-11, wherein the modified target cell comprises an intracellular or cell-surface marker.

[0222] Clause 13. The method of clause 12, wherein the intracellular or cell-surface marker is labelled by immunofluorescence staining.

[0223] Clause 14. The method of any one of clauses 1-13, wherein the modified target cell comprises at least one target gene comprising a reporter gene inserted within the coding region of the target gene.

[0224] Clause 15. The method of clause 14, wherein the reporter gene encodes a fluorescent protein.

[0225] Clause 16. The method of clause 13, wherein the cell-surface marker is labelled by a fluorescently-labelled specific binding protein for the cell-surface marker.

[0226] Clause 17. The method of any one of clauses 1-16, further comprising enriching cells expressing differential levels of the reporter gene by fluorescence-activated cell sorting or magnetic-activated cell sorting.

[0227] Clause 18. The method of any one of clauses 1-17, wherein the modulated phenotype is a change in gene expression of at least one target gene.

[0228] Clause 19. The method of any one of clauses 1-18, further comprising comparing the gene expression of at least one target gene in the test cells with the gene expression of the target gene in a plurality of control cells.

[0229] Clause 20. The method of any one of clauses 1-19, wherein the change in gene expression of the target gene is an increase or decrease in gene expression compared to a control plurality of cells.

[0230] Clause 21. The method of any one of clauses 1-20, wherein the change in gene expression is determined by a change in protein expression, RNA expression, or protein activity.

[0231] Clause 22. The method of any one of clauses 1-21, wherein identifying and characterizing the sgRNAs within the population of selected test cells comprises high-throughput sequencing, qRT-PCR, or RNA-seq.

[0232] Clause 23. The method of any one of clauses 1-22, wherein the library of sgRNAs targets one or more gene regulatory elements in the entire genome

[0233] Clause 24. The method of clause 9, wherein the genomic region is between about 0 bp to about 150 Mb upstream and/or downstream of the transcription start site of at least one target gene.

[0234] Clause 25. The method of clause 24, wherein the genomic region is the β-globin locus.

[0235] Clause 26. The method of clause 25, wherein the modified target cell comprises an endogenous gene encoding hemoglobin subunit epsilon (HBE1) chain tagged with a reporter gene.

[0236] Clause 27. The method of clause 26, wherein the reporter gene is red fluorescent protein.

[0237] Clause 28. The method of clause 23, wherein the at least one target gene is HER2 gene.

[0238] Clause 29. The method of any one of clauses 1-28, wherein the fusion protein comprises a dCas9 domain and a transcriptional activator.

[0239] Clause 30. The method of clause 29, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 2.

[0240] Clause 31. The method of any one of clauses 1-30, wherein the fusion protein comprises a dCas9 domain and a transcriptional repressor.

[0241] Clause 32. The method of clause 31, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO:3.

[0242] Clause 33. The method of any one of clauses 1-32, wherein the fusion protein comprises a dCas9 domain and a site-specific nuclease.

[0243] Clause 34. The method of any one of clauses 1-33, wherein the library of sgRNAs comprises between at least about 1,000 to about 100,000,000 structurally distinct sgRNAs.

[0244] Clause 35. The method of any one of clauses 1-34, wherein the sgRNAs are selected to inhibit transcription of the target gene.

[0245] Clause 36. The method of any one of clauses 1-35, wherein the sgRNAs are selected to activate transcription of the target gene.

[0246] Clause 37. The method of any one of clauses 1-36, wherein the sgRNAs are selected to knockout at least one target gene.

[0247] Clause 38. The method of any one of clauses 1-37, wherein identifying and characterizing the sgRNAs within the population of selected test cells comprises deep sequencing of the genomic DNA of the test cell.

[0248] Clause 39. The method of any one of clauses 1-38, wherein selecting the test cells for a change in gene expression of the target gene further comprises culturing the cells and selecting the cells on the basis of cellular proliferation or survival.

[0249] Clause 40. The method of clause 39, wherein the culturing the cells is performed in the presence of a selection agent.

[0250] Clause 41. The method of clause 40, wherein the selection agent a chemotherapeutic, a DNA damaging agent, a cytotoxic agent, a growth factor, a transcription factor, a kinase, a drug, an exogenous gene under the control of a heterologous promoter, or a hormone.

[0251] Clause 42. The method of clause 40, wherein the plurality of control cells is cultured in the same conditions as the plurality of test cell but without the presence of the selection agent.

[0252] Clause 43. The method of any one of clauses 1-42, wherein the sgRNA is encoded by a polynucleotide sequence and packaged into a lentiviral vector, thereby generating a gRNA library lentiviral pool.

[0253] Clause 44. The method of clause 43, wherein the lentiviral vector comprises an expression cassette comprising a promoter operably linked to the polynucleotide sequence encoding the sgRNA.

[0254] Clause 45. The method of clause 44, wherein the promoter operably linked to the polynucleotide encoding the sgRNA is inducible.

[0255] Clause 46. The method of clause 43, wherein the target cells are transduced with the gRNA library lentiviral pool at a low multiplicity of infection such that >90% of the target cells express a single gRNA.

[0256] Clause 47. The method of clause 43, wherein the target cells are transduced with the gRNA library lentiviral pool at a high multiplicity of infection such that >90% of the target cells expression more than one gRNAs.

[0257] Clause 48. The method of any one of clauses 1-47, wherein the at least one target gene is a disease-relevant gene.

[0258] Clause 49. The method of any one of clauses 1-48, wherein the unmodified target cell is a eukaryotic cell.

[0259] Clause 50. The method of any one of clauses 1-49, wherein the unmodified target cell is a mammalian cell.

[0260] Clause 51. The method of any one of clauses 1-50, wherein the unmodified target cell is a A431, 293T, SKBR3, or K562 cell.

[0261] Clause 52. The method of any one of clauses 1-51, wherein the genome comprises a human genome.

Appendix – Sequences

Streptococcus pyogenes Cas 9 (with D10A, H849A) (SEQ ID NO: 1)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL
KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
YDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKK
YGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

dCas9$^{p300\ Core}$: (Addgene Plasmid 61357) amino acid sequence; 3X "Flag" Epitope, **Nuclear Localization Sequence, *Streptococcus pyogenes* Cas9 (D10A, H840A), p300 Core Effector, "HA" Epitope (SEQ ID NO: 2)**

MDYKDHDGDYKDHDIDYKDDDDKMA**PKKKRKVG**RG**MDKKYSIGLAIGTNSVGWA**
VITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILL
SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVK
YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQL
IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR

**QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV
LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**PIA
GSKAS**PKKKRKVG**RA**IFKPEELRQALMPTLEALYRQDPESLPFRQPVDPQLLGIPD
YFDIVKSPMDLSTIKRKLDTGQYQEPWQYVDDIWLMFNNAWLYNRKTSRVYKYCS
KLSEVFEQEIDPVMQSLGYCCGRKLEFSPQTLCCYGKQLCTIPRDATYYSYQNRYH
FCEKCFNEIQGESVSLGDDPSQPQTTINKEQFSKRKNDTLDPELFVECTECGRKMH
QICVLHHEIIWPAGFVCDGCLKKSARTRKENKFSAKRLPSTRLGTFLENRVNDFLR
RQNHPESGEVTVRVVHASDKTVEVKPGMKARFVDSGEMAESFPYRTKALFAFEEI
DGVDLCFFGMHVQEYGSDCPPPNQRRVYISYLDSVHFFRPKCLRTAVYHEILIGYL
EYVKKLGYTTGHIWACPPSEGDDYIFHCHPPDQKIPKPKRLQEWYKKMLDKAVSE
RIVHDYKDIFKQATEDRLTSAKELPYFEGDFWPNVLEESIKELEQEEEERKREENTS
NESTDVTKGDSKNAKKKNNKKTSKNKSSLSRGNKKKPGMPNVSNDLSQKLYATM
EKHKEVFFVIRLIAGPAANSLPPIVDPDPLIPCDLMDGRDAFLTLARDKHLEFSSLRR
AQWSTMCMLVELHTQSQD**<u>YPYDVPDYAS</u>**

<u>dCas9KRAB (SEQ ID NO: 3)</u>
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGMDKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE
KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE
KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID
RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSRADPKKKRKVASDAKSLTAWSR

TLVTFKDVFVDFTREEWKLLDTAQQILYRNVMLENYKNLVSLGYQLTKPDVILRLEKGE
EPWLVEREIHQETHPDSETAFEIKSSVPKKKRKVAS

Nm-dCas9$^{p300\ Core}$: (Addgene Plasmid 61365) amino acid sequence; ***Neisseria meningitidis*** **Cas9 (D16A, D587A, H588A, N611A), Nuclear Localization Sequence, p300 Core Effector, "HA" Epitope (SEQ ID NO: 5)**

MAAFKPNPINYILGLAIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTGDS
LAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQ
LRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNA
HALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEF
GNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERF
IWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFK
GLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLF
KTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACA
EIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIE
TAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYE
QQHGKCLYSGKEINLGRLNEKGYVEIAAALPFSRTWDDSFNNKVLVLGSEAQNKG
NQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDT
RYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHA
LDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFF
AQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPN
RKMSGQGHMETVKSAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALK
ARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIA
DNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNF
KFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKT
ALSFQKYQIDELGKEIRPCRLKKRPPVRSRADPKKKRKVEASGRA**IFKPEELRQAL**
MPTLEALYRQDPESLPFRQPVDPQLLGIPDYFDIVKSPMDLSTIKRKLDTGQYQEP
WQYVDDIWLMFNNAWLYNRKTSRVYKYCSKLSEVFEQEIDPVMQSLGYCCGRKL
EFSPQTLCCYGKQLCTIPRDATYYSYQNRYHFCEKCFNEIQGESVSLGDDPSQPQTT
INKEQFSKRKNDTLDPELFVECTECGRKMHQICVLHHEIIWPAGFVCDGCLKKSAR
TRKENKFSAKRLPSTRLGTFLENRVNDFLRRQNHPESGEVTVRVVHASDKTVEVKP
GMKARFVDSGEMAESFPYRTKALFAFEEIDGVDLCFFGMHVQEYGSDCPPPNQRR
VYISYLDSVHFFRPKCLRTAVYHEILIGYLEYVKKLGYTTGHIWACPPSEGDDYIFH
CHPPDQKIPKPKRLQEWYKKMLDKAVSERIVHDYKDIFKQATEDRLTSAKELPYF
EGDFWPNVLEESIKELEQEEEERKREENTSNESTDVTKGDSKNAKKKNNKKTSKN
KSSLSRGNKKKPGMPNVSNDLSQKLYATMEKHKEVFFVIRLIAGPAANSLPPIVDPD
PLIPCDLMDGRDAFLTLARDKHLEFSSLRRAQWSTMCMLVELHTQSQDYPYDVPD
YAS

CLAIMS

What is claimed:

1. A method of high-throughput screening for one or more putative gene regulatory elements in a genome that modulate a phenotype, the method comprising:

a) contacting a plurality of modified target cells with a library of single guide RNAs (sgRNAs) that target a plurality of gene regulatory elements within the genome, thereby generating a plurality of test cells,

b) selecting a population of test cells or an organism having a modulated phenotype;

c) quantitating the frequency of the sgRNAs within the population of selected cells or the organism, wherein the sgRNAs that target gene regulatory elements that modulate the phenotype are overrepresented or underrepresented in the selected cells; and

d) identifying and characterizing the sgRNAs within the population of selected test cells or the organism thereby identifying the gene regulatory elements that modulate the phenotype,

wherein the modified target cell or organism comprises a fusion protein, the fusion protein comprising a first polypeptide domain comprising a nuclease-deficient Cas9 and a second polypeptide domain having an activity selected from the group consisting of transcription activation activity, transcription repression activity, nuclease activity, transcription release factor activity, histone modification activity, nucleic acid association activity, DNA methylase activity, and direct or indirect DNA demethylase activity.

2. The method of claim 1, wherein the modulated phenotype is a change in gene expression of at least one target gene or a change in cell or organismal phenotype and the gene regulatory element modulates the gene expression or the cell or organismal phenotype.

3. The method of claim 2, wherein the cell phenotype is at least one of cell response to stimuli, cell death, cell growth, drug resistance, or drug sensitivity.

4. The method of claim 3, wherein the stimuli comprises at least one of a physical signal, an environmental signal, a hormone, a growth factor, an inflammatory cytokine, an anti-inflammatory cytokine, a small molecule drug, or a transcription factor.

5. The method of claim 2, wherein the cell phenotype is T-cell phenotype, cell differentiation, oncogenesis, immunomodulation, cell response to stimuli, cell death, cell growth, cell motility, cell metabolism, cell immunogenicity, drug resistance, or drug sensitivity.

6. The method of claim 1, wherein the library of sgRNAs targets a plurality of target sites in the genome.

7. The method of claim 6, wherein the comprising promoters, DNAse I hypersensitivity sites, Transposase-Accessible Chromatin sites, DNA methylation, and/or epigenetic marks.

8. The method of claim 7, wherein the target sites are determined by DNase-sequencing (DNase-seq), Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq), ChIP-sequencing, self-transcribing active regulatory region sequencing, single molecule real time sequencing (SMRT), or Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE–seq), or micrococcal nuclease sequencing (MNase-seq).

9. The method of claim 1, wherein the library of sgRNAs targets one or more gene regulatory elements in a genomic region of a target gene.

10. The method of claim 9, wherein the target gene comprises an endogenous target gene.

11. The method of claim 1, wherein between 5 and 50 gRNAs are generated per gene regulatory element.

12. The method of claim 1, wherein the modified target cell comprises an intracellular or cell-surface marker.

13. The method of claim 12, wherein the intracellular or cell-surface marker is labelled by immunofluorescence staining.

14. The method of claim 1, wherein the modified target cell comprises at least one target gene comprising a reporter gene inserted within the coding region of the target gene.

15. The method of claim 14, wherein the reporter gene encodes a fluorescent protein.

16. The method of claim 13, wherein the cell-surface marker is labelled by a fluorescently-labelled specific binding protein for the cell-surface marker.

17. The method of claim 1, further comprising enriching cells expressing differential levels of the reporter gene by fluorescence-activated cell sorting or magnetic-activated cell sorting.

18. The method of claim 1, wherein the modulated phenotype is a change in gene expression of at least one target gene.

19. The method of claim 1, further comprising comparing the gene expression of at least one target gene in the test cells with the gene expression of the target gene in a plurality of control cells.

20. The method of claim 1, wherein the change in gene expression of the target gene is an increase or decrease in gene expression compared to a control plurality of cells.

21. The method of claim 1, wherein the change in gene expression is determined by a change in protein expression, RNA expression, or protein activity.

22. The method of claim 1, wherein identifying and characterizing the sgRNAs within the population of selected test cells comprises high-throughput sequencing, qRT-PCR, or RNA-seq.

23. The method of claim 1, wherein the library of sgRNAs targets one or more gene regulatory elements in the entire genome.

24. The method of claim 11, wherein the genomic region is between about 0 bp to about 150 Mb upstream and/or downstream of the transcription start site of at least one target gene.

25. The method of claim 24, wherein the genomic region is the β-globin locus.

26. The method of claim 25, wherein the modified target cell comprises an endogenous gene encoding hemoglobin subunit epsilon (HBE1) chain tagged with a reporter gene.

27. The method of claim 26, wherein the reporter gene is red fluorescent protein.

28. The method of claim 23, wherein the at least one target gene is HER2 gene.

29. The method of claim 1, wherein the fusion protein comprises a dCas9 domain and a transcriptional activator.

30. The method of claim 29, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 2.

31. The method of claim 1, wherein the fusion protein comprises a dCas9 domain and a transcriptional repressor.

32. The method of claim 31, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO:3.

33. The method of claim 1, wherein the fusion protein comprises a dCas9 domain and a site-specific nuclease.

34. The method of claim 1, wherein the library of sgRNAs comprises between at least about 1,000 to about 100,000,000 structurally distinct sgRNAs.

35. The method of claim 1, wherein the sgRNAs are selected to inhibit transcription of the target gene.

36. The method of claim 1, wherein the sgRNAs are selected to activate transcription of the target gene.

37. The method of claim 1, wherein the sgRNAs are selected to knockout at least one target gene.

38. The method of claim 1, wherein identifying and characterizing the sgRNAs within the population of selected test cells comprises deep sequencing of the genomic DNA of the test cell.

39. The method of claim 1, wherein selecting the test cells for a change in gene expression of the target gene further comprises culturing the cells and selecting the cells on the basis of cellular proliferation or survival.

40. The method of claim 39, wherein the culturing the cells is performed in the presence of a selection agent.

41. The method of claim 40, wherein the selection agent a chemotherapeutic, a DNA damaging agent, a cytotoxic agent, a growth factor, a transcription factor, a kinase, a drug, an exogenous gene under the control of a heterologous promoter, or a hormone.

42. The method of claim 40, wherein the plurality of control cells is cultured in the same conditions as the plurality of test cell but without the presence of the selection agent.

43. The method of claim 1, wherein the sgRNA is encoded by a polynucleotide sequence and packaged into a lentiviral vector, thereby generating a gRNA library lentiviral pool.

44. The method of claim 43, wherein the lentiviral vector comprises an expression cassette comprising a promoter operably linked to the polynucleotide sequence encoding the sgRNA.

45. The method of claim 44, wherein the promoter operably linked to the polynucleotide encoding the sgRNA is inducible.

46. The method of claim 43, wherein the target cells are transduced with the gRNA library lentiviral pool at a low multiplicity of infection such that >90% of the target cells express a single gRNA.

47. The method of claim 43, wherein the target cells are transduced with the gRNA library lentiviral pool at a high multiplicity of infection such that >90% of the target cells expression more than one gRNAs.

48. The method of claim 1, wherein the at least one target gene is a disease-relevant gene.

49. The method of claim 1, wherein the unmodified target cell is a eukaryotic cell.

50. The method of claim 1, wherein the unmodified target cell is a mammalian cell.

51. The method of claim 1, wherein the unmodified target cell is a A431, 293T, SKBR3, or K562 cell.

52. The method of claim 1, wherein the genome comprises a human genome.

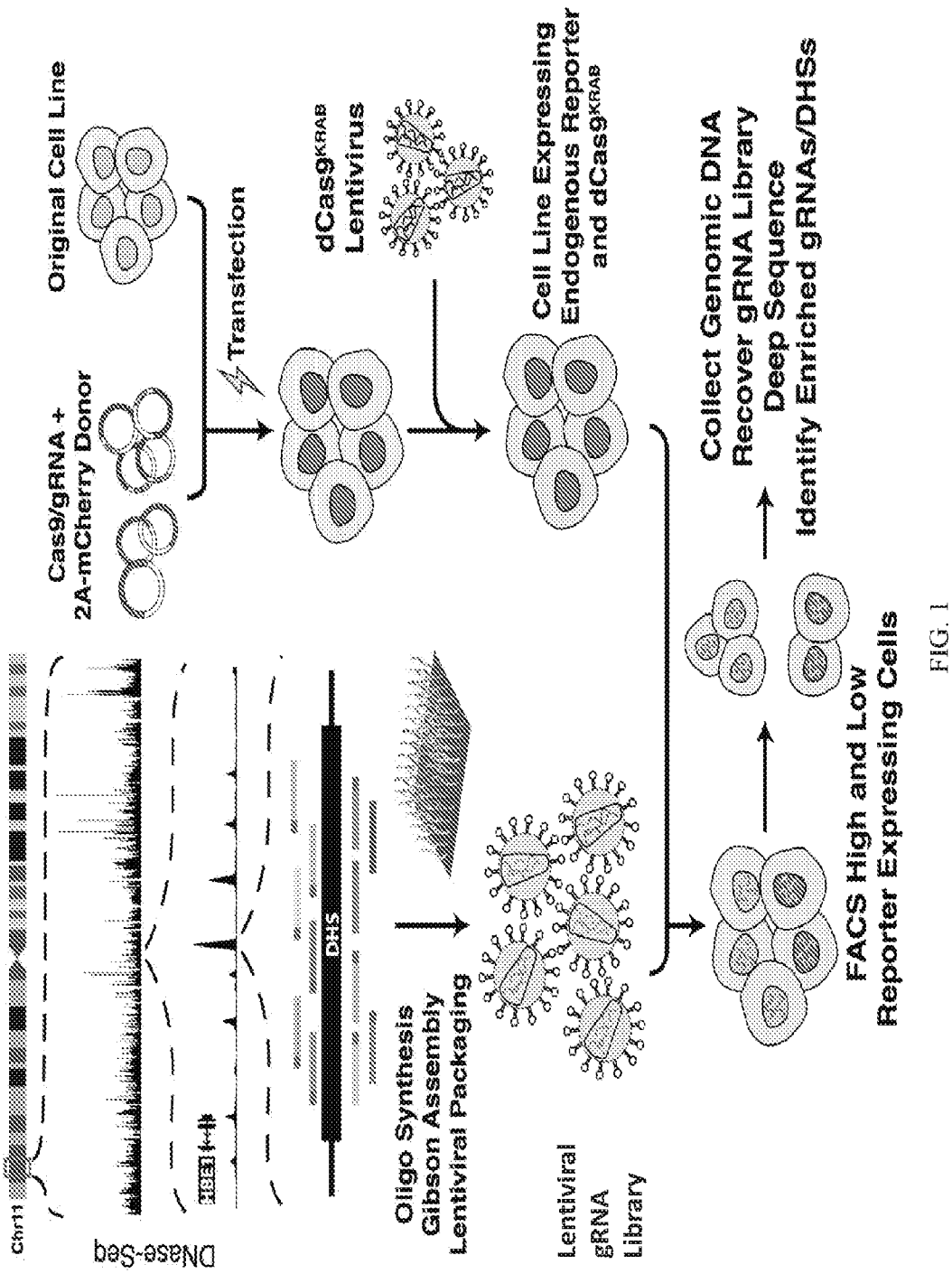

Original Cell Line
Cas9/gRNA +
2A-mCherry Donor
Transfection
dCas9KRAB
Lentivirus
Cell Line Expressing
Endogenous Reporter
and dCas9KRAB
Collect Genomic DNA
Recover gRNA Library
Deep Sequence
Identify Enriched gRNAs/DHSs
FACS High and Low
Reporter Expressing Cells
Lentiviral
gRNA
Library
Oligo Synthesis
Gibson Assembly
Lentiviral Packaging
DHS
DNase-Seq
Chr11

FIG. 1

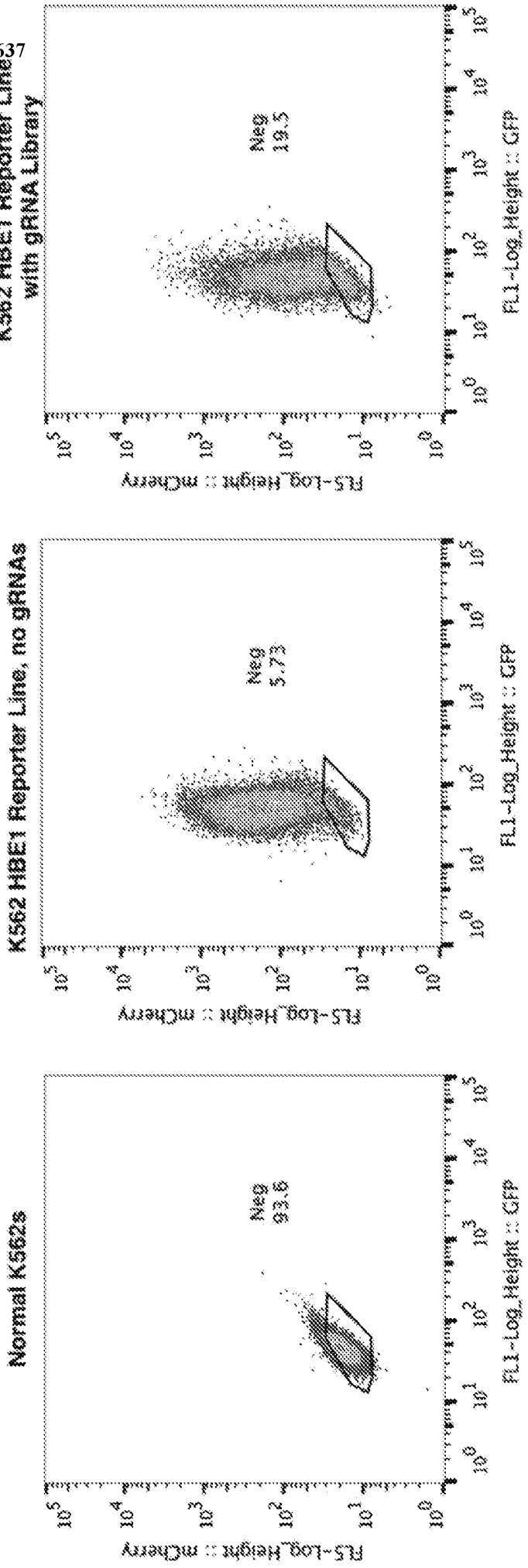

Normal K562s
Neg
93.6
K562 HBE1 Reporter Line, no gRNAs
Neg
5.73
K562 HBE1 Reporter Line with gRNA Library
Neg
19.5
FL1-Log_Height :: GFP
FL5-Log_Height :: mCherry

FIG. 2A

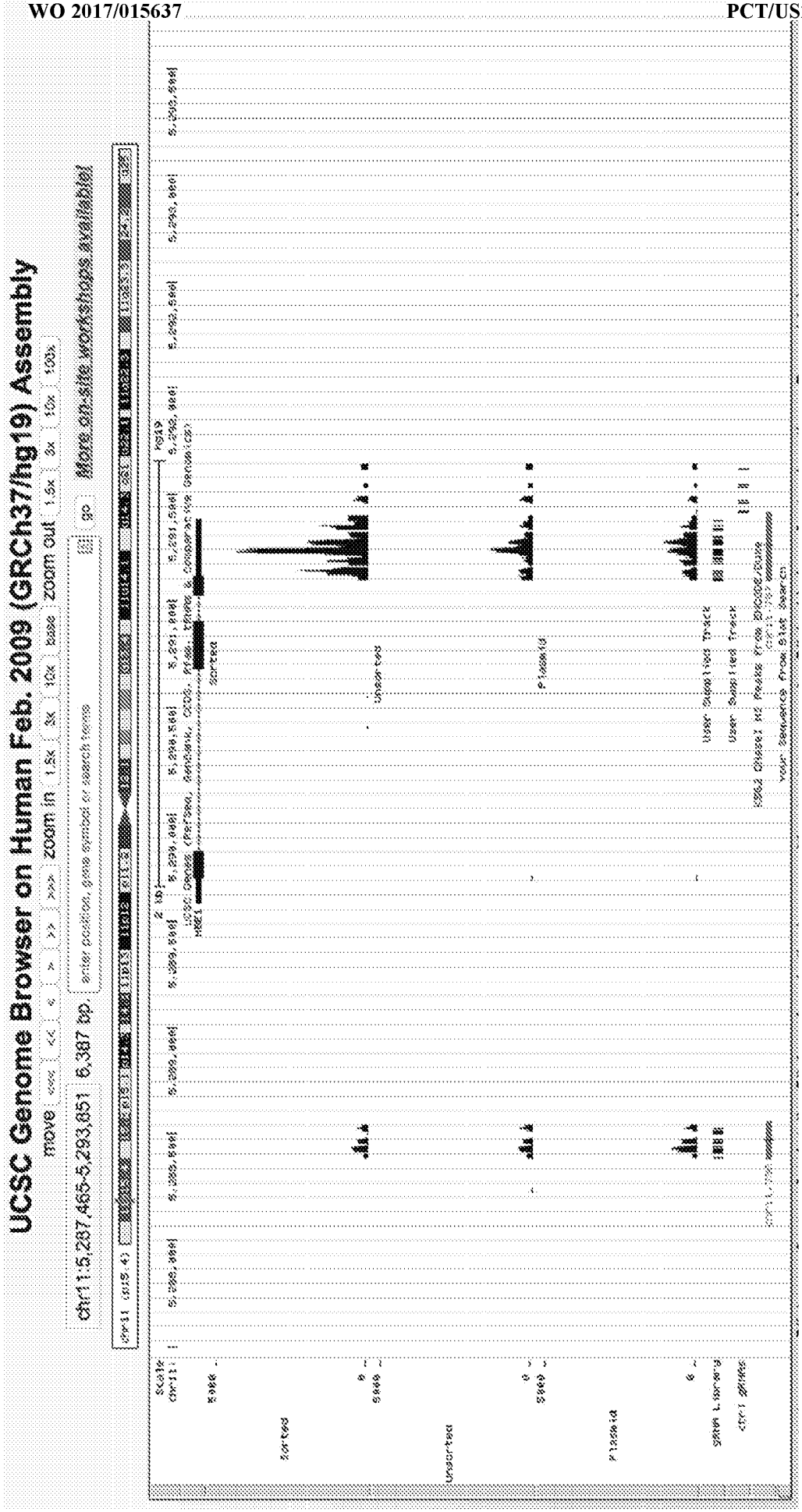

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move
zoom in
zoom out
base
chr11:5,287,465-5,293,851 6,387 bp.
go
More on-site workshops available!

FIG. 2B

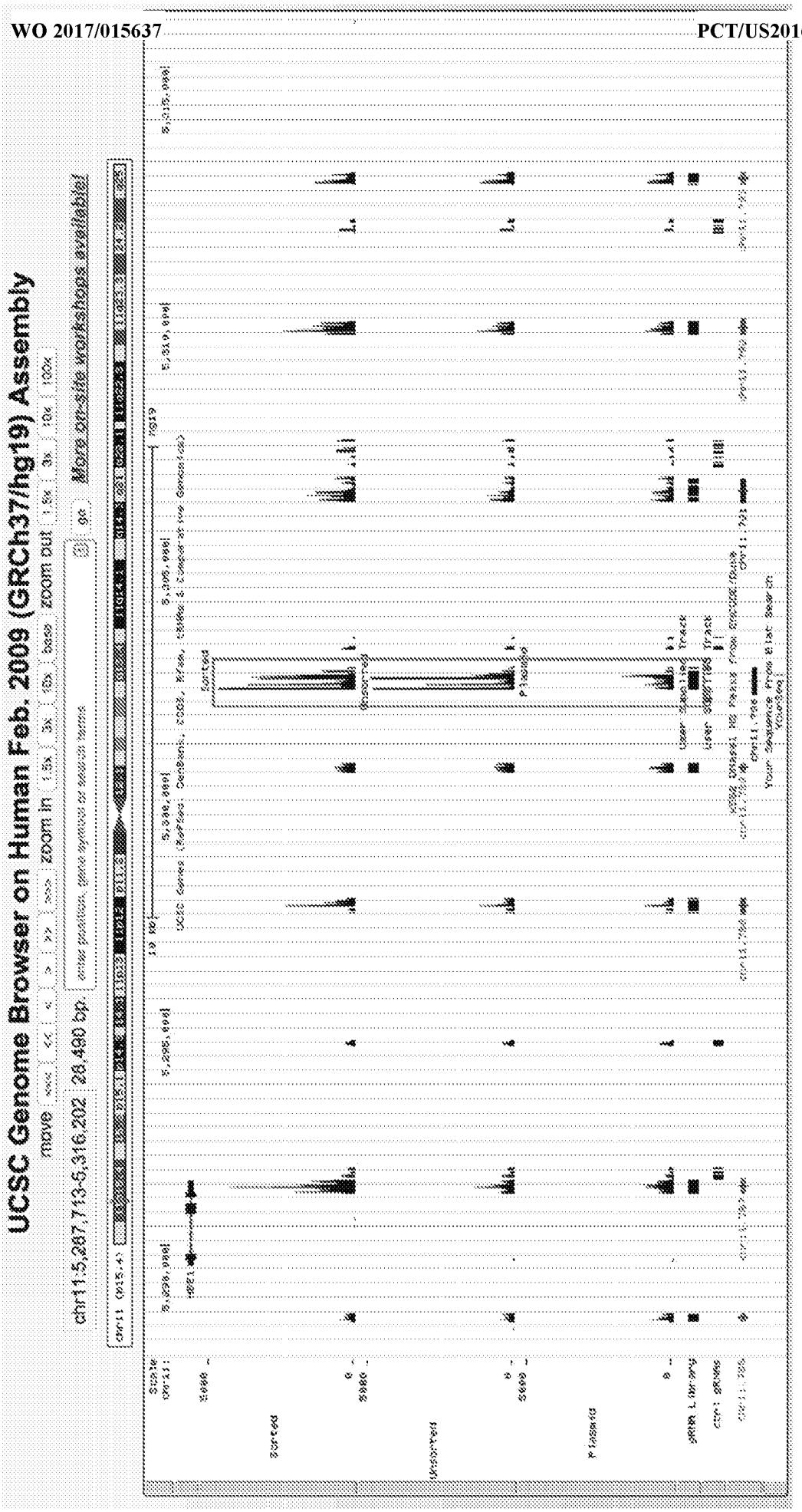

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move
zoom in
zoom out
chr11:5,287,713-5,316,202 28,490 bp
go
More on-site workshops available!

FIG. 2C

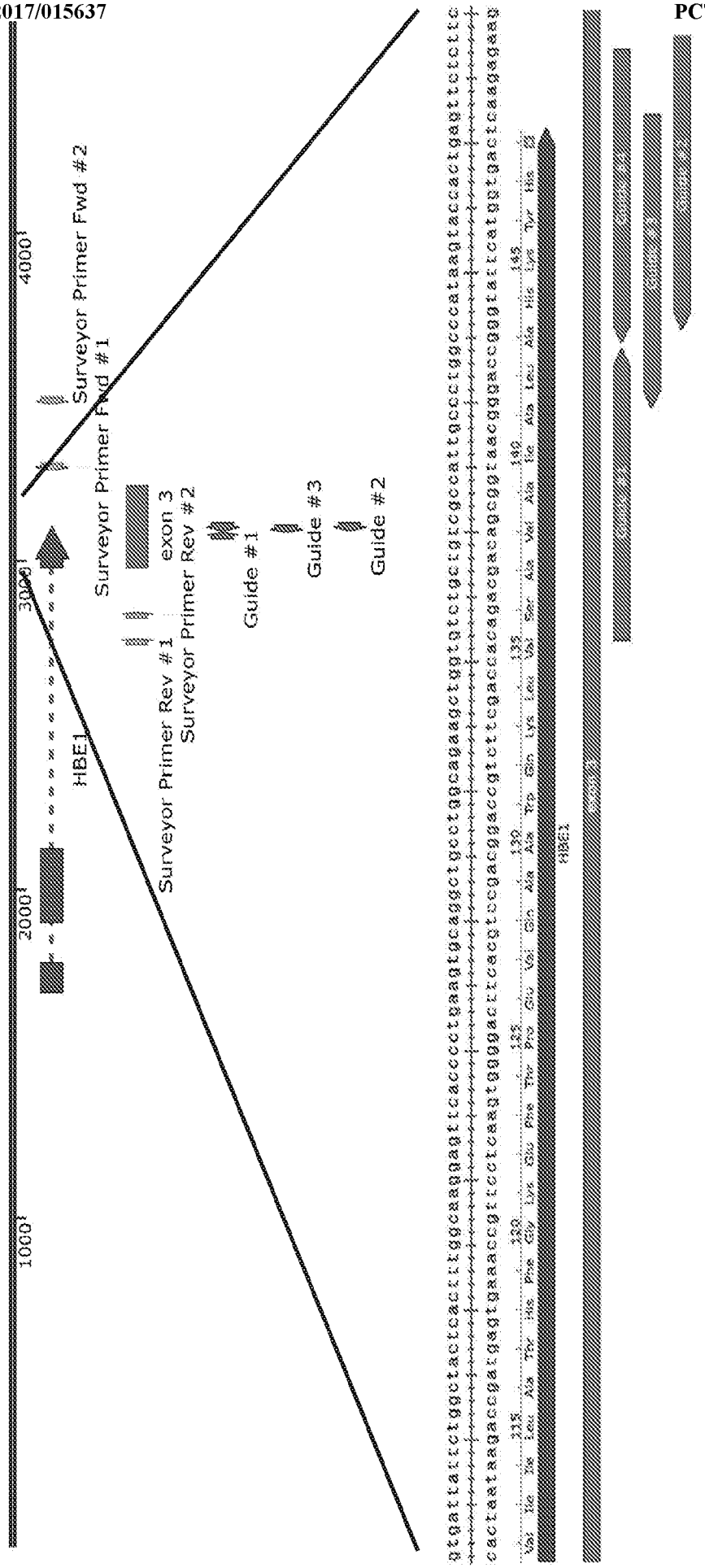

1000
2000
3000
4000
HBE1
Surveyor Primer Fwd #1
Surveyor Primer Fwd #2
Surveyor Primer Rev #1
exon 3
Surveyor Primer Rev #2
Guide #1
Guide #3
Guide #2

FIG. 3A

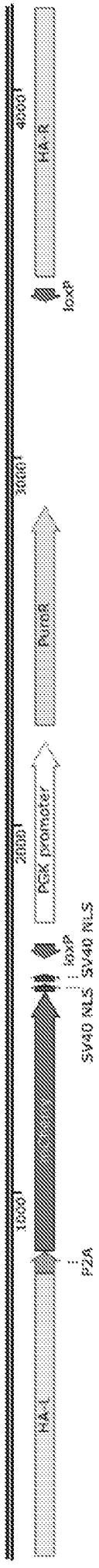

HBE1 Donor
4224 bp
HA-L
T2A
SV40 NLS
loxP
PGK promoter
SV40 NLS
Puro
loxP
HA-R

FIG. 3B

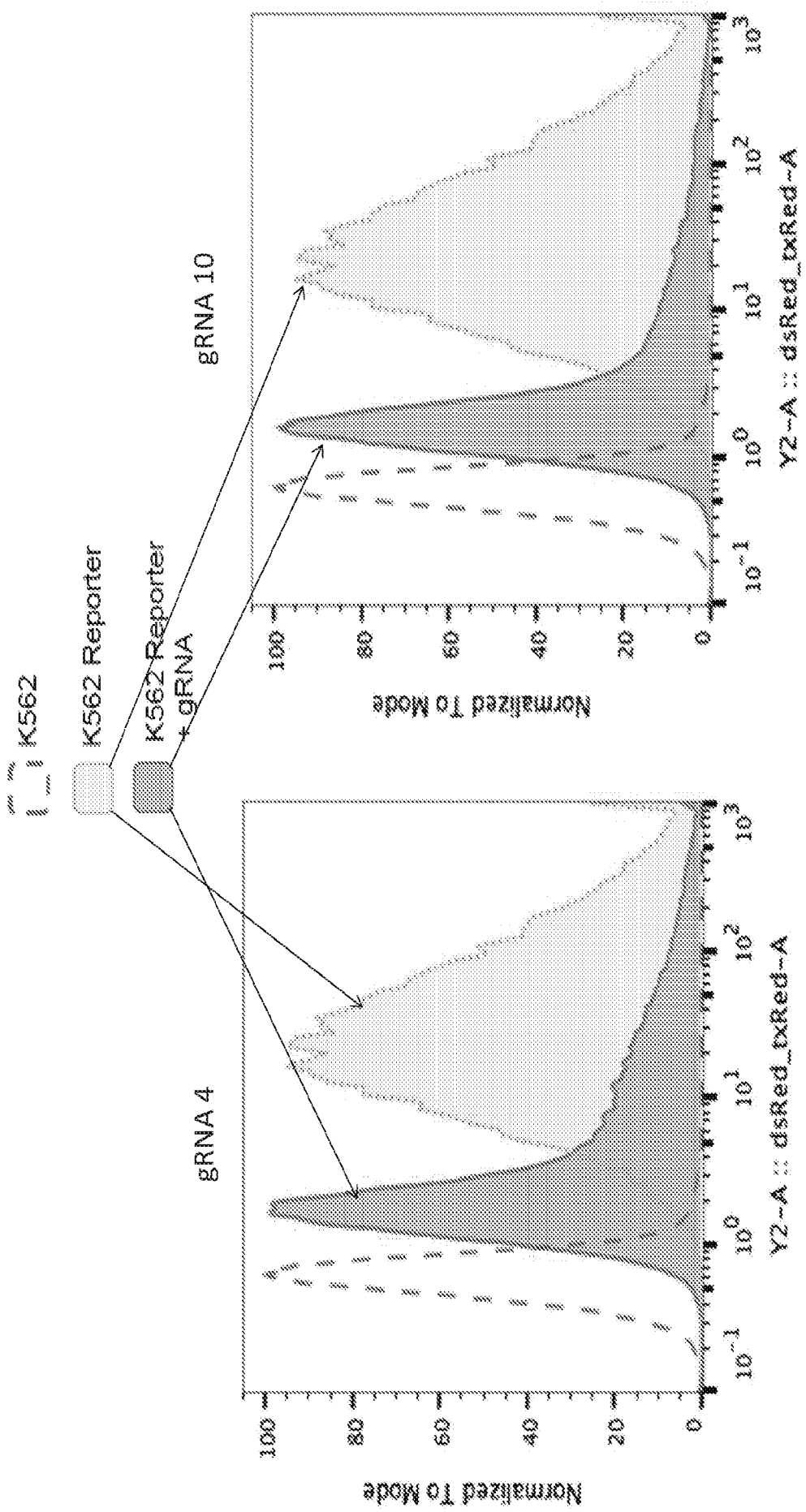

K562
K562 Reporter
K562 Reporter + gRNA
gRNA 10
gRNA 4
Normalized To Mode
Y2-A :: dsRed_txRed-A

FIG. 3C

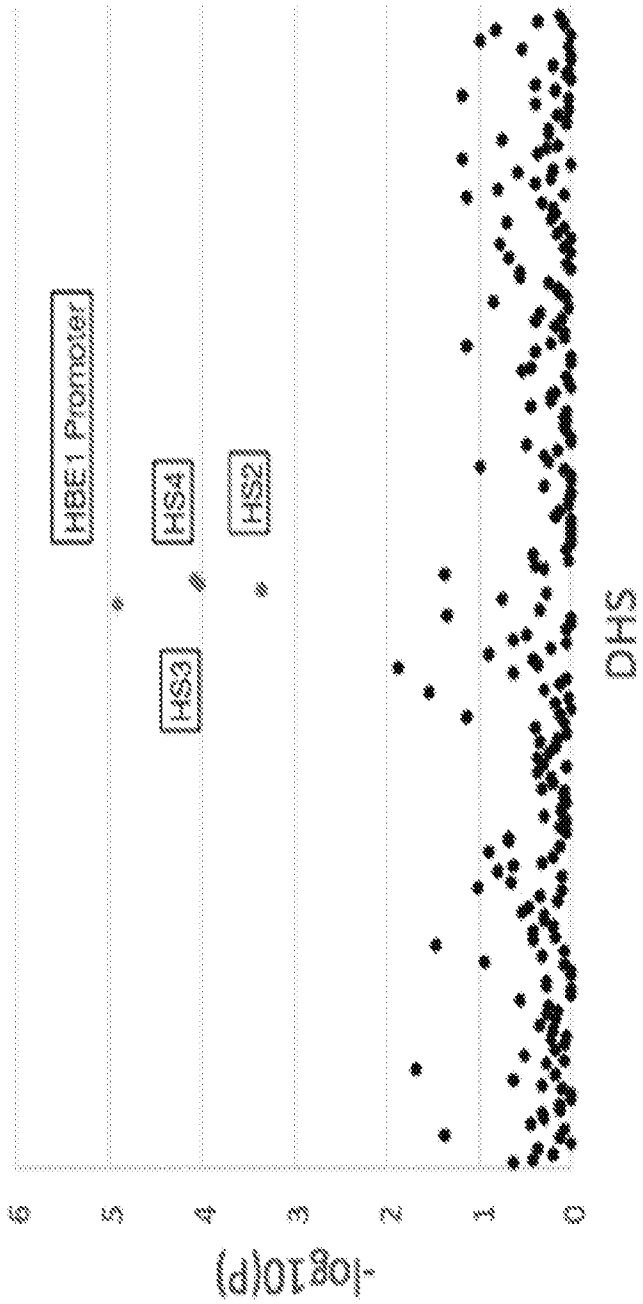


FIG. 4

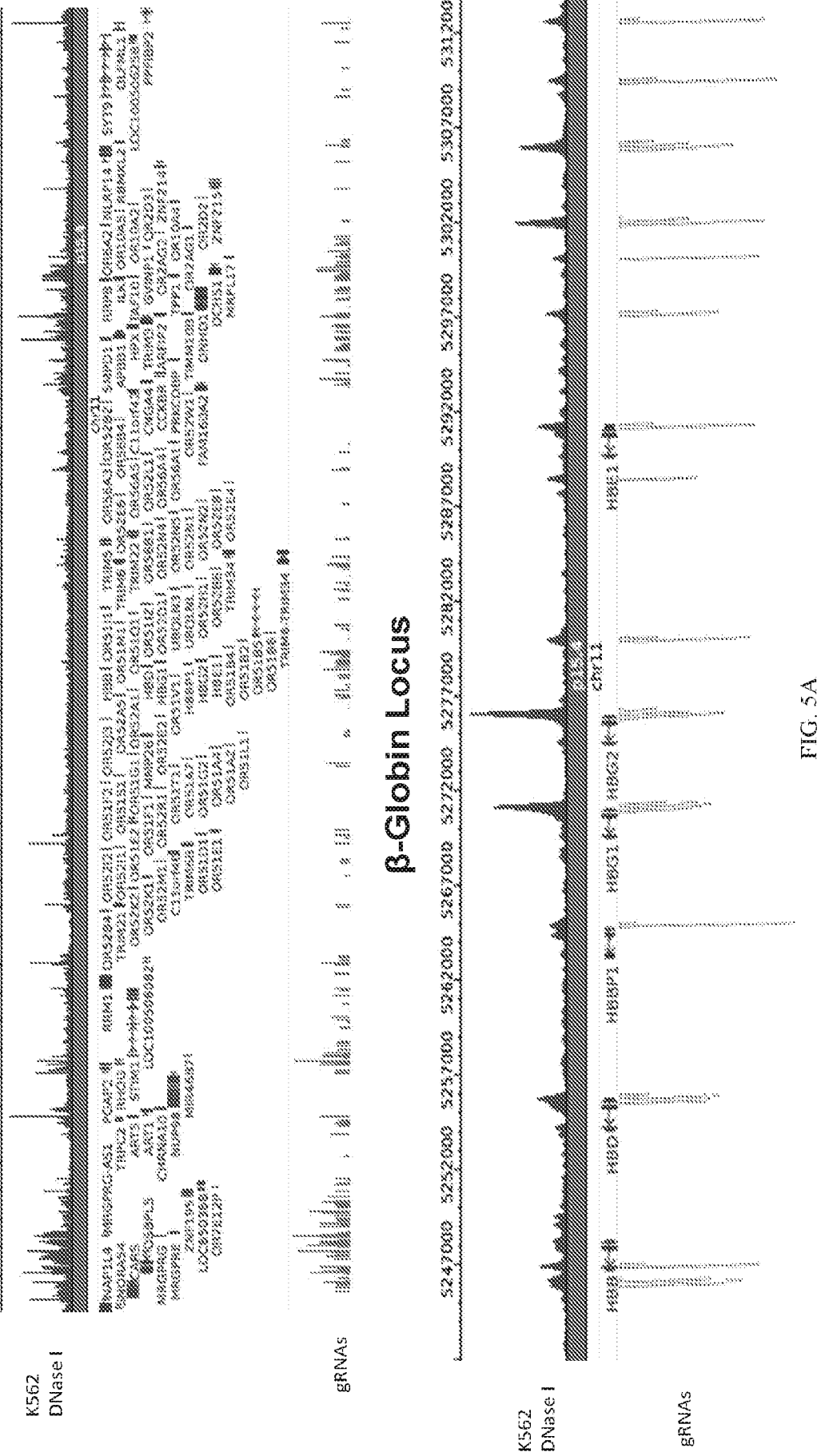

HBE1 Full Library
K562
DNase I
gRNAs
β-Globin Locus
K562
DNase I
gRNAs

FIG. 5A

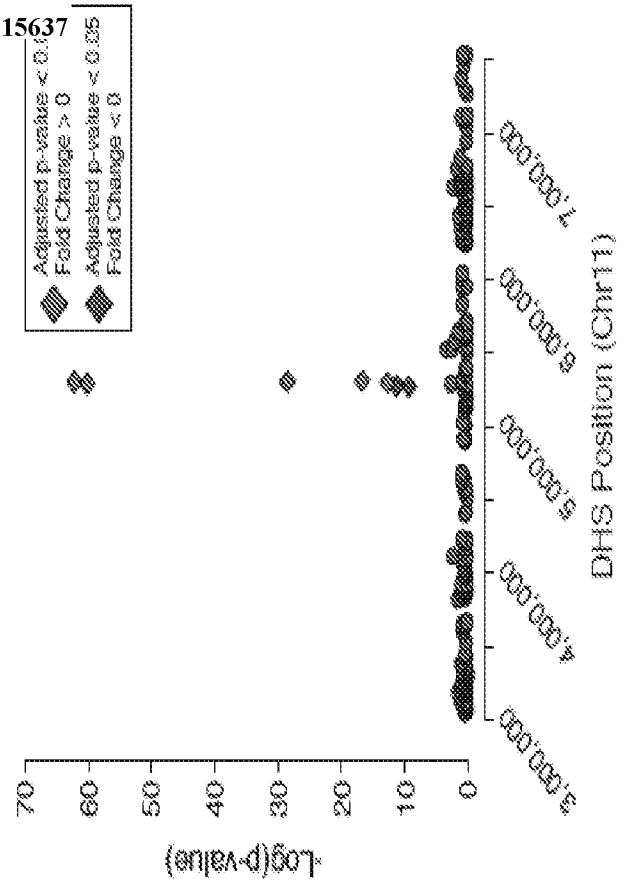

Adjusted p-value < 0.05
Fold Change > 0
Adjusted p-value < 0.05
Fold Change < 0
-Log(p-value)
70
60
50
40
30
20
10
0
3,000,000
4,000,000
5,000,000
6,000,000
7,000,000
DHS Position (Chr11)

FIG. 5C

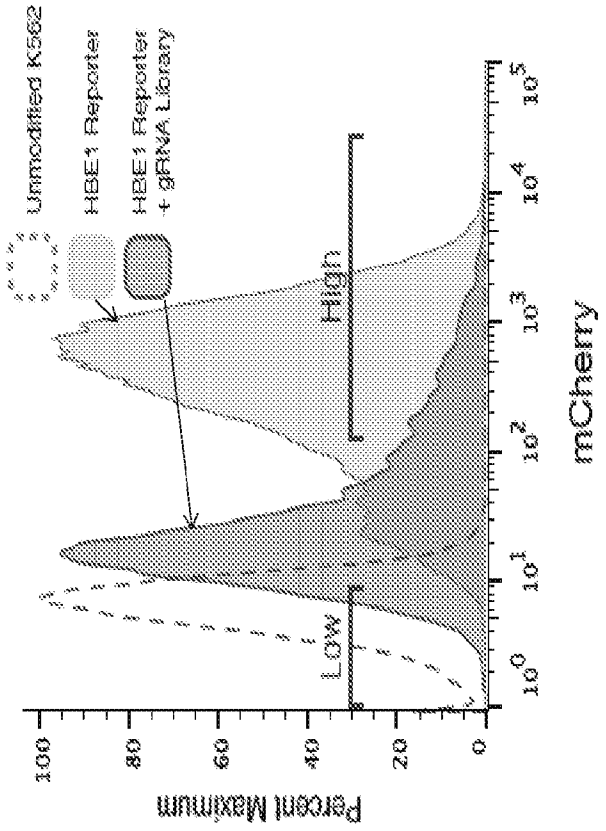

Unmodified K562
HBE1 Reporter
HBE1 Reporter + gRNA Library
Low
High
Percent Maximum
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
10^5
mCherry

FIG. 5B

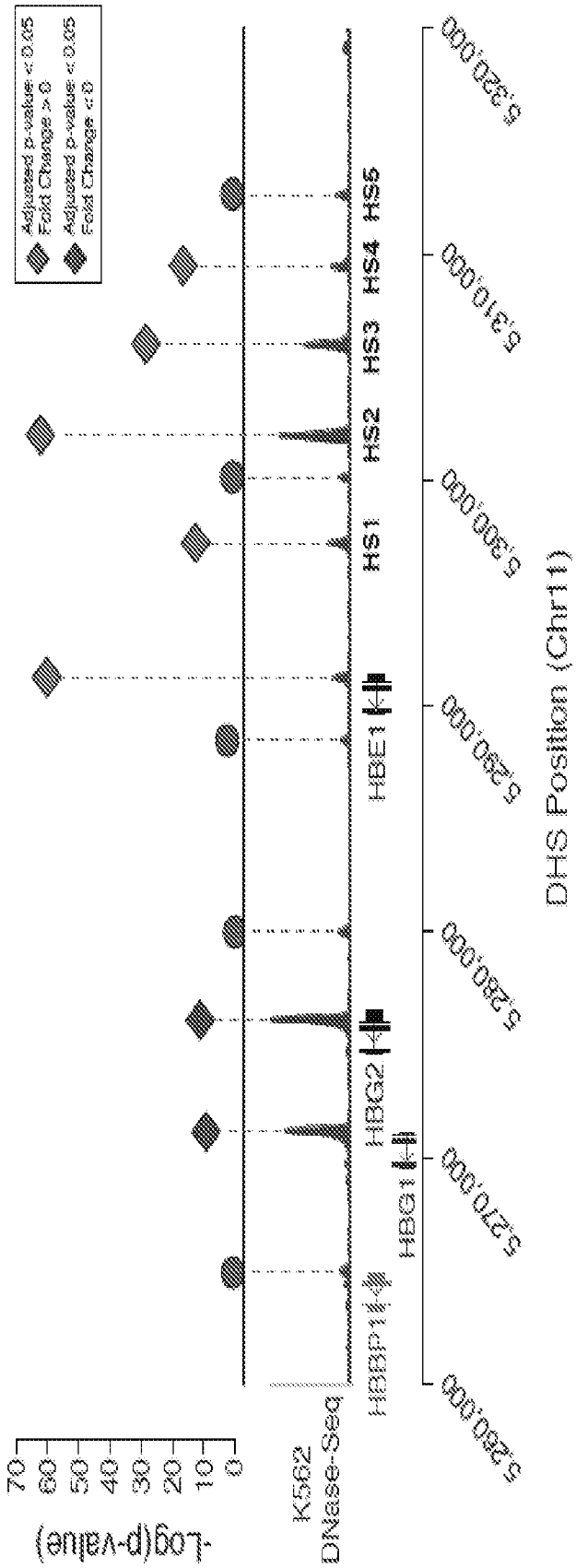

Adjusted p-value <= 0.05
Fold Change >= 0
Adjusted p-value <= 0.05
Fold Change <= 0
HS5
HS4
HS3
HS2
HS1
HBE1
HBG2
HBG1
HBBP1
K562
DNase-Seq
-Log(p-value)
70
60
50
40
30
20
10
0
5,320,000
5,310,000
5,300,000
5,290,000
5,280,000
5,270,000
5,260,000
DHS Position (Chr11)

FIG. 5D

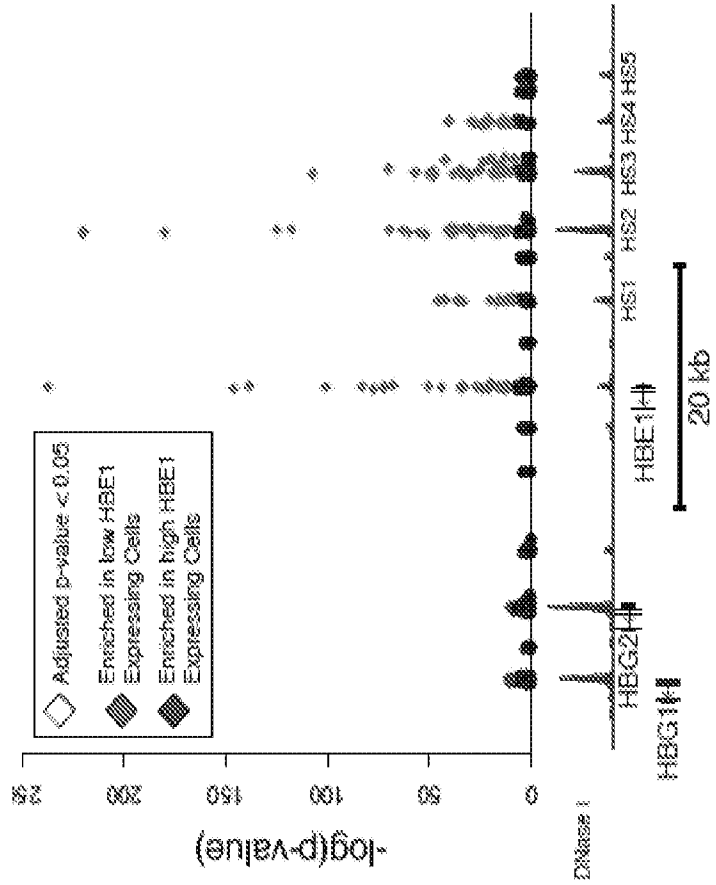

Adjusted p-value < 0.05
Enriched in low HBE1 Expressing Cells
Enriched in high HBE1 Expressing Cells
-log(p-value)
0
50
100
150
200
250
DNaseI
HBG2
HBG1
HBE1
20 kb
HS1
HS2
HS3
HS4
HS5

FIG. 6B

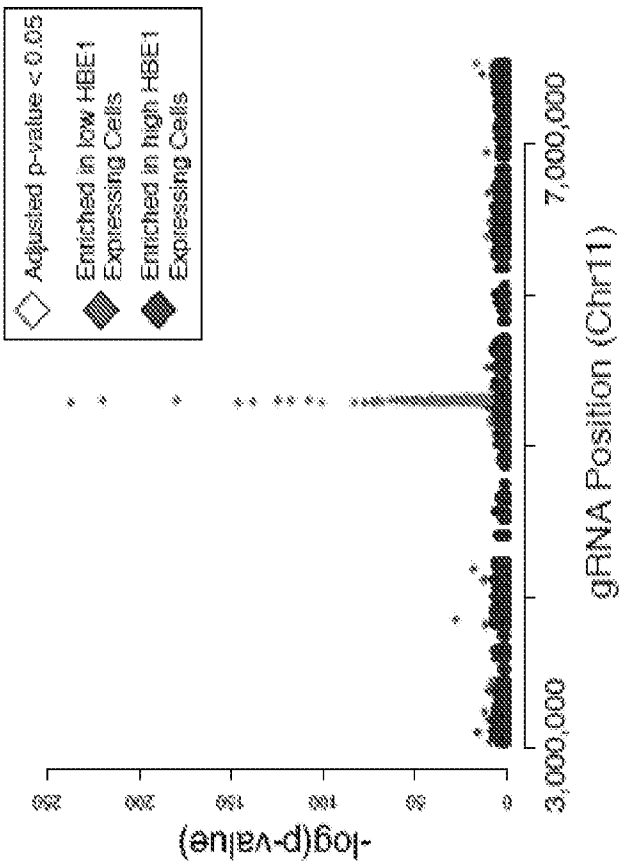

Adjusted p-value < 0.05
Enriched in low HBE1 Expressing Cells
Enriched in high HBE1 Expressing Cells
-log(p-value)
0
50
100
150
200
3,000,000
7,000,000
gRNA Position (Chr11)

FIG. 6A

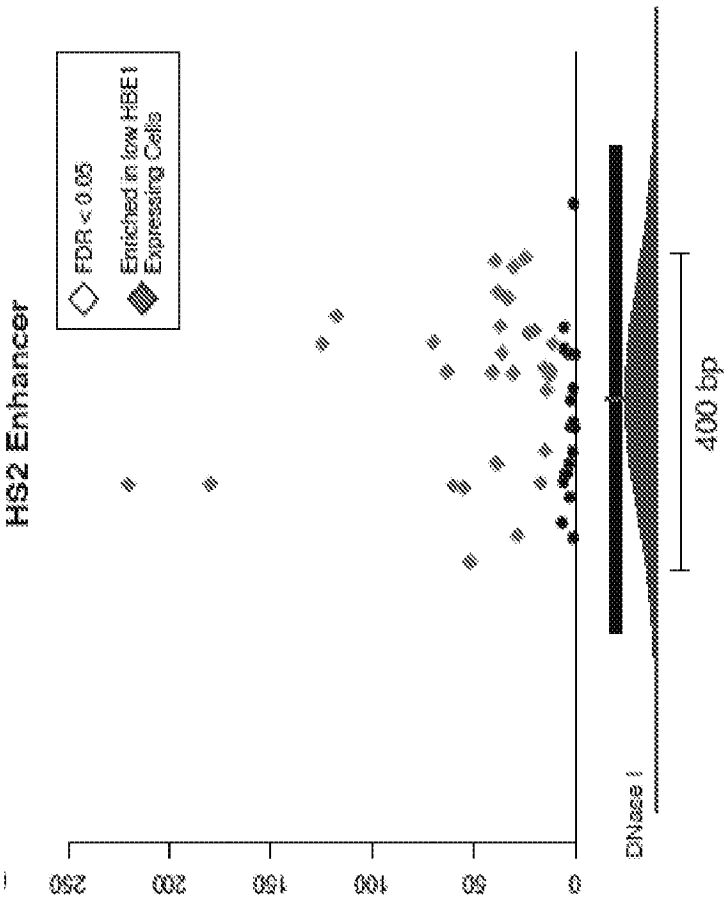

HS2 Enhancer
FDR < 0.05
Enriched in low HBE1 Expressing Cells
250
200
150
100
50
0
DNase I
400 bp

FIG. 6C

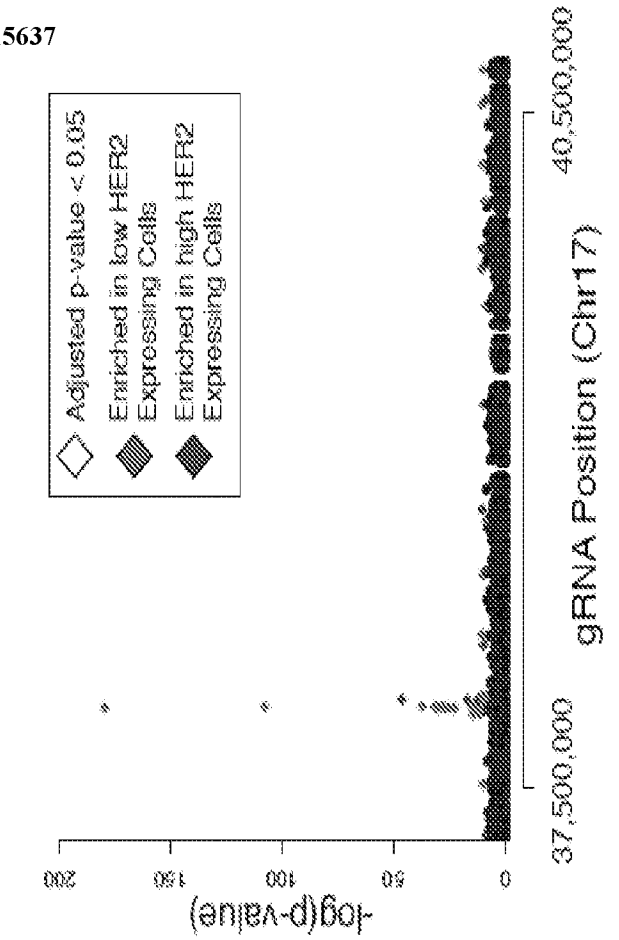

Adjusted p-value < 0.05
Enriched in low HER2 Expressing Cells
Enriched in high HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
37,500,000
40,500,000
gRNA Position (Chr17)

FIG. 7B

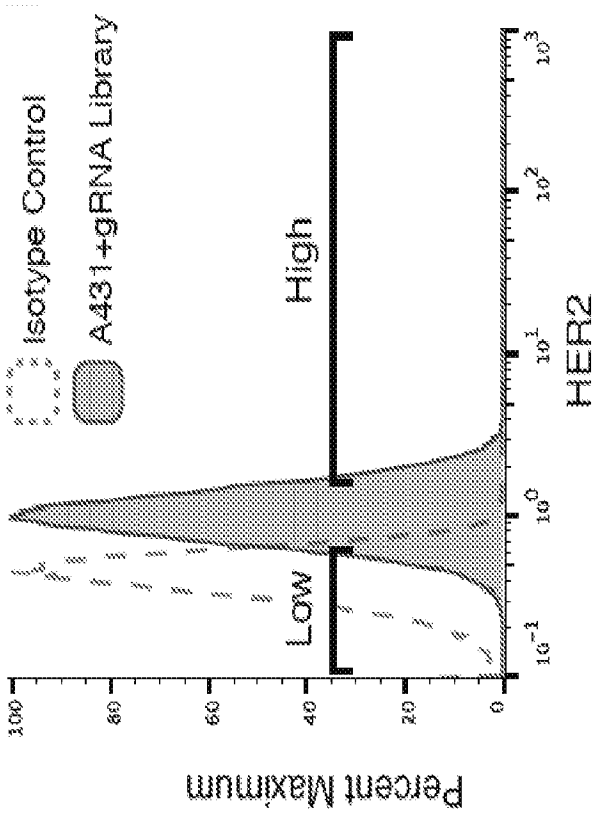

Isotype Control
A431+gRNA Library
Low
High
Percent Maximum
0
20
40
60
80
100
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
HER2

FIG. 7A

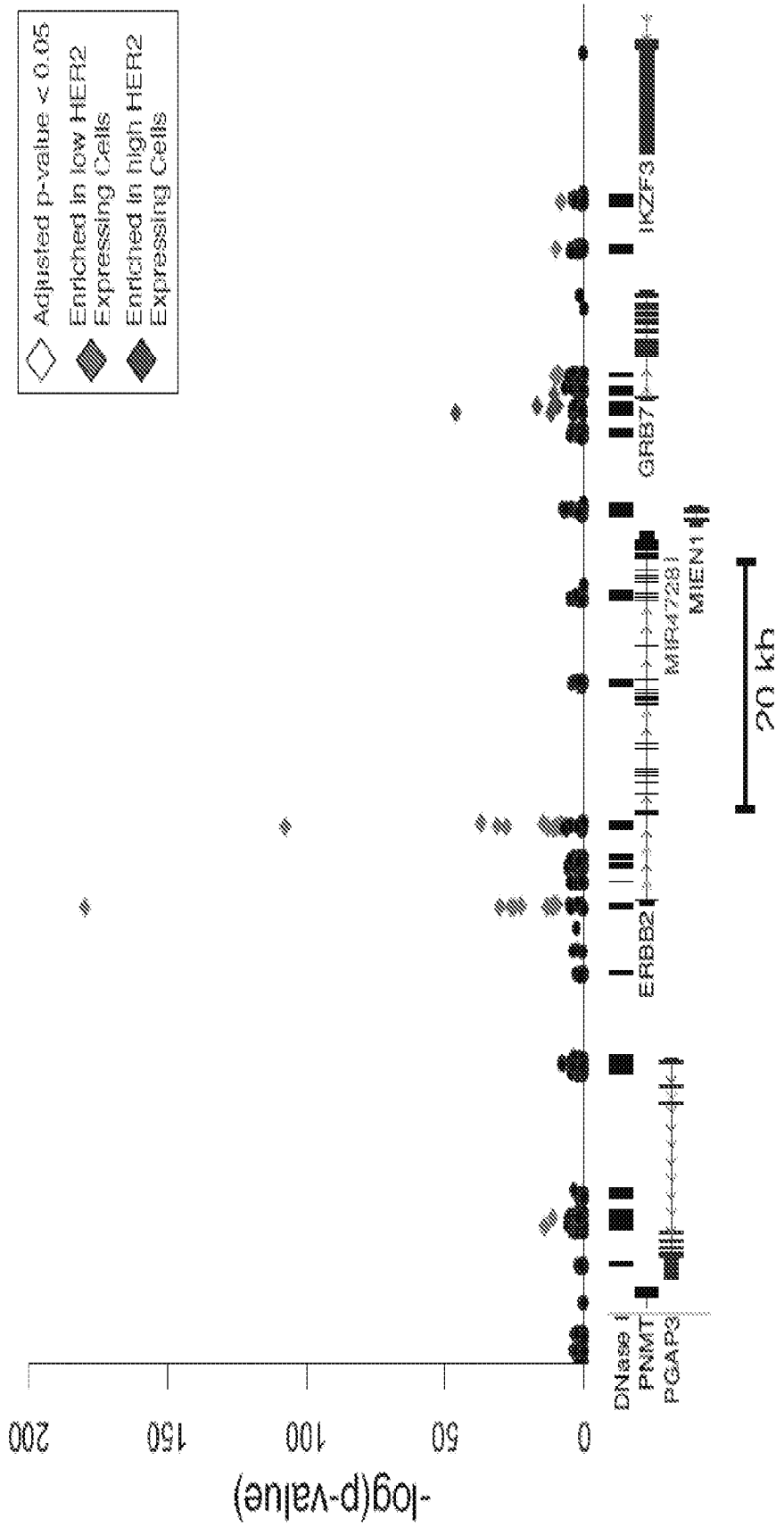

Adjusted p-value < 0.05
Enriched in low HER2 Expressing Cells
Enriched in high HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
DNase I
PNMT
PGAP3
ERBB2
MIR4728
MIEN1
GRB7
IKZF3
20 kb

FIG. 7C

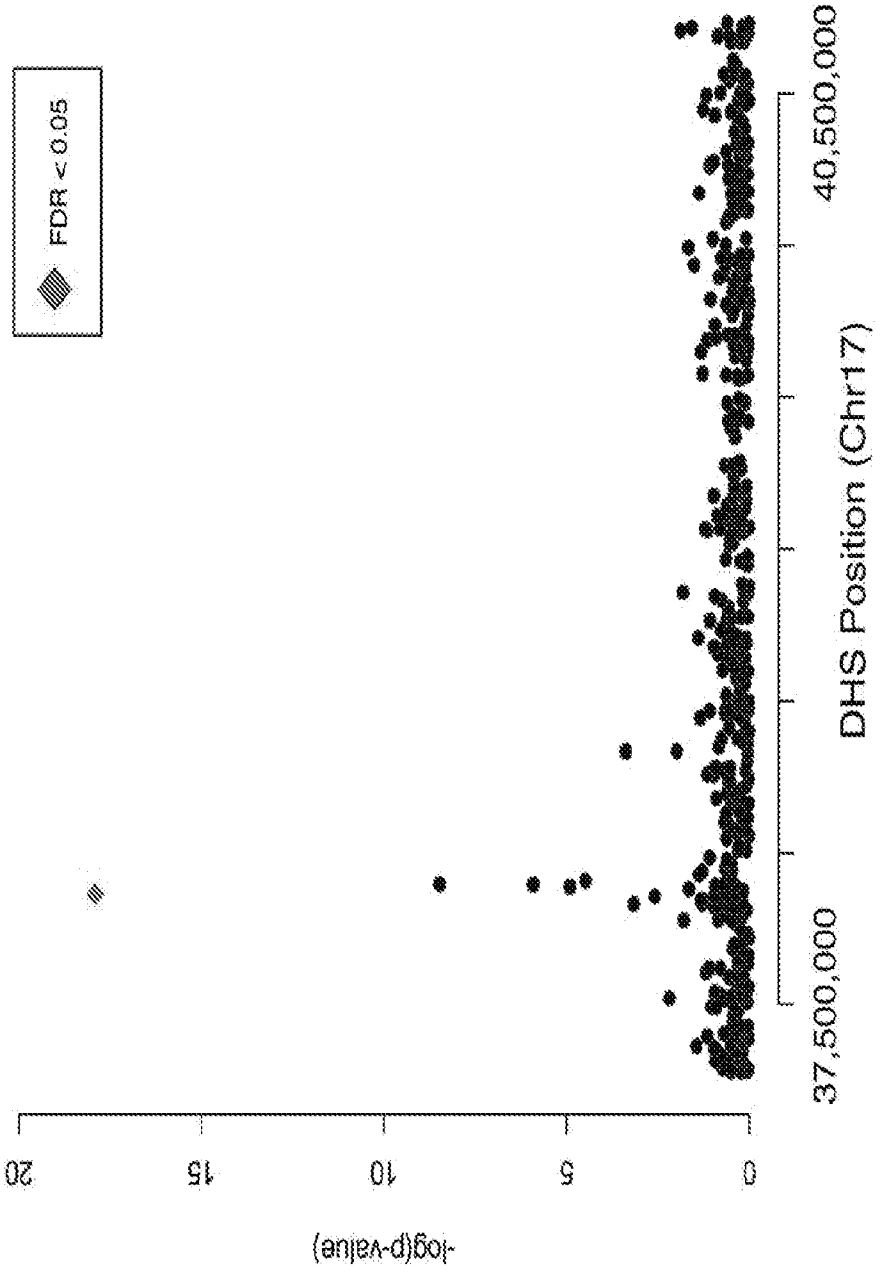

FDR < 0.05
-log(p-value)
0
5
10
15
20
37,500,000
40,500,000
DHS Position (Chr17)

FIG. 7D

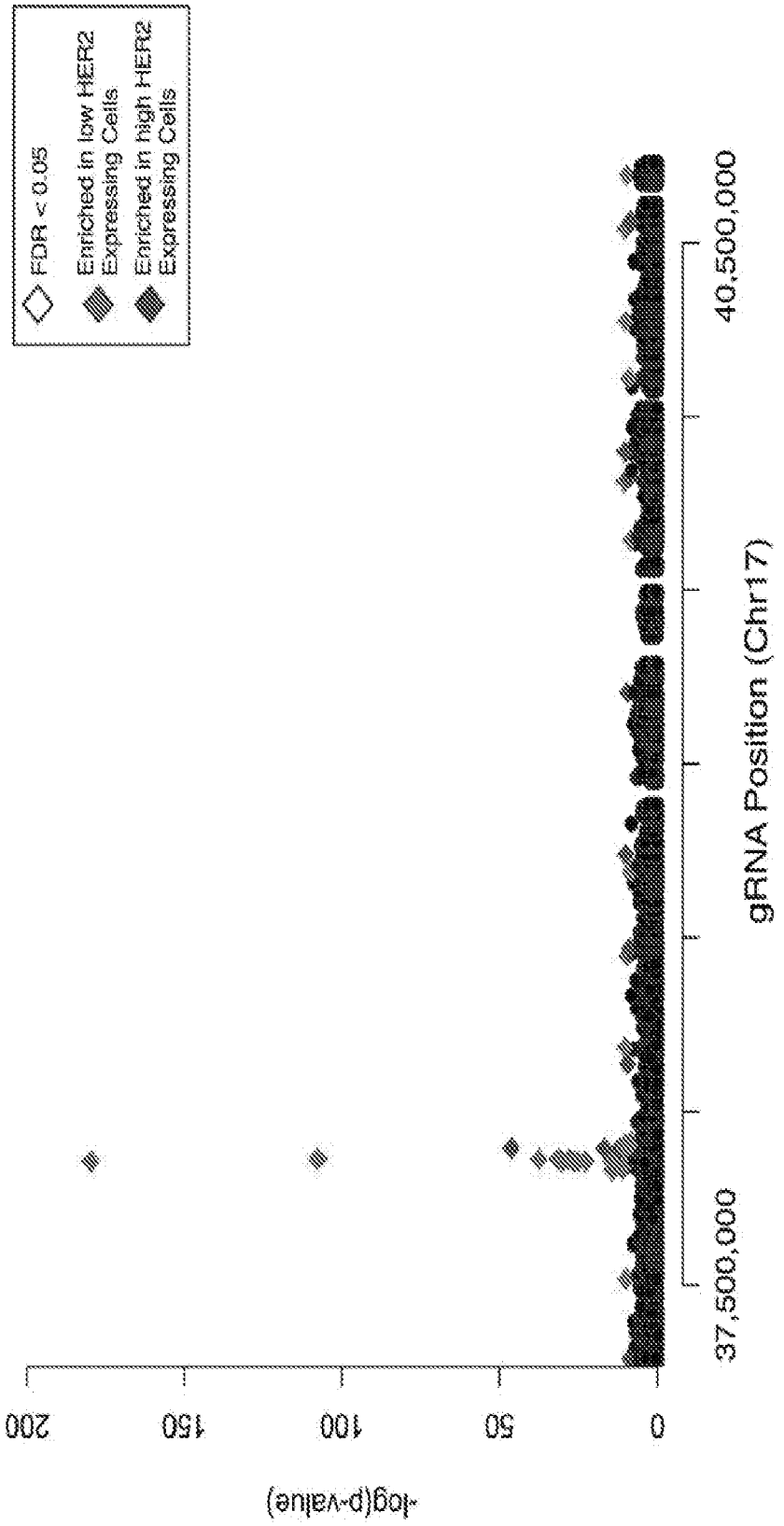

FDR < 0.05
Enriched in low HER2 Expressing Cells
Enriched in high HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
37,500,000
40,500,000
gRNA Position (Chr17)

FIG. 8A

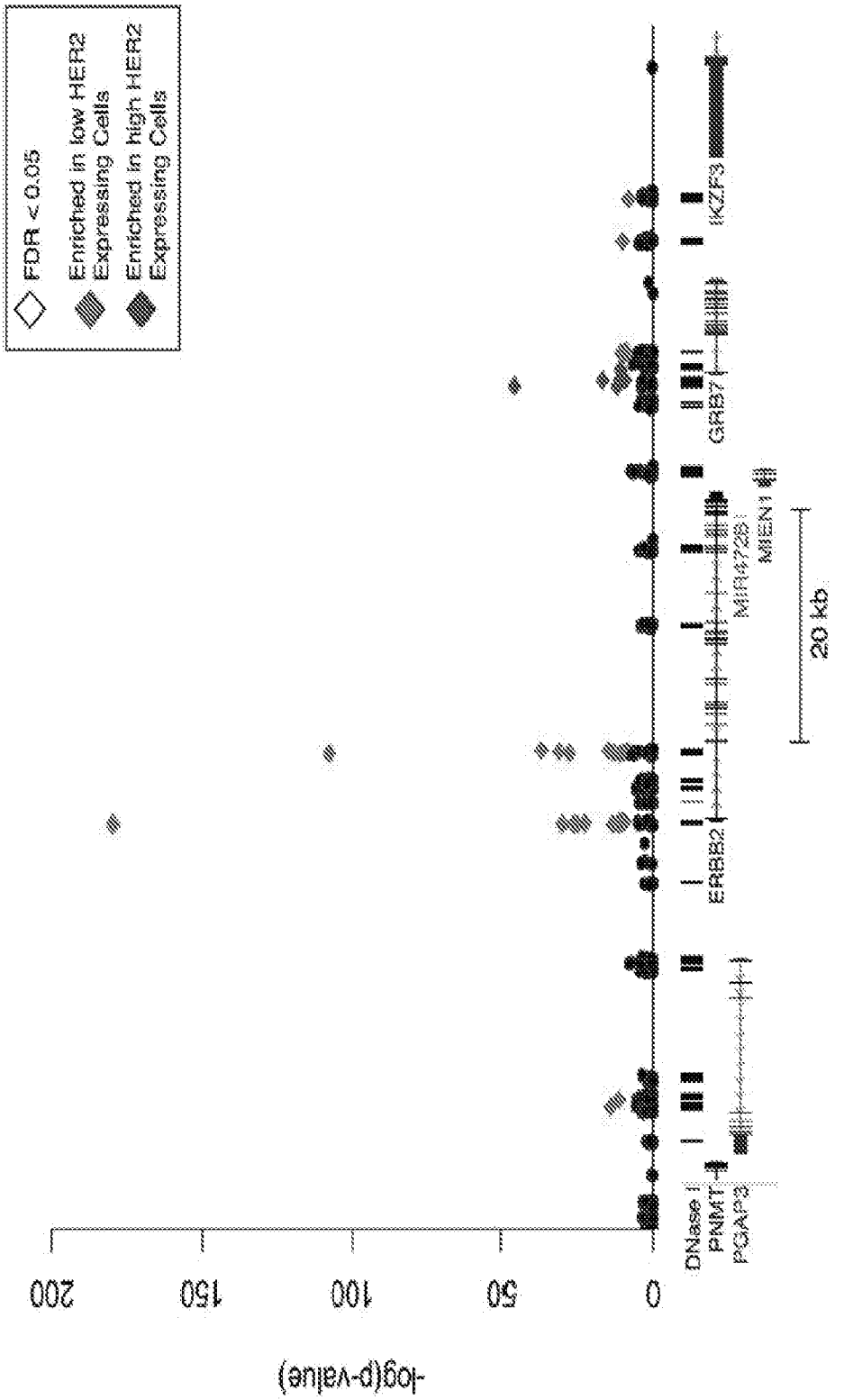

FDR < 0.05
Enriched in low HER2 Expressing Cells
Enriched in high HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
DNase I
PNMT
PGAP3
ERBB2
MIR4728
MIEN1
GRB7
IKZF3
20 kb

FIG. 8B

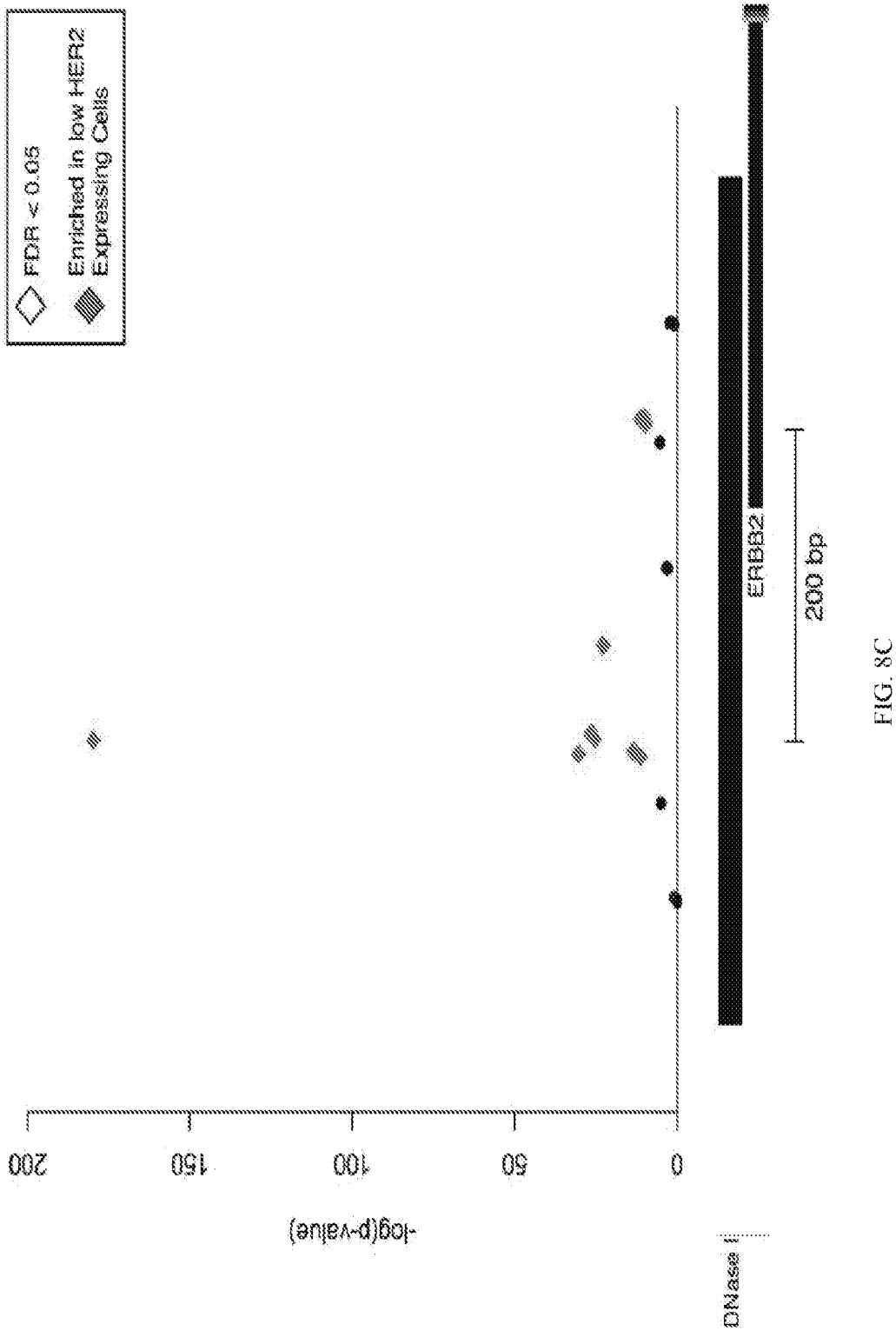

FDR < 0.05
Enriched in low HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
ERBB2
200 bp
DNase I

FIG. 8C

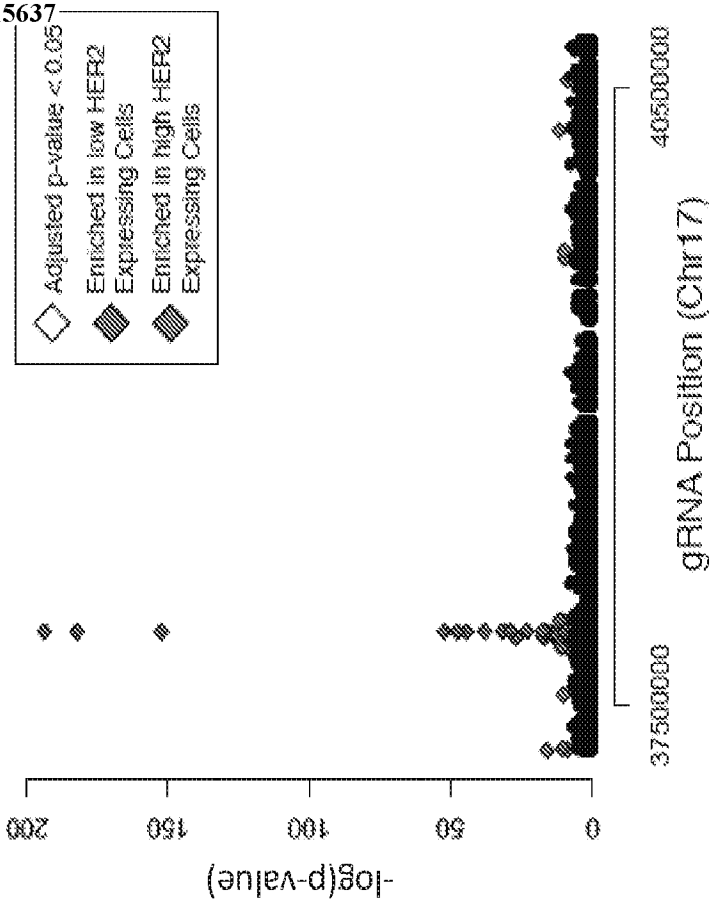

Adjusted p-value < 0.05
Enriched in low HER2 Expressing Cells
Enriched in high HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
37500000
40500000
gRNA Position (Chr17)

FIG. 9B

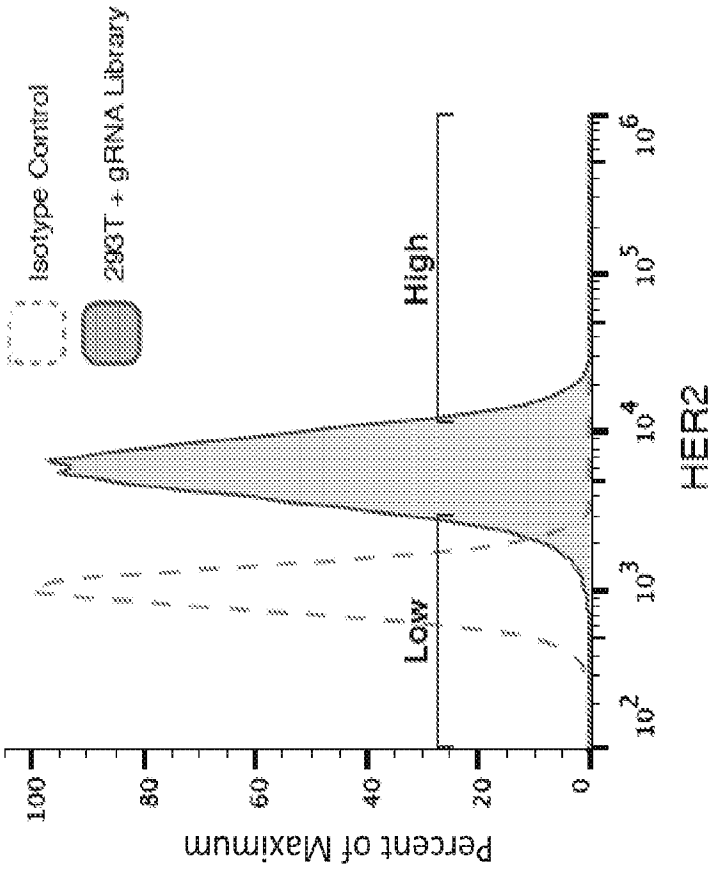

Isotype Control
293T + gRNA Library
Low
High
Percent of Maximum
0
20
40
60
80
100
10^2
10^3
10^4
10^5
10^6
HER2

FIG. 9A

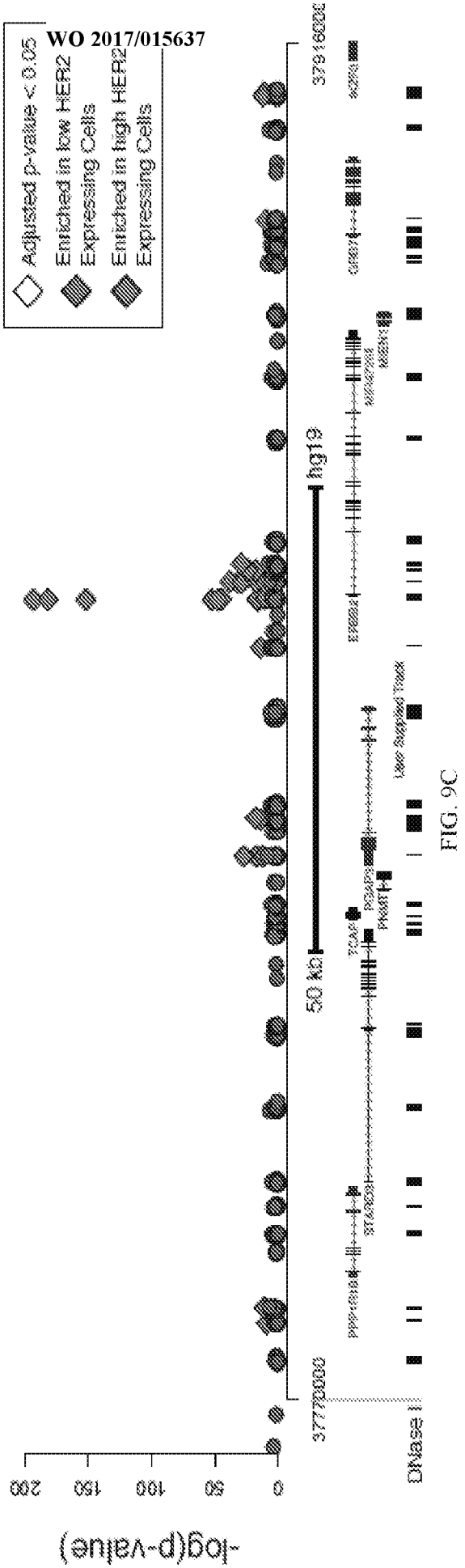

Adjusted p-value < 0.05
Enriched in low HER2 Expressing Cells
Enriched in high HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
37700000
37716000
hg19
50 kb
DNase I

FIG. 9C

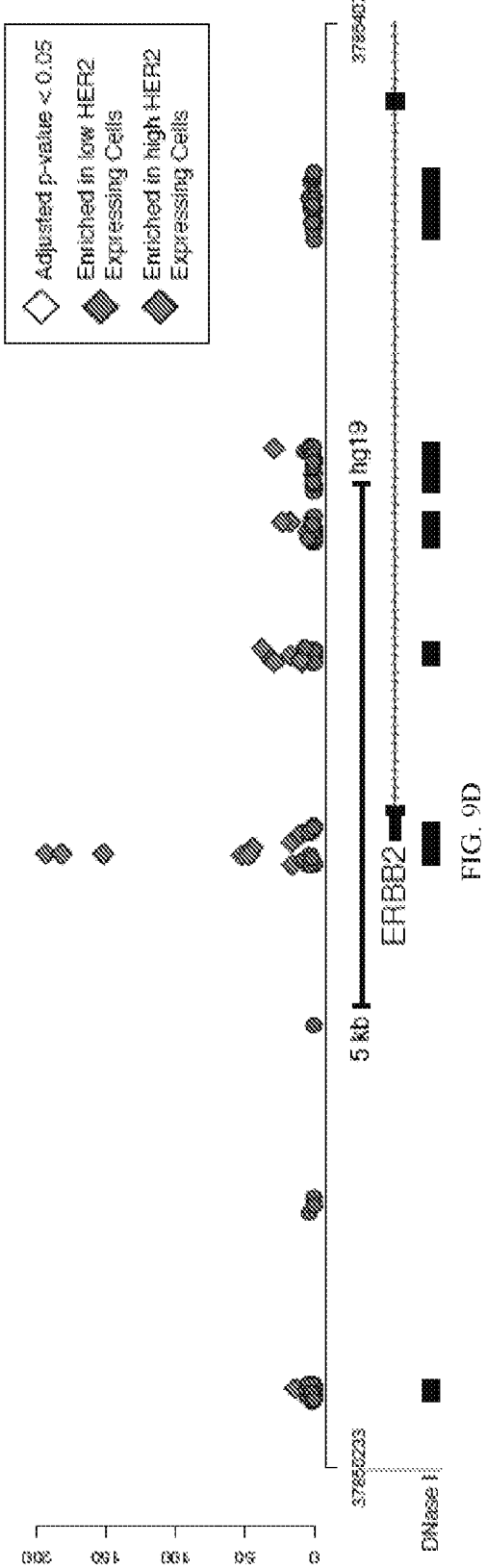

Adjusted p-value < 0.05
Enriched in low HER2 Expressing Cells
Enriched in high HER2 Expressing Cells
-log(p-value)
0
50
100
150
200
37655233
37664074
hg19
5 kb
ERBB2
DNase I

FIG. 9D

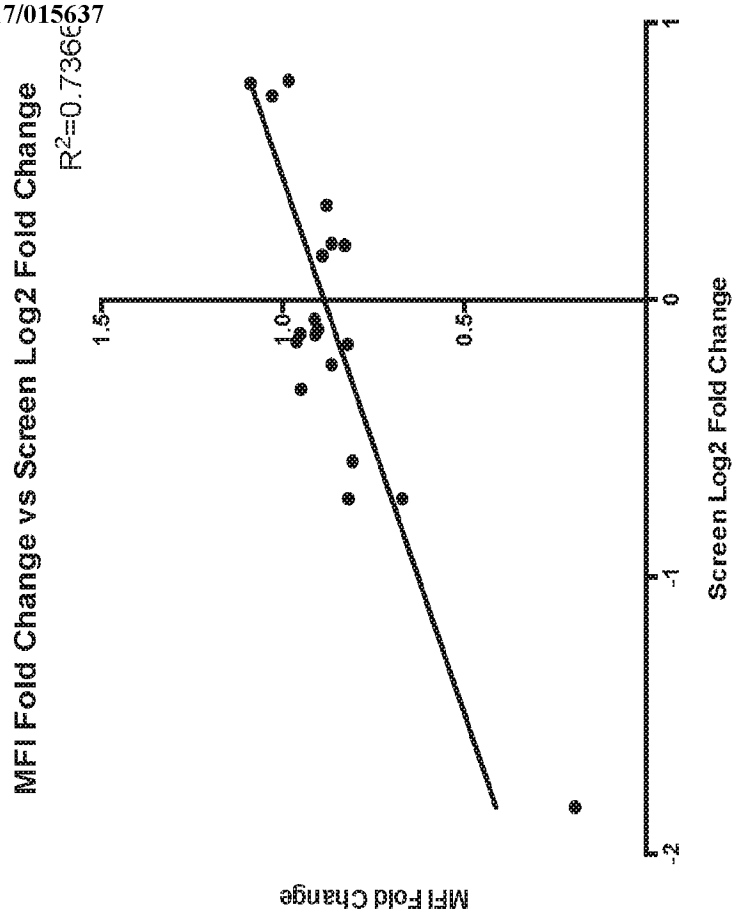


FIG. 10B

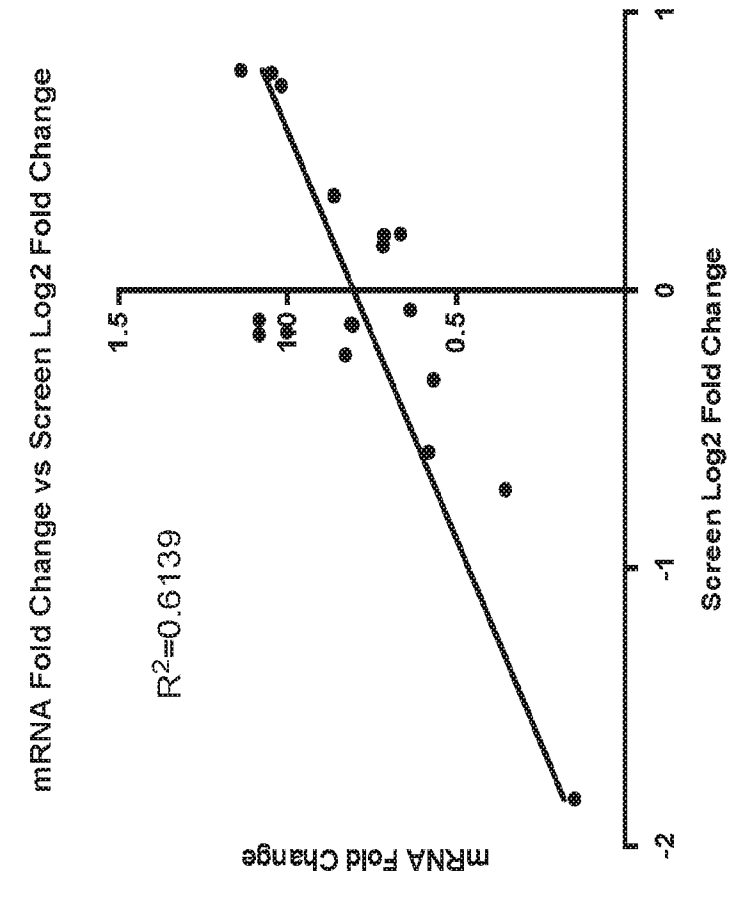


FIG. 10A

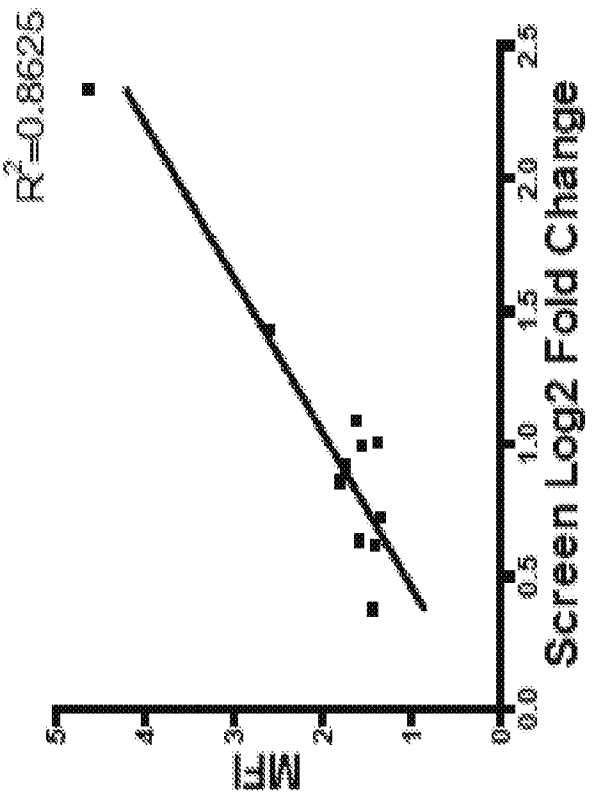


FIG. 10D

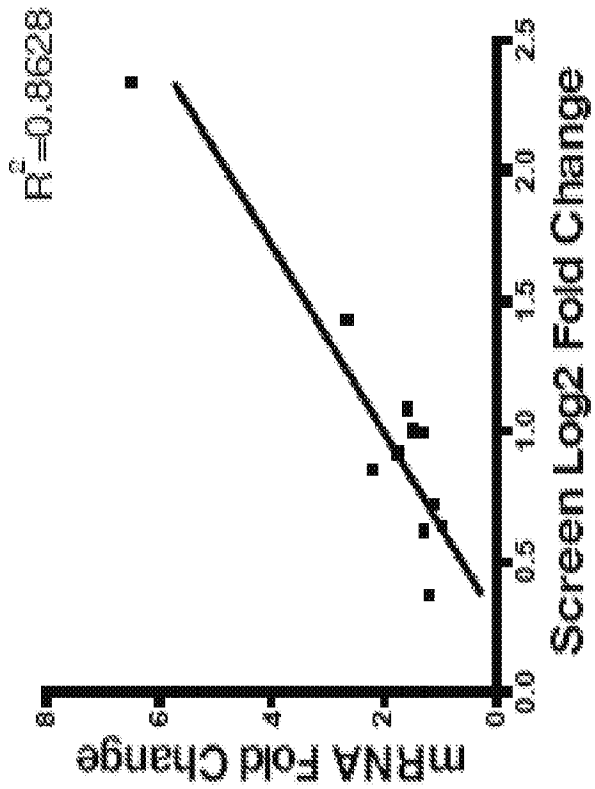


FIG. 10C

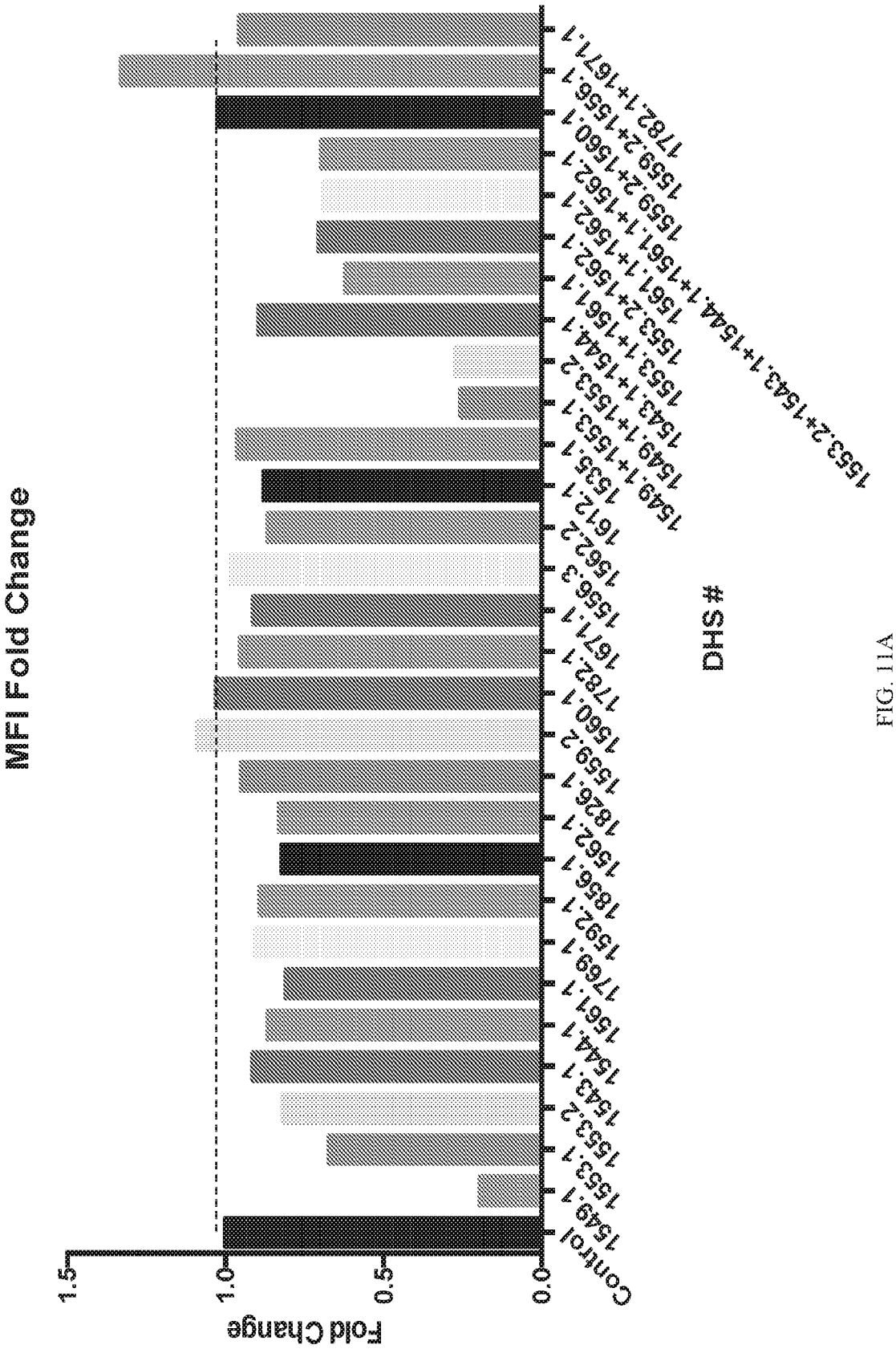

MFI Fold Change
Fold Change
1.5
1.0
0.5
0.0
DHS #
Control
1549.1
1553.1
1553.2
1543.1
1554.1
1571.1
1769.1
1592.1
1856.1
1562.1
1826.1
1559.2
1560.1
1782.1
1671.1
1556.3
1562.2
1612.1
1535.1
1549.1+1553.1
1543.1+1553.2
1553.1+1554.1
1553.2+1561.1
1561.1+1562.1
1559.1+1562.1
1559.2+1560.1
1782.1+1556.1
1553.2+1543.1+1554.1+1561.1+1562.1

FIG. 11A

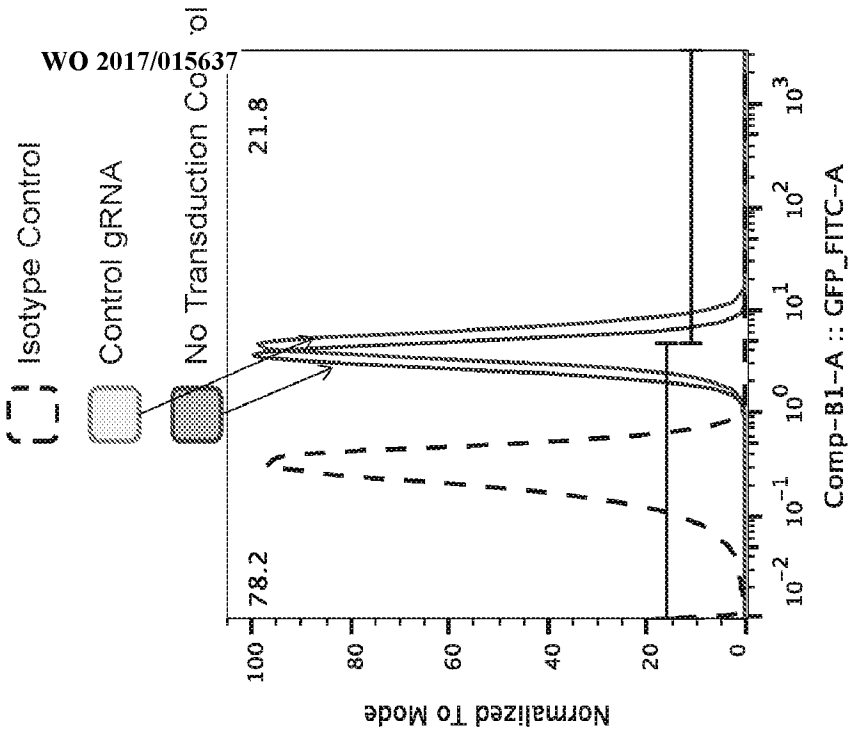


FIG. 11C

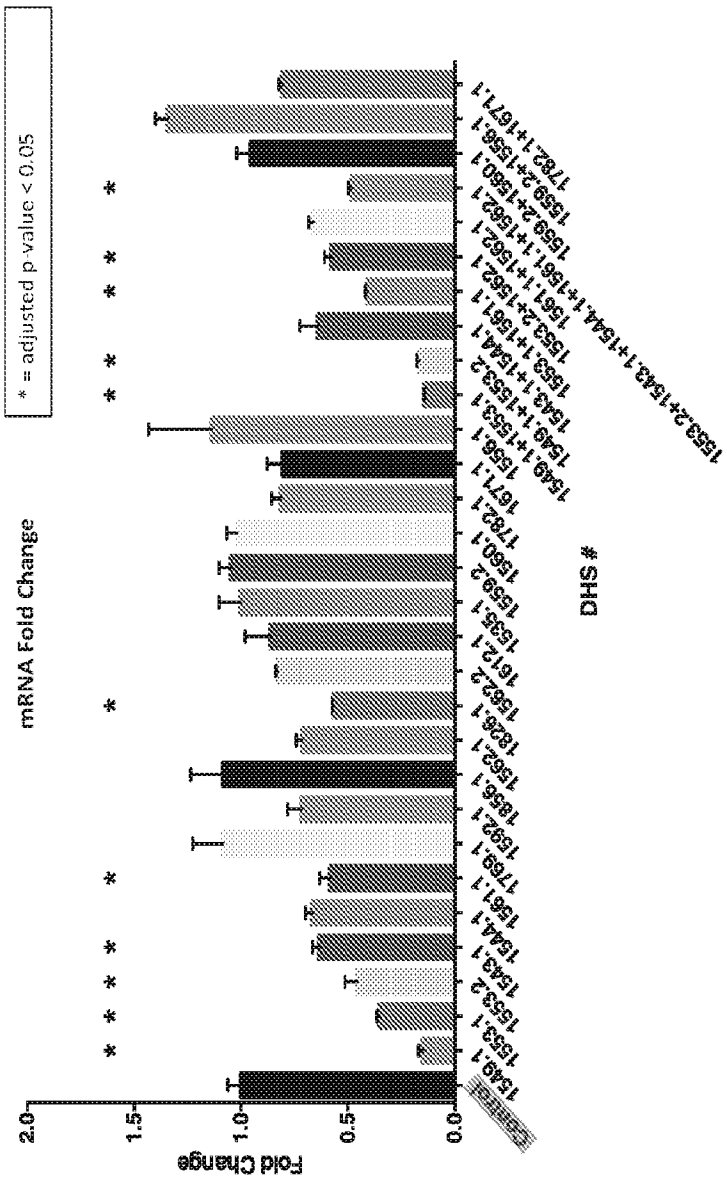


FIG. 11B

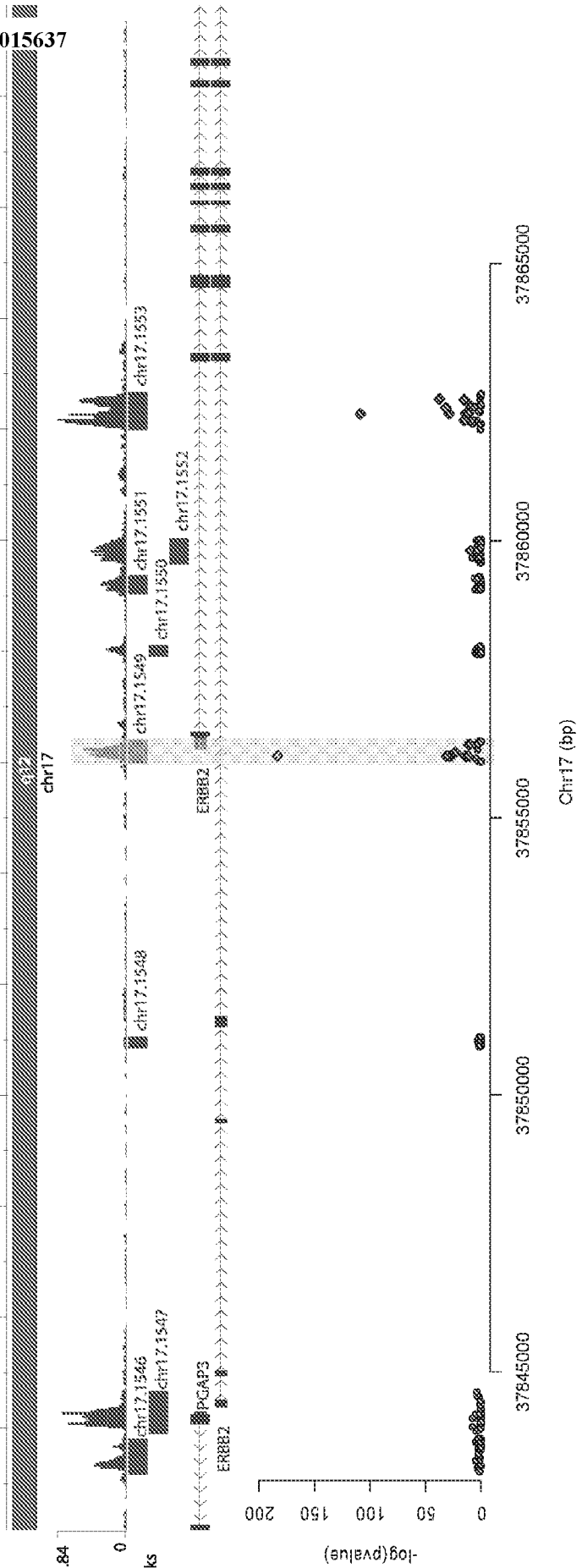

37844000
37846000
37848000
37850000
37852000
37854000
37856000
37858000
37860000
37862000
37864000
37866000
37868000
chr17
113.84
0
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
chr17.1546
chr17.1547
chr17.1548
chr17.1549
chr17.1550
chr17.1551
chr17.1552
chr17.1553
PGAP3
ERBB2
ERBB2
200
150
100
50
0
-log(pvalue)
37845000
37850000
37855000
37860000
37865000
Chr17 (bp)

FIG. 12A

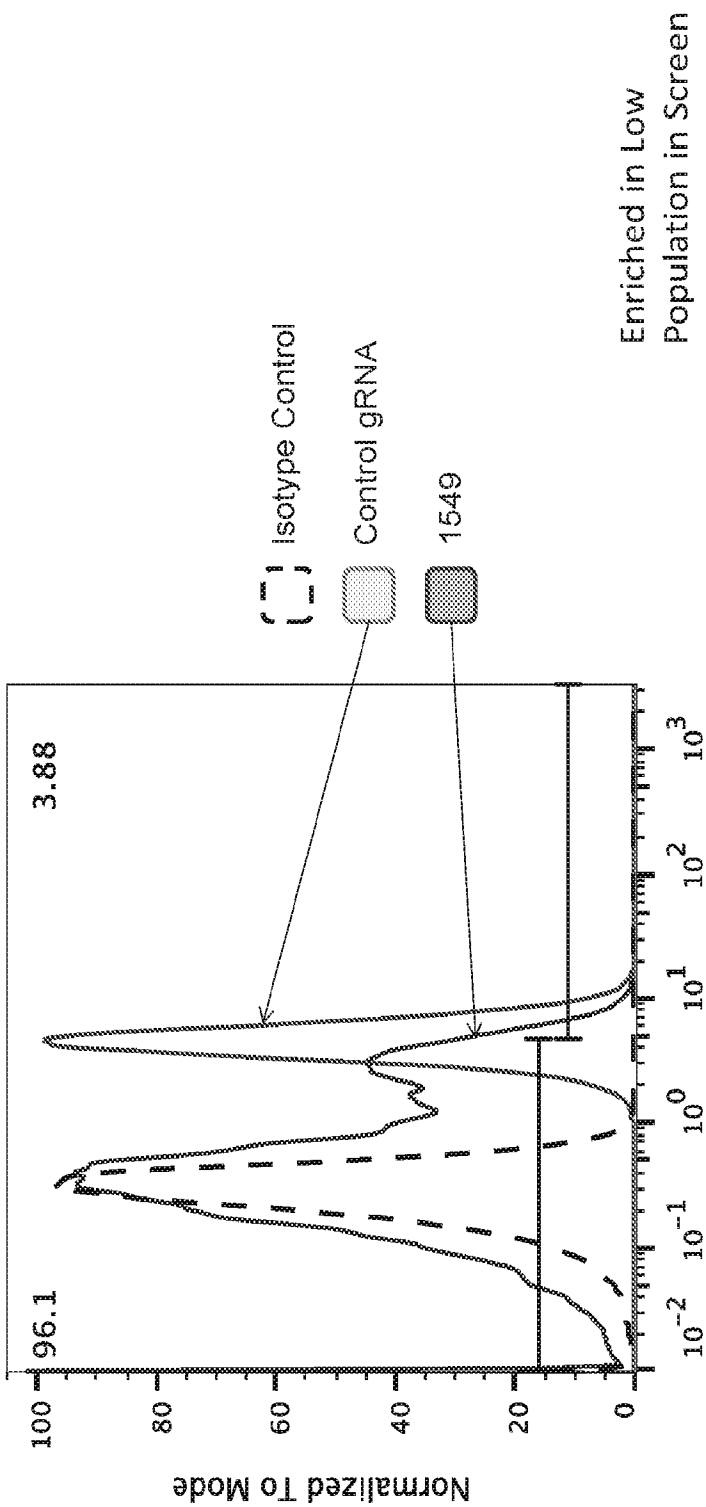

Enriched in Low
Population in Screen
Isotype Control
Control gRNA
1549
3.88
96.1
Normalized To Mode
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-A :: GFP_FITC-A

FIG. 12B

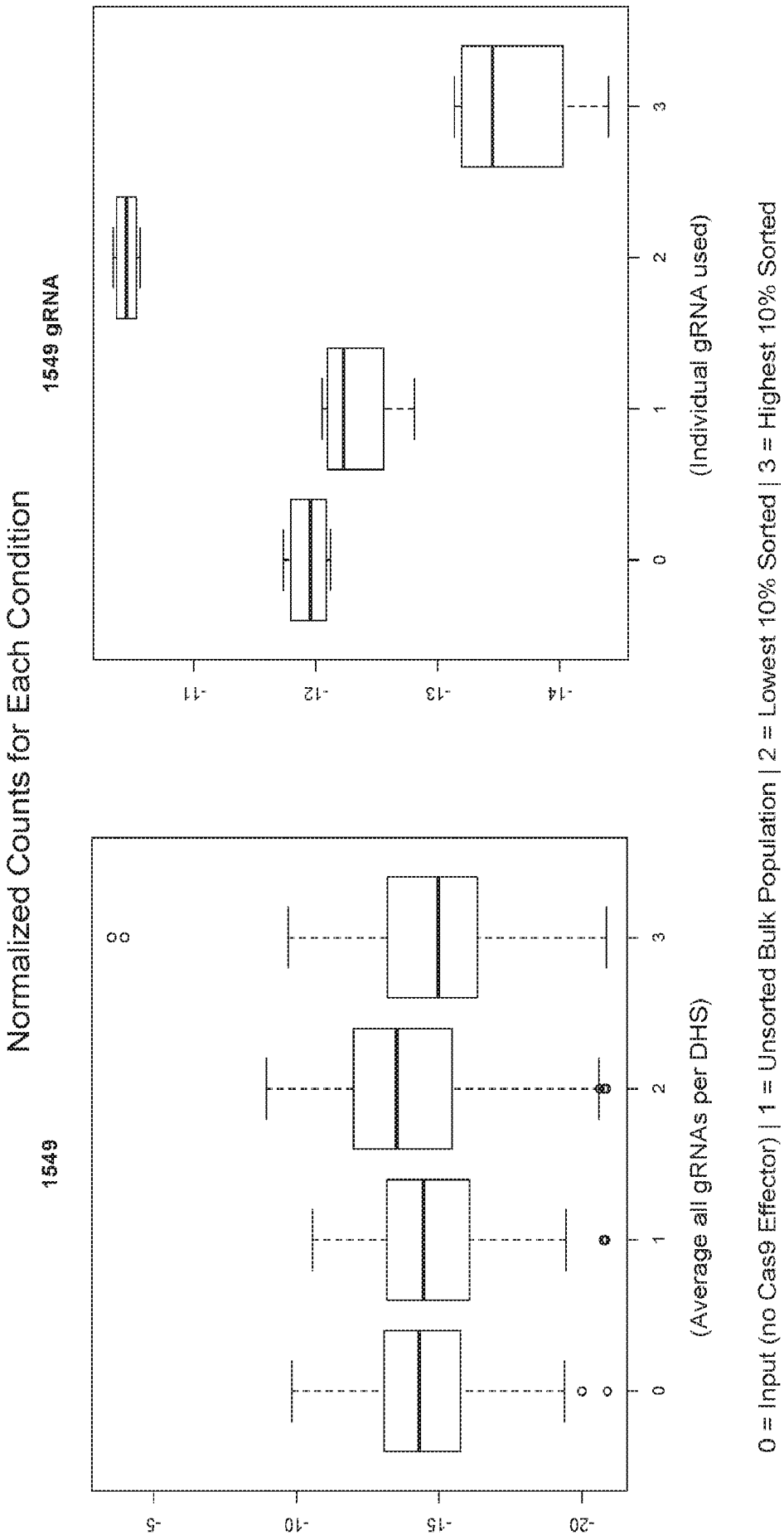

Normalized Counts for Each Condition
1549
1549 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 12C

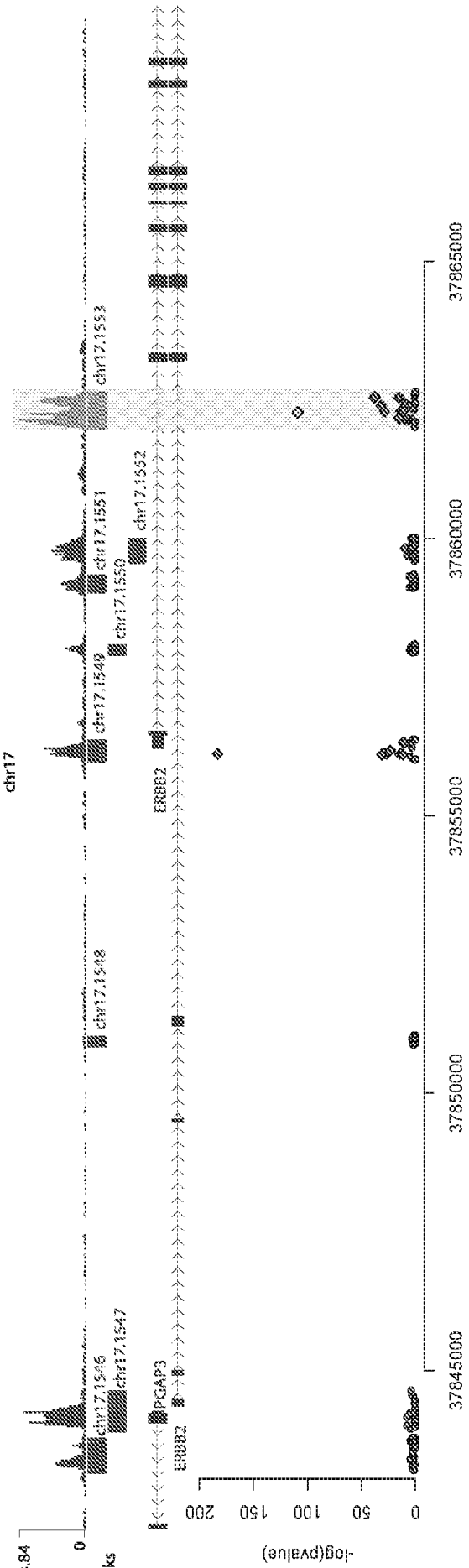

37844000
37846000
37848000
37850000
37852000
37854000
37856000
37858000
37860000
37862000
37864000
37866000
37868000
chr17
113.84
0
SK8R3 DNase I
SK8R3 DHS Peaks
RefSeq genes
chr17:1546
chr17:1547
chr17:1548
chr17:1549
chr17:1550
chr17:1551
chr17:1552
chr17:1553
PGAP3
ERBB2
ERBB2
0
50
100
150
200
-log(pvalue)
37845000
37850000
37855000
37860000
37865000
Chr17 (bp)

FIG. 13A

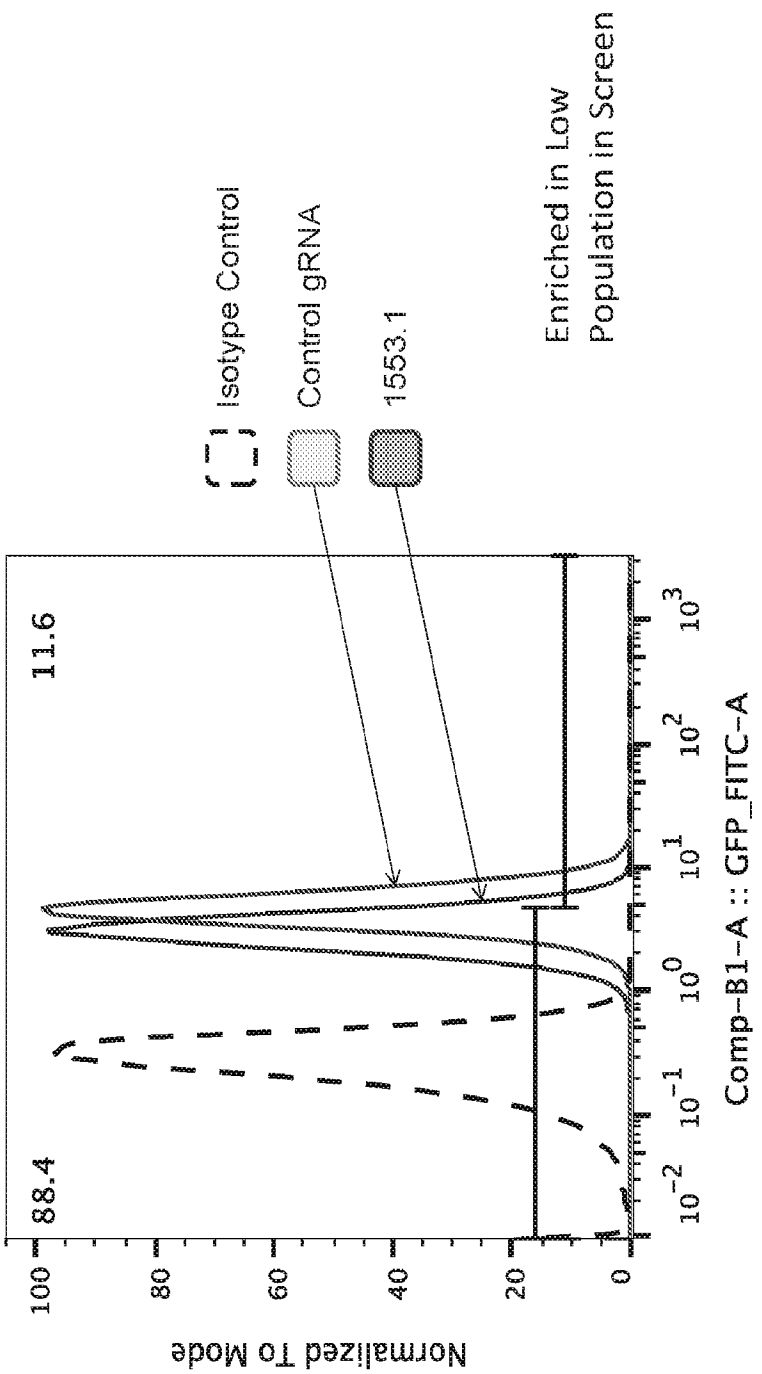

Isotype Control
Control gRNA
1553.1
Enriched in Low
Population in Screen
11.6
88.4
Normalized To Mode
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-A :: GFP_FITC-A

FIG. 13B

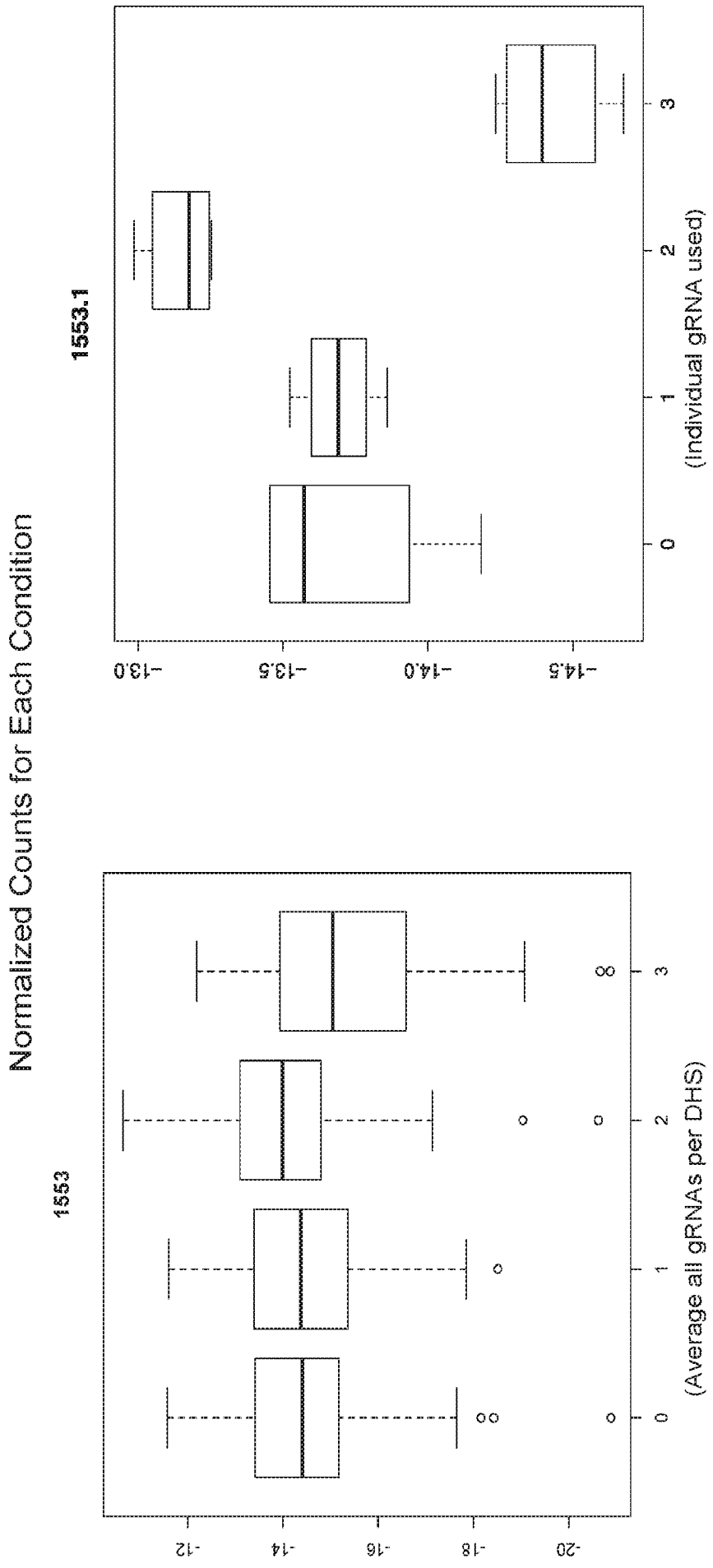

Normalized Counts for Each Condition
1553
-12
-14
-16
-18
-20
0
1
2
3
(Average all gRNAs per DHS)
1553.1
-13.0
-13.5
-14.0
-14.5
0
1
2
3
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 13C

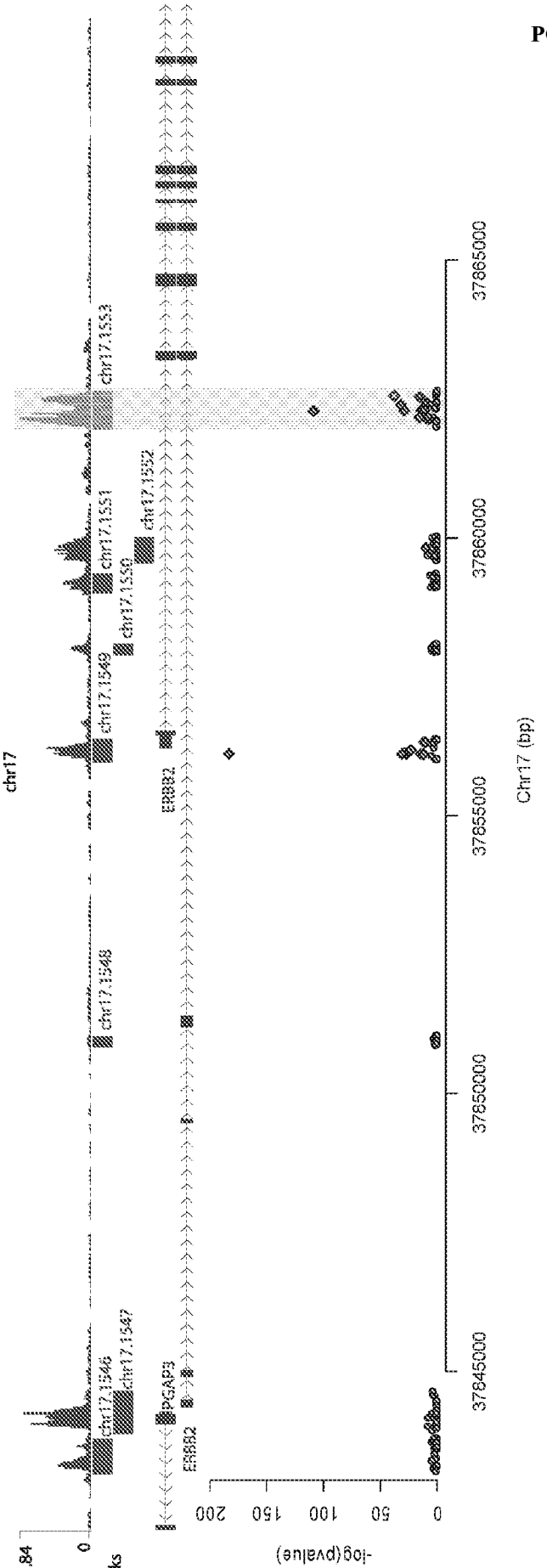

37844000
37846000
37848000
37850000
37852000
37854000
37856000
37858000
37860000
37862000
37864000
37866000
37868000
chr17
q12
SKBR3 DNase I
113,84
0
SKBR3 DHS Peaks
RefSeq genes
chr17:1546
chr17:1547
chr17:1548
chr17:1549
chr17:1550
chr17:1551
chr17:1552
chr17:1553
PGAP3
ERBB2
ERBB2
-log(pvalue)
0
50
100
150
200
37845000
37850000
37855000
37860000
37865000
Chr17 (bp)

FIG. 13D

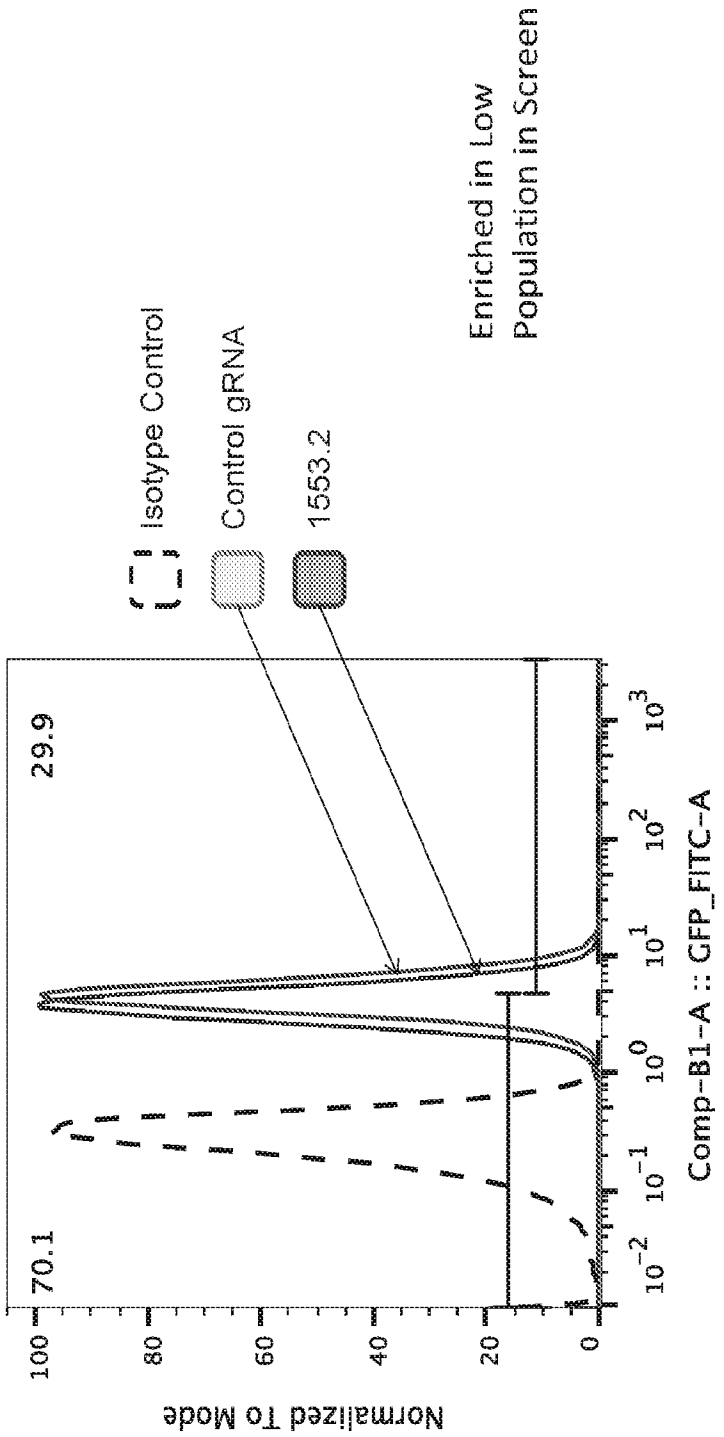

Enriched in Low
Population in Screen
Isotype Control
Control gRNA
1553.2
29.9
70.1
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 13E

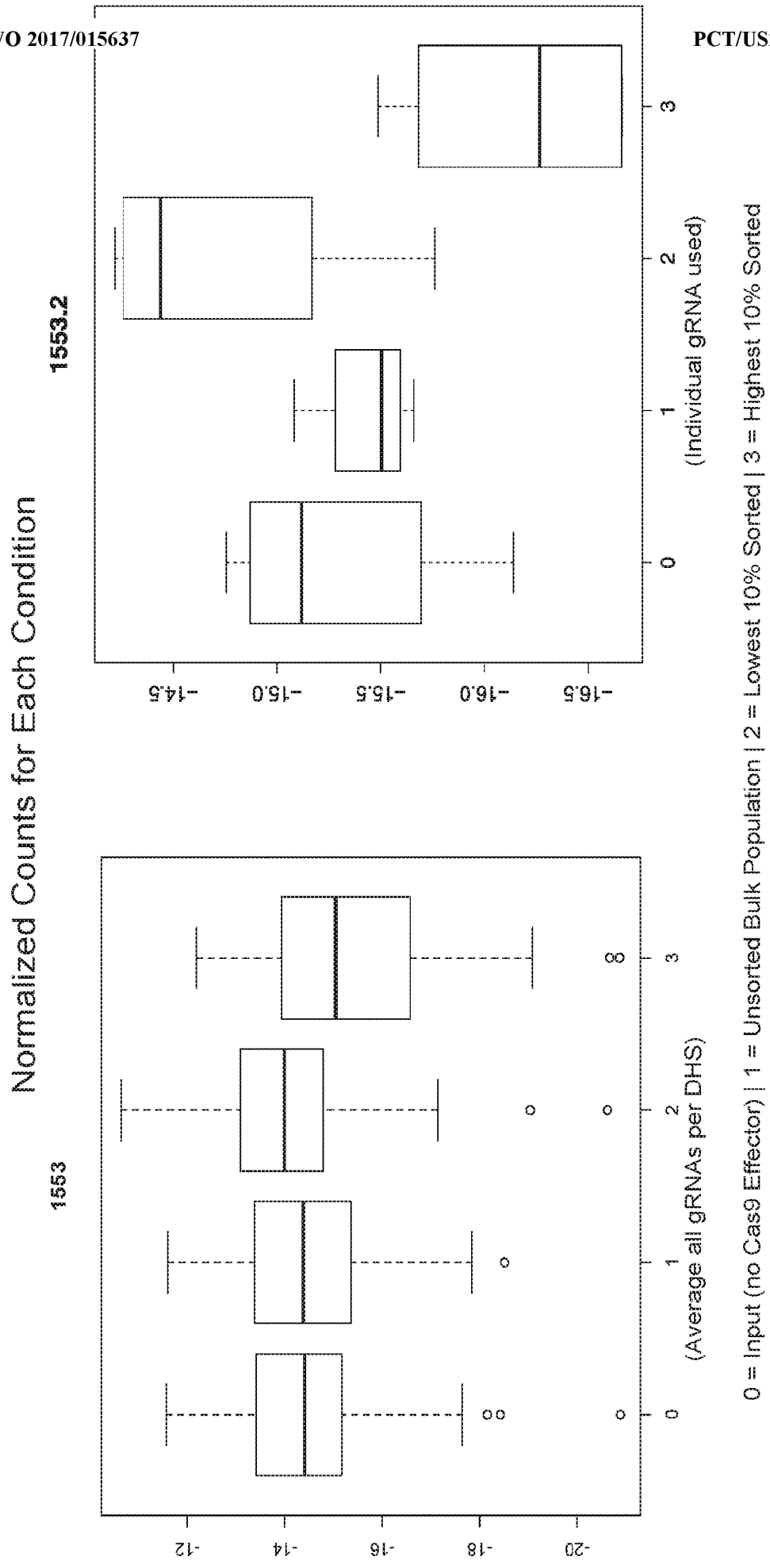

Normalized Counts for Each Condition
1553
1553.2
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 13F

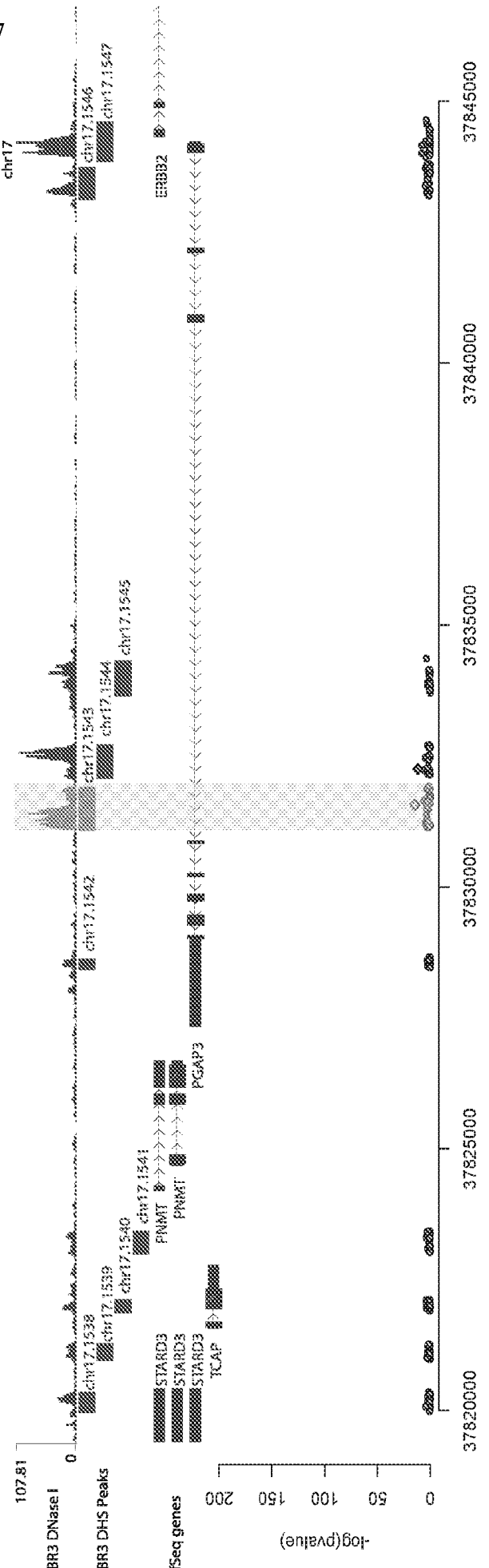

37822000
37824000
37826000
37828000
37830000
37832000
37834000
37836000
37838000
37840000
37842000
37844000
37846000
chr17
chr17:1538
chr17:1539
chr17:1540
chr17:1541
chr17:1542
chr17:1543
chr17:1544
chr17:1545
chr17:1546
chr17:1547
STARD3
STARD3
STARD3
TCAP
PNMT
PNMT
PGAP3
ERBB2
107.81
0
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
200
150
100
50
0
-log(pvalue)
37820000
37825000
37830000
37835000
37840000
37845000
Chr17 (bp)

FIG. 14A

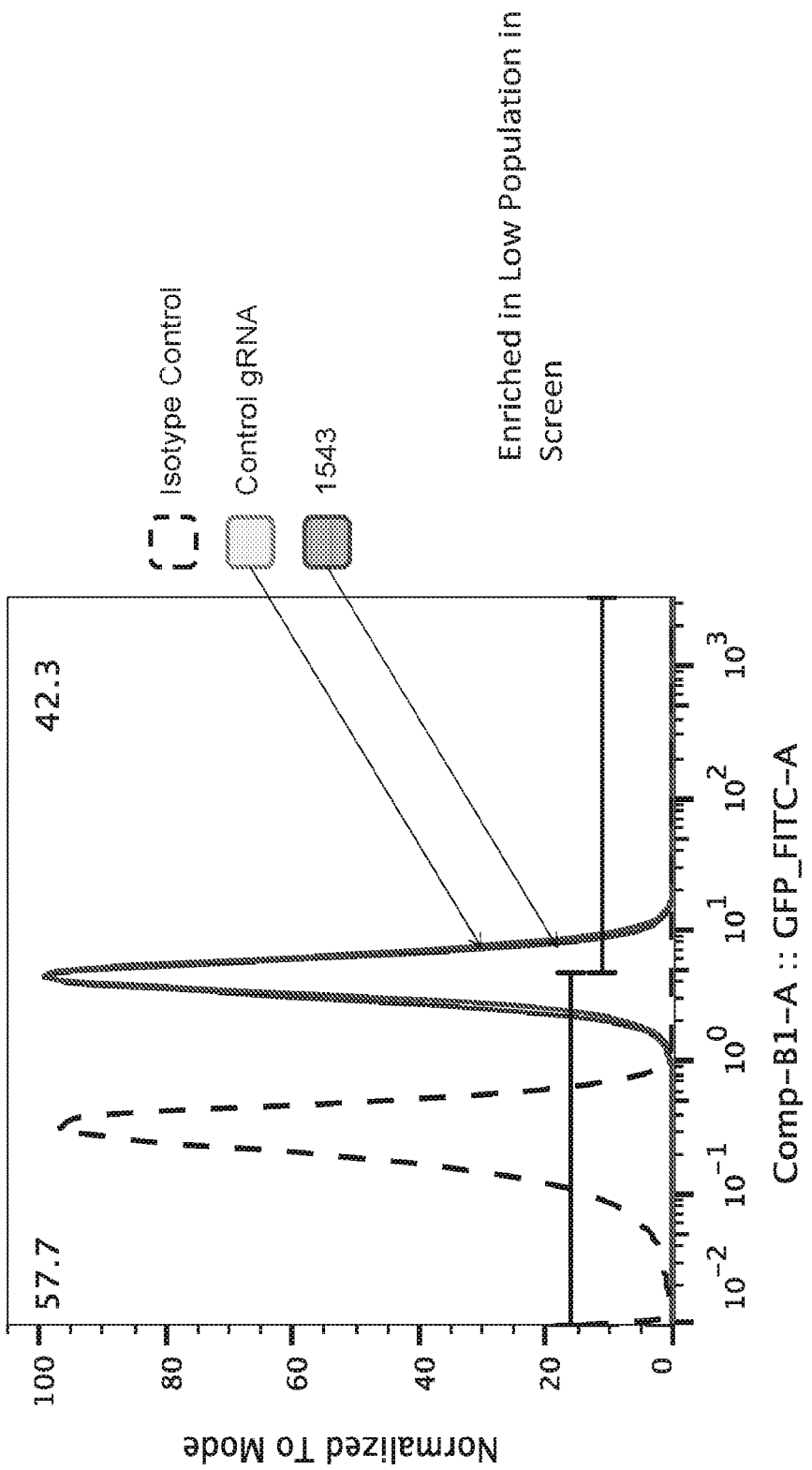

Isotype Control
Control gRNA
1543
Enriched in Low Population in Screen
42.3
57.7
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 14B

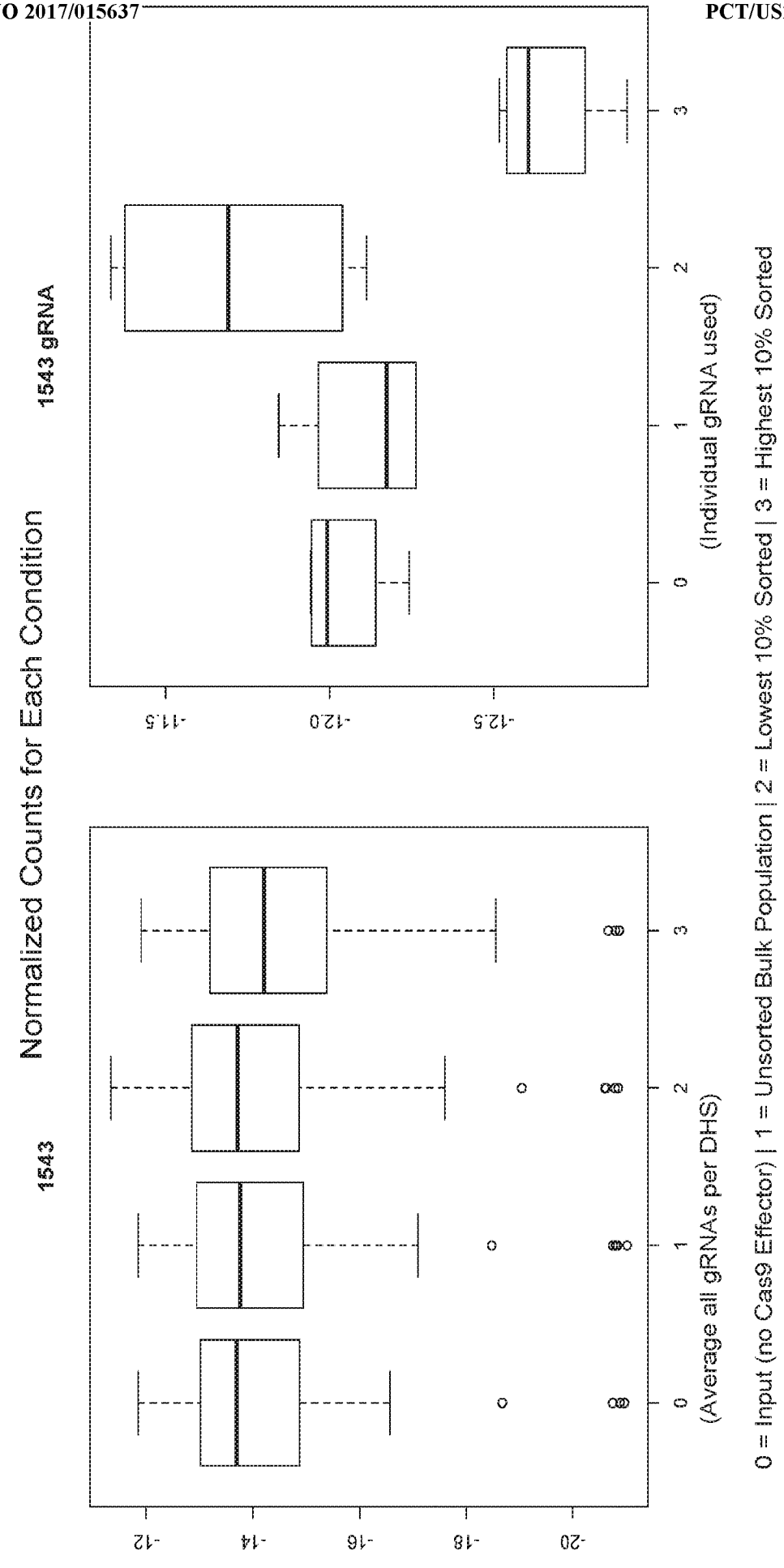

Normalized Counts for Each Condition
1543
1543 gRNA
-12
-14
-16
-18
-20
-11.5
-12.0
-12.5
0
1
2
3
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 14C

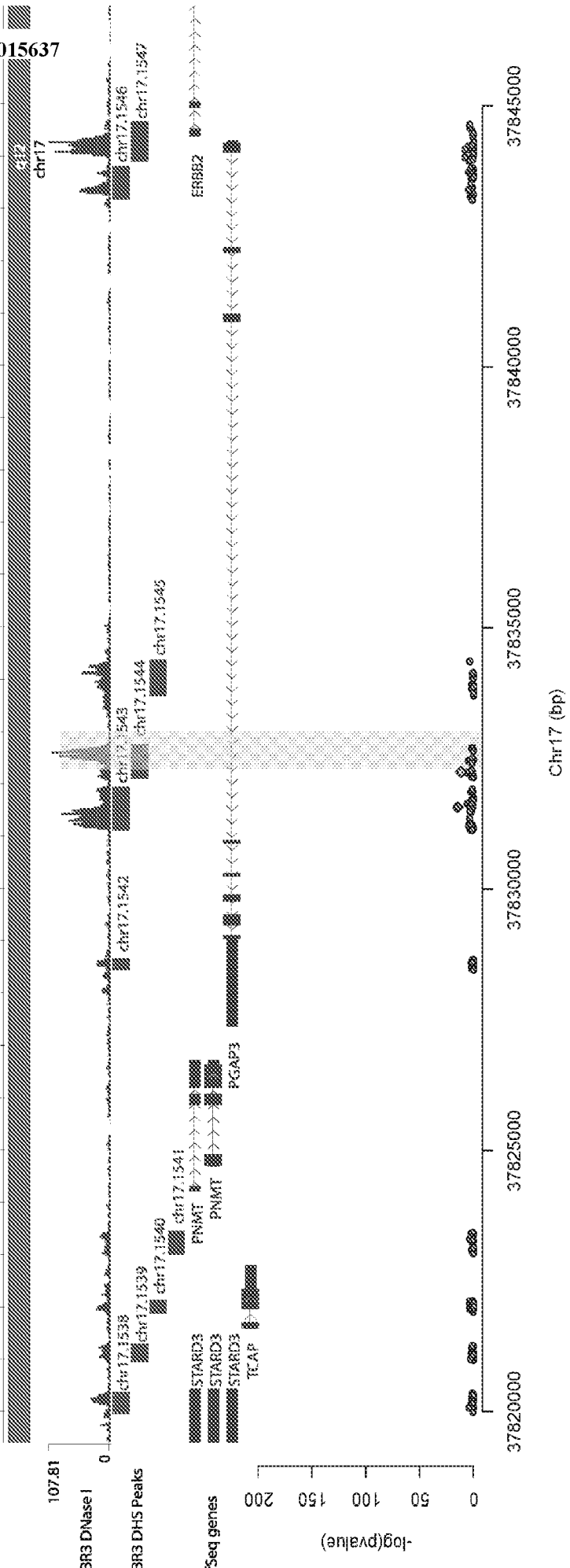

chr17
ERBB2
chr17.1546
chr17.1547
chr17.1545
chr17.1544
chr17.1543
chr17.1542
chr17.1541
chr17.1540
chr17.1539
chr17.1538
PNMT
PGAP3
STARD3
TCAP
107.81
0
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
200
150
100
50
0
-log(pvalue)
37820000
37825000
37830000
37835000
37840000
37845000
Chr17 (bp)

FIG. 15A

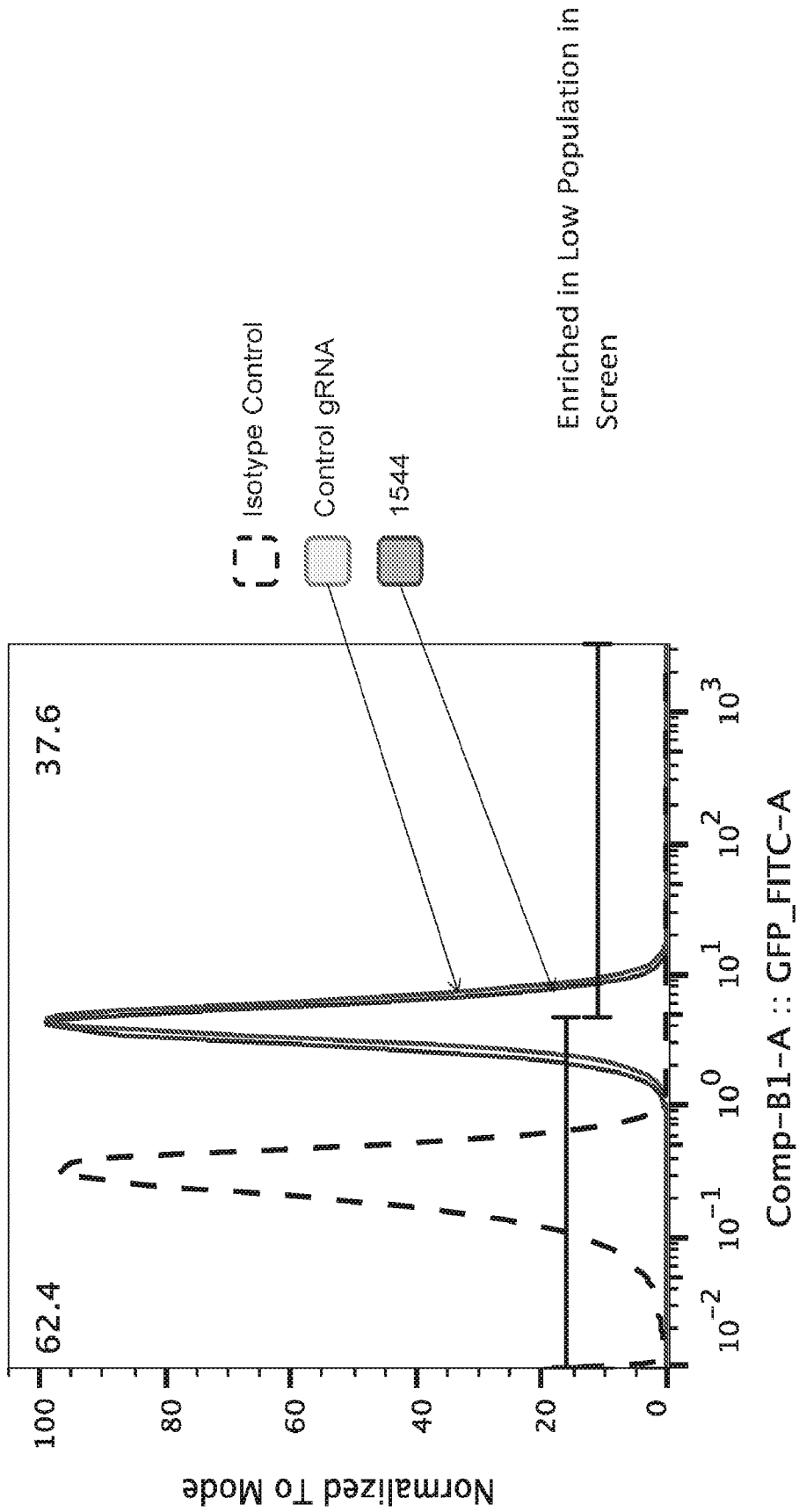

Isotype Control
Control gRNA
1544
Enriched in Low Population in Screen
37.6
62.4
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 15B

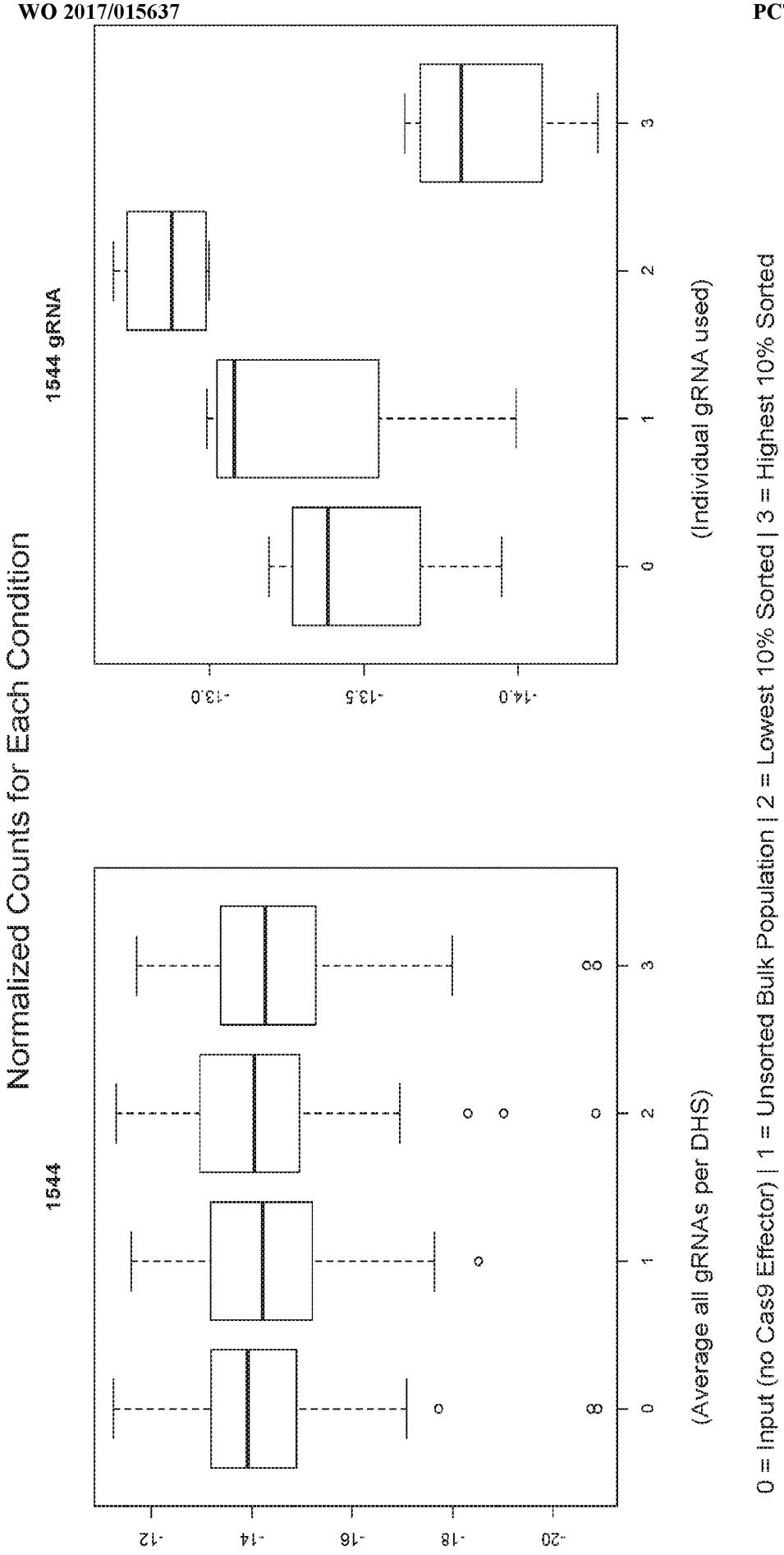

Normalized Counts for Each Condition
1544
1544 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 15C

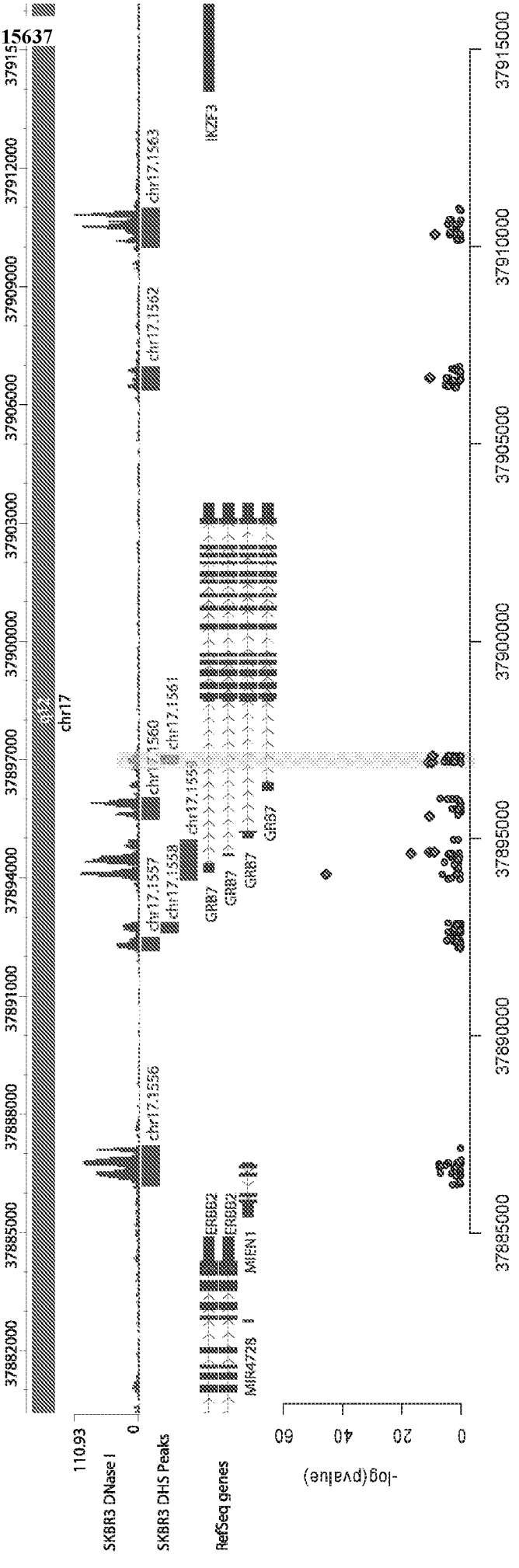

37882000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915
chr17
SKBR3 DNase I
110.93
0
SKBR3 DHS Peaks
RefSeq genes
chr17:1556
chr17:1557
chr17:1558
chr17:1559
chr17:1560
chr17:1561
chr17:1562
chr17:1563
ERBB2
ERBB2
MIEN1
MIR4728
GRB7
GRB7
GRB7
GRB7
IKZF3
-log(pvalue)
0
20
40
60
37885000
37890000
37895000
37900000
37905000
37910000
37915000
Chr17 (bp)

FIG. 16A

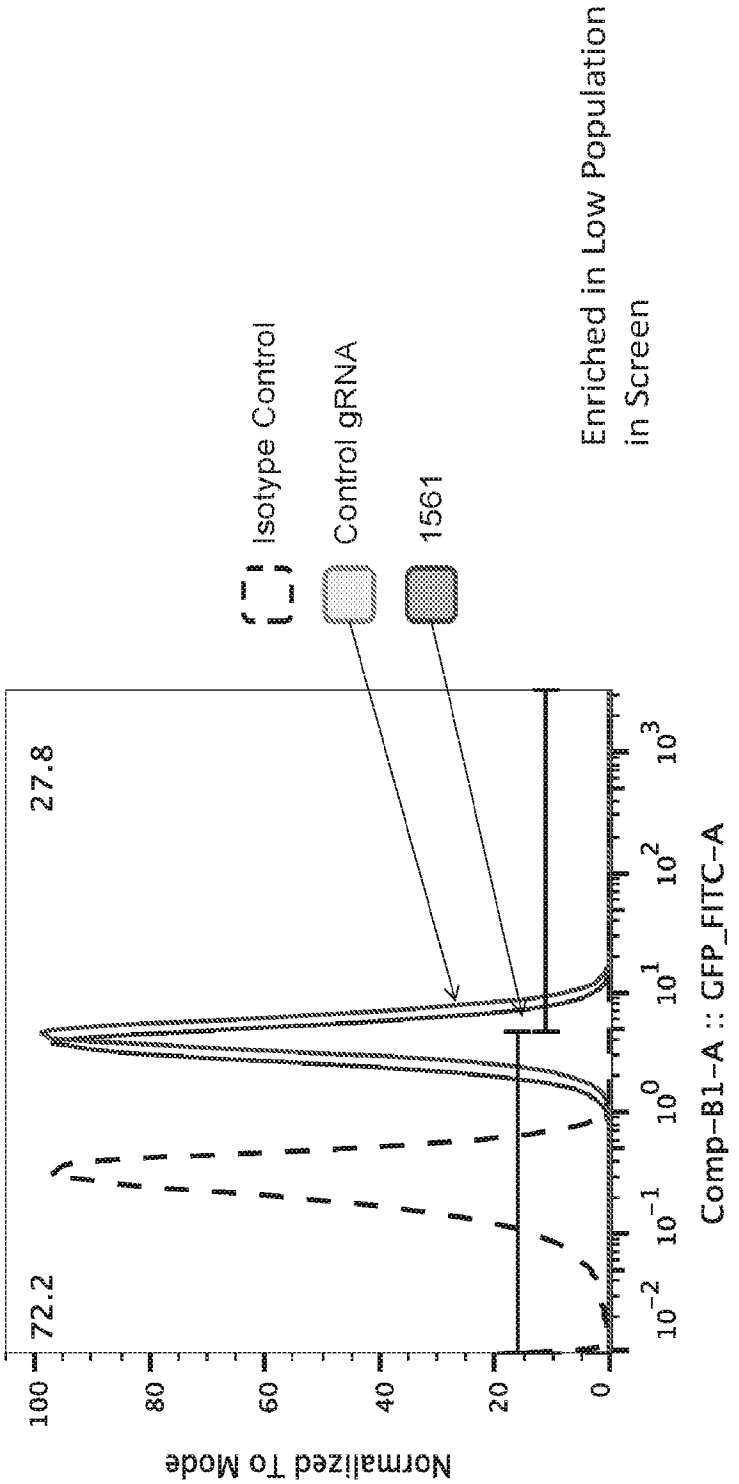

Isotype Control
Control gRNA
1561
Enriched in Low Population in Screen
27.8
72.2
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 16B

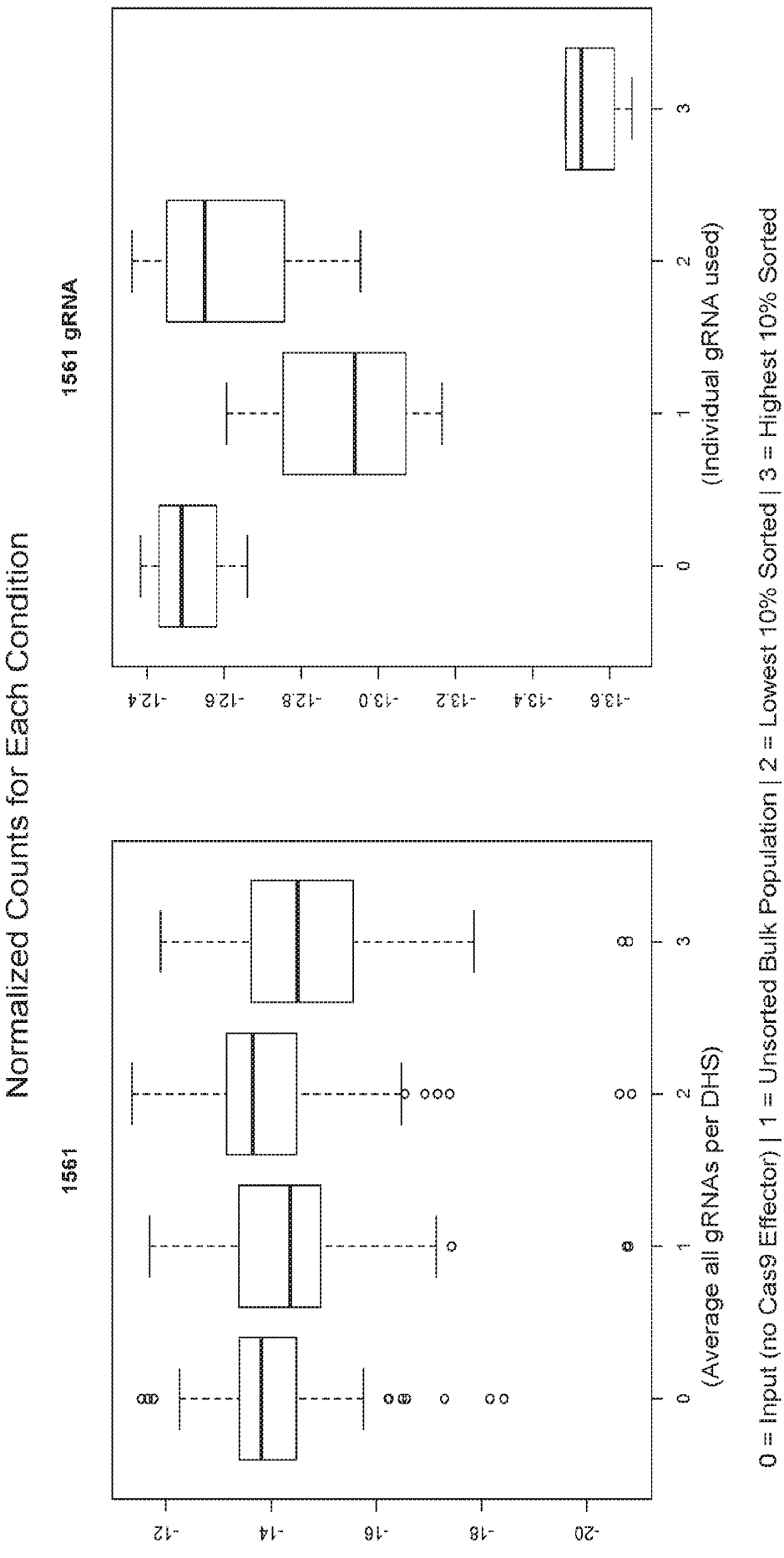

Normalized Counts for Each Condition
1561
1561 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 16C

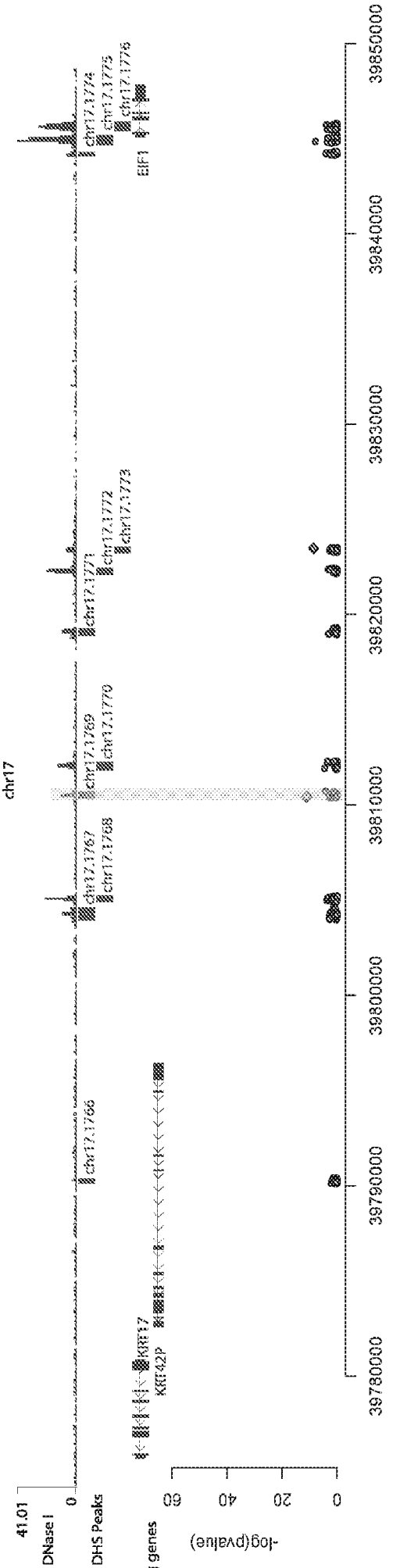

39778000
39783000
39788000
39793000
39798000
39803000
39808000
39813000
39818000
39823000
39828000
39833000
39838000
39843000
chr17
q23.2
41.01
0
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
KRT42P
KRT17
chr17:1766
chr17:1767
chr17:1768
chr17:1769
chr17:1770
chr17:1771
chr17:1772
chr17:1773
chr17:1774
chr17:1775
chr17:1776
EIF1
60
40
20
0
-log(pvalue)
39780000
39790000
39800000
39810000
39820000
39830000
39840000
39850000
Chr17 (bp)

FIG. 17A

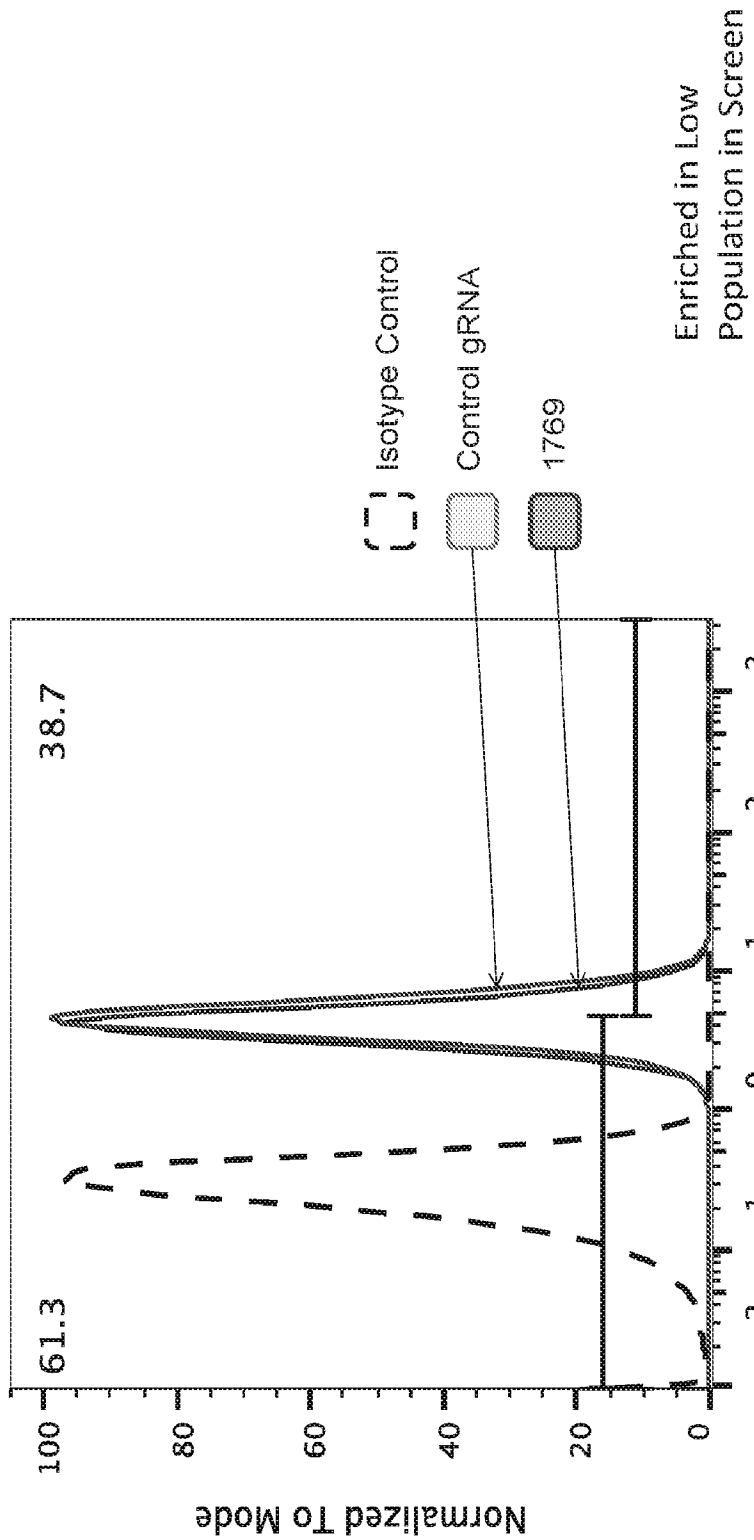

Isotype Control
Control gRNA
1769
Enriched in Low
Population in Screen
38.7
61.3
Normalized To Mode
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-A :: GFP_FITC-A

FIG. 17B

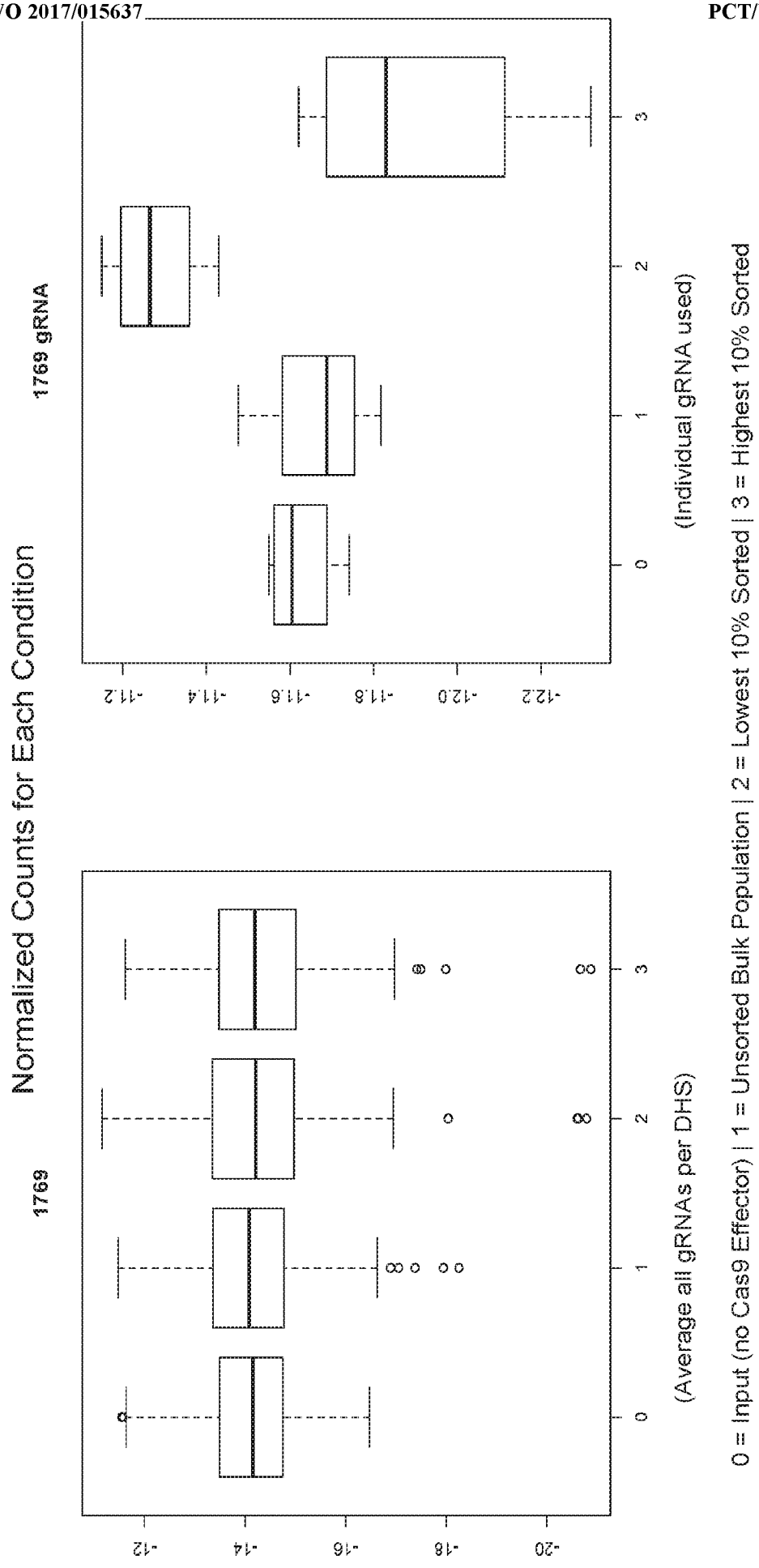

Normalized Counts for Each Condition
1769
1769 gRNA
-12
-14
-16
-18
-20
-11.2
-11.4
-11.6
-11.8
-12.0
-12.2
0
1
2
3
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 17C

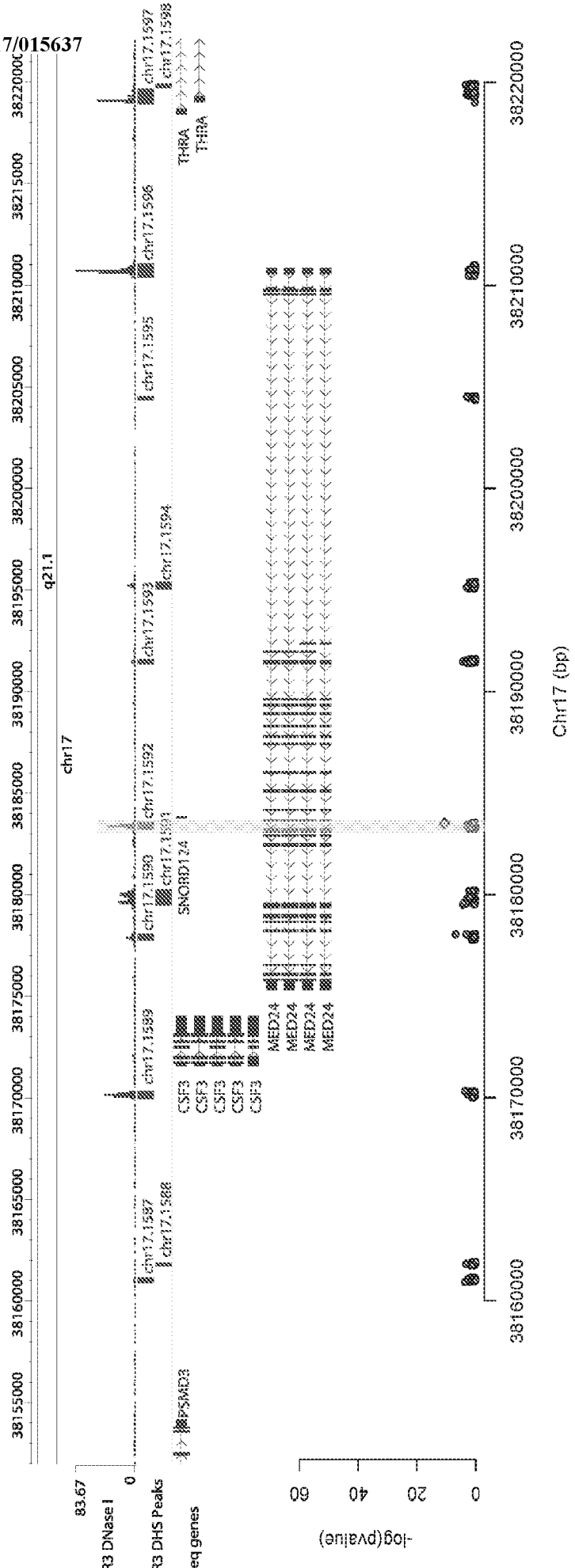

chr17
q21.1
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
38155000
38160000
38165000
38170000
38175000
38180000
38185000
38190000
38195000
38200000
38205000
38210000
38215000
38220000
PSMD3
CSF3
SNORD124
MED24
THRA

FIG. 18A

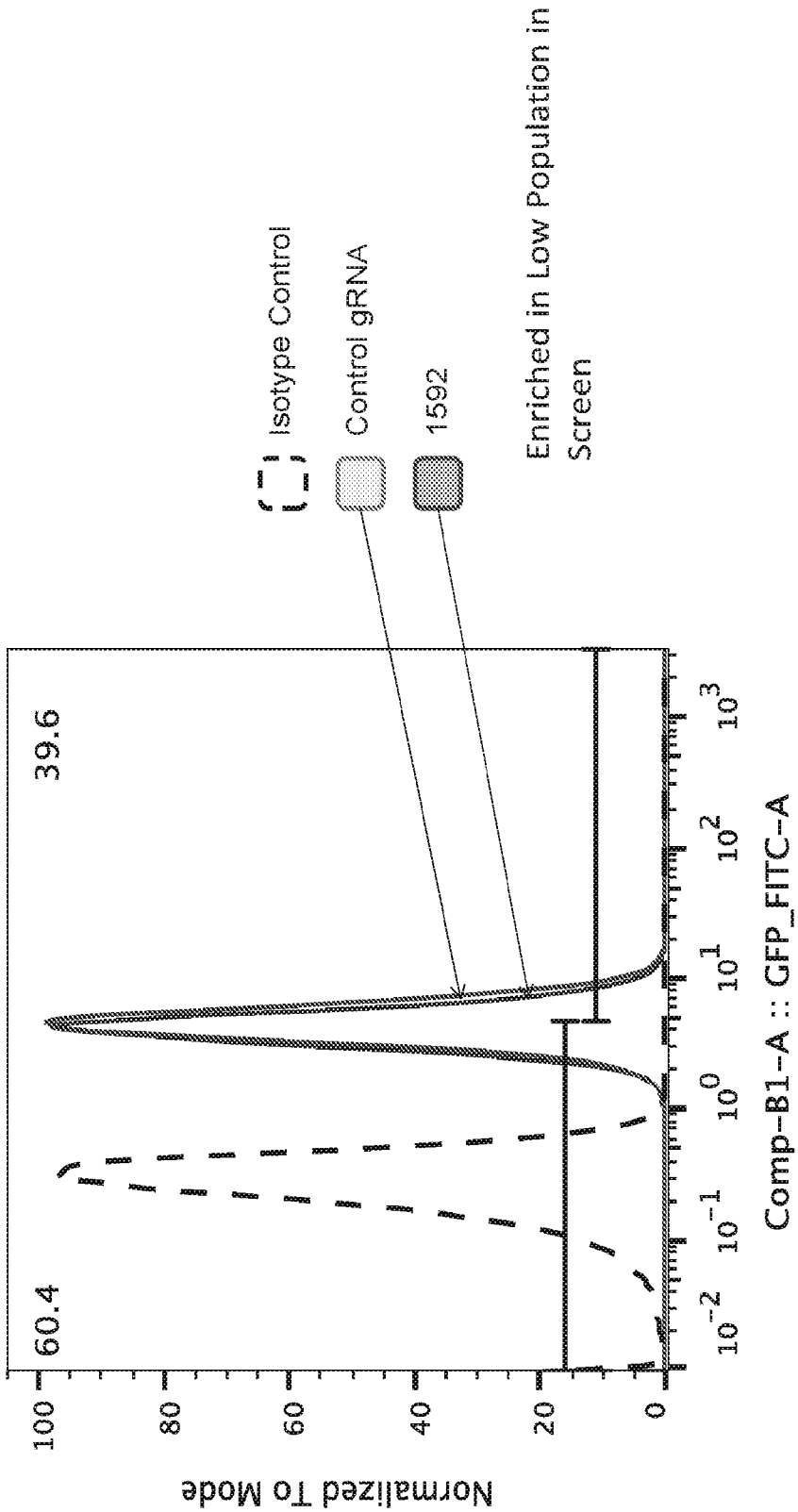

Isotype Control
Control gRNA
1592
Enriched in Low Population in Screen
39.6
60.4
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 18B

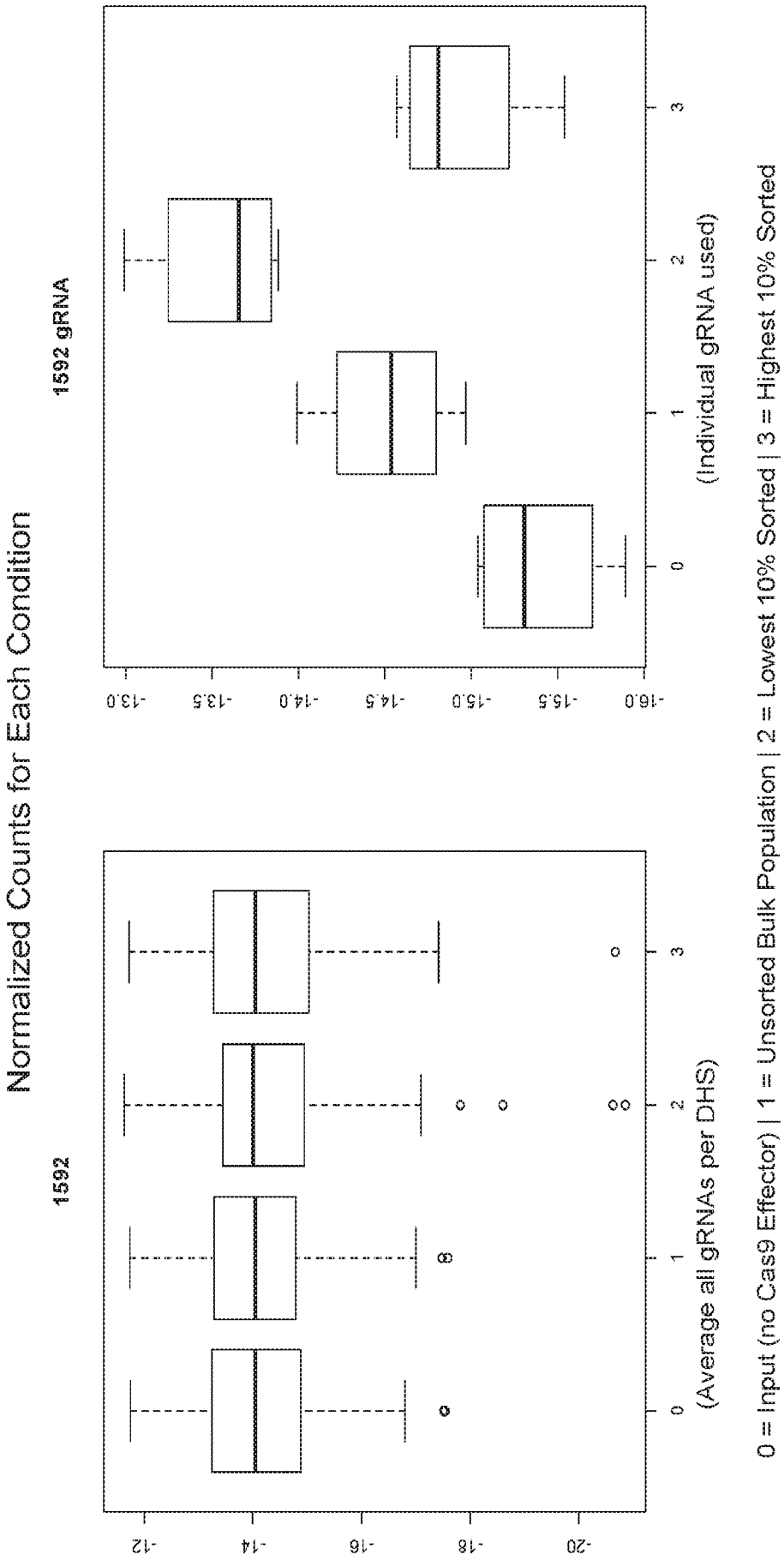

Normalized Counts for Each Condition
1592
-12
-14
-16
-18
-20
0
1
2
3
(Average all gRNAs per DHS)
1592 gRNA
-13.0
-13.5
-14.0
-14.5
-15.0
-15.5
-16.0
0
1
2
3
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 18C

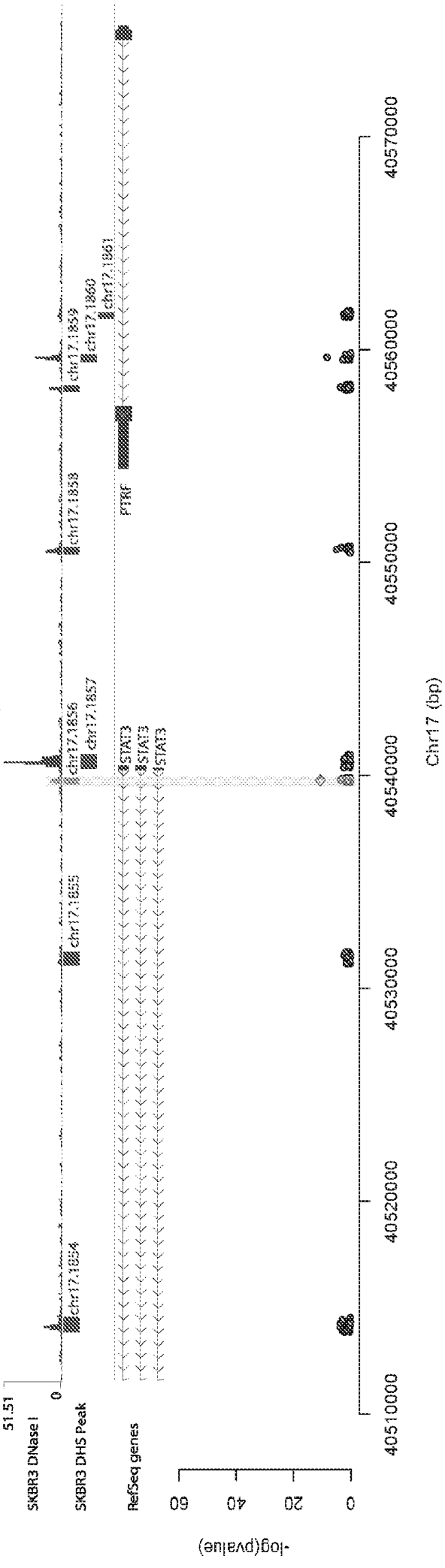

40515000
40520000
40525000
40530000
40535000
40540000
40545000
40550000
40555000
40560000
40565000
40570000
chr17
chr17.1854
chr17.1855
chr17.1856
chr17.1857
chr17.1858
chr17.1859
chr17.1860
chr17.1861
STAT3
STAT3
STAT3
PTRF
51.51
0
SKBR3 DNase I
SKBR3 DHS Peak
RefSeq genes
60
40
20
0
-log(pvalue)
40510000
40520000
40530000
40540000
40550000
40560000
40570000
Chr17 (bp)

FIG. 19A

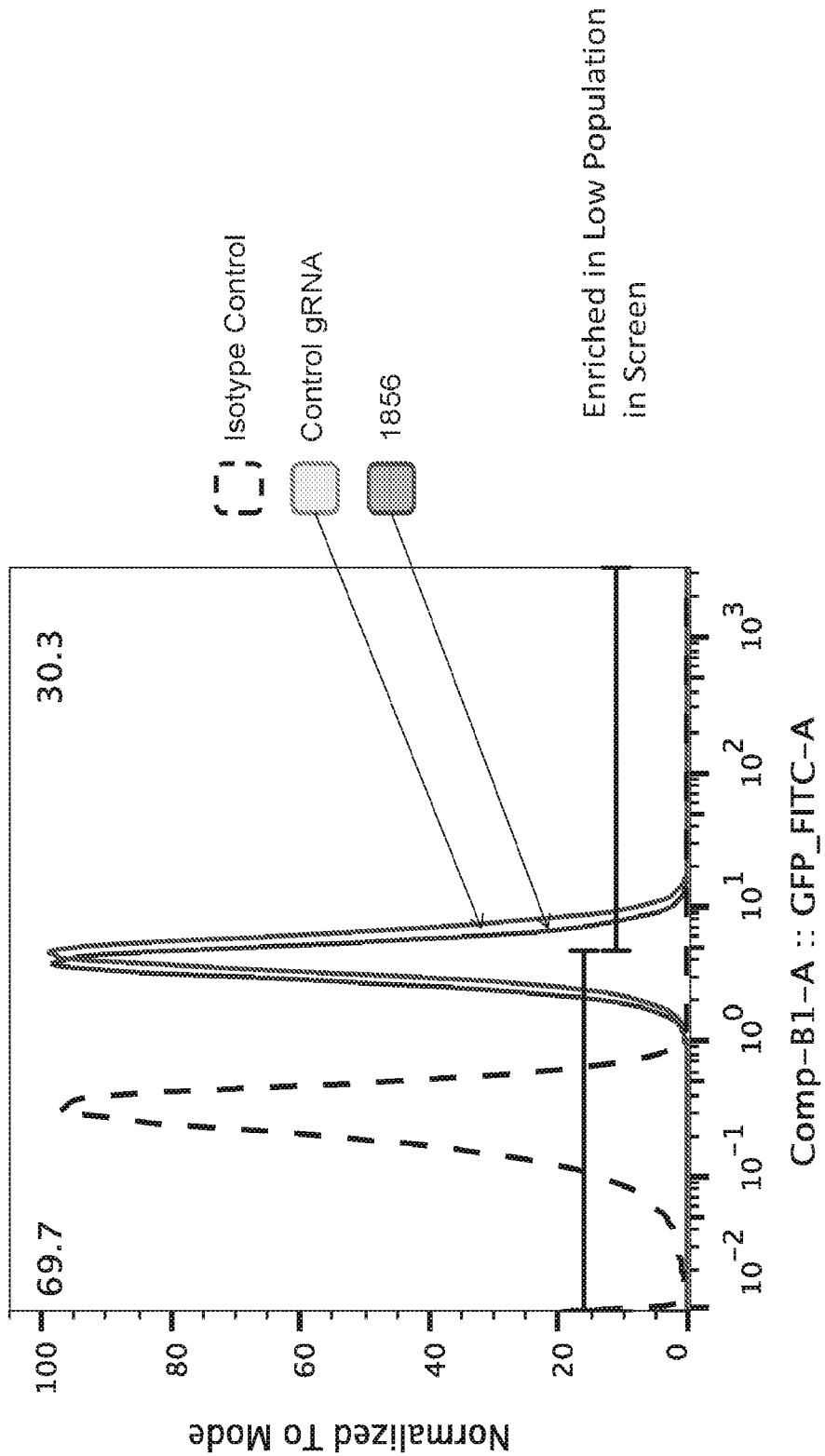

Isotype Control
Control gRNA
1856
Enriched in Low Population in Screen
30.3
69.7
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 19B

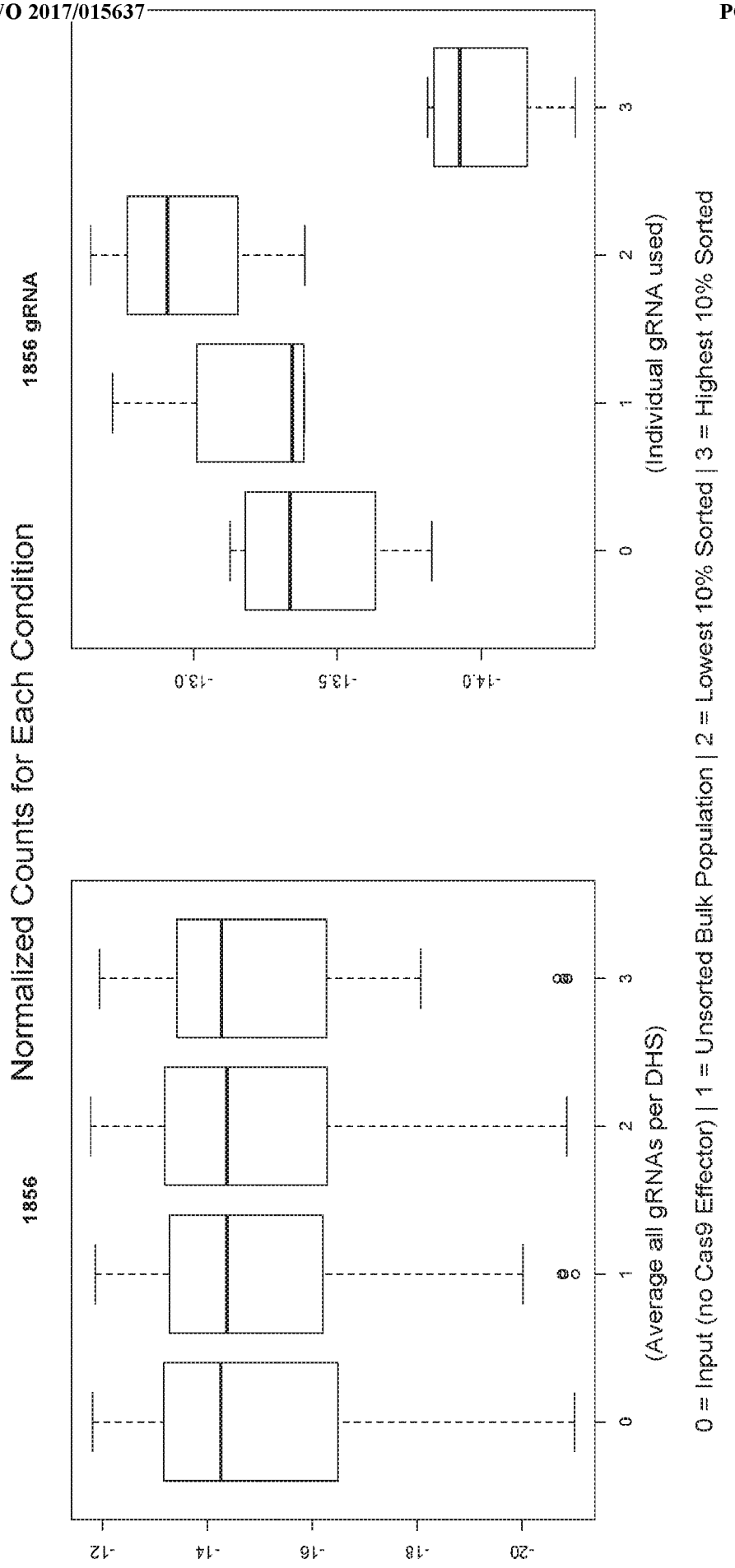

Normalized Counts for Each Condition
1856
1856 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 19C

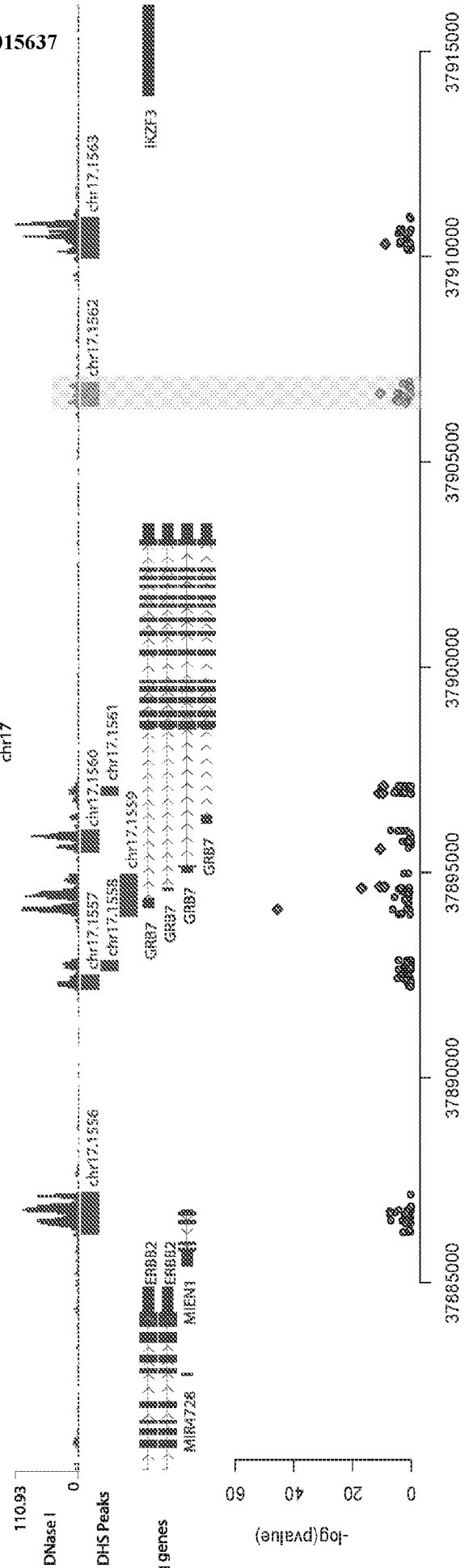

chr17
37882000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915000
SKBR3 DNase I
110.93
0
SKBR3 DHS Peaks
RefSeq genes
chr17.1556
chr17.1557
chr17.1558
chr17.1559
chr17.1560
chr17.1561
chr17.1562
chr17.1563
IKZF3
GRB7
ERBB2
MIEN1
MIR4728
-log(pvalue)
0
20
40
60
Chr17 (bp)
37885000
37890000
37895000
37900000
37905000
37910000
37915000

FIG. 20A

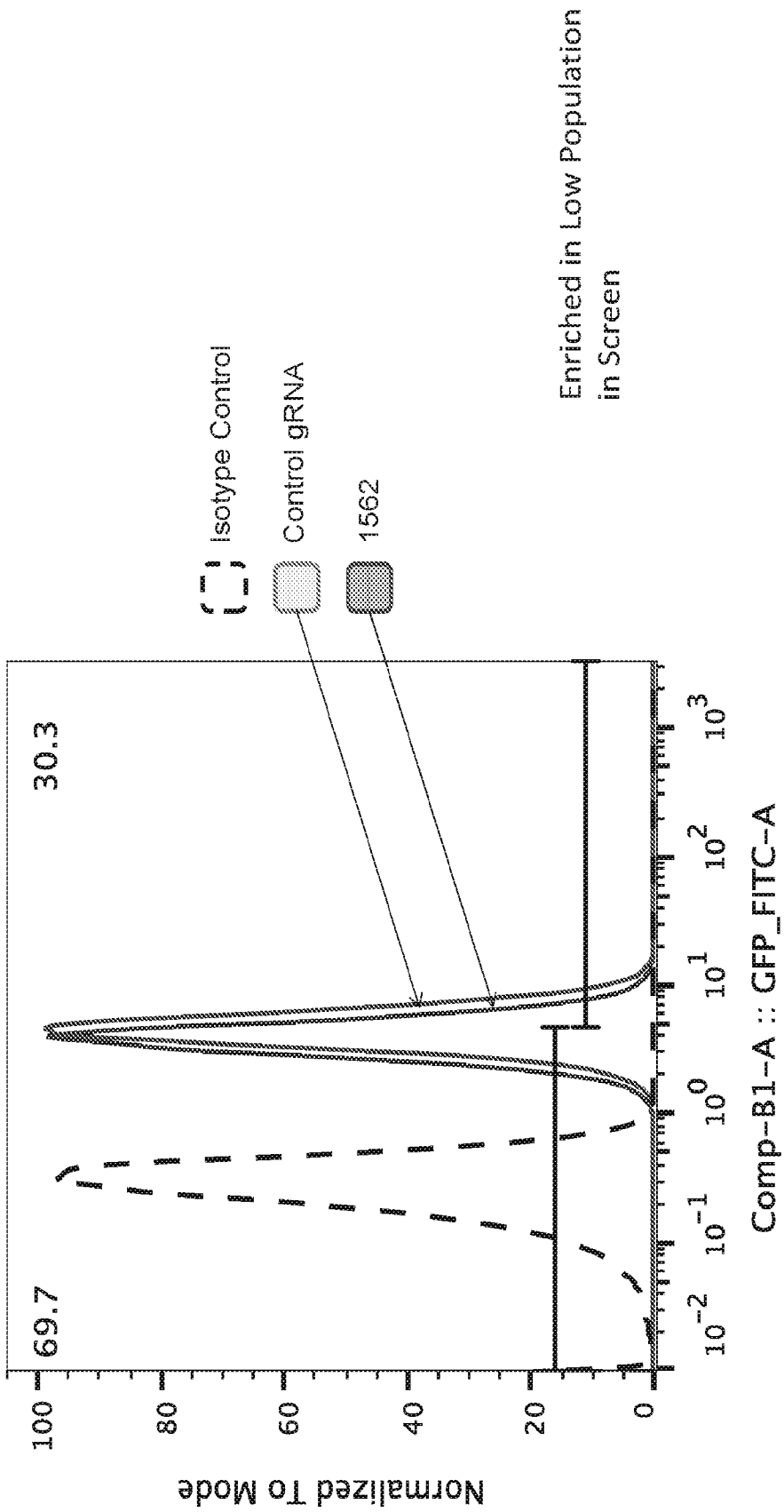

Enriched in Low Population
in Screen
Isotype Control
Control gRNA
1562
30.3
69.7
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 20B

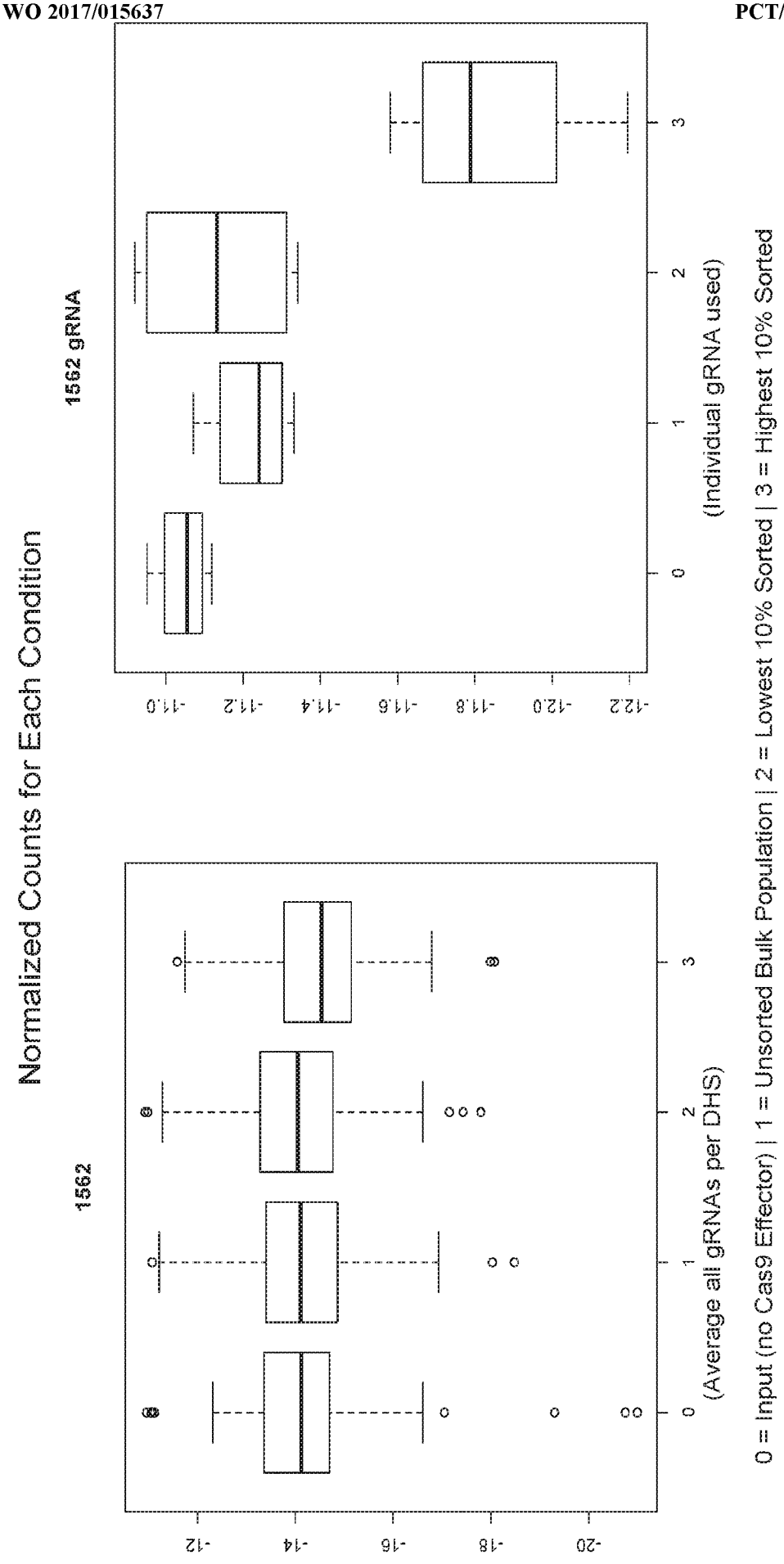

Normalized Counts for Each Condition
1562
-12
-14
-16
-18
-20
0
1
2
3
(Average all gRNAs per DHS)
1562 gRNA
-11.0
-11.2
-11.4
-11.6
-11.8
-12.0
-12.2
0
1
2
3
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 20C

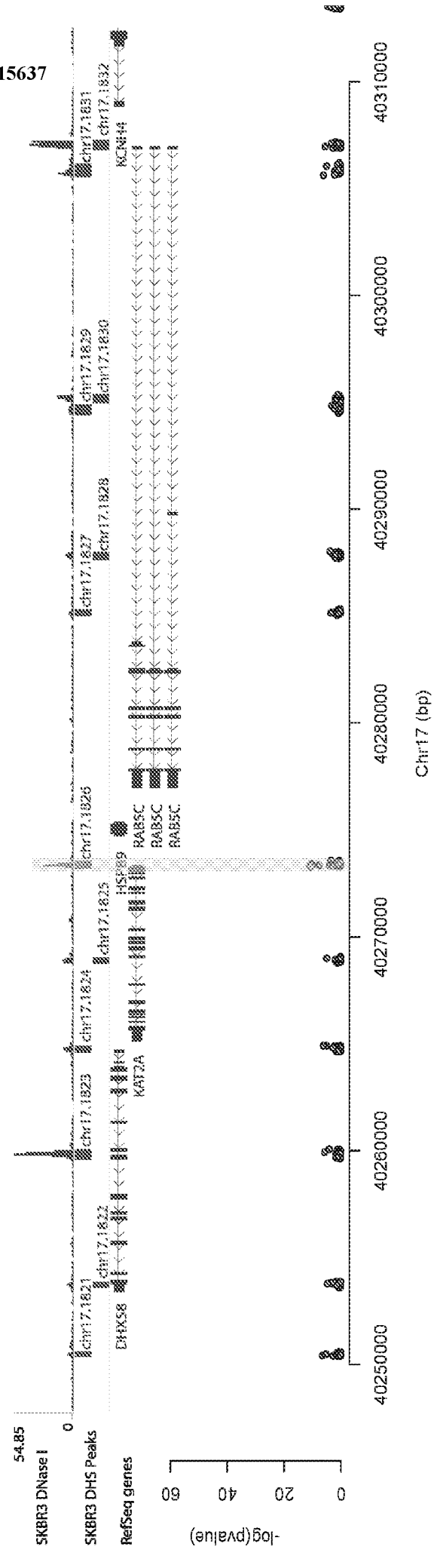

SKBR3 DNase I
54.85
0
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
0
20
40
60
40250000
40260000
40270000
40280000
40290000
40300000
40310000
Chr17 (bp)
40251000
40256000
40261000
40266000
40271000
40276000
40281000
40286000
40291000
40296000
40301000
40306000
chr17.1821
chr17.1822
chr17.1823
chr17.1824
chr17.1825
chr17.1826
chr17.1827
chr17.1828
chr17.1829
chr17.1830
chr17.1831
chr17.1832
DHX58
KAT2A
HSPB9
RABSC
KCNH4

FIG. 21A

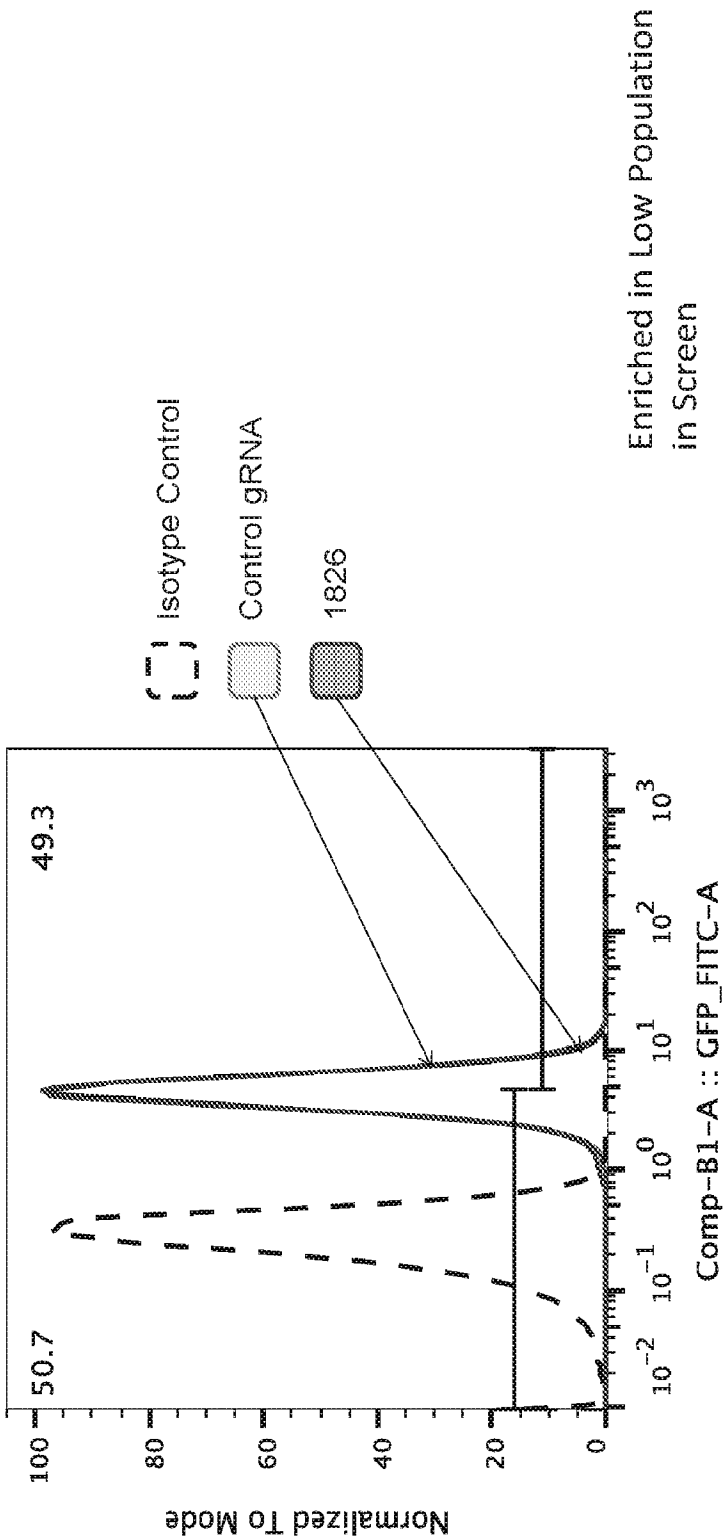

Enriched in Low Population
in Screen
Isotype Control
Control gRNA
1826
49.3
50.7
Normalized To Mode
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-A :: GFP_FITC-A

FIG. 21B

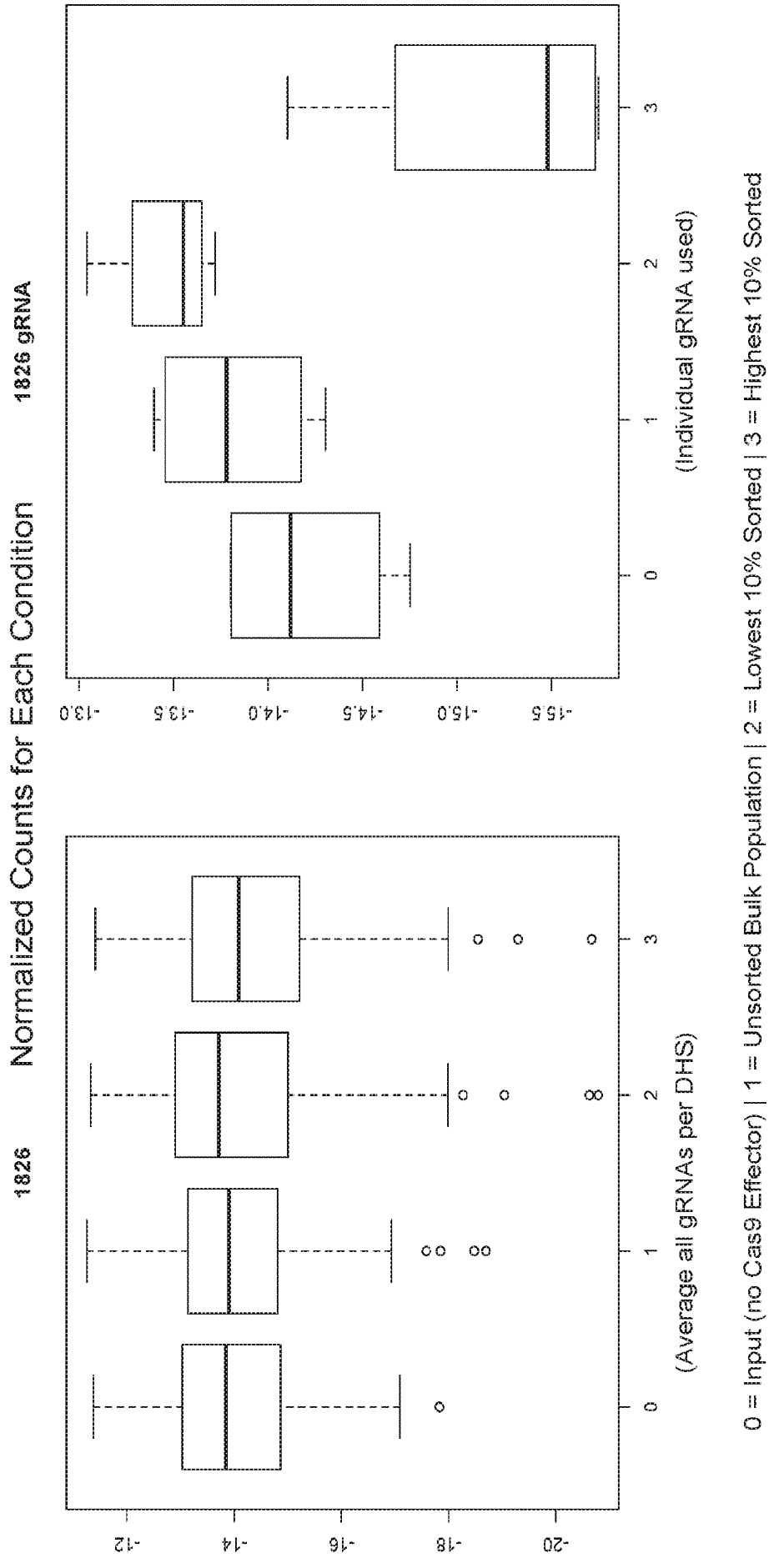

Normalized Counts for Each Condition
1826
1826 gRNA
-12
-14
-16
-18
-20
-13.0
-13.5
-14.0
-14.5
-15.0
-15.5
0
1
2
3
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 21C

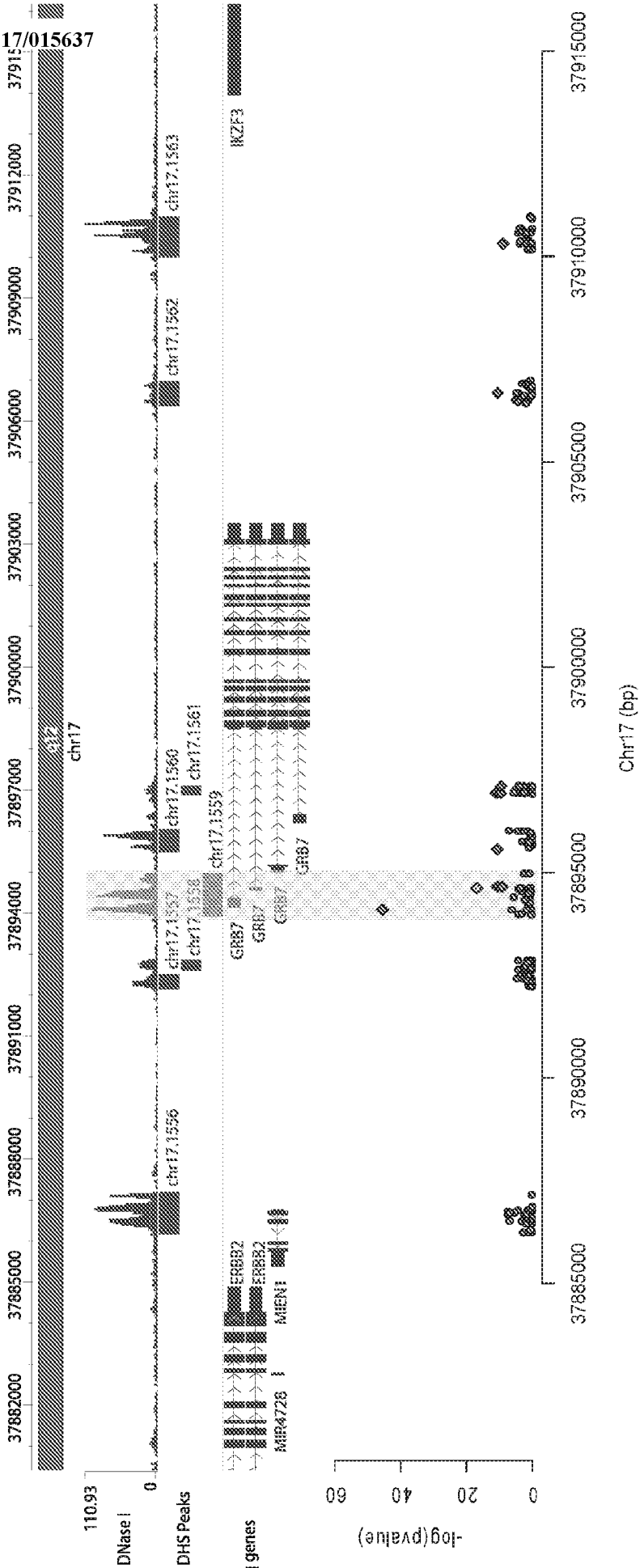

chr17
37862000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915
Chr17 (bp)
37885000
37890000
37895000
37900000
37905000
37910000
37915000
SKBR3 DNase I
110.93
0
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
0
20
40
60
chr17.1556
chr17.1557
chr17.1558
chr17.1559
chr17.1560
chr17.1561
chr17.1562
chr17.1563
ERBB2
ERBB2
MIEN1
MIR4728
GRB7
GRB7
GRB7
GRB7
KIZF3

FIG. 22A

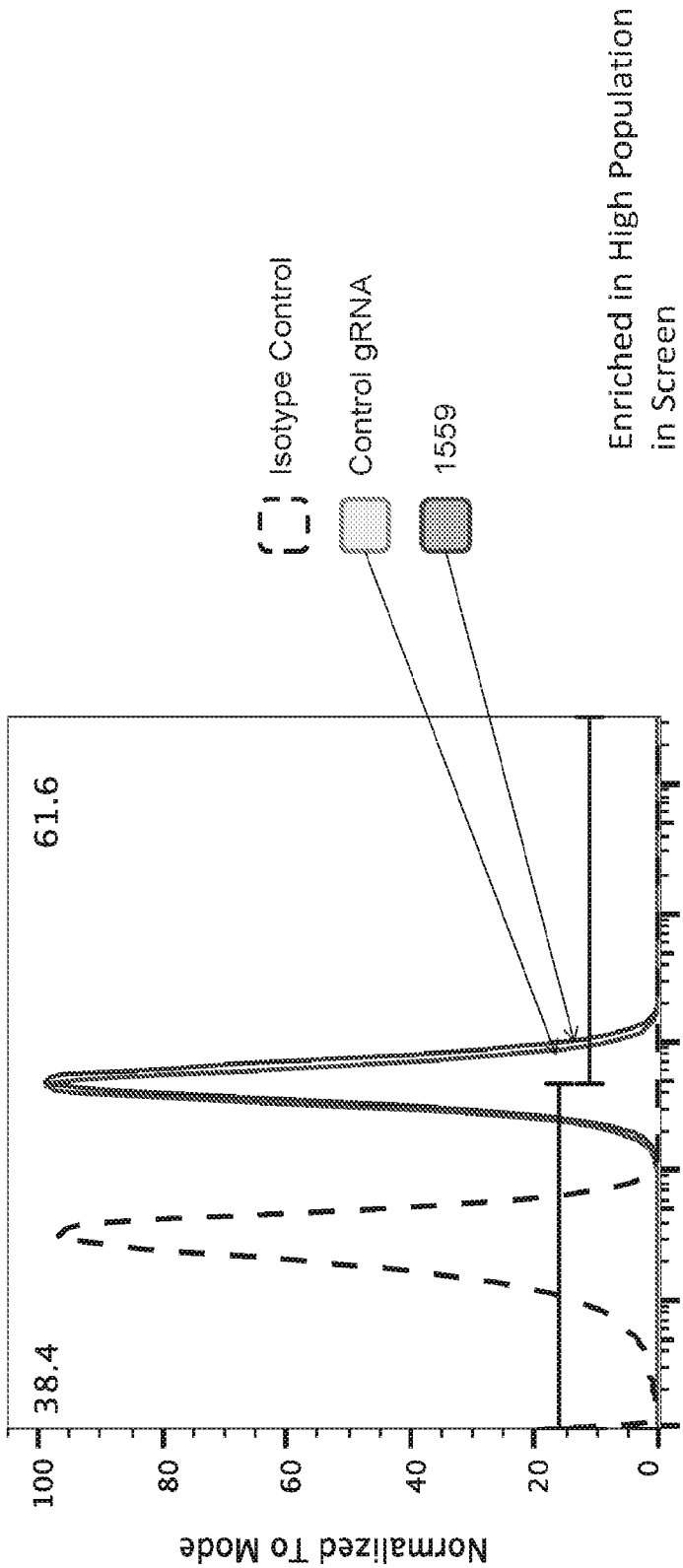

Isotype Control
Control gRNA
1559
Enriched in High Population in Screen
61.6
38.4
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 22B

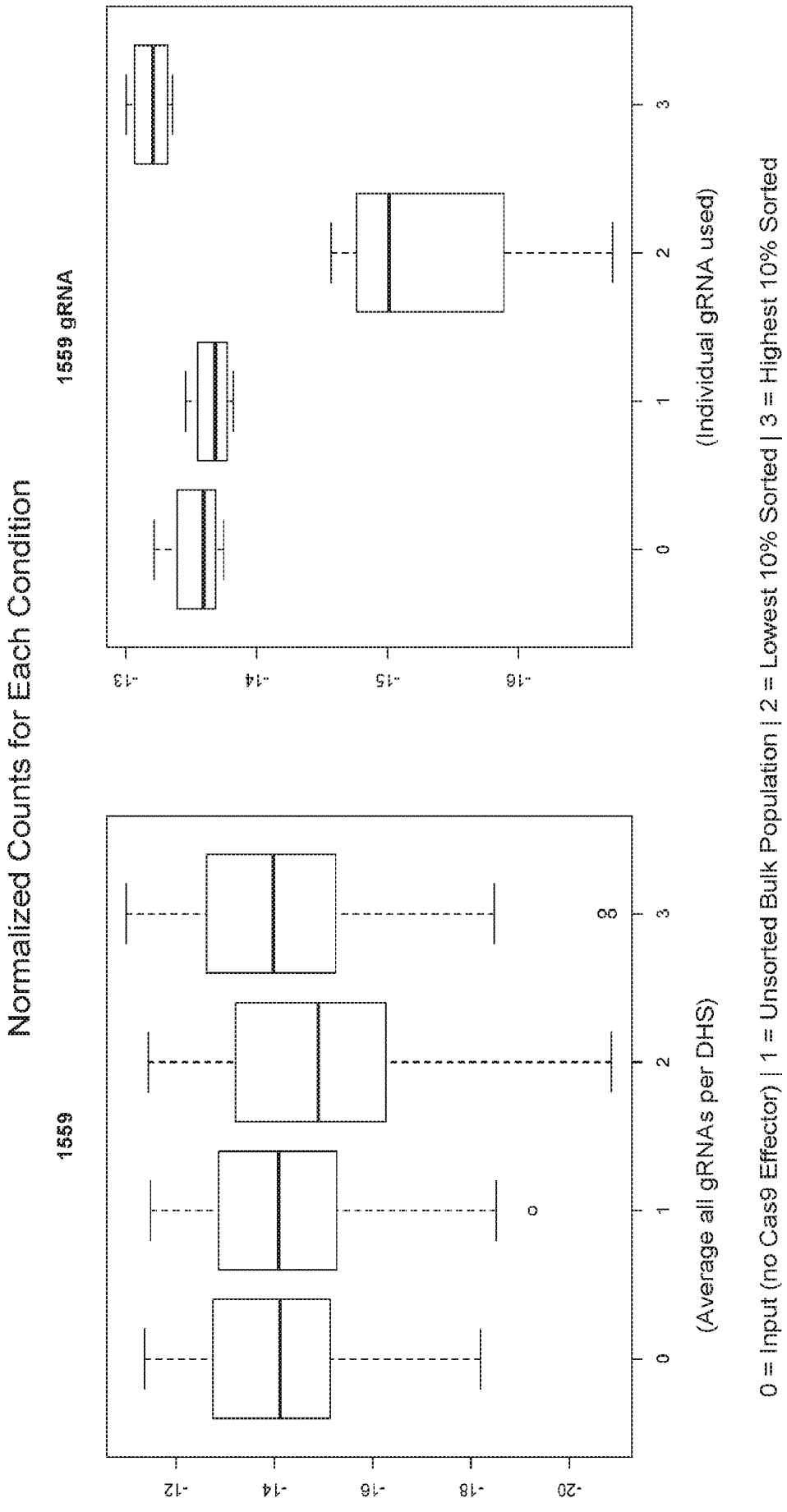

Normalized Counts for Each Condition
1559
1559 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 22C

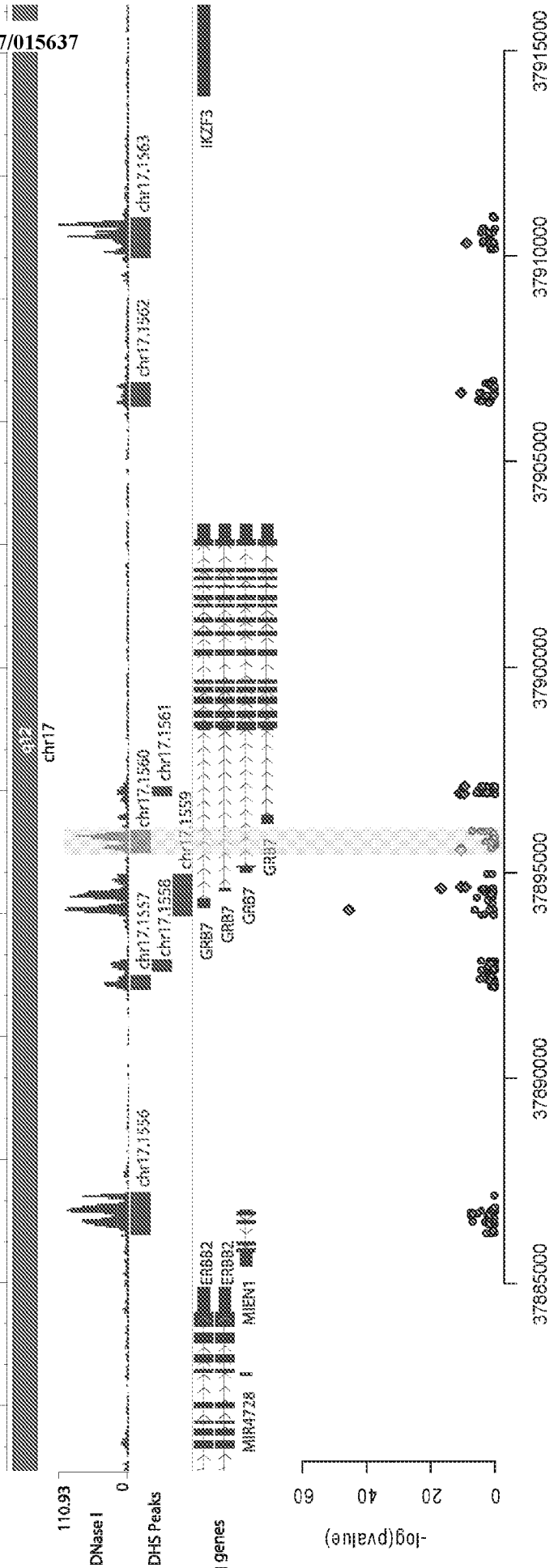

37882000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915000
chr17
110.93
0
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
chr17:1556
chr17:1557
chr17:1558
chr17:1559
chr17:1560
chr17:1561
chr17:1562
chr17:1563
ERBB2
MIEN1
MIR4728
GRB7
IKZF3
60
40
20
0
-log(pvalue)
37885000
37890000
37895000
37900000
37905000
37910000
37915000
Chr17 (bp)

FIG. 23A

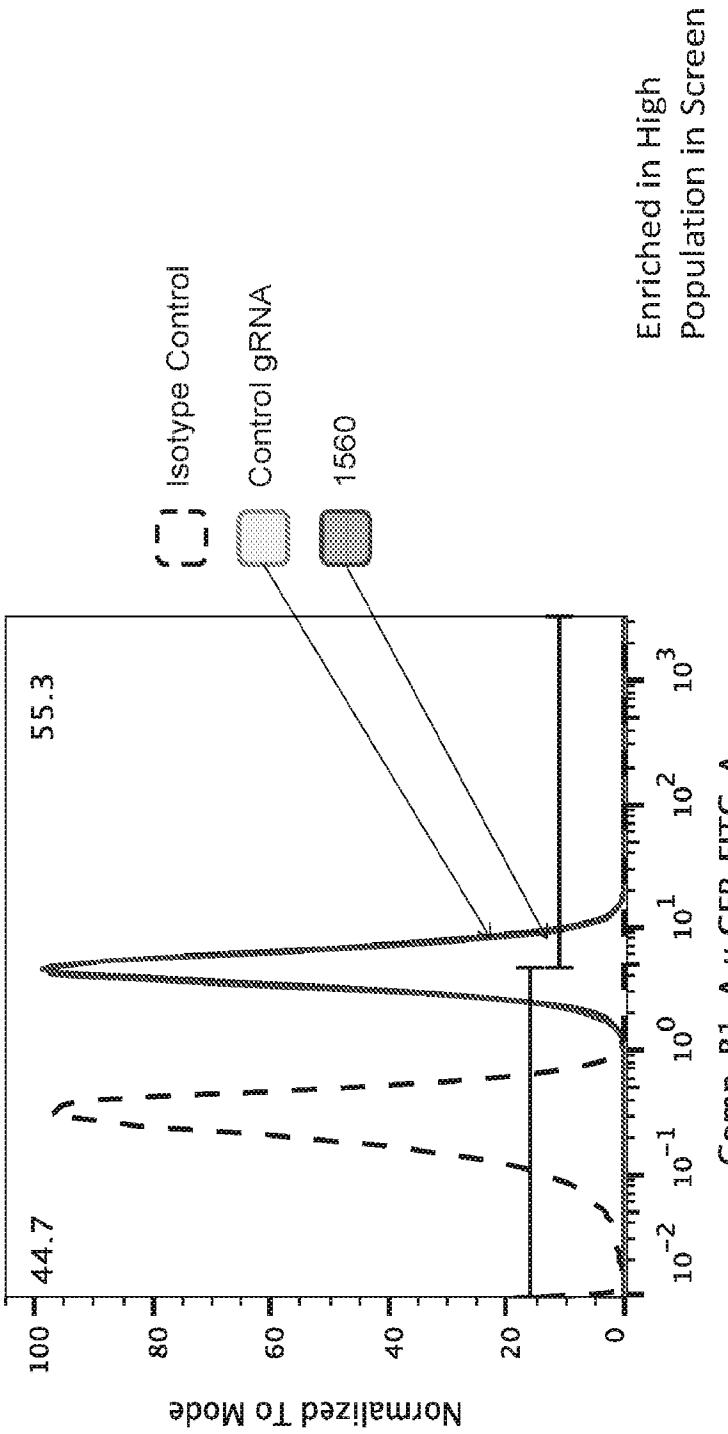

Enriched in High
Population in Screen
Isotype Control
Control gRNA
1560
55.3
44.7
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 23B

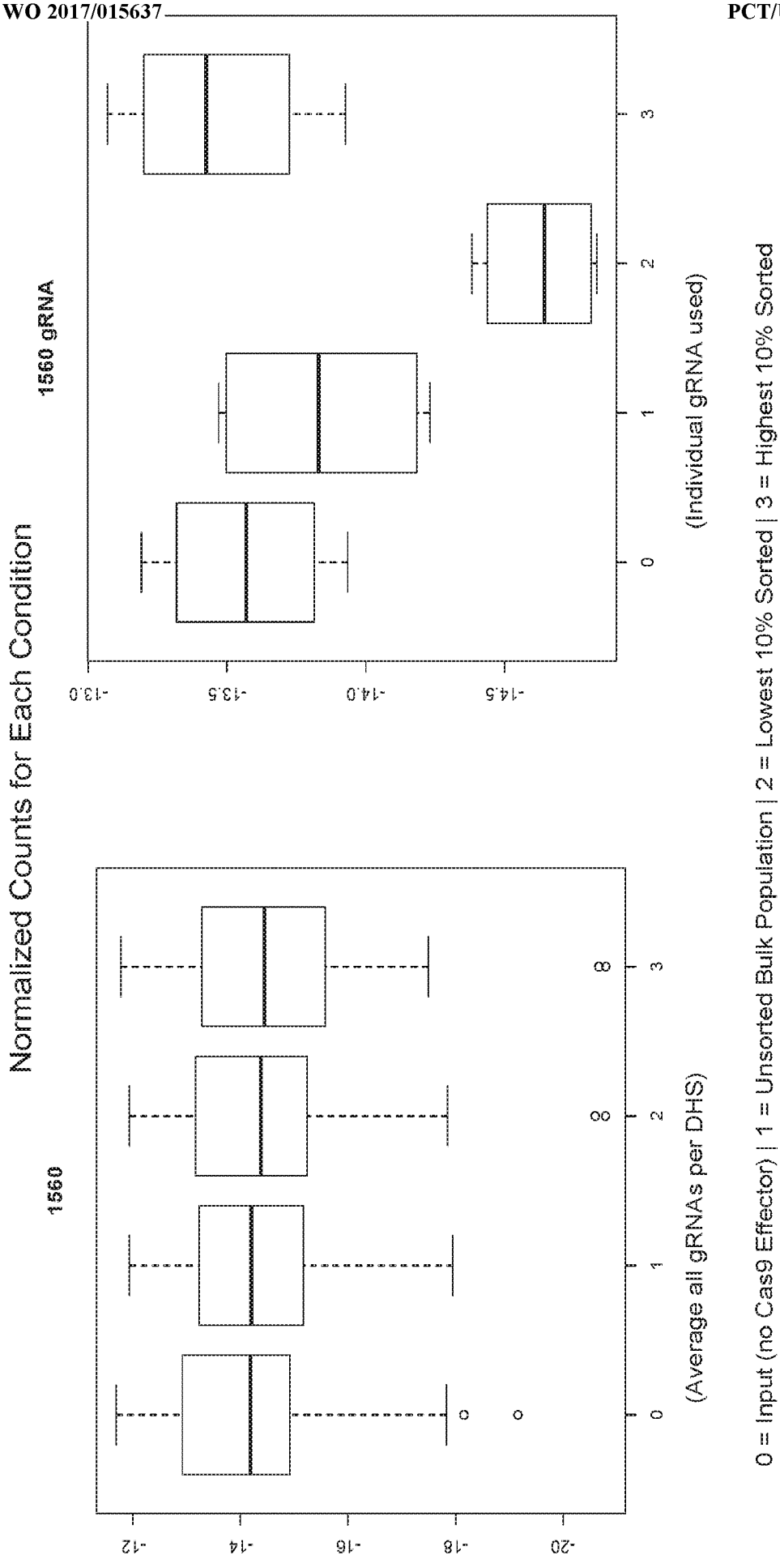

Normalized Counts for Each Condition
1560 gRNA
-13.0
-13.5
-14.0
-14.5
0
1
2
3
(Individual gRNA used)
1560
-12
-14
-16
-18
-20
0
1
2
3
(Average all gRNAs per DHS)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 23C

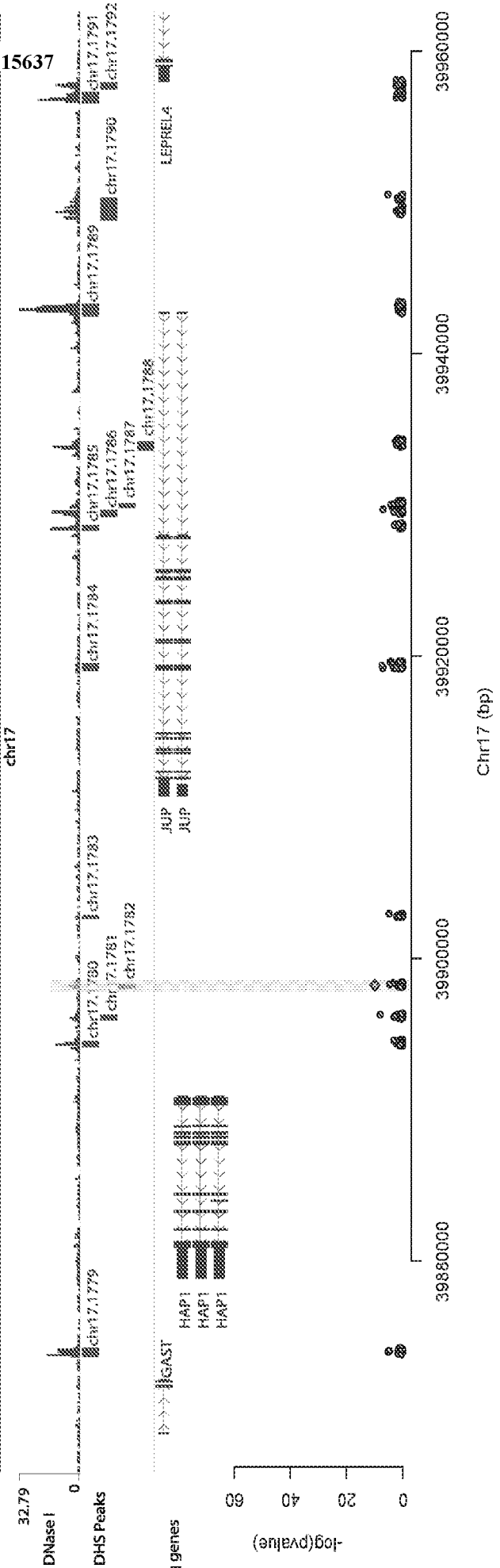

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
q21.2
LEPREL4
JUP
HAP1
BGAST

FIG. 24A

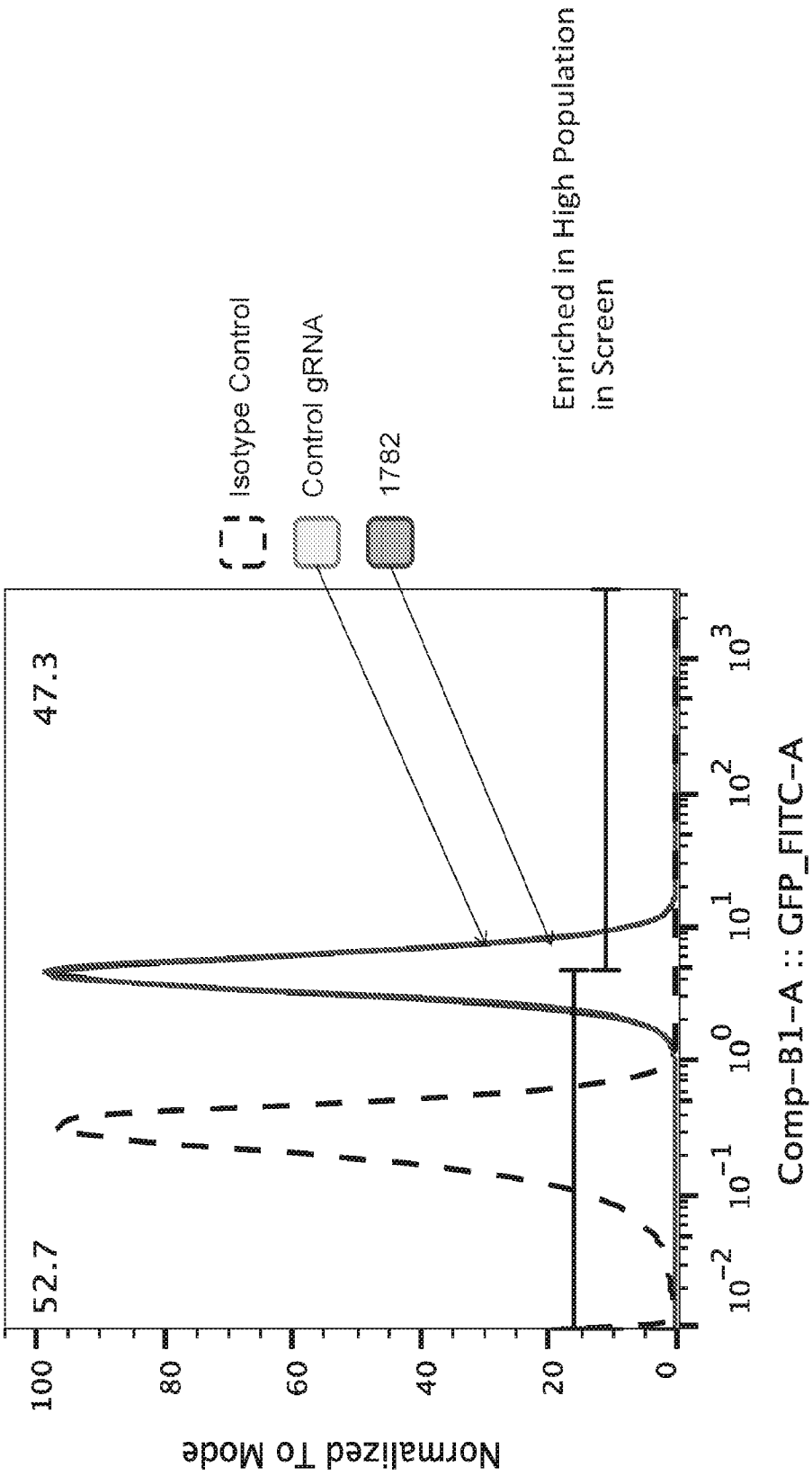

Isotype Control
Control gRNA
1782
Enriched in High Population in Screen
47.3
52.7
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 24B

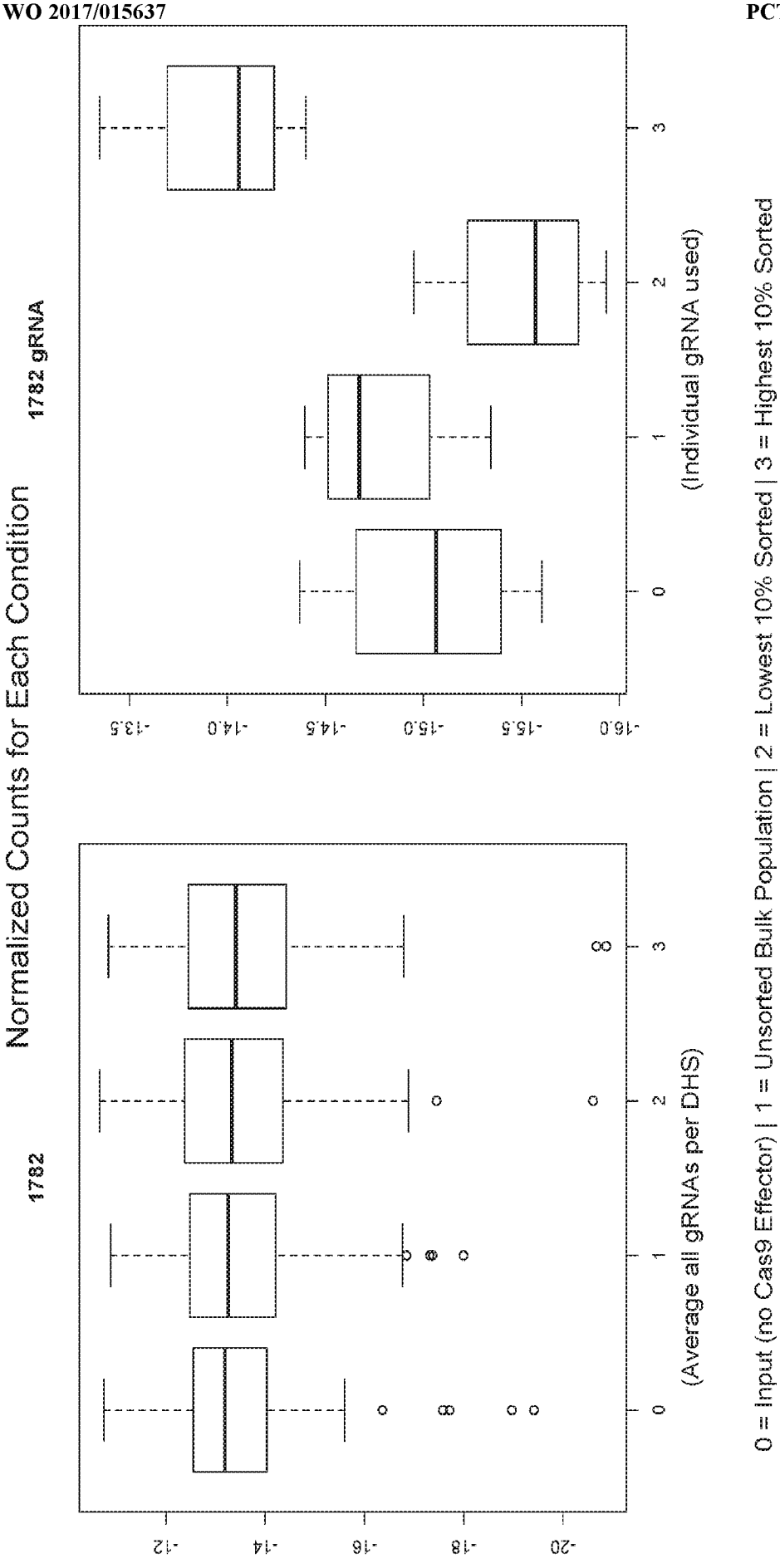

Normalized Counts for Each Condition
1782
1782 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 24C

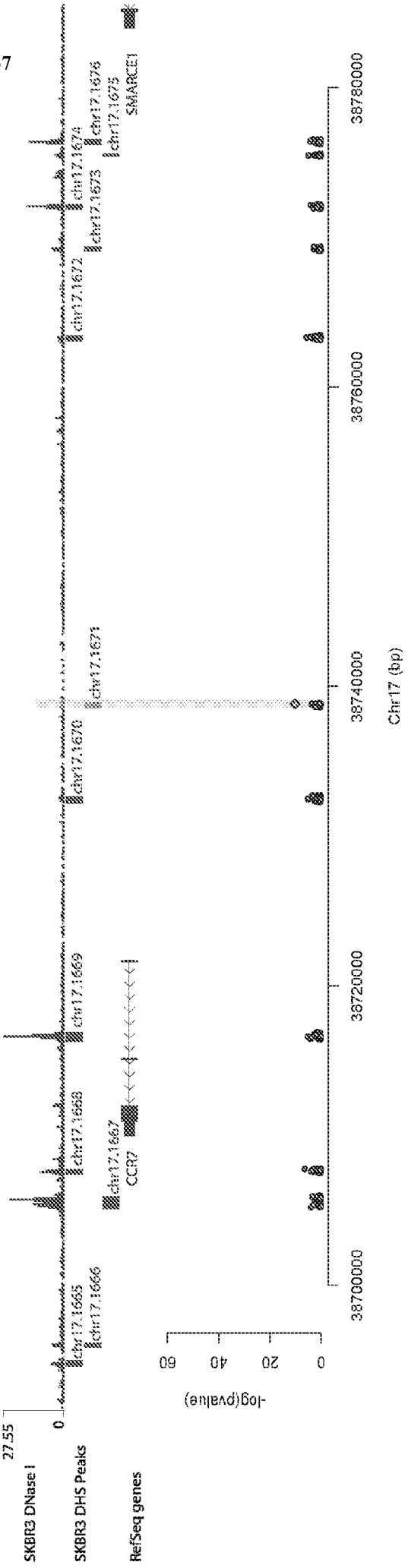

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
27.55
0
-log(pvalue)
0
20
40
60
chr17
q21.3
38700000
38710000
38720000
38730000
38740000
38750000
38760000
38770000
38780000
Chr17 (bp)
chr17.1665
chr17.1666
chr17.1667
chr17.1668
chr17.1669
chr17.1670
chr17.1671
chr17.1672
chr17.1673
chr17.1674
chr17.1675
chr17.1676
CCR7
SMARCE1

FIG. 25A

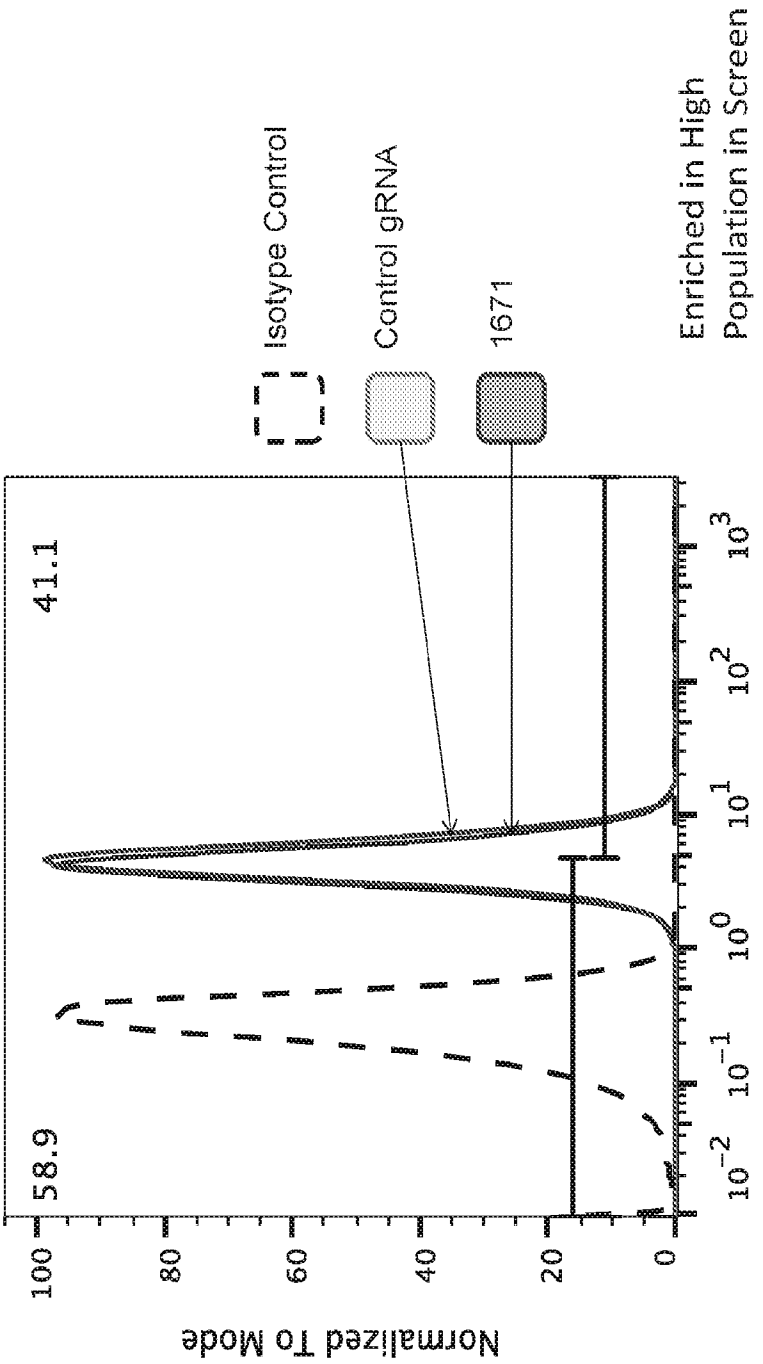

Isotype Control
Control gRNA
1671
Enriched in High Population in Screen
41.1
58.9
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 25B

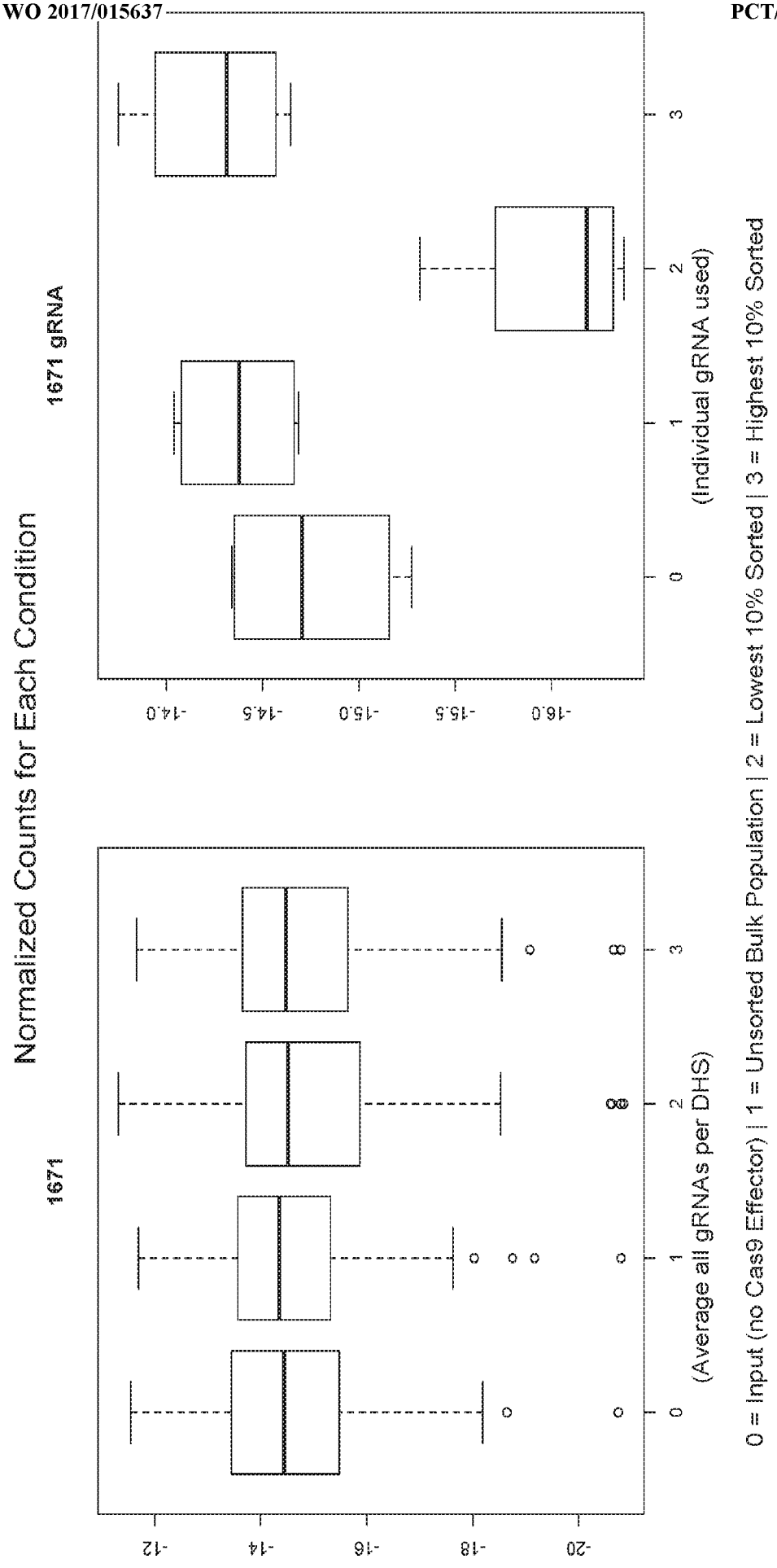

Normalized Counts for Each Condition
1671
1671 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 25C

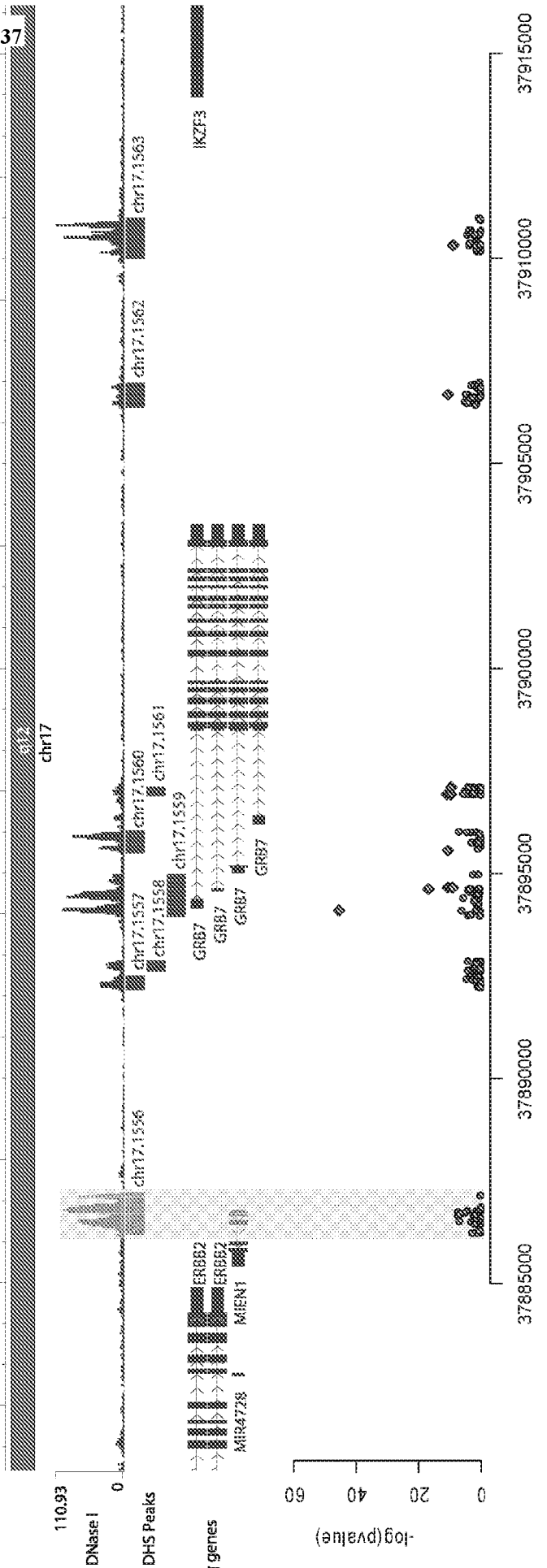

chr17
37882000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915000
SKBR3 DNase I
110.93
0
SKBR3 DHS Peaks
RefSeq genes
chr17.1556
chr17.1557
chr17.1558
chr17.1559
chr17.1560
chr17.1561
chr17.1562
chr17.1563
ERBB2
ERBB2
MIEN1
MIR4728
GRB7
GRB7
GRB7
GRB7
IKZF3
-log(pvalue)
0
20
40
60
37885000
37890000
37895000
37900000
37905000
37910000
37915000
Chr17 (bp)

FIG. 26A

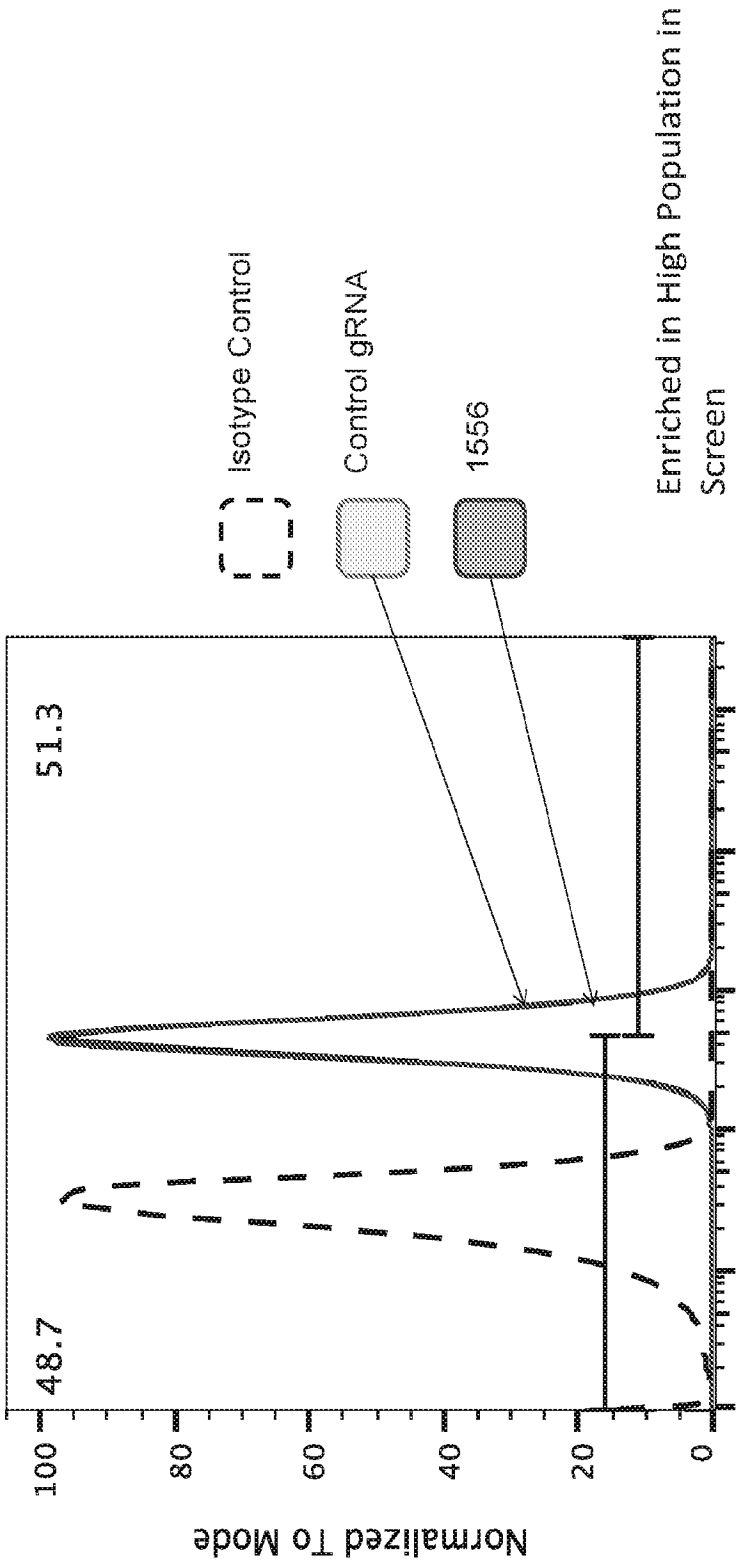

Isotype Control
Control gRNA
1556
Enriched in High Population in Screen
51.3
48.7
Normalized To Mode
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-A :: GFP_FITC-A

FIG. 26B

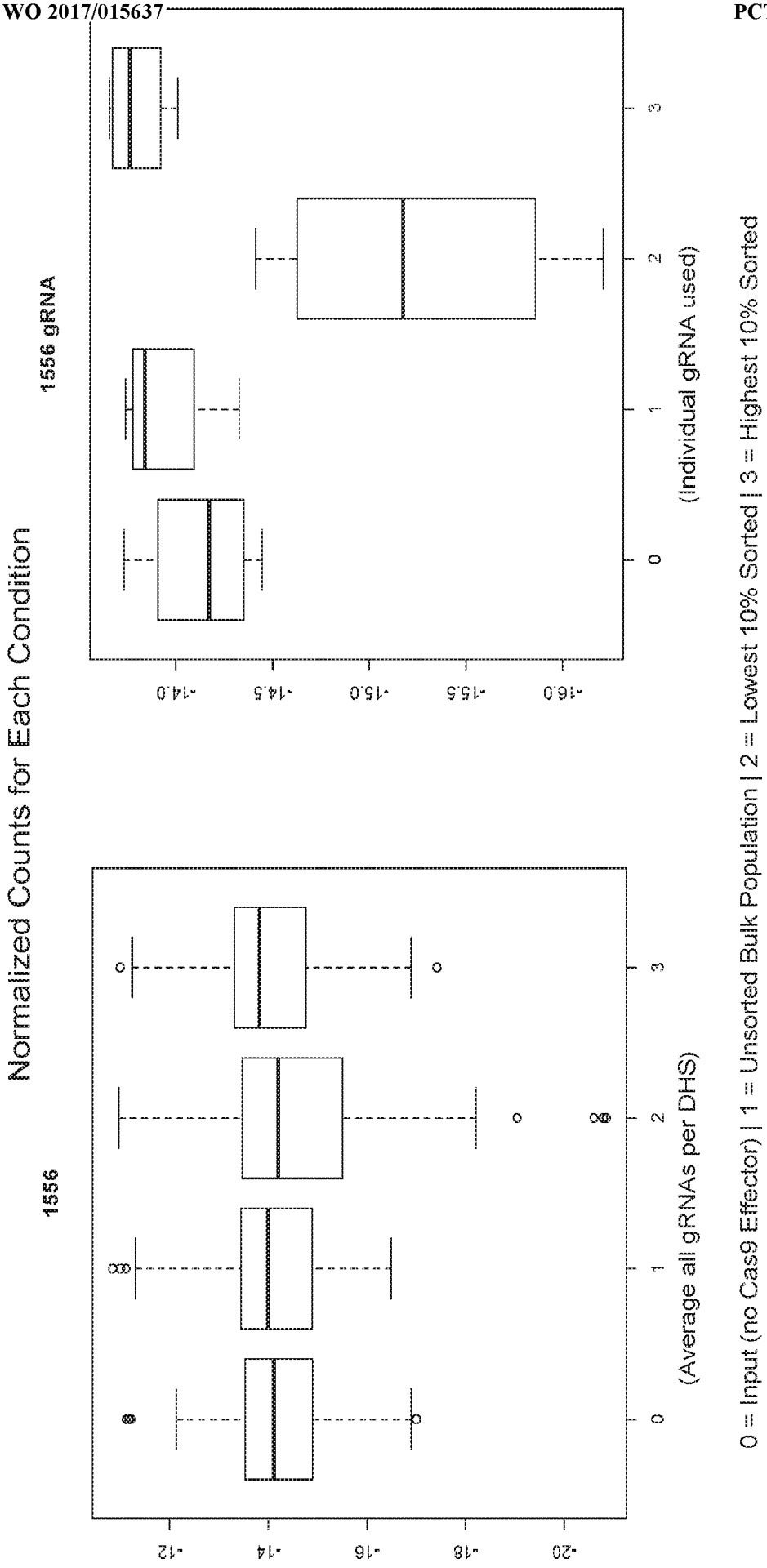

Normalized Counts for Each Condition
1556
1556 gRNA
-12
-14
-16
-18
-20
-14.0
-14.5
-15.0
-15.5
-16.0
0
1
2
3
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 26C

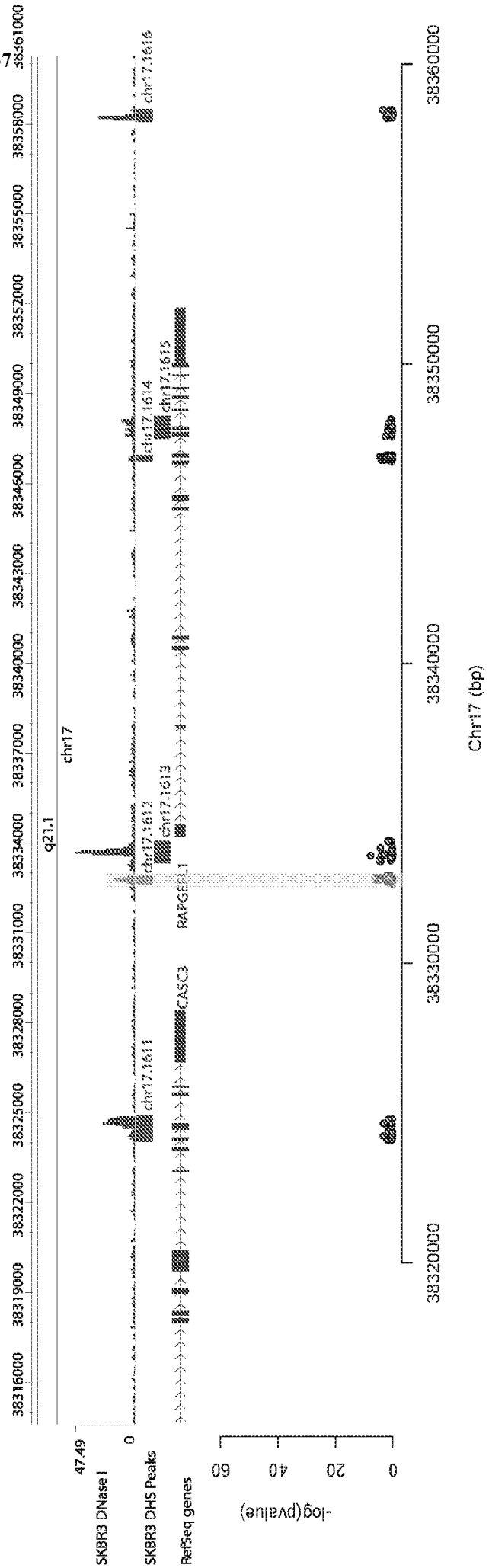

38316000
38319000
38322000
38325000
38328000
38331000
38334000
38337000
38340000
38343000
38346000
38349000
38352000
38355000
38358000
38361000
chr17
q21.1
chr17.1611
chr17.1612
chr17.1613
chr17.1614
chr17.1615
chr17.1616
CASC3
RAPGEFL1
47.49
0
SK8R3 DNase I
SK8R3 DHS Peaks
RefSeq genes
60
40
20
0
-log(pvalue)
38320000
38330000
38340000
38350000
38360000
Chr17 (bp)

FIG. 27A

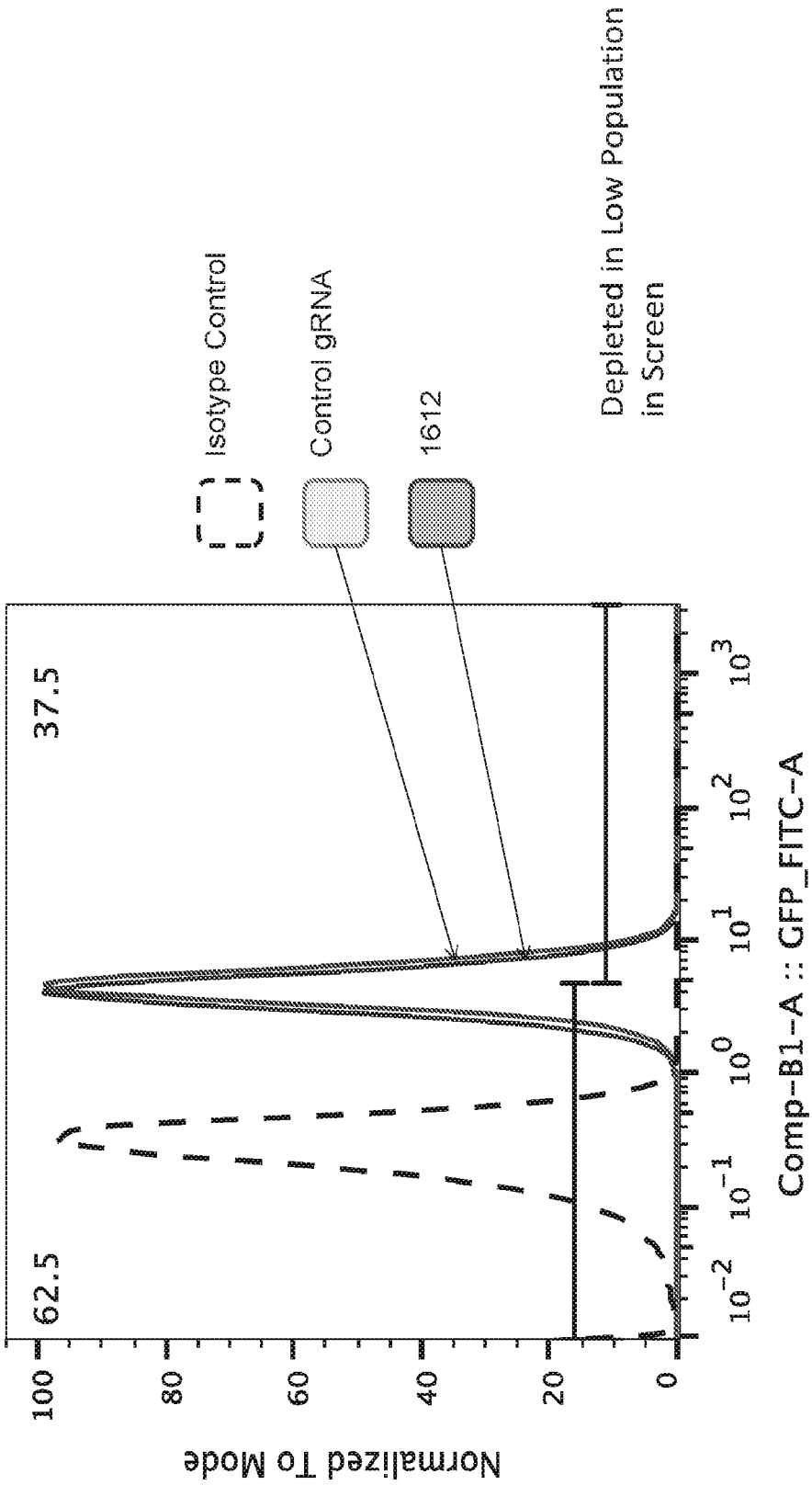

Isotype Control
Control gRNA
1612
Depleted in Low Population in Screen
37.5
62.5
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 27B

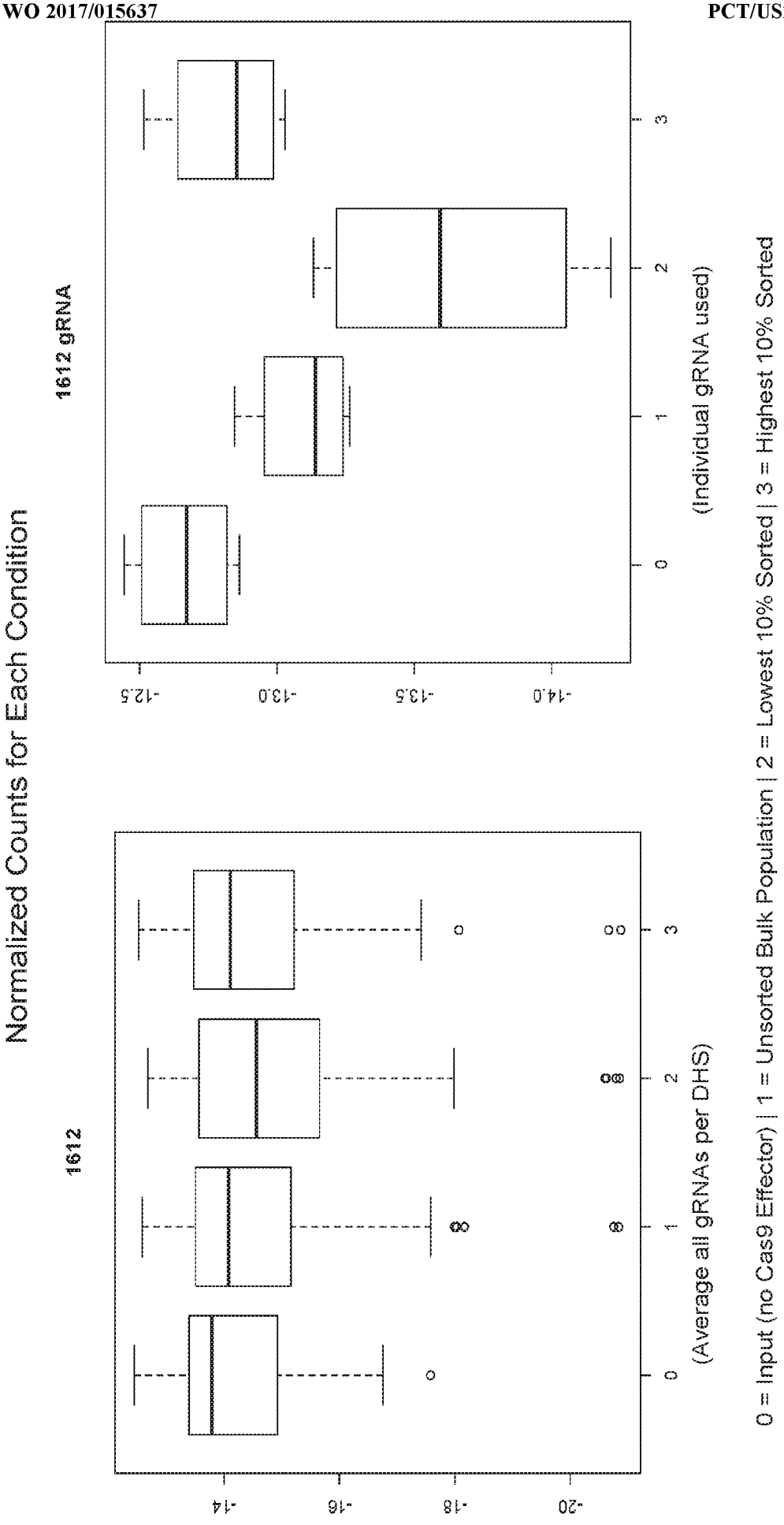

Normalized Counts for Each Condition
1612
1612 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 27C

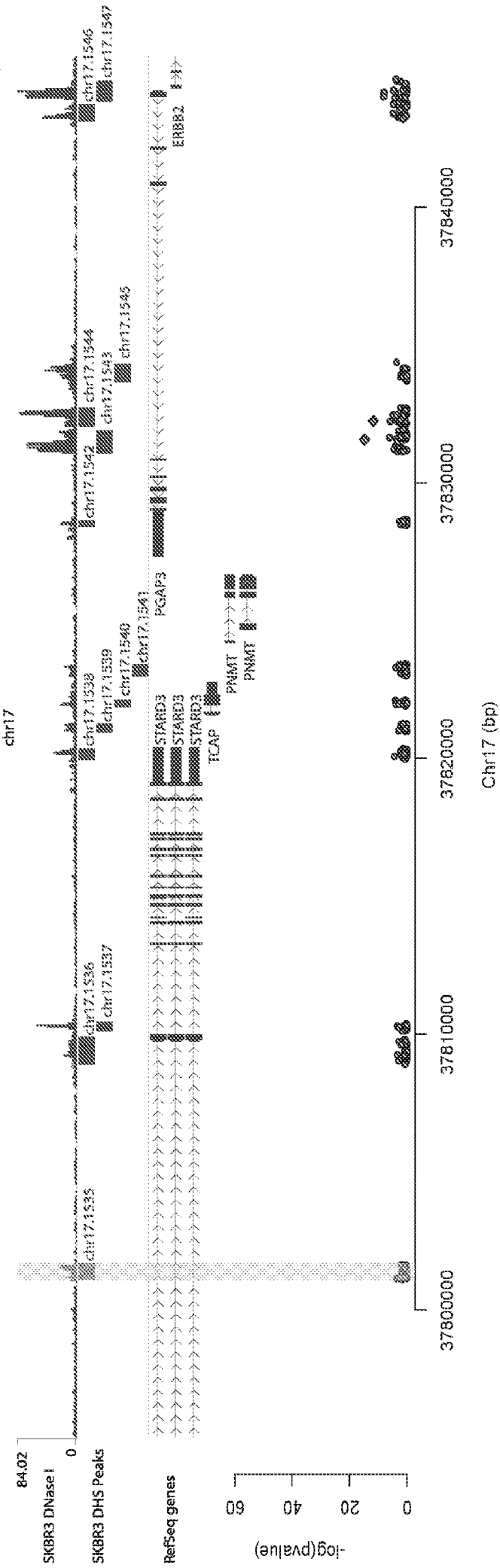

chr17
37800000
37804000
37808000
37812000
37816000
37820000
37824000
37828000
37832000
37836000
37840000
37844000
chr17:1535
chr17:1536
chr17:1537
chr17:1538
chr17:1539
chr17:1540
chr17:1541
chr17:1542
chr17:1543
chr17:1544
chr17:1545
chr17:1546
chr17:1547
PGAP3
STARD3
STARD3
STARD3
TCAP
PNMT
PNMT
ERBB2
84.02
0
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
0
20
40
60
-log(pvalue)
37800000
37810000
37820000
37830000
37840000
Chr17 (bp)

FIG. 28A

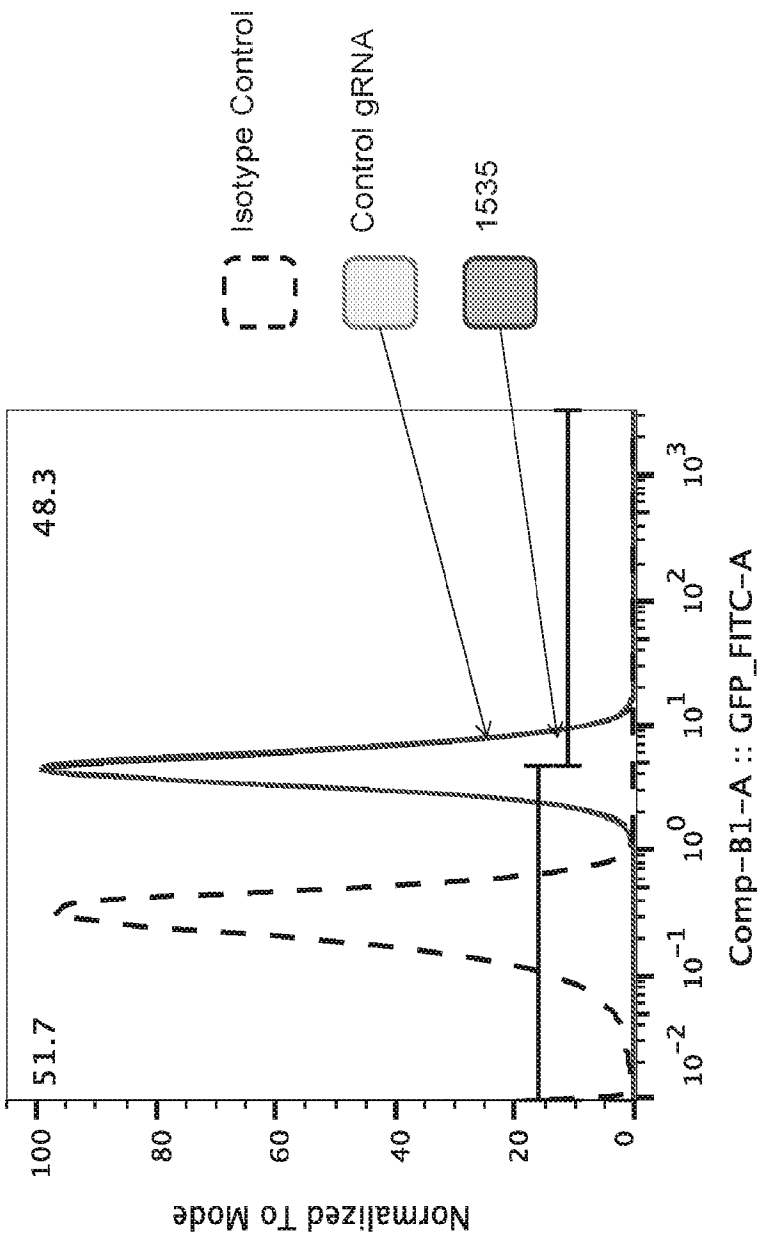

Isotype Control
Control gRNA
1535
48.3
51.7
Normalized To Mode
100
80
60
40
20
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-A :: GFP_FITC-A

FIG. 28B

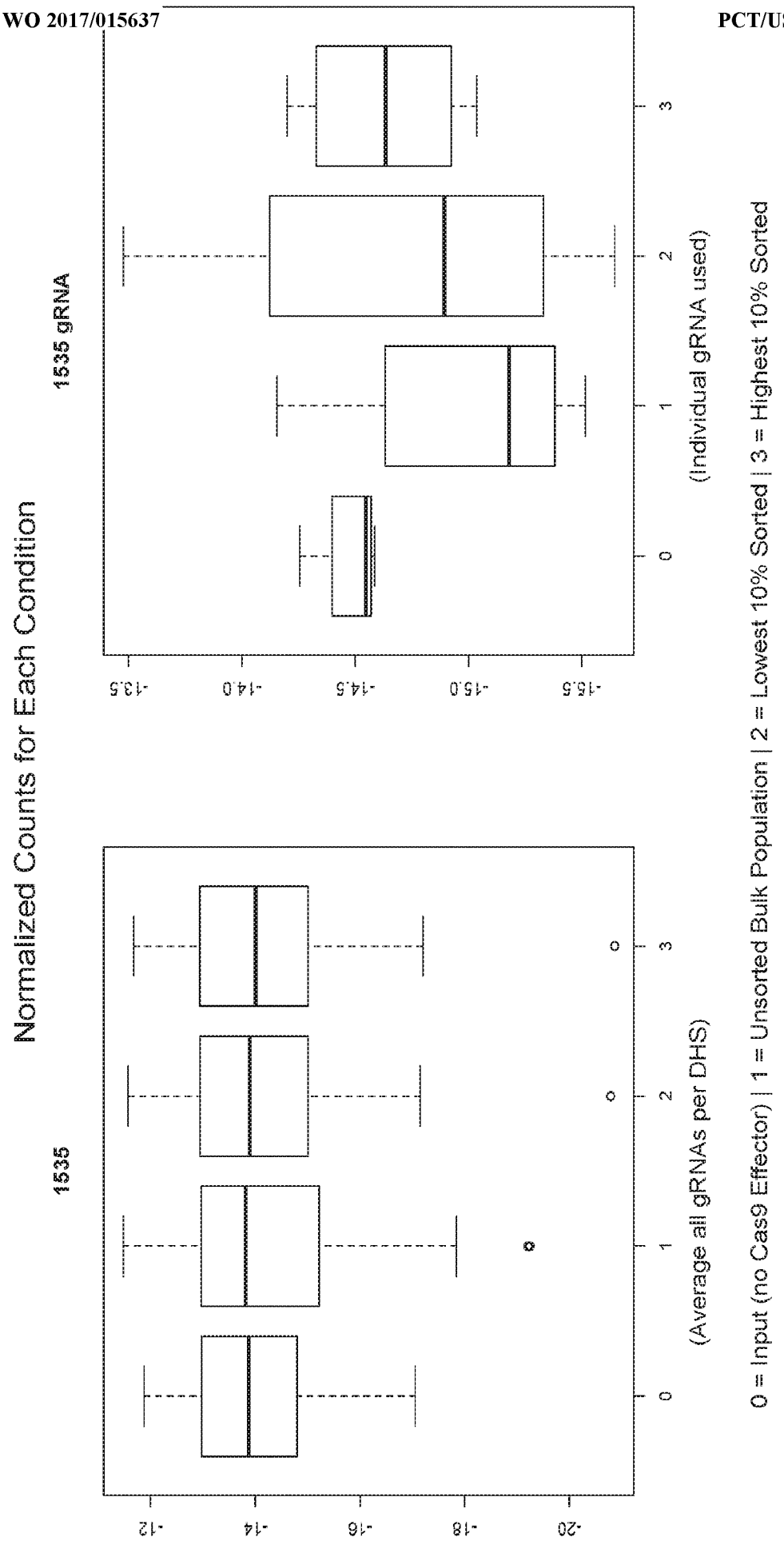

Normalized Counts for Each Condition
1535
1535 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 28C

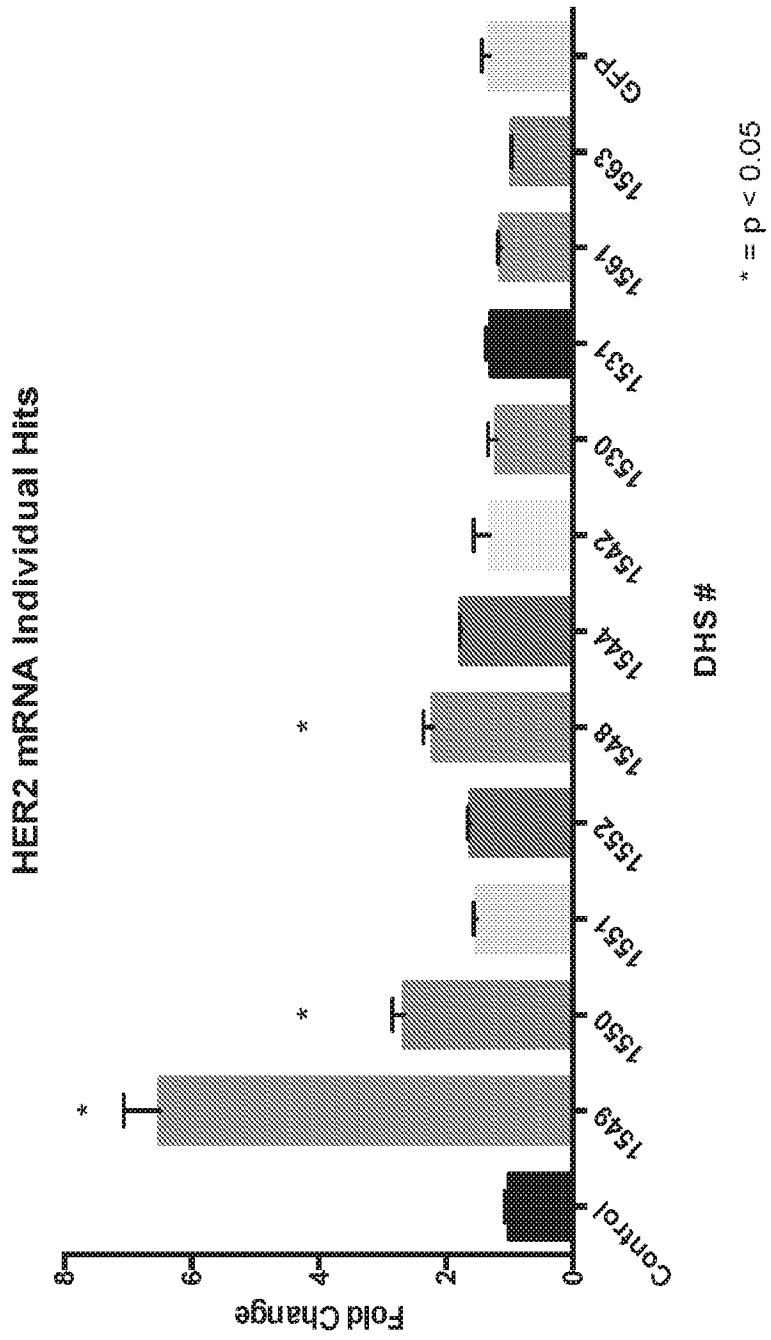

HER2 mRNA Individual Hits
Fold Change
0
2
4
6
8
Control
1549
1550
1551
1552
1548
1544
1542
1530
1531
1561
1563
GFP
DHS #
* = p < 0.05

FIG. 29

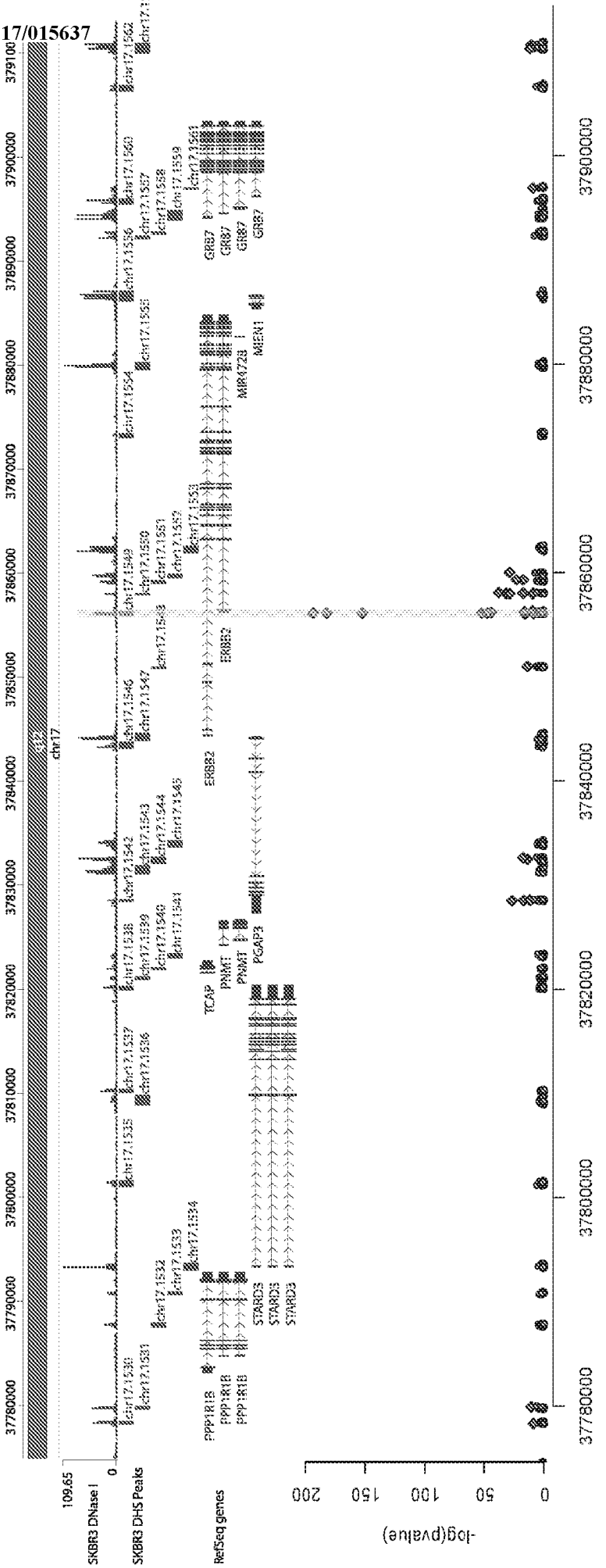

chr17
SKBR3 DNase I
SKBR3 DH5 Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
37780000
37790000
37800000
37810000
37820000
37830000
37840000
37850000
37860000
37870000
37880000
37890000
37900000
37910000
ERBB2
PNMT
PGAP3
STARD3
TCAP
PPP1R1B
MIEN1
GRB7

FIG. 30A

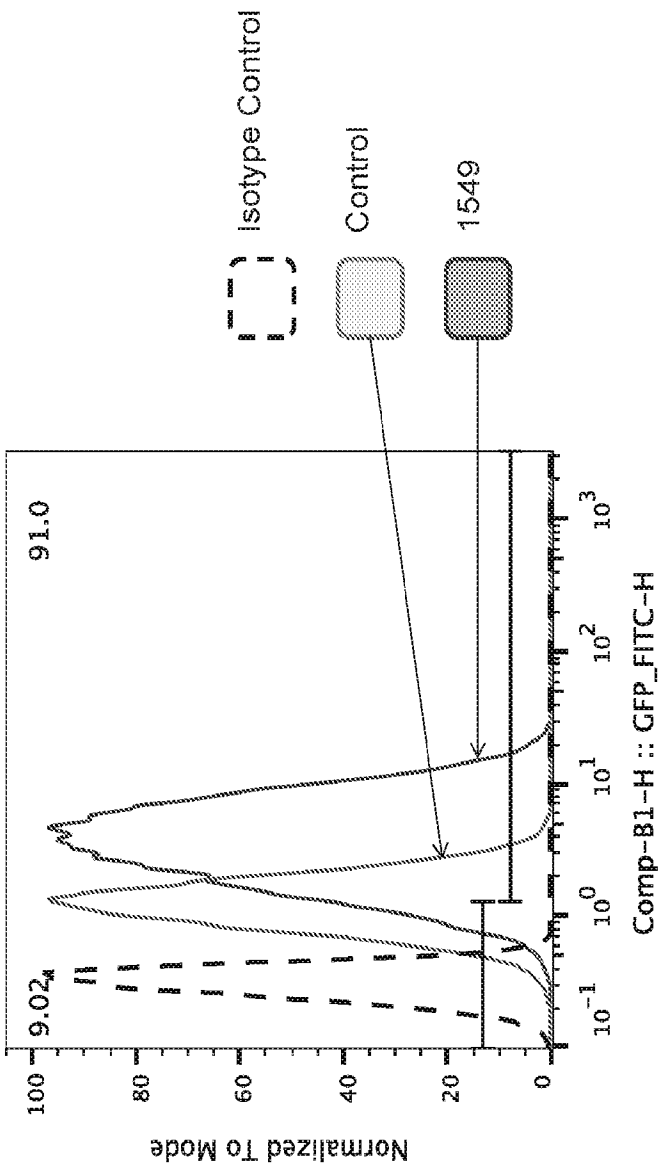

Isotype Control
Control
1549
91.0
9.02
Normalized To Mode
Comp-B1-H :: GFP_FITC-H

FIG. 30B

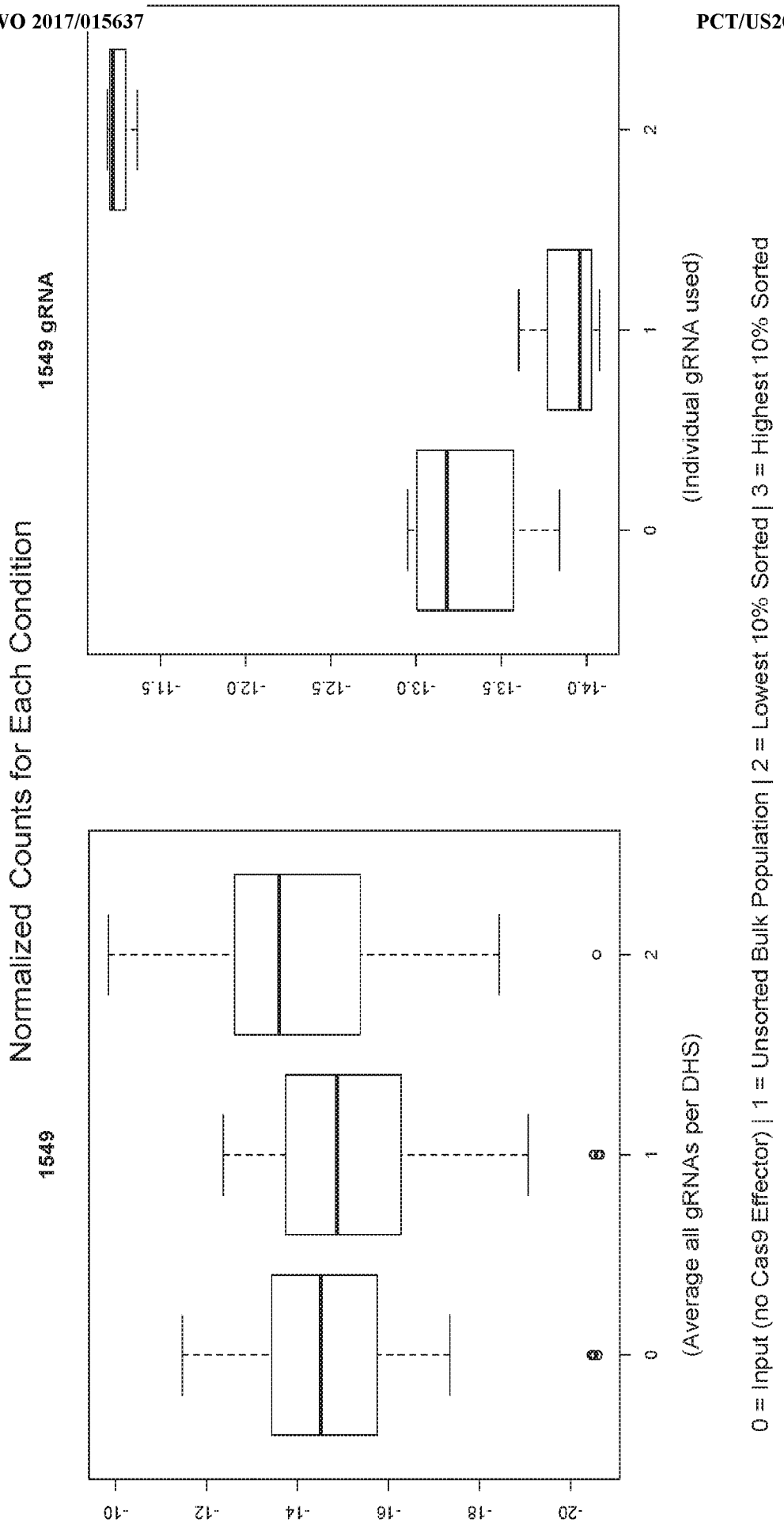

Normalized Counts for Each Condition
1549
1549 gRNA
-10
-12
-14
-16
-18
-20
-11.5
-12.0
-12.5
-13.0
-13.5
-14.0
0
1
2
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 30C

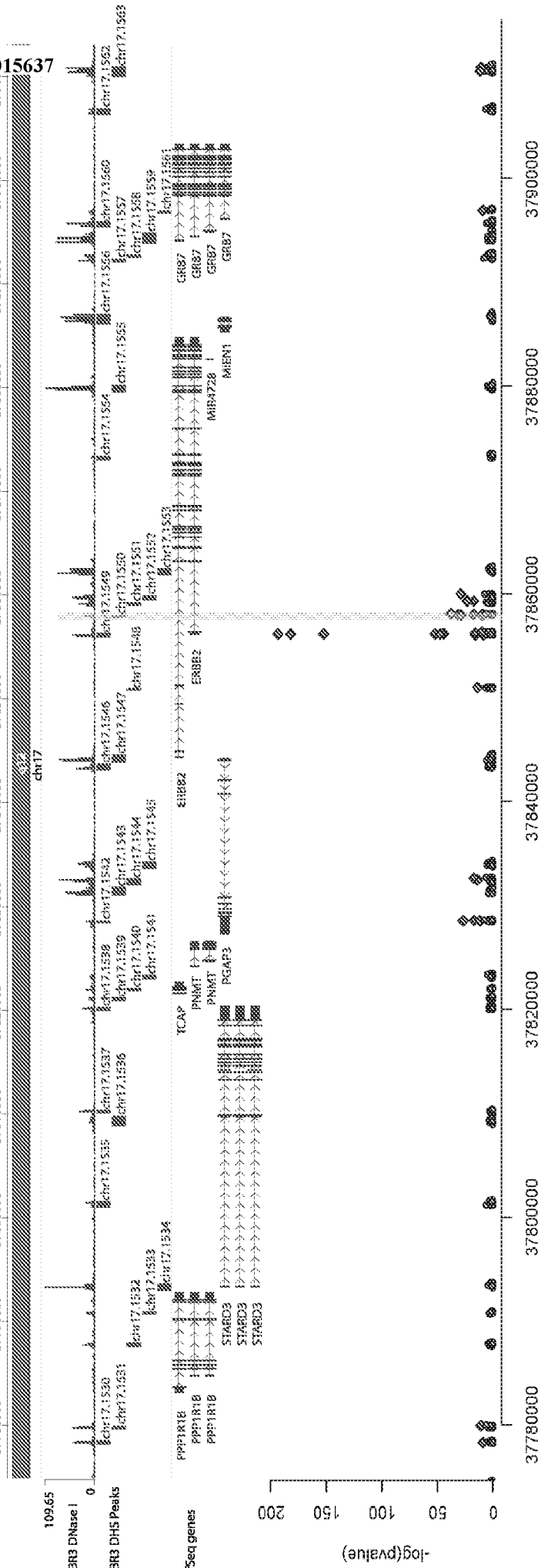

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
37780000
37790000
37800000
37810000
37820000
37830000
37840000
37850000
37860000
37870000
37880000
37890000
37900000
ERBB2
PNMT
PGAP3
TCAP
STARD3
PPP1R1B
MIEN1
GRB7

FIG. 31A

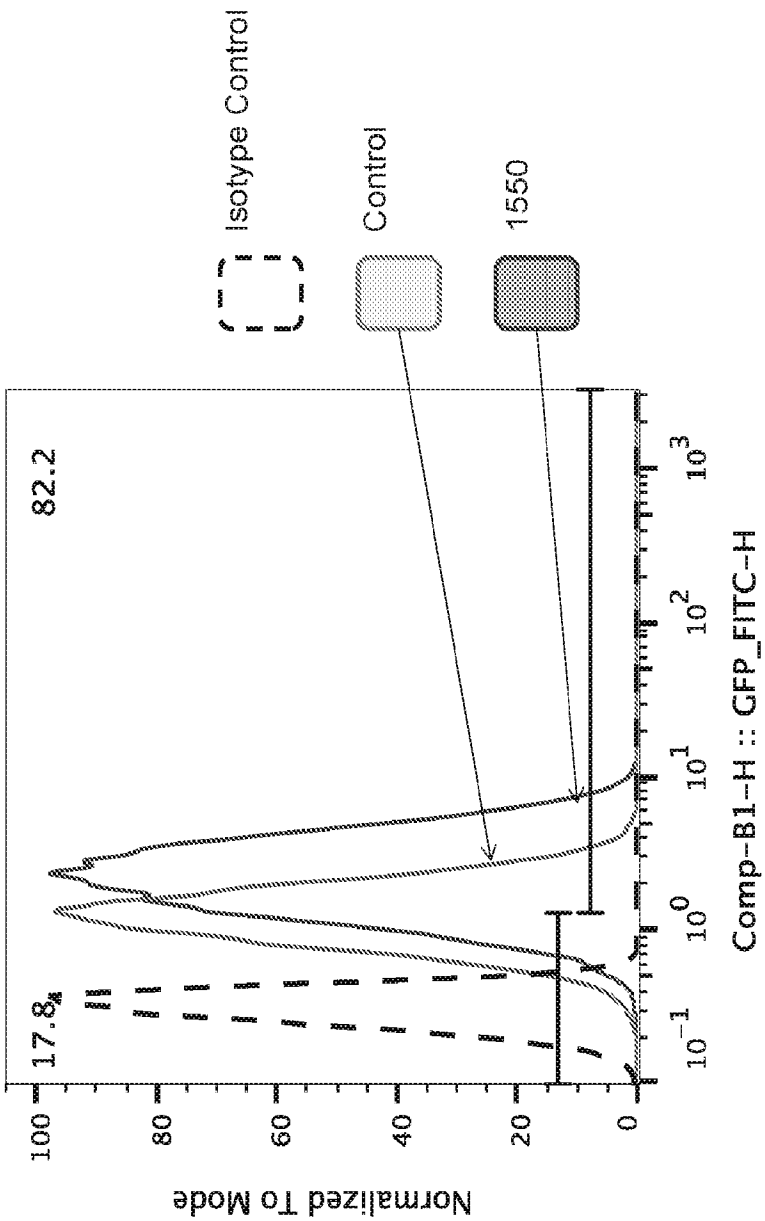

Isotype Control
Control
1550
82.2
17.8
Normalized To Mode
Comp-B1-H :: GFP_FITC-H
0
20
40
60
80
100
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}

FIG. 31B

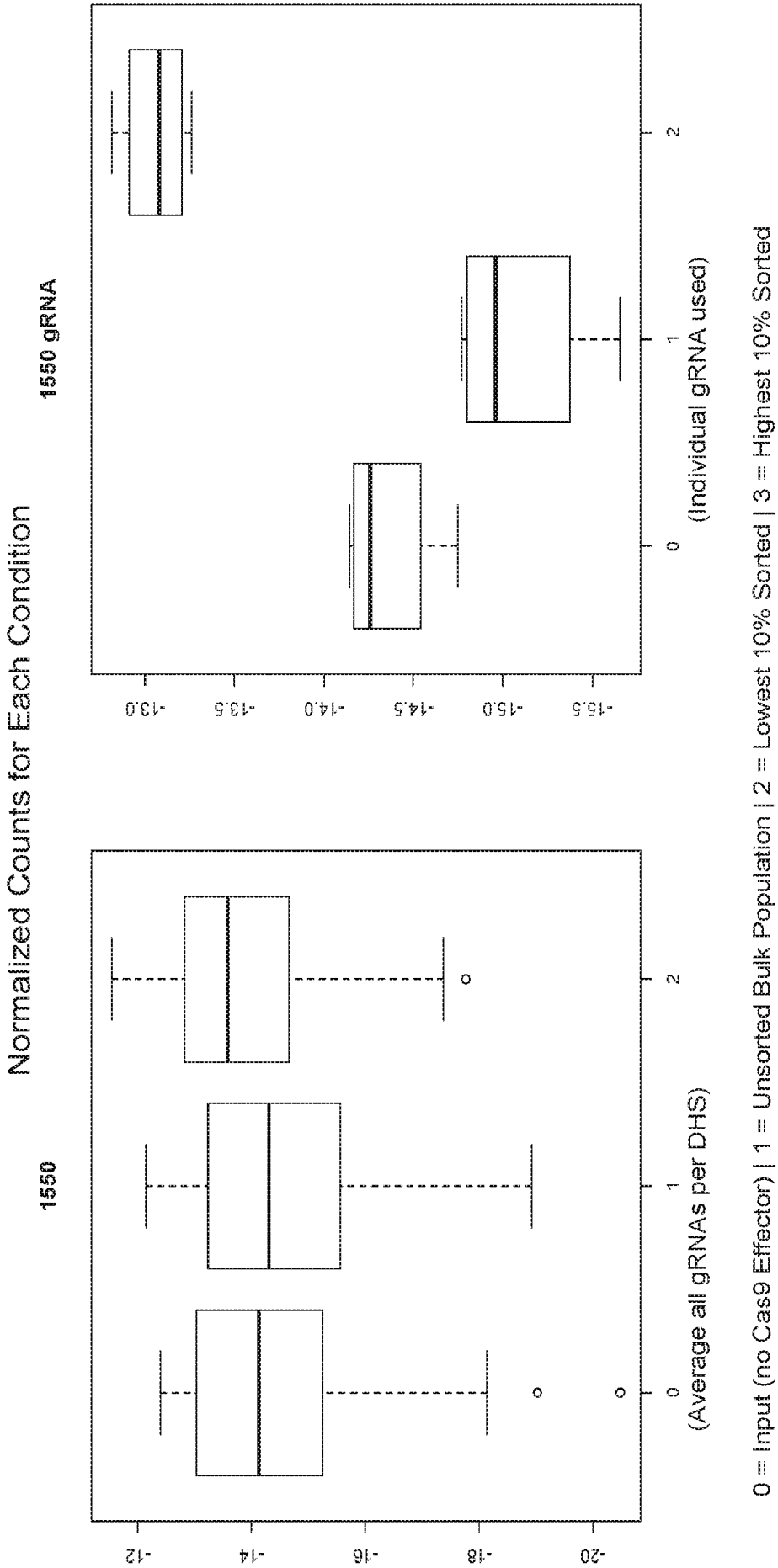

Normalized Counts for Each Condition
1550
1550 gRNA
-12
-14
-16
-18
-20
-13.0
-13.5
-14.0
-14.5
-15.0
-15.5
0
1
2
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 31C

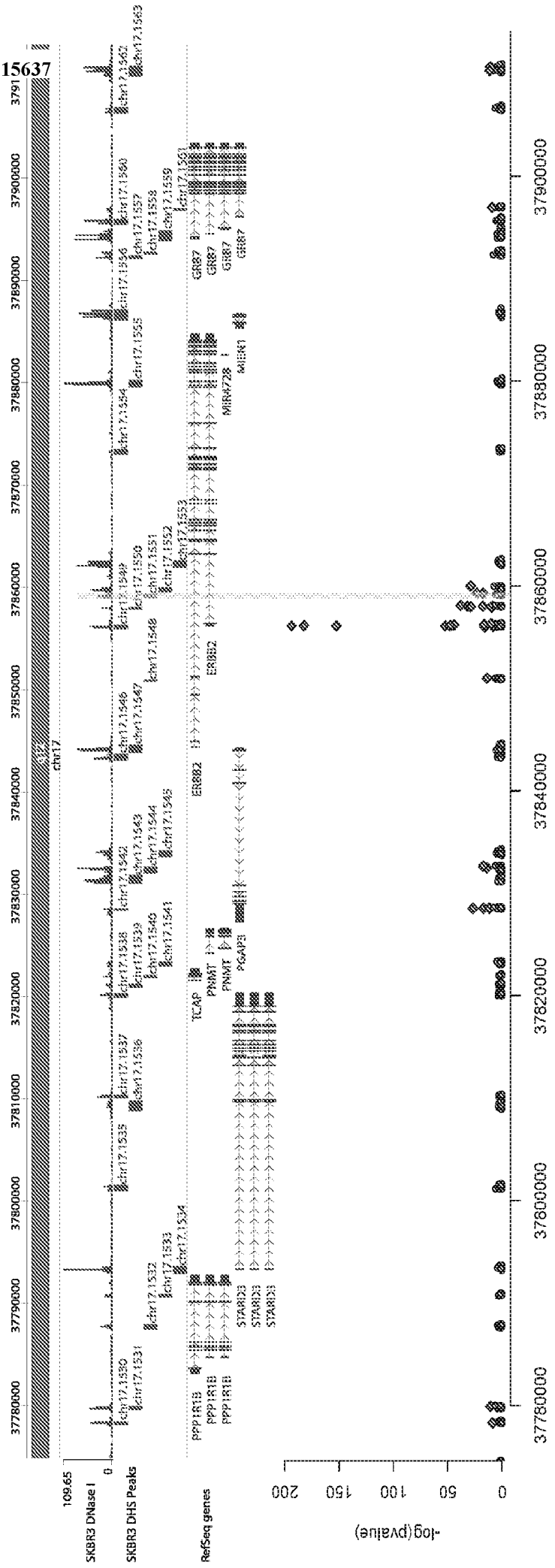

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
37780000
37800000
37820000
37840000
37860000
37880000
37900000

FIG. 32A

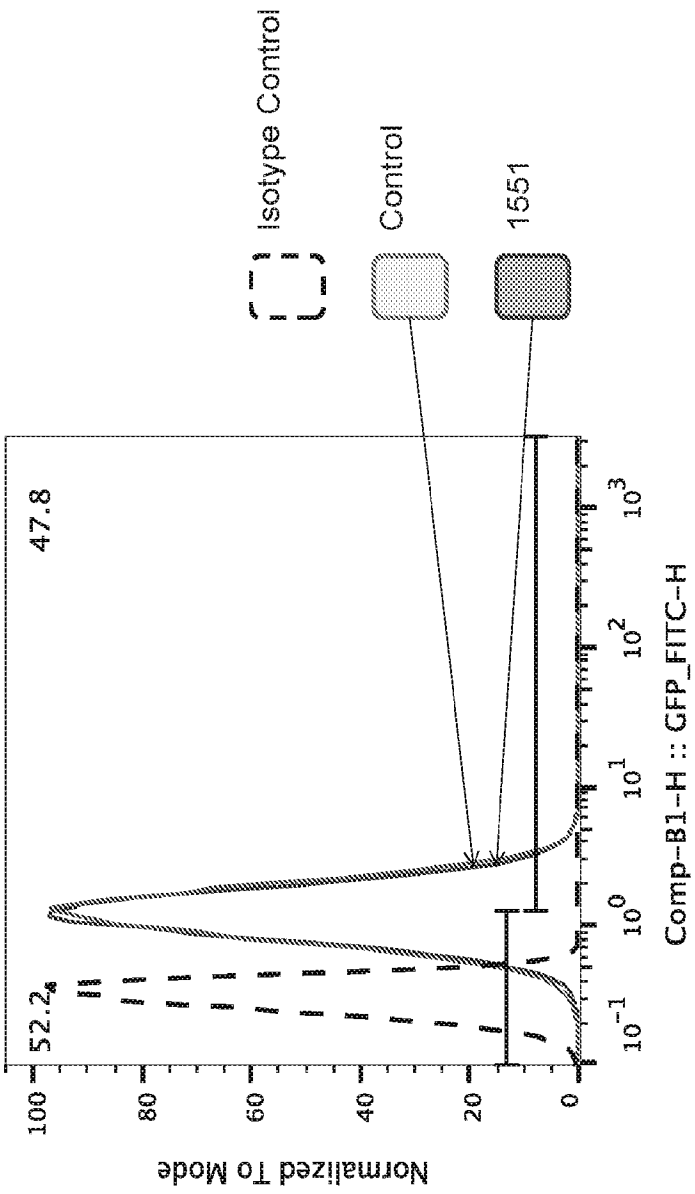

Isotype Control
Control
1551
47.8
52.2
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 32B

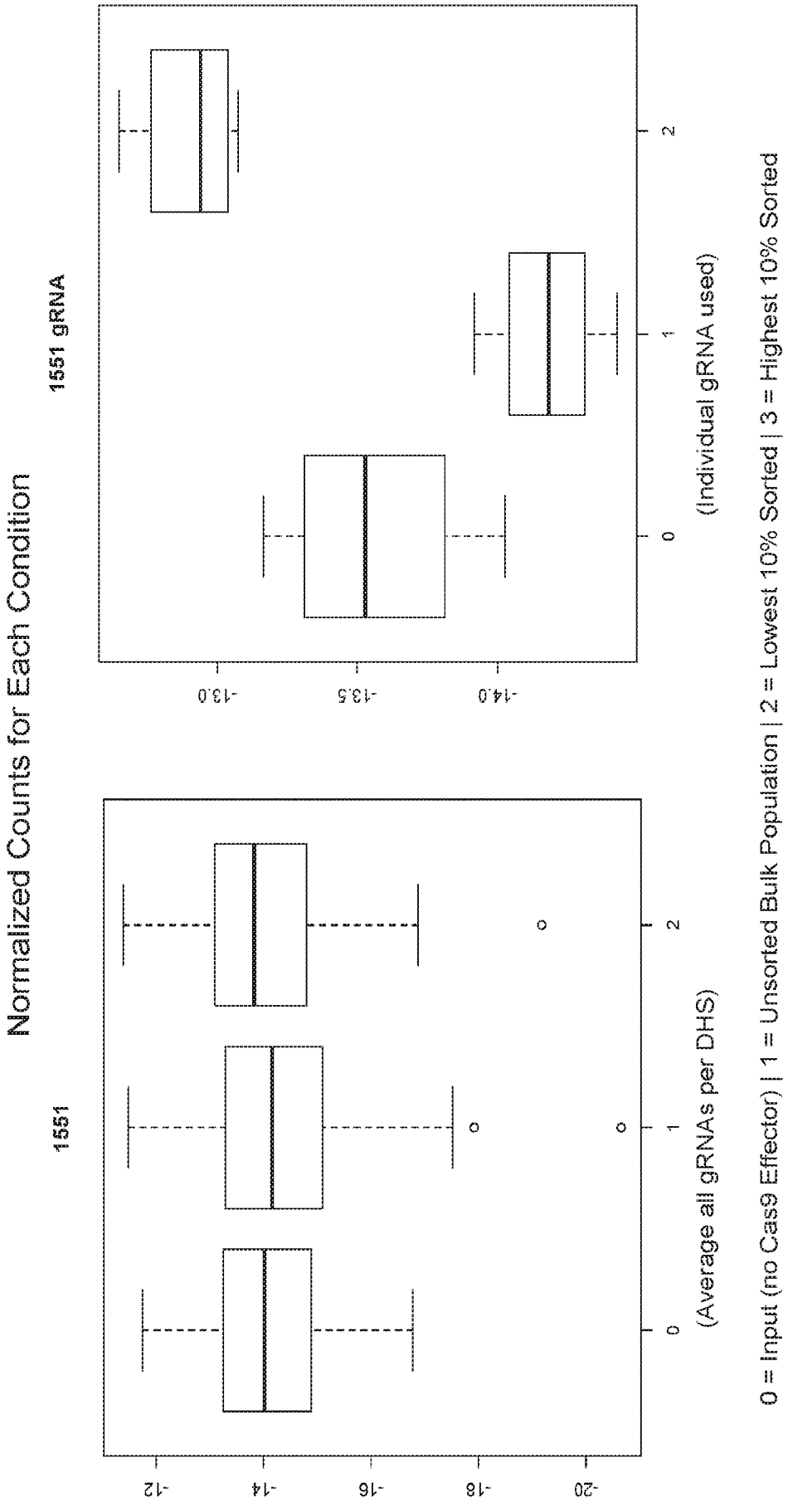

Normalized Counts for Each Condition
1551
-12
-14
-16
-18
-20
0
1
2
(Average all gRNAs per DHS)
1551 gRNA
-13.0
-13.5
-14.0
0
1
2
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 32C

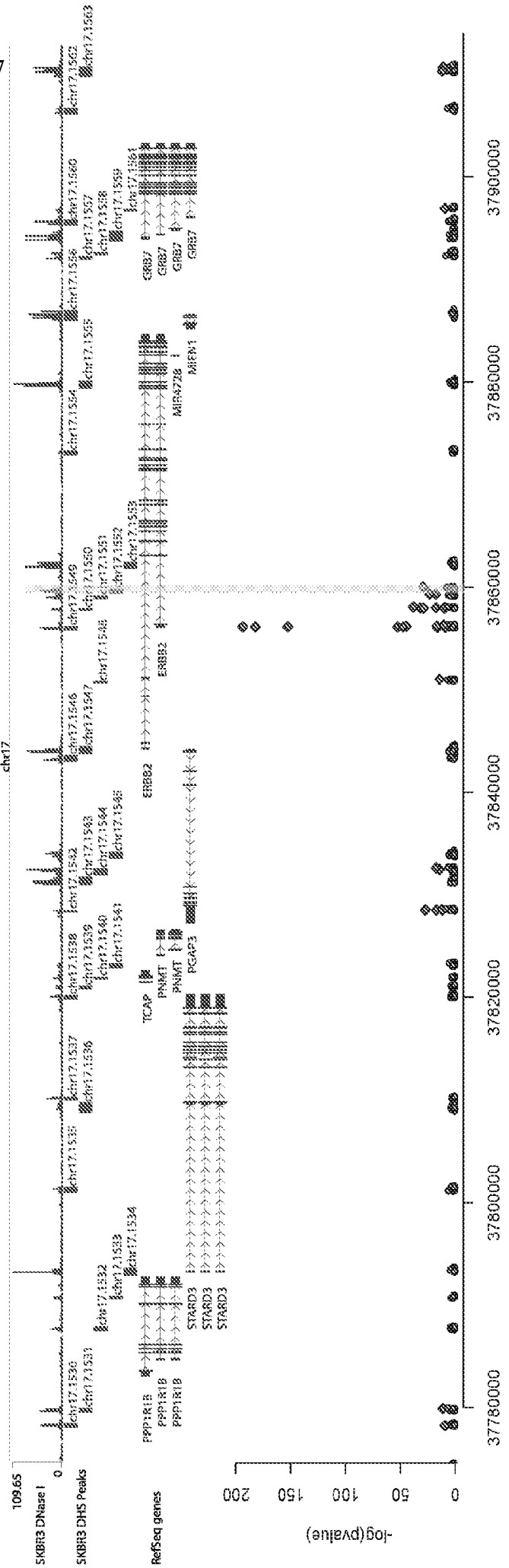

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
ERBB2
PNMT
PGAP3
TCAP
STARD3
PPP1R1B
MIEN1
GRB7

FIG. 33A

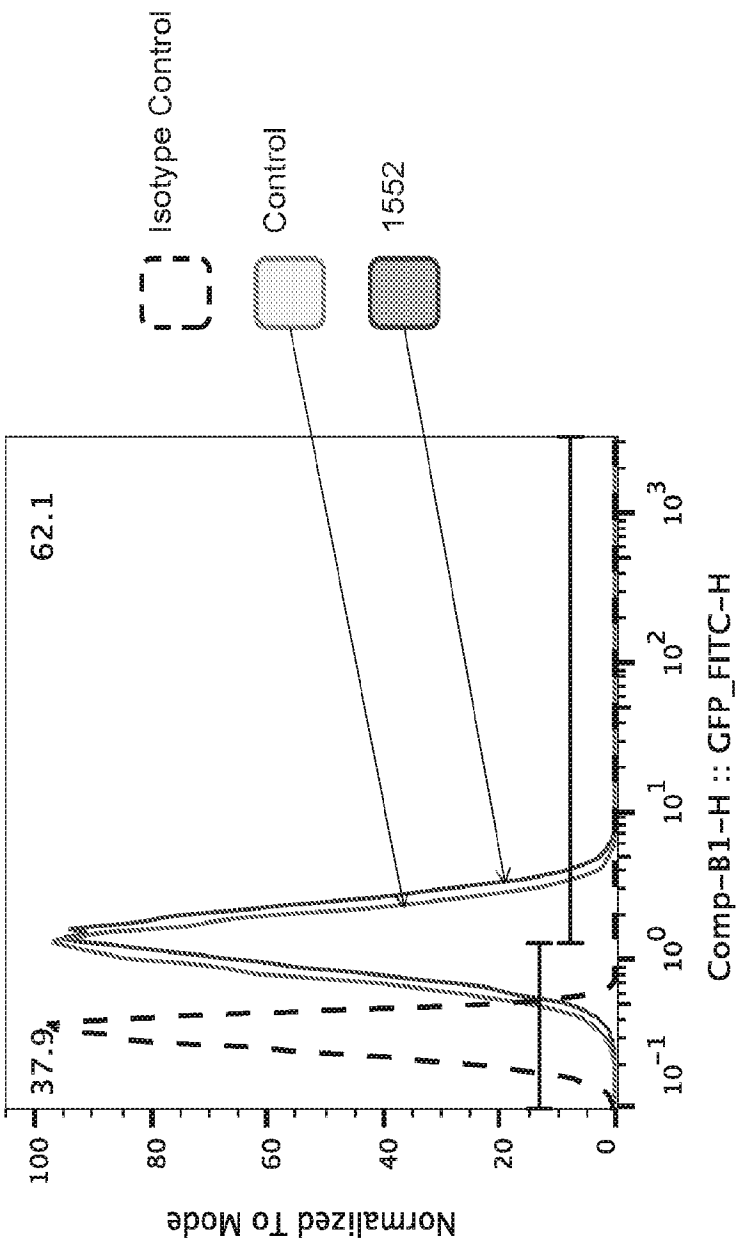

Isotype Control
Control
1552
62.1
37.9
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 33B

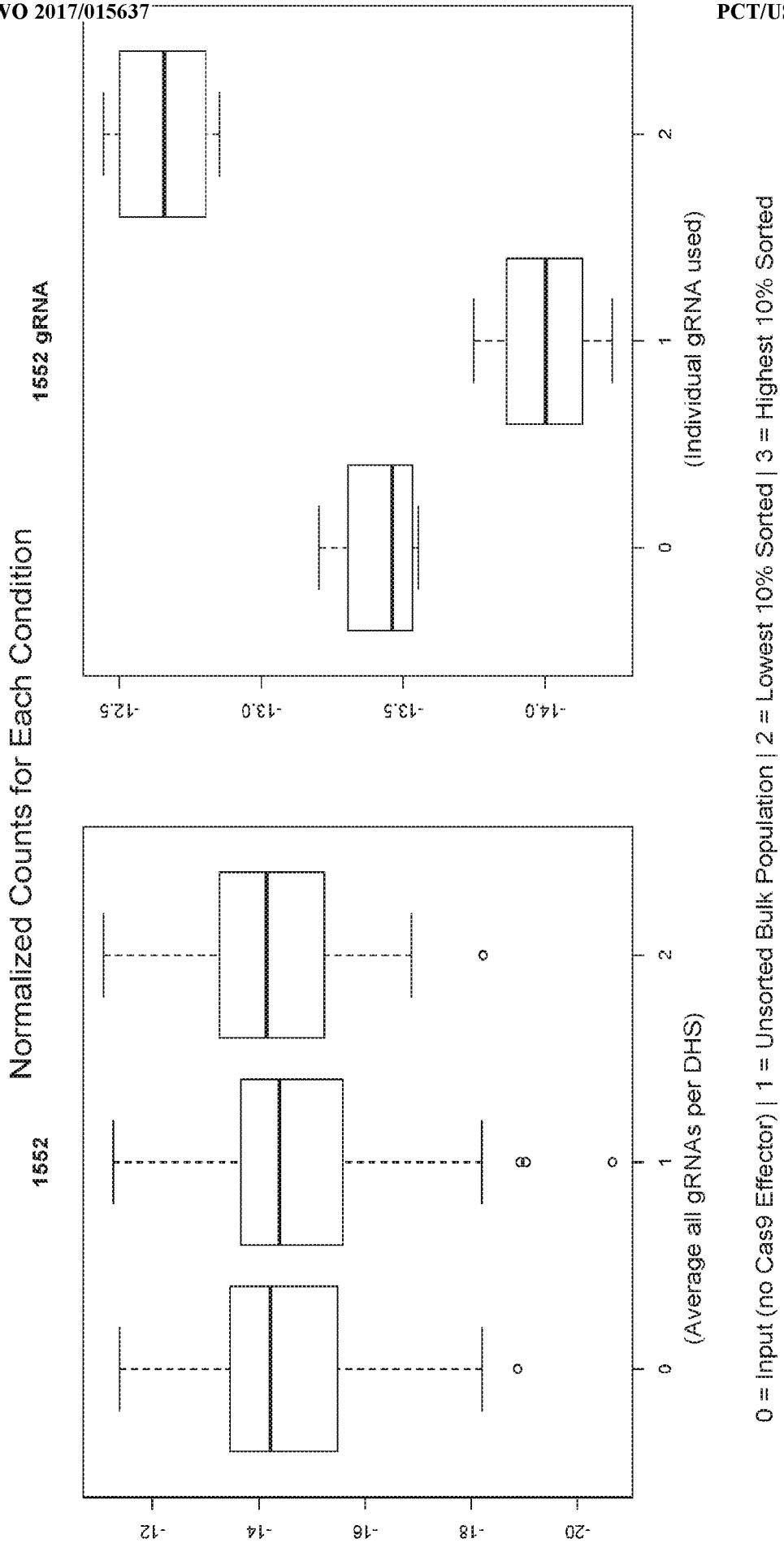

Normalized Counts for Each Condition
1552
1552 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 33C

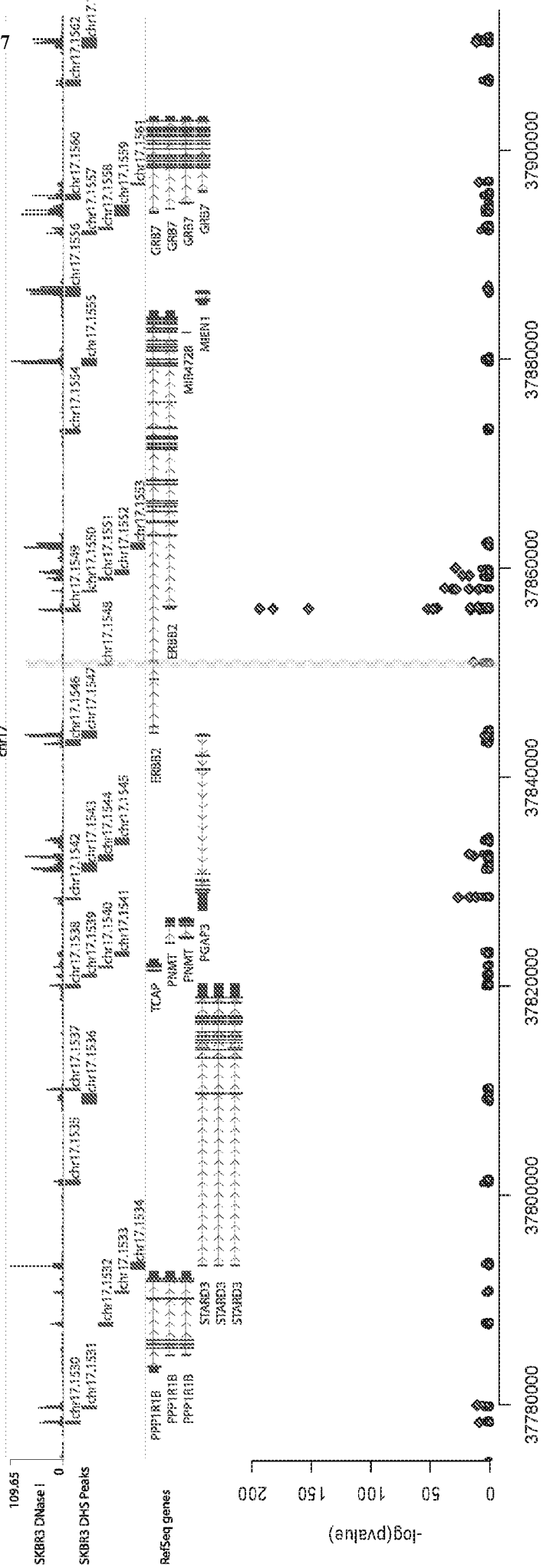

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
37780000
37790000
37800000
37810000
37820000
37830000
37840000
37850000
37860000
37870000
37880000
37890000
37900000
ERBB2
PNMT
PGAP3
TCAP
STARD3
PPP1R1B
MIR4728
MIEN1
GRB7

FIG. 34A

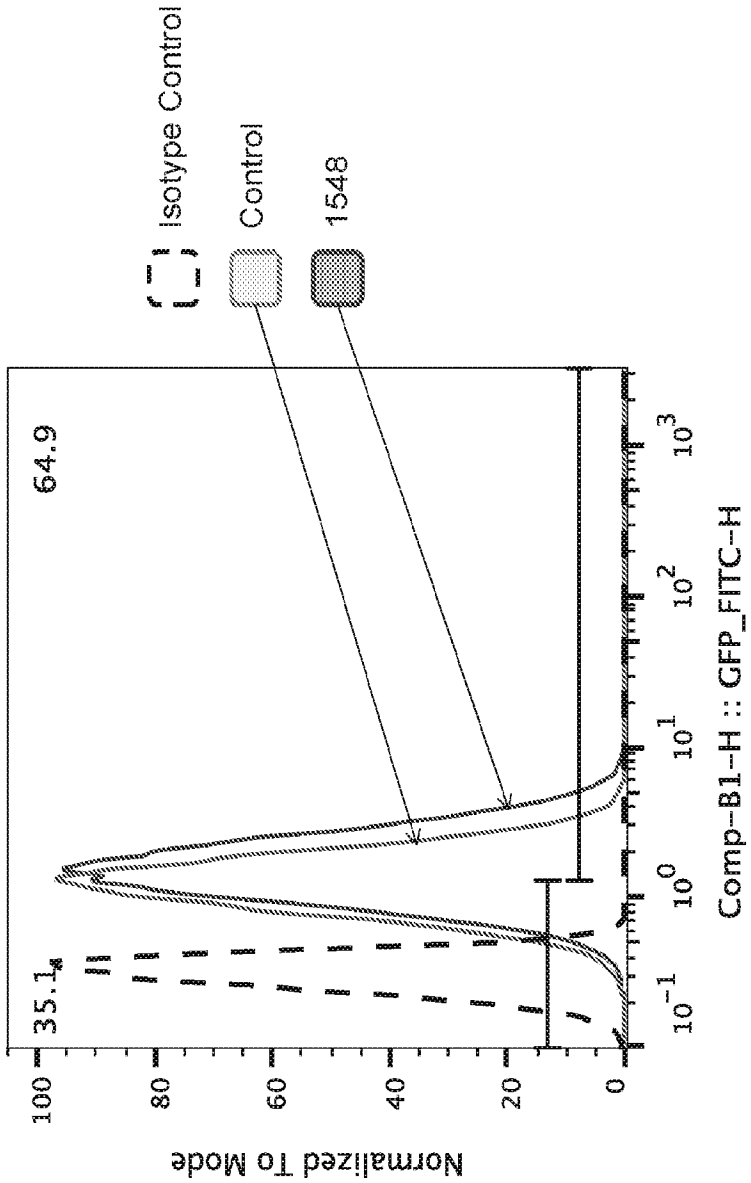

Isotype Control
Control
1548
64.9
35.1
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 34B

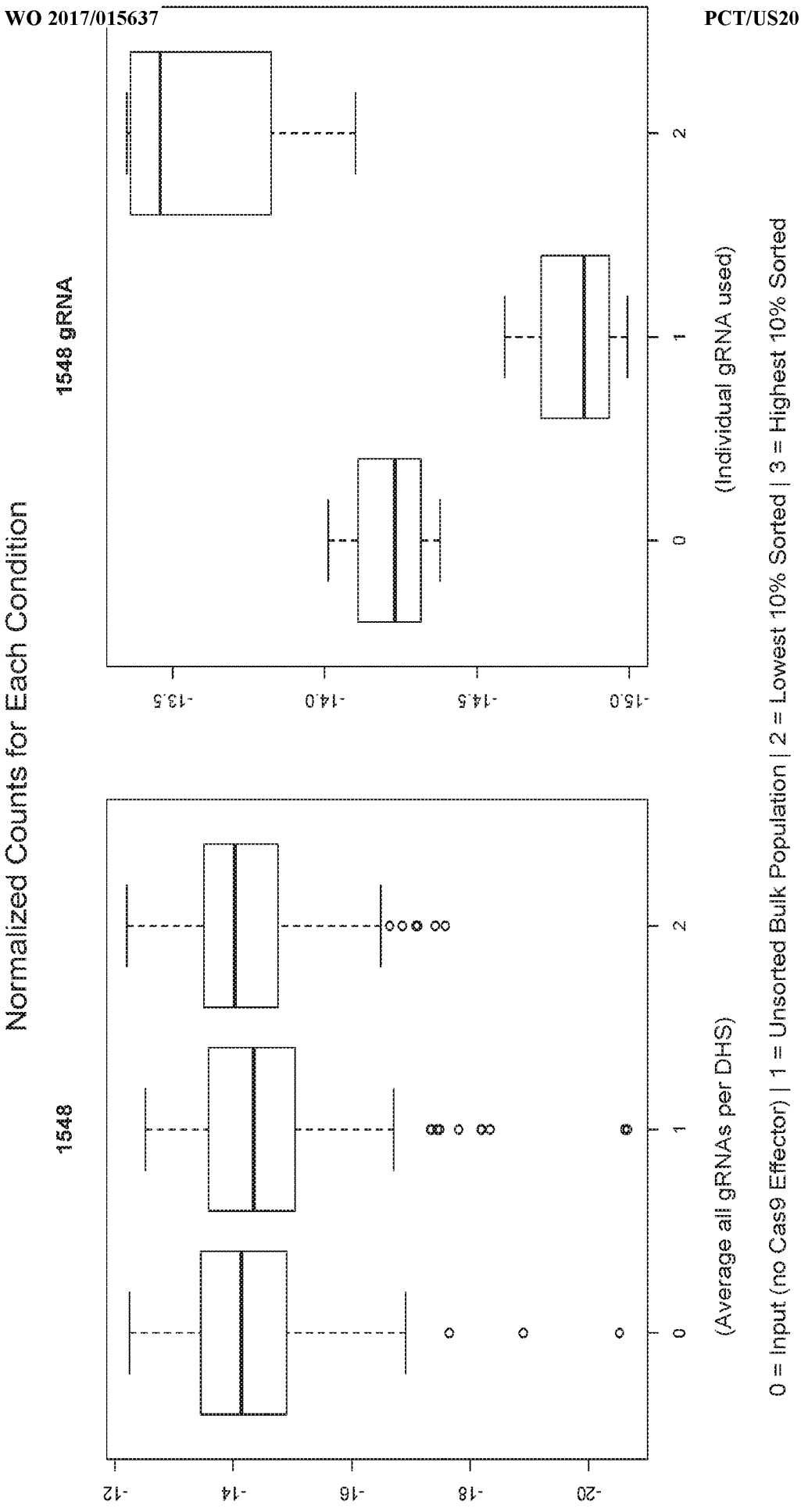

Normalized Counts for Each Condition
1548
-12
-14
-16
-18
-20
0
1
2
(Average all gRNAs per DHS)
1548 gRNA
-13.5
-14.0
-14.5
-15.0
0
1
2
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 34C

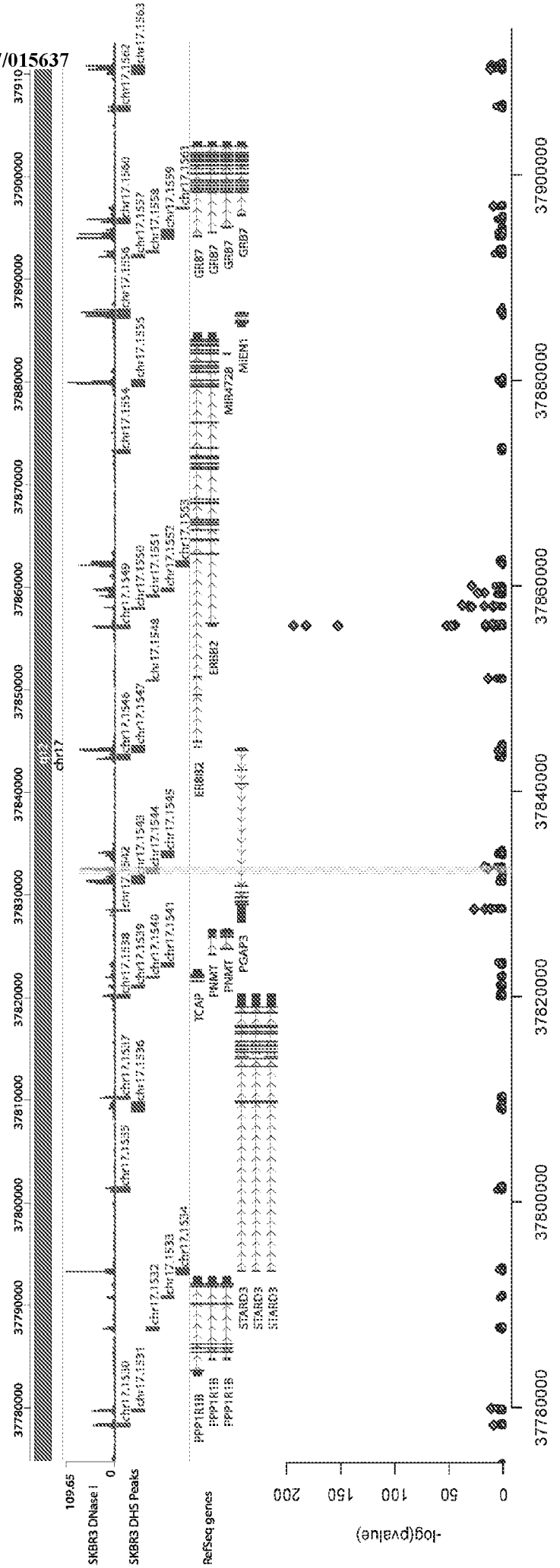

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
ERBB2
PGAP3
STARD3
PPP1R1B
MIEN1
GRB7

FIG. 35A

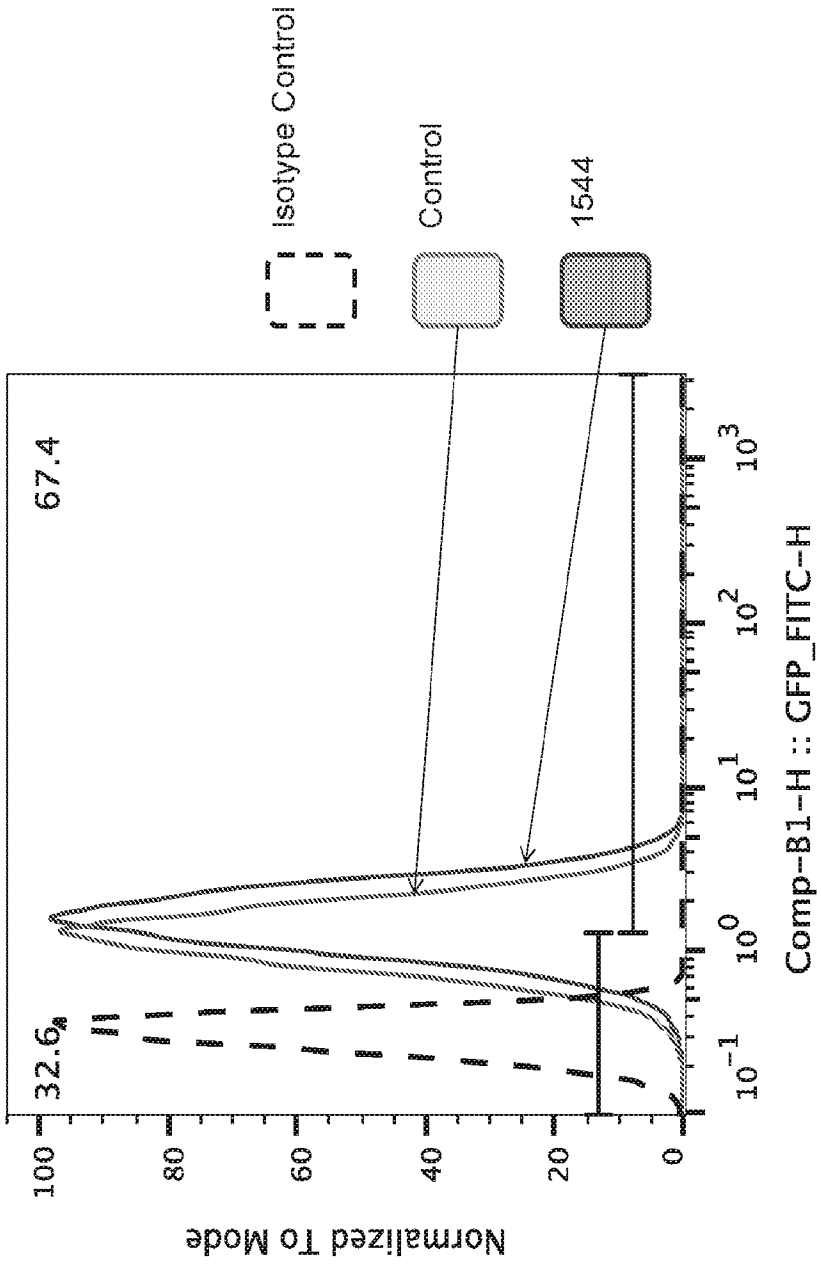

Isotype Control
Control
1544
67.4
32.6
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 35B

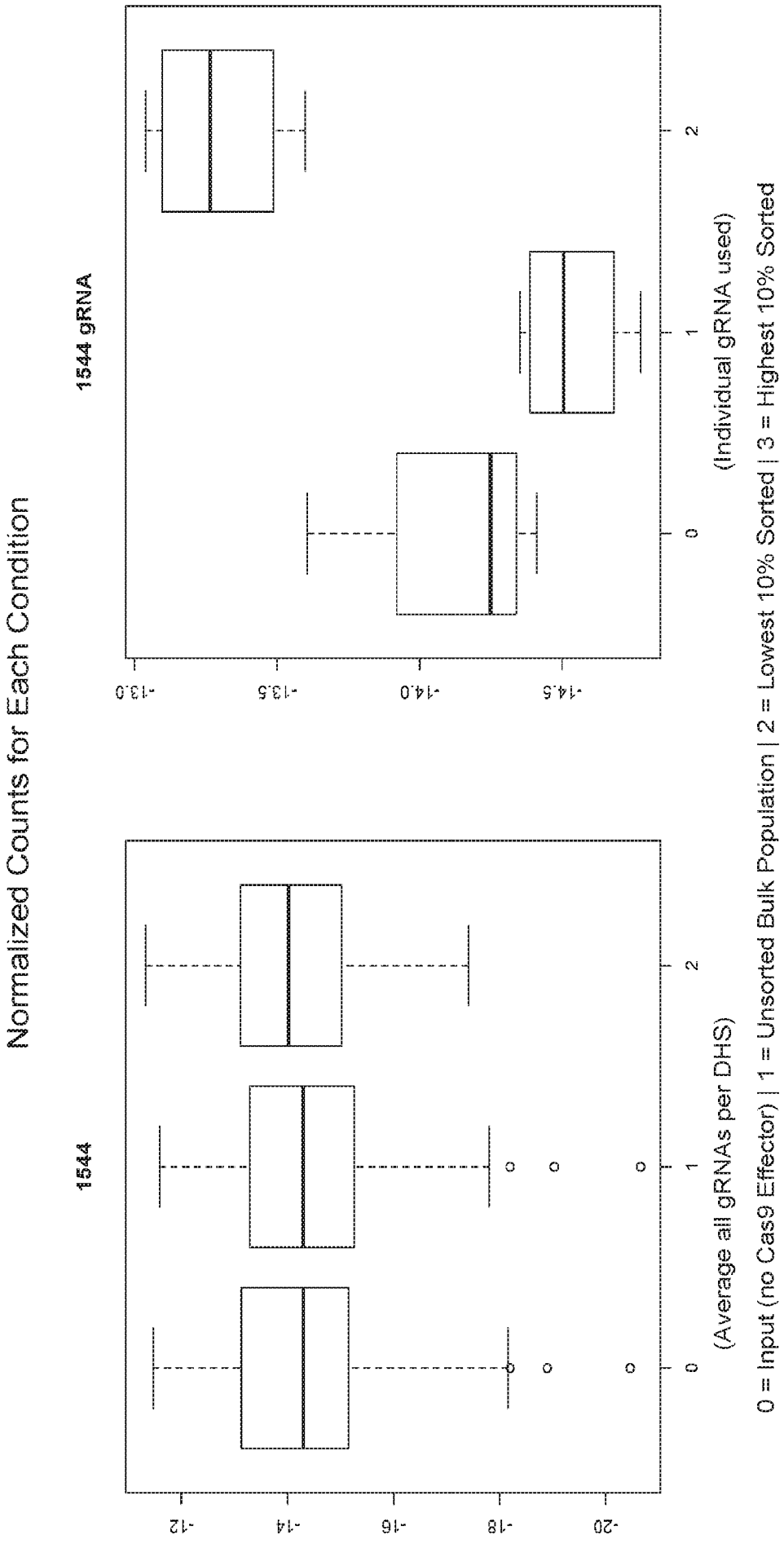

Normalized Counts for Each Condition
1544
1544 gRNA
-12
-14
-16
-18
-20
-13.0
-13.5
-14.0
-14.5
0
1
2
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 35C

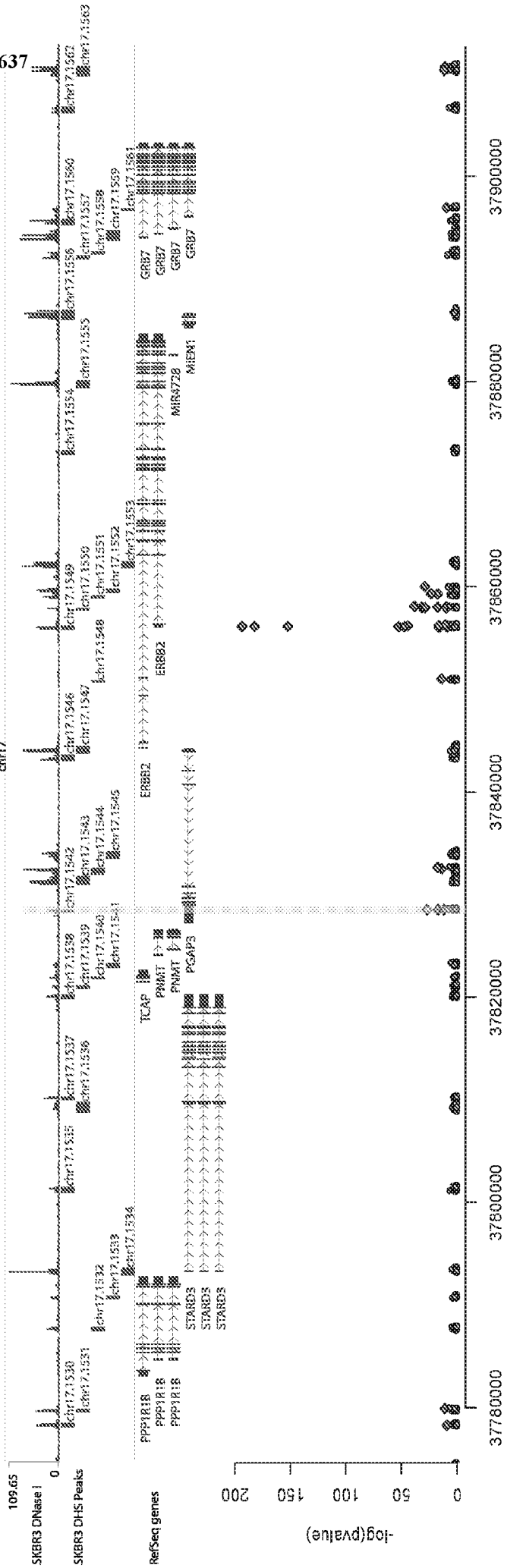

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
ERBB2
GRB7
MIEN1
MIR4728
PPP1R1B
STARD3
TCAP
PNMT
PGAP3

FIG. 36A

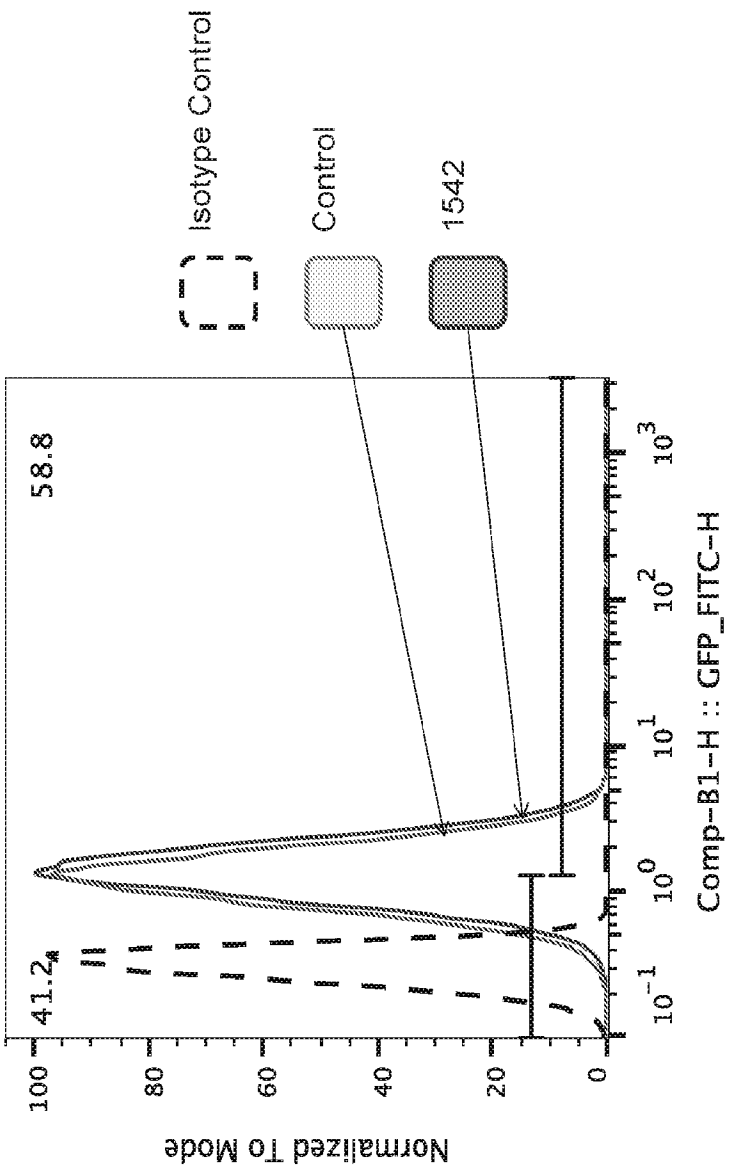

Isotype Control
Control
1542
58.8
41.2
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 36B

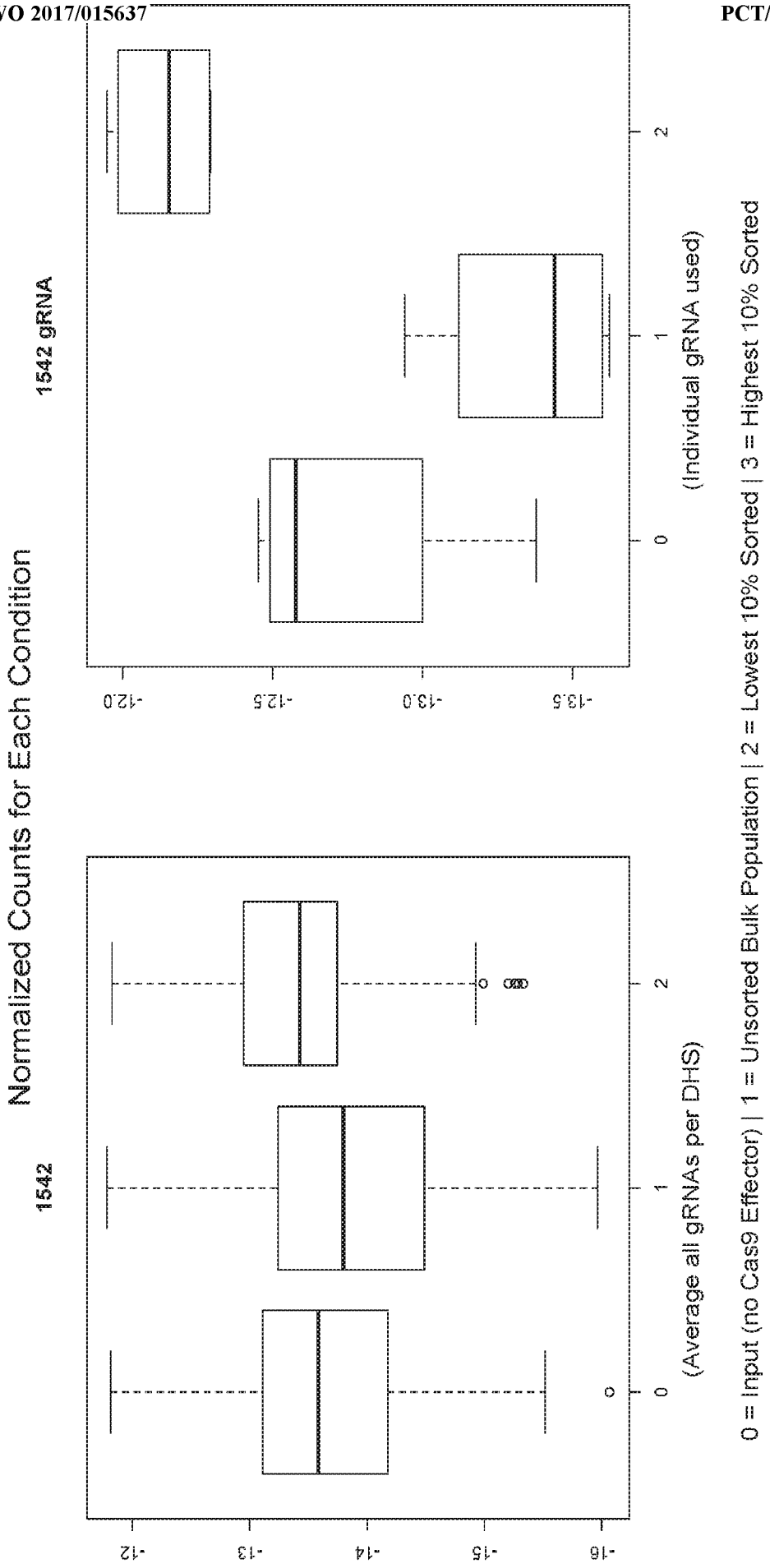

Normalized Counts for Each Condition
1542
1542 gRNA
-12
-13
-14
-15
-16
-12.0
-12.5
-13.0
-13.5
0
1
2
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 36C

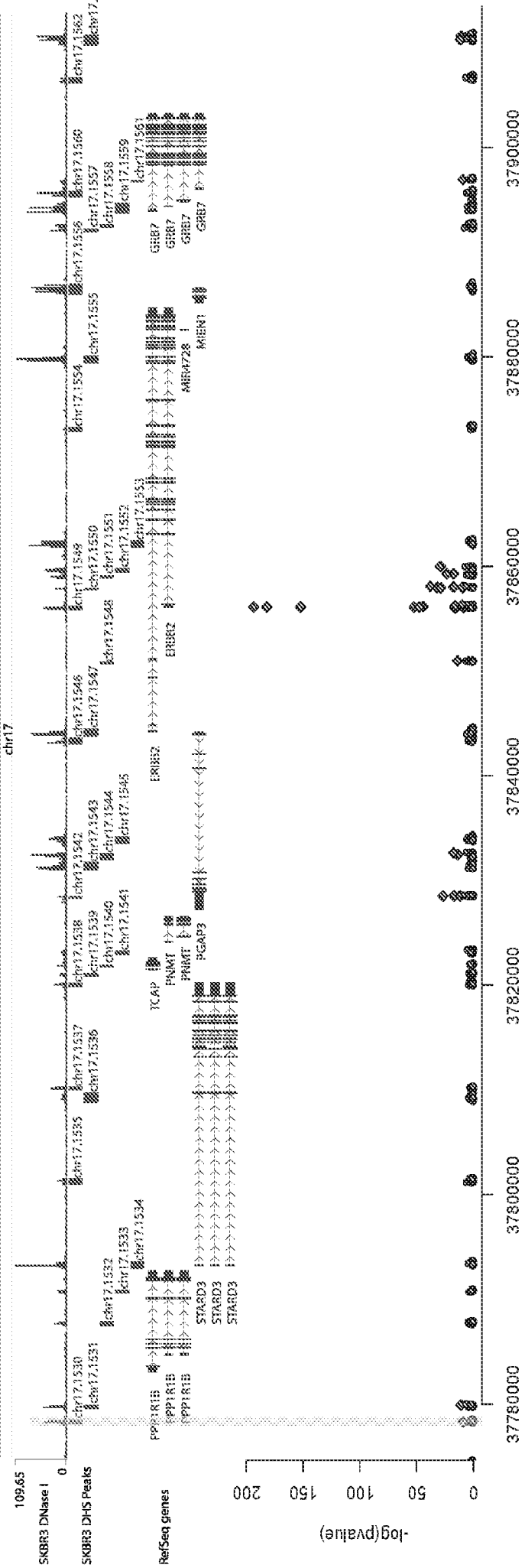

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
37780000
37790000
37800000
37810000
37820000
37830000
37840000
37850000
37860000
37870000
37880000
37890000
37900000
STARD3
PPP1R1B
TCAP
PNMT
PGAP3
ERBB2
GRB7
MIEN1
MIR4728

FIG. 37A

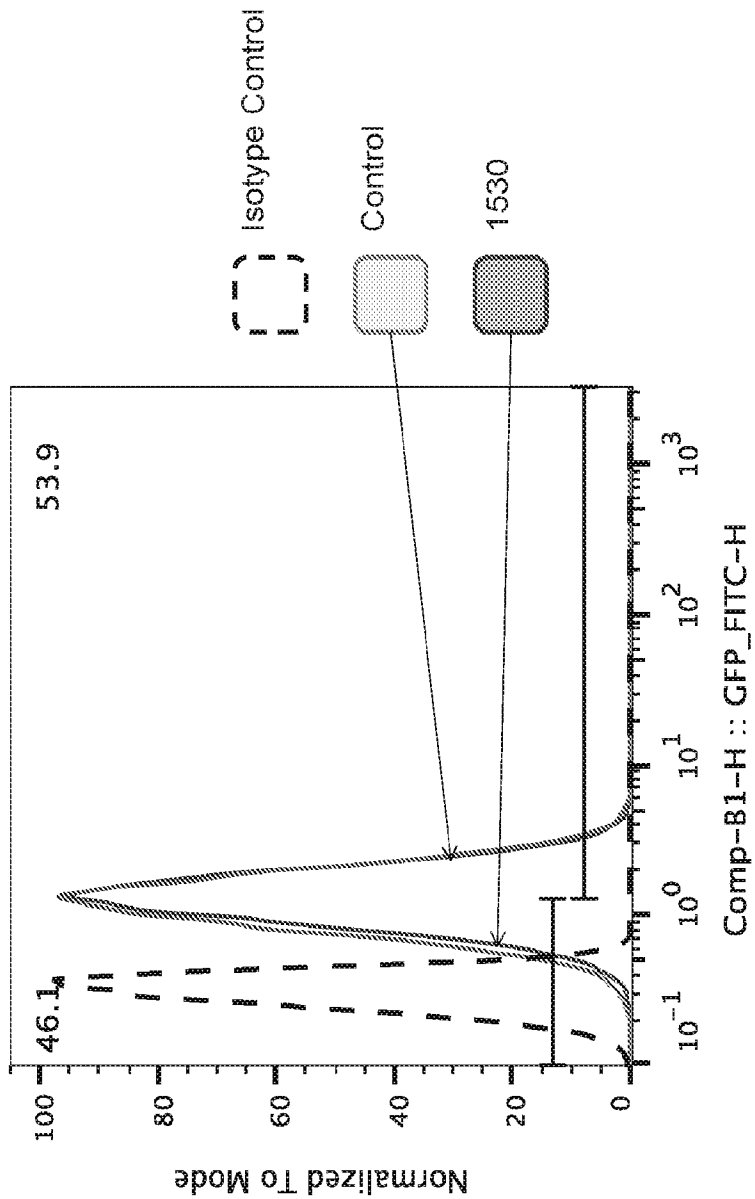

Isotype Control
Control
1530
53.9
46.1
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 37B

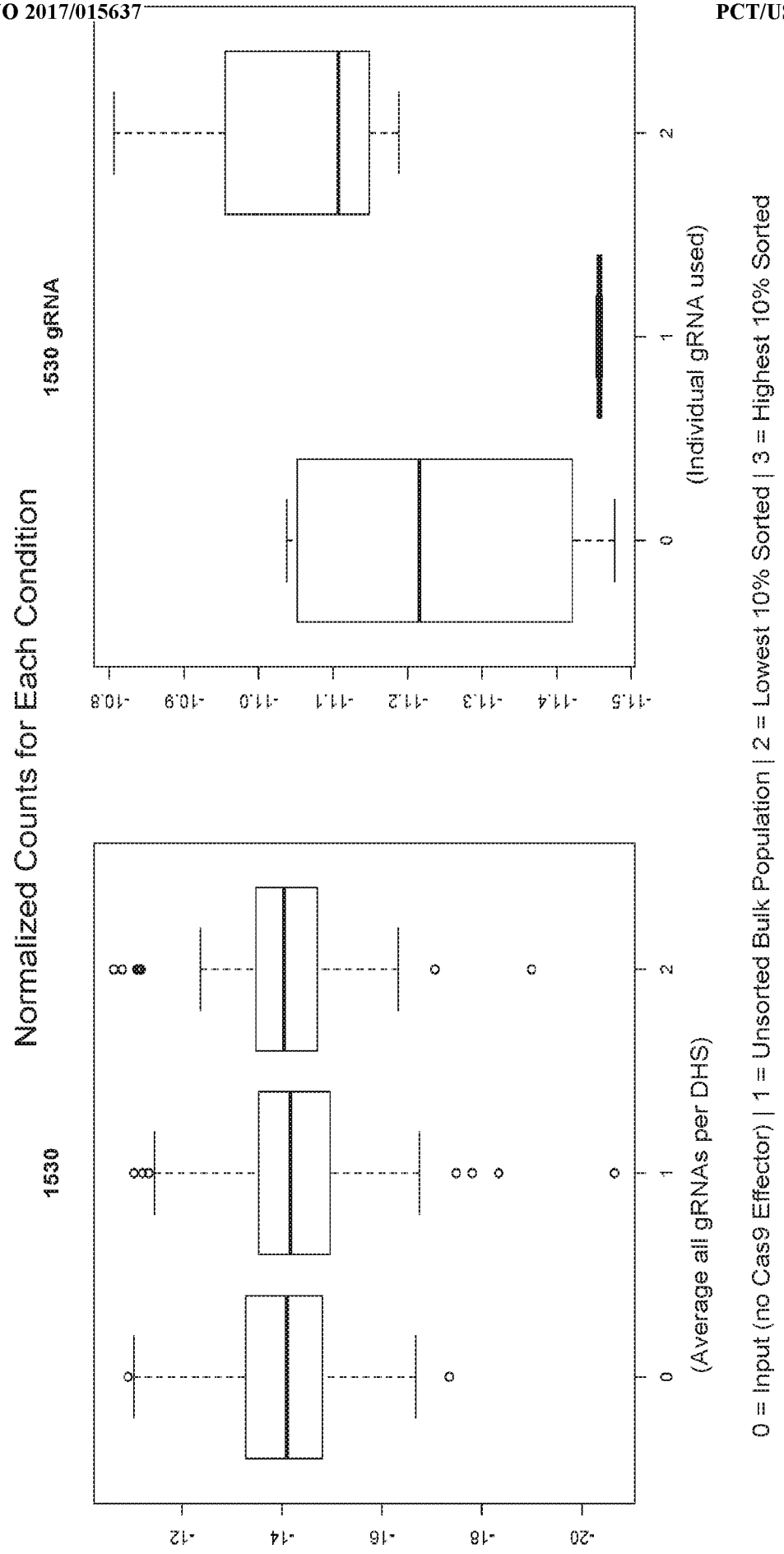

Normalized Counts for Each Condition
1530
1530 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 37C

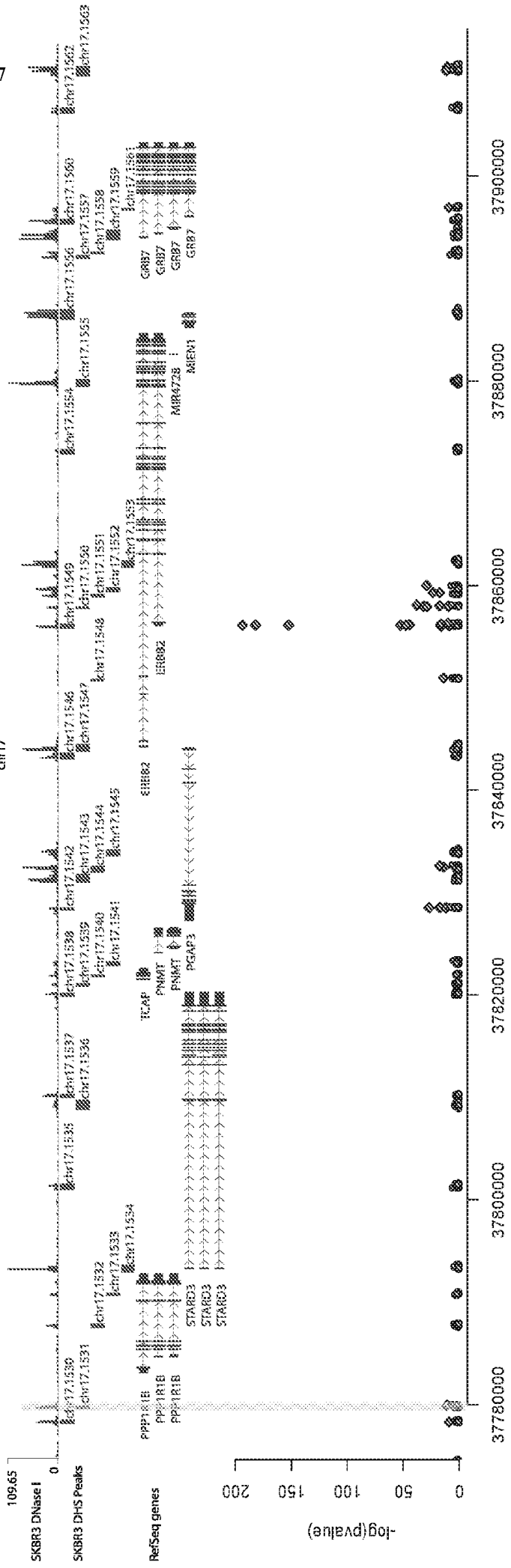

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
37780000
37790000
37800000
37810000
37820000
37830000
37840000
37850000
37860000
37870000
37880000
37890000
37900000
STARD3
PPP1R1B
TCAP
PNMT
PGAP3
ERBB2
GRB7
MIR4728
MIEN1

FIG. 38A

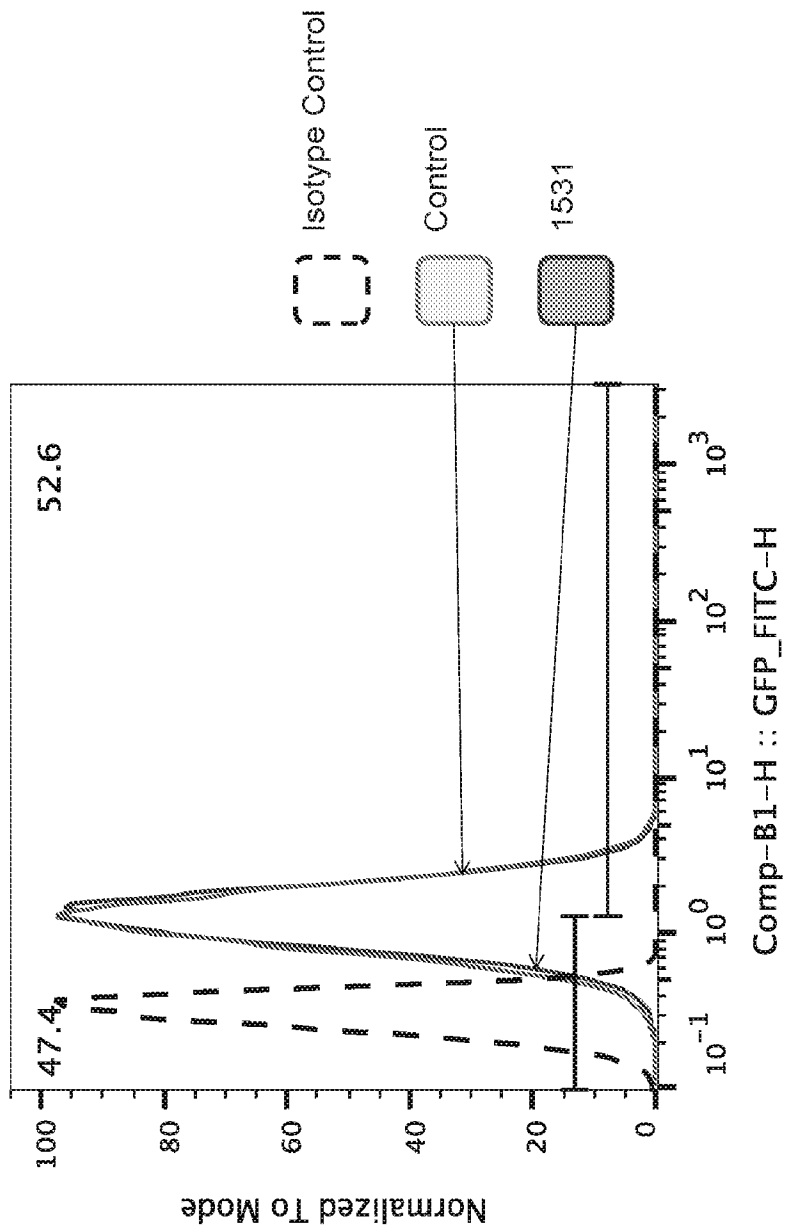

Isotype Control
Control
1531
52.6
47.4
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 38B

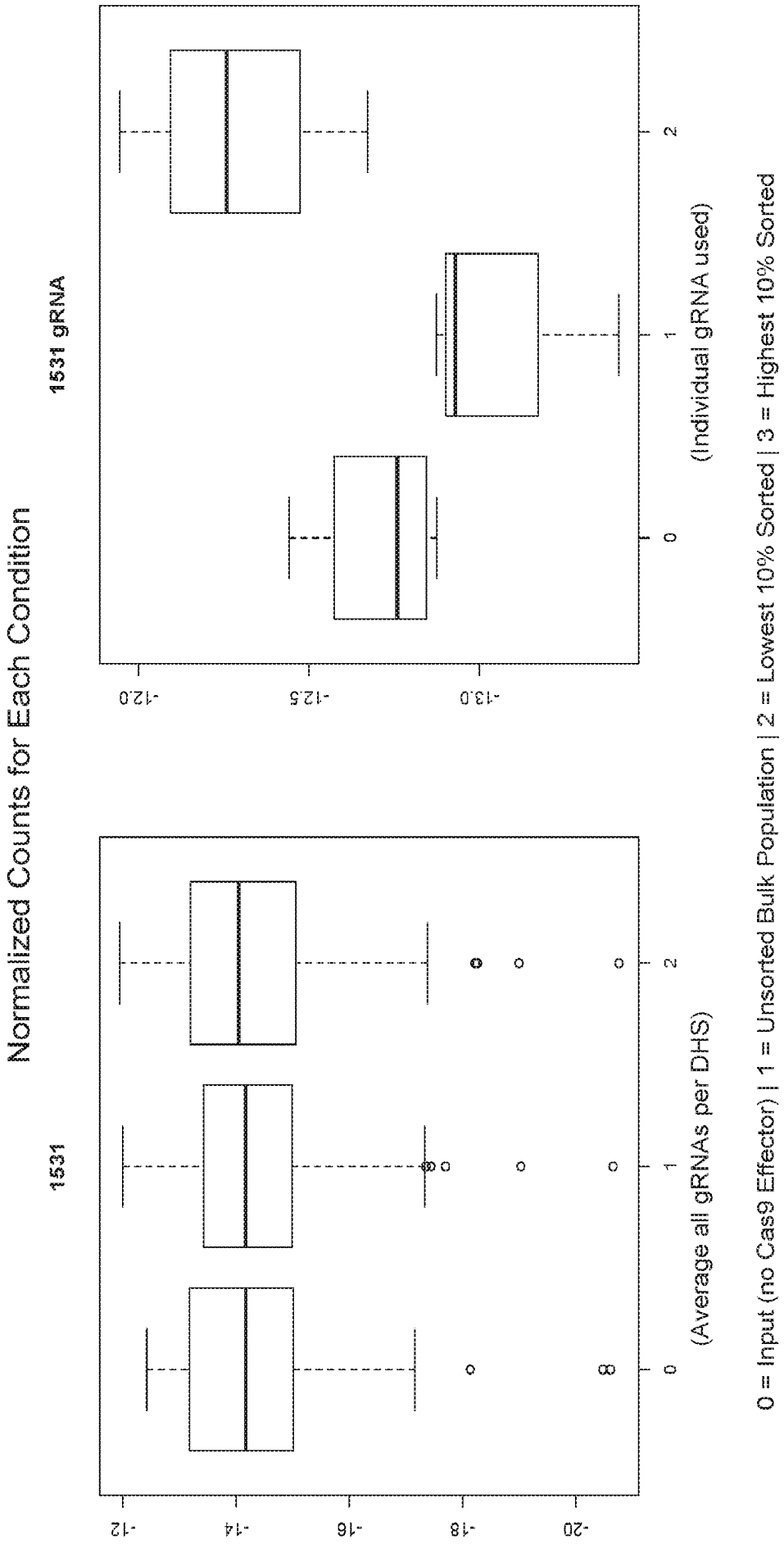

Normalized Counts for Each Condition
1531
1531 gRNA
-12
-14
-16
-18
-20
-12.0
-12.5
-13.0
0
1
2
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 38C

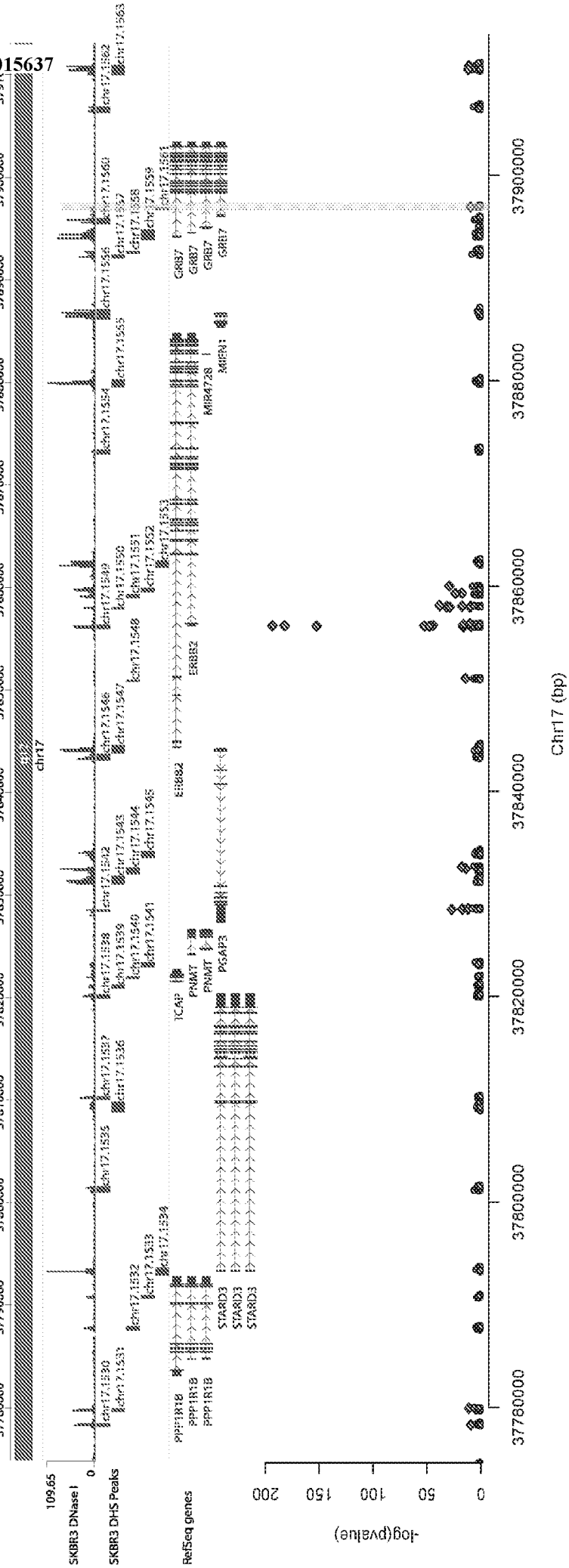

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
ERBB2
PNMT
TCAP
PGAP3
STARD3
PPP1R1B
GRB7
MIR4728
MIEN1

FIG. 39A

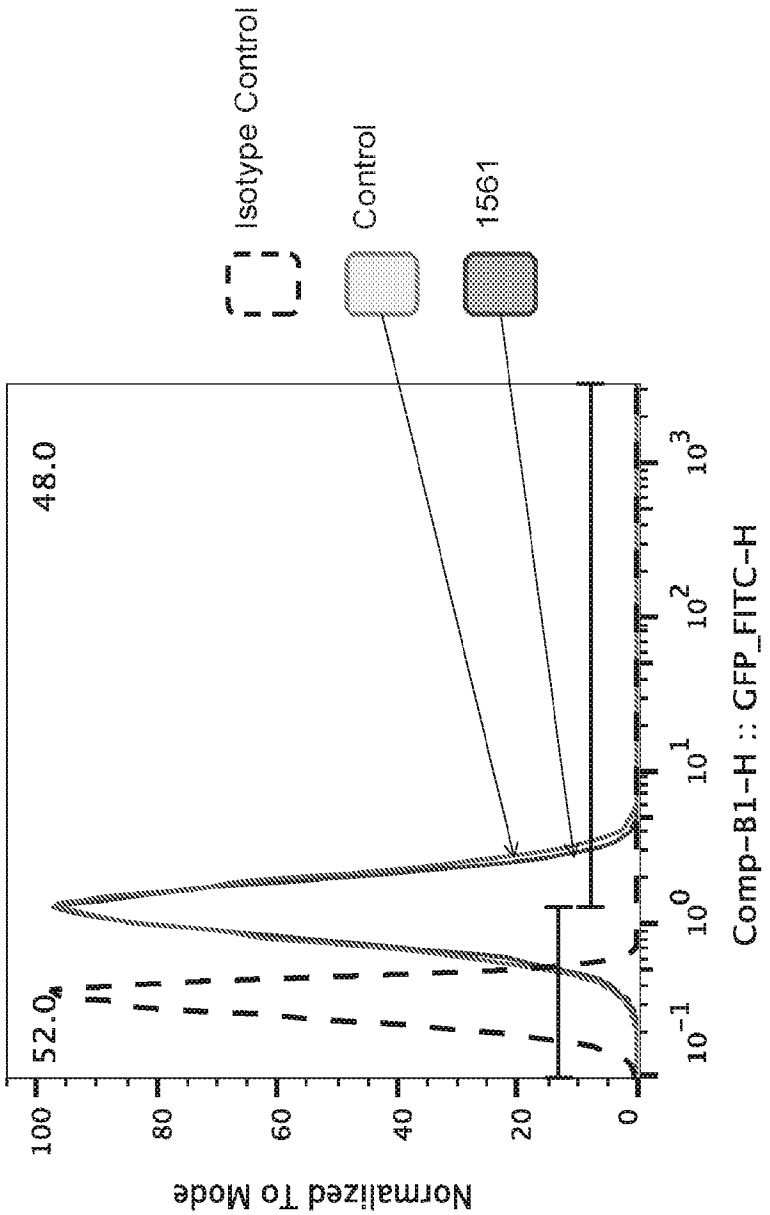

Isotype Control
Control
1561
48.0
52.0
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 39B

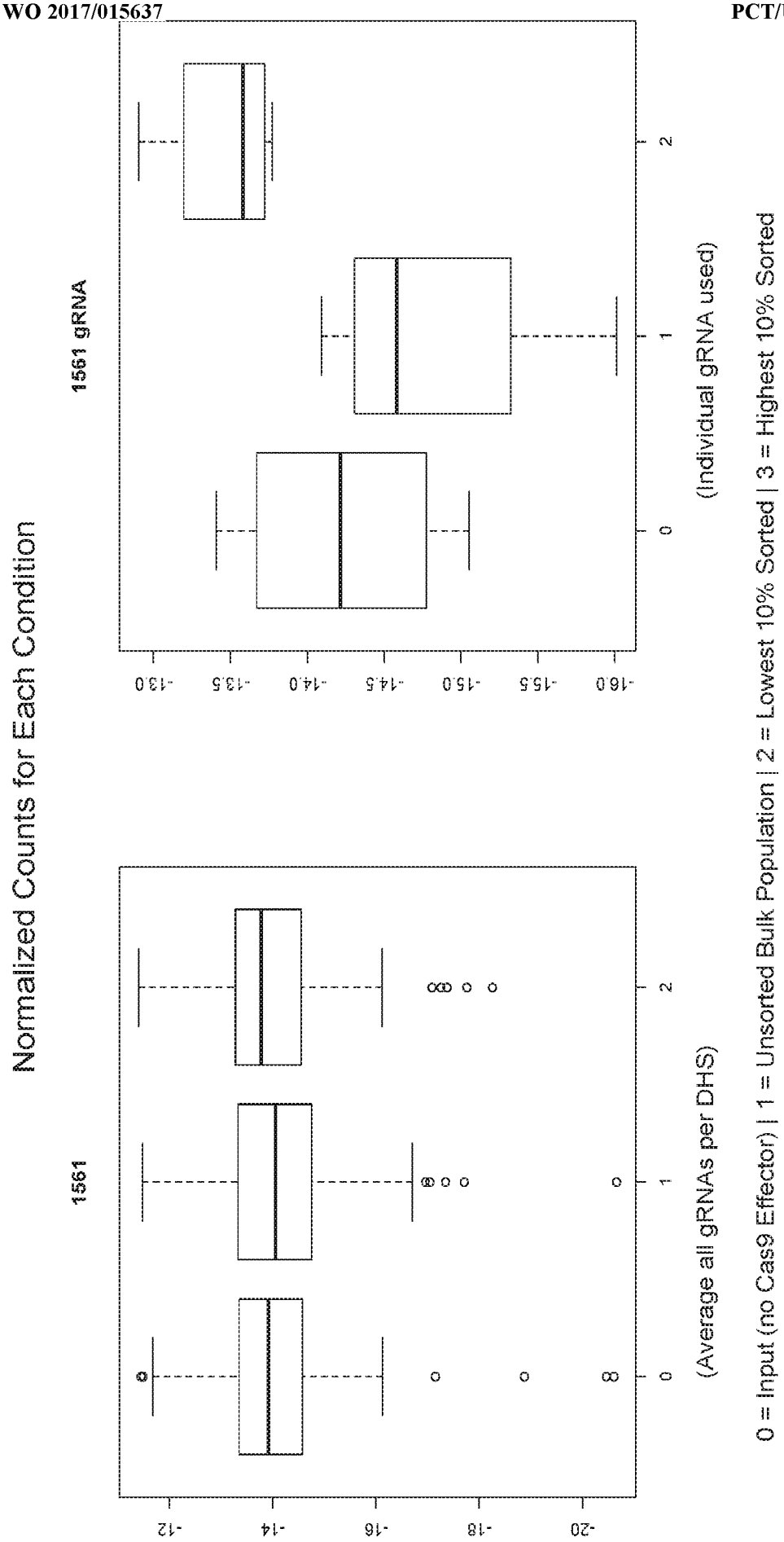

Normalized Counts for Each Condition
1561
1561 gRNA
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 39C

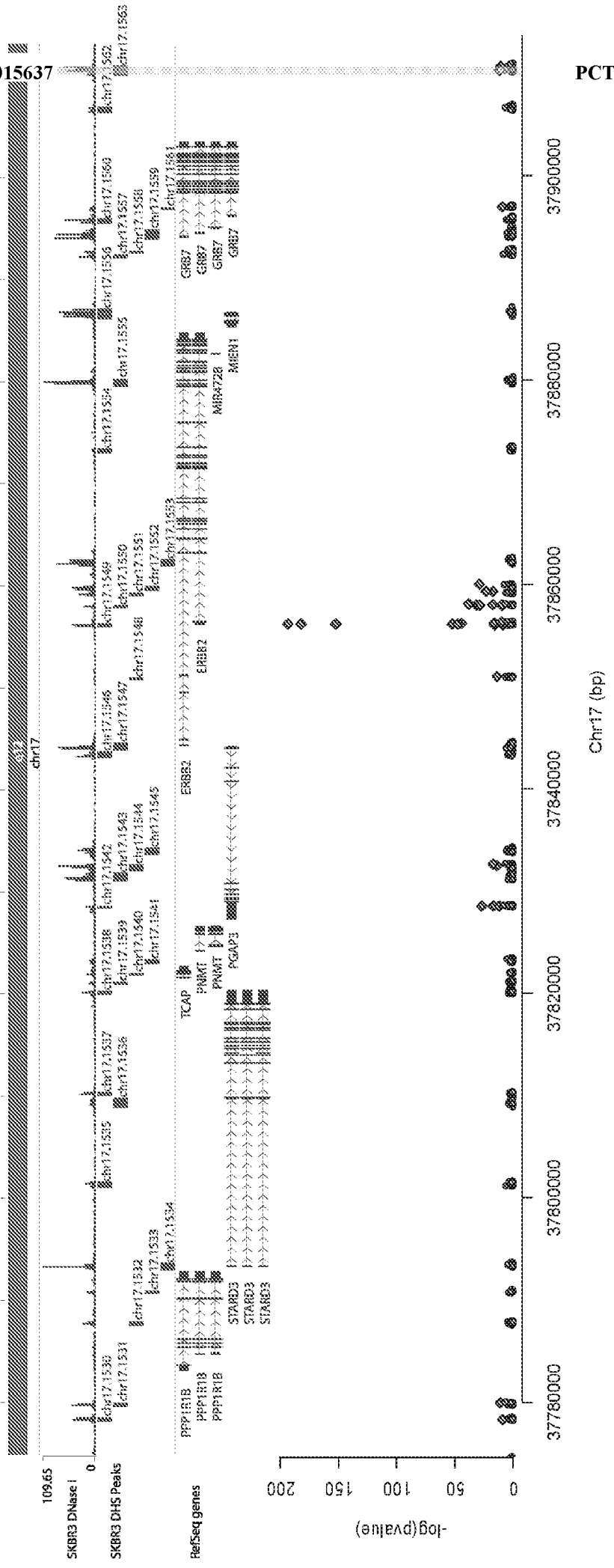

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
ERBB2
TCAP
PNMT
PGAP3
STARD3
PPP1R1B
MIEN1
GRB7
37780000
37800000
37820000
37840000
37860000
37880000
37900000

FIG. 40A

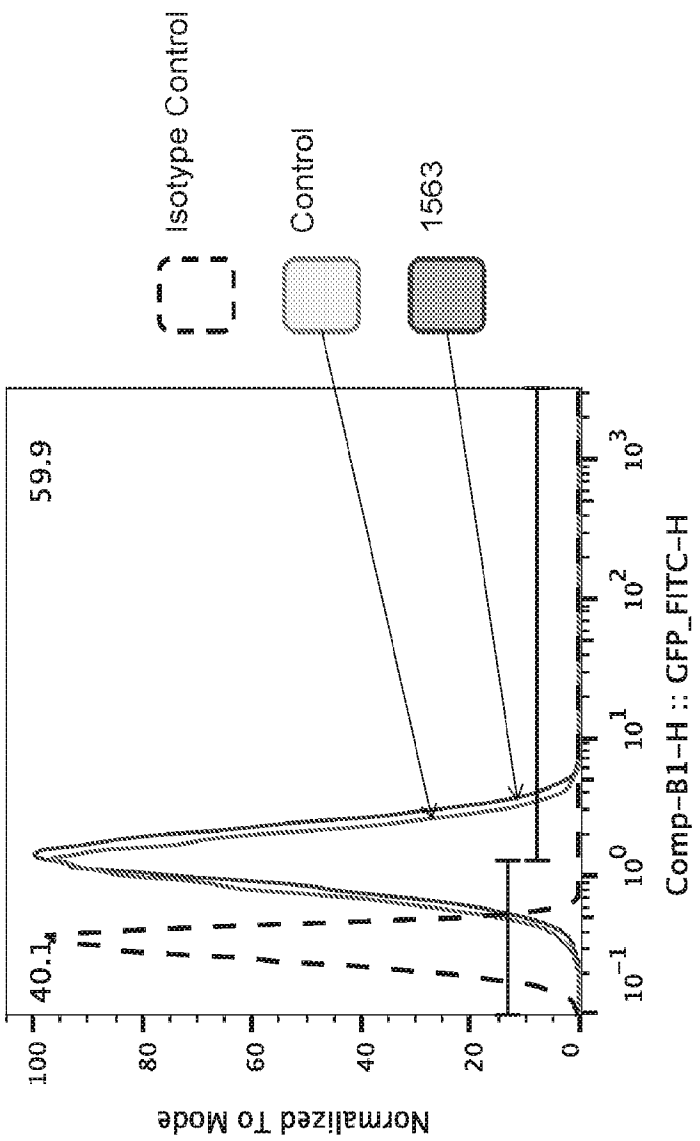

Isotype Control
Control
1563
59.9
40.1
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Comp-B1-H :: GFP_FITC-H

FIG. 40B

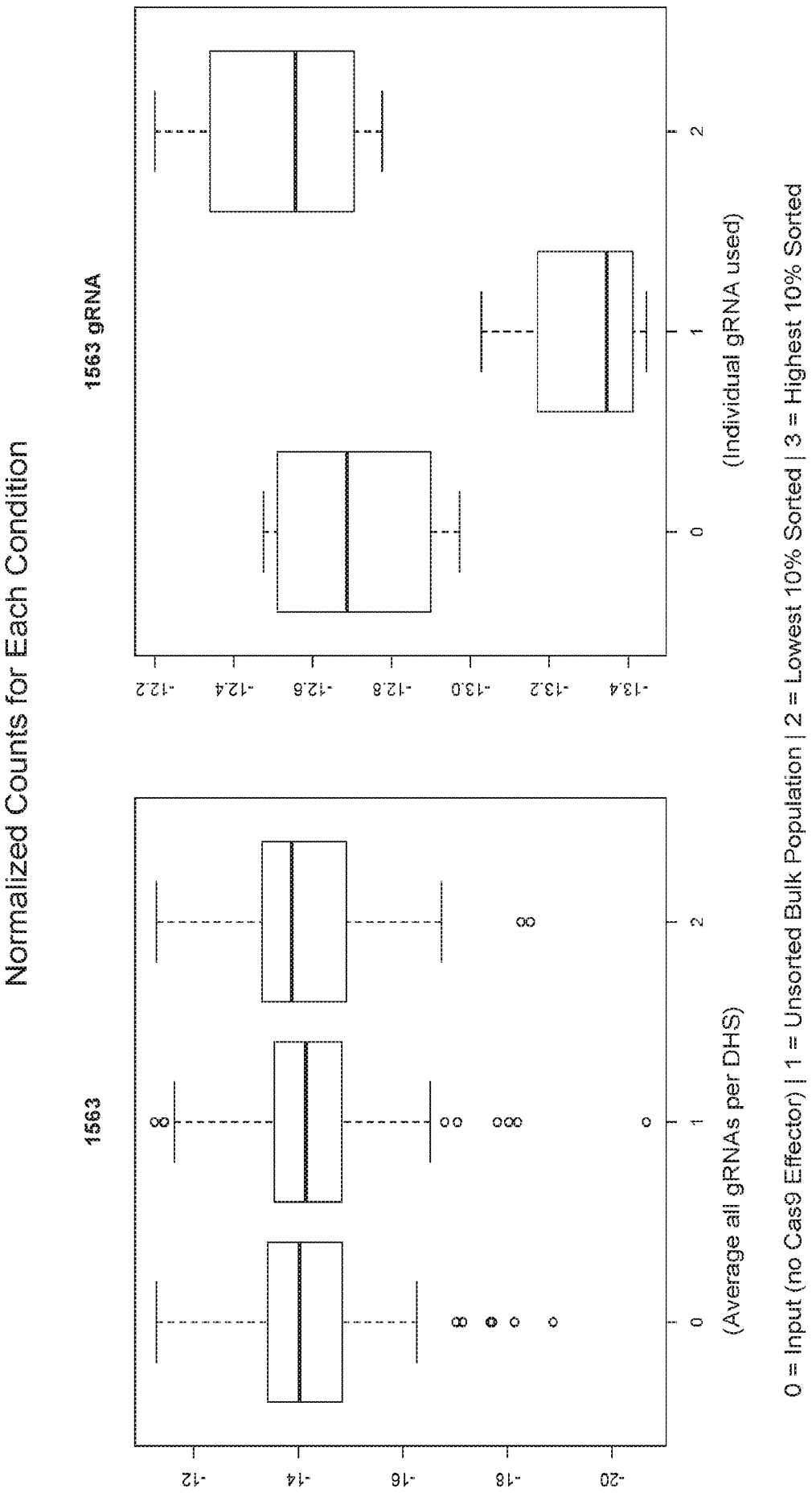

Normalized Counts for Each Condition
1563
1563 gRNA
-12
-14
-16
-18
-20
-12.2
-12.4
-12.6
-12.8
-13.0
-13.2
-13.4
0
1
2
(Average all gRNAs per DHS)
(Individual gRNA used)
0 = Input (no Cas9 Effector) | 1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 40C

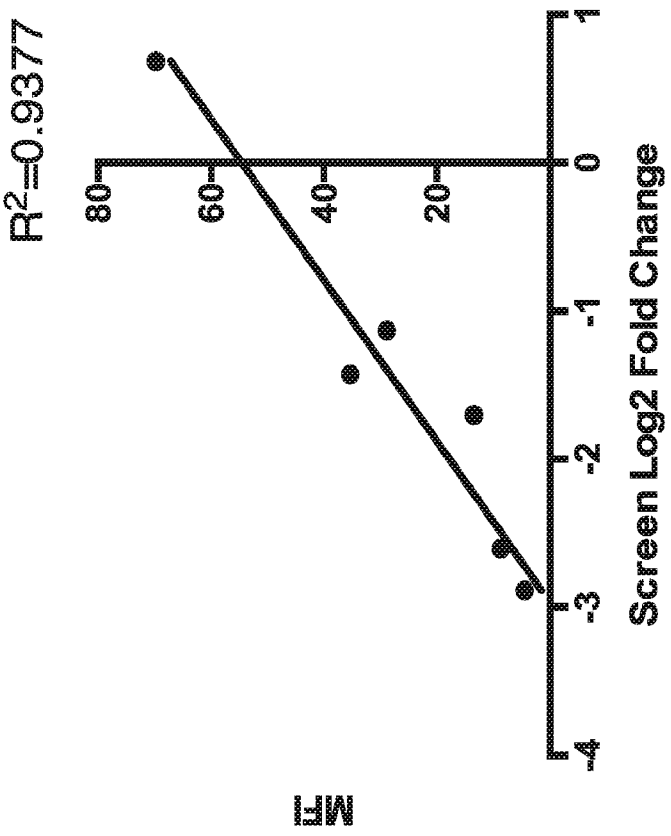


FIG. 41B

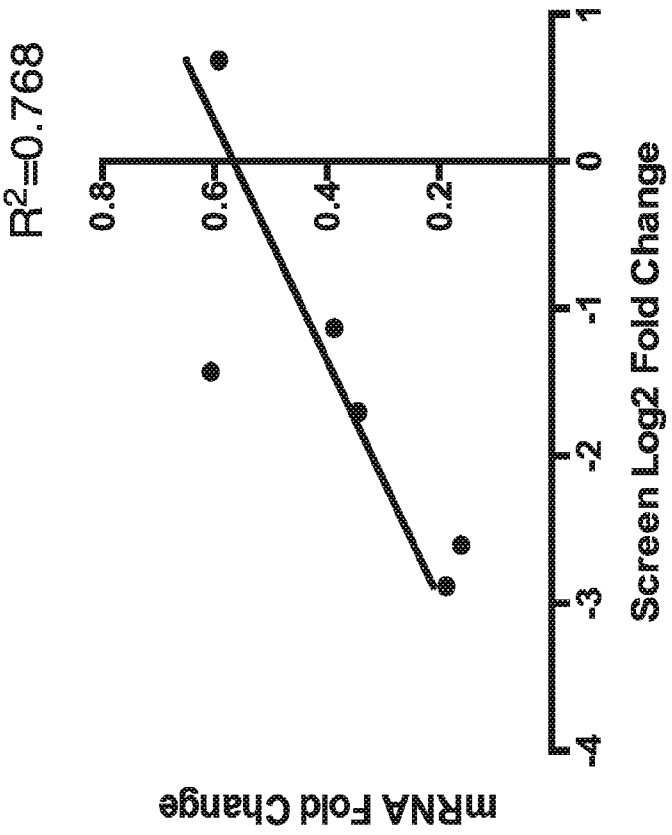


FIG. 41A

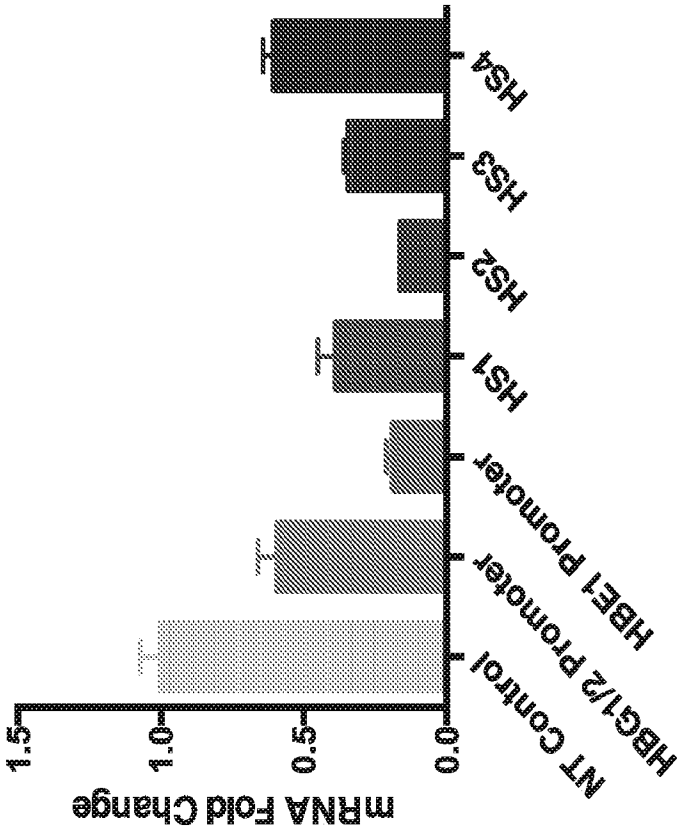


FIG. 42B

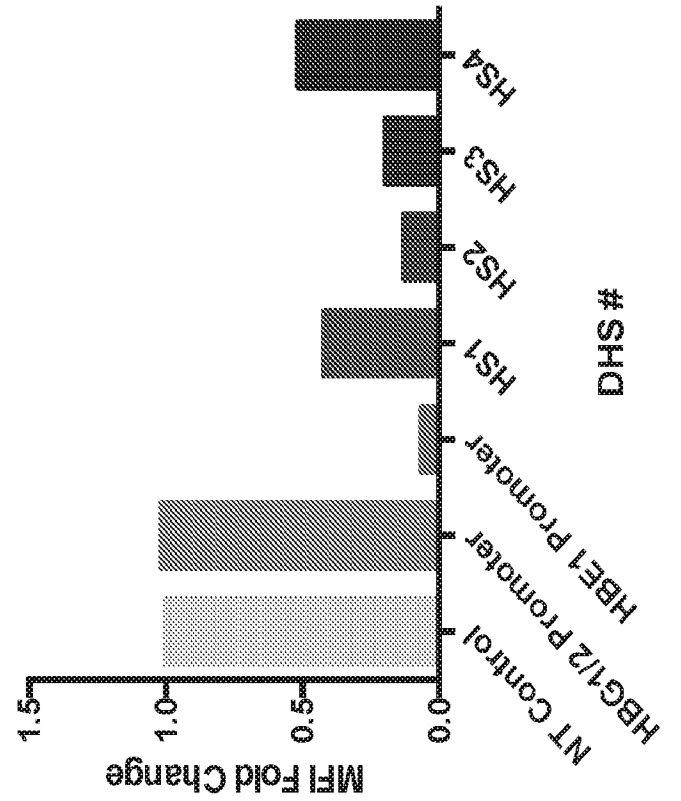


FIG. 42A

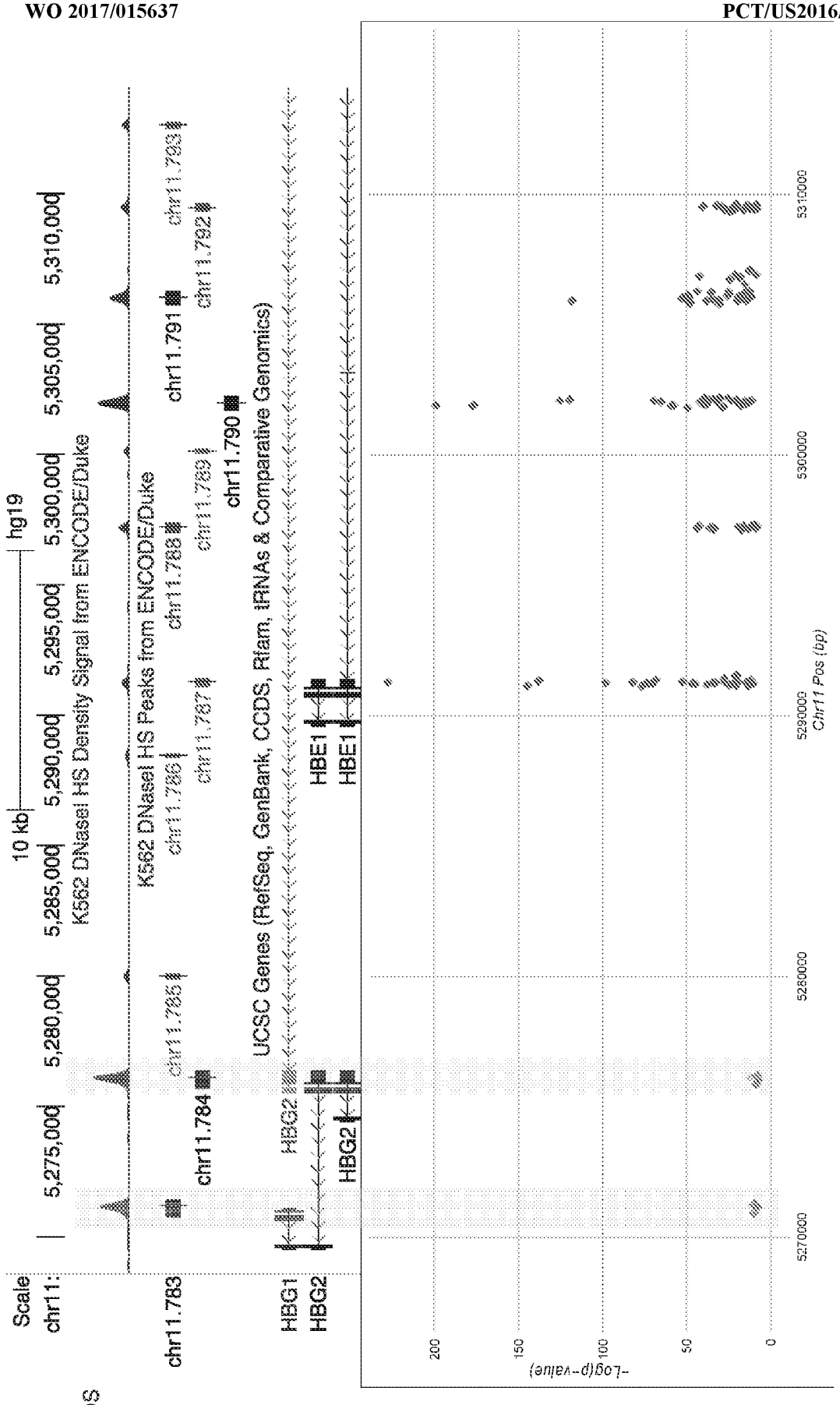

Scale
chr11:
5,275,000
5,280,000
5,285,000
10 kb
5,290,000
5,295,000
5,300,000
5,305,000
5,310,000
hg19
K562 DNaseI HS Density Signal from ENCODE/Duke
K562 DS
K562 DNaseI HS Peaks from ENCODE/Duke
chr11.783
chr11.785
chr11.786
chr11.788
chr11.791
chr11.793
chr11.784
chr11.787
chr11.789
chr11.792
chr11.790
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
HBG1
HBG2
HBE1
-Log(p-value)
Chr11 Pos (bp)

FIG. 43A

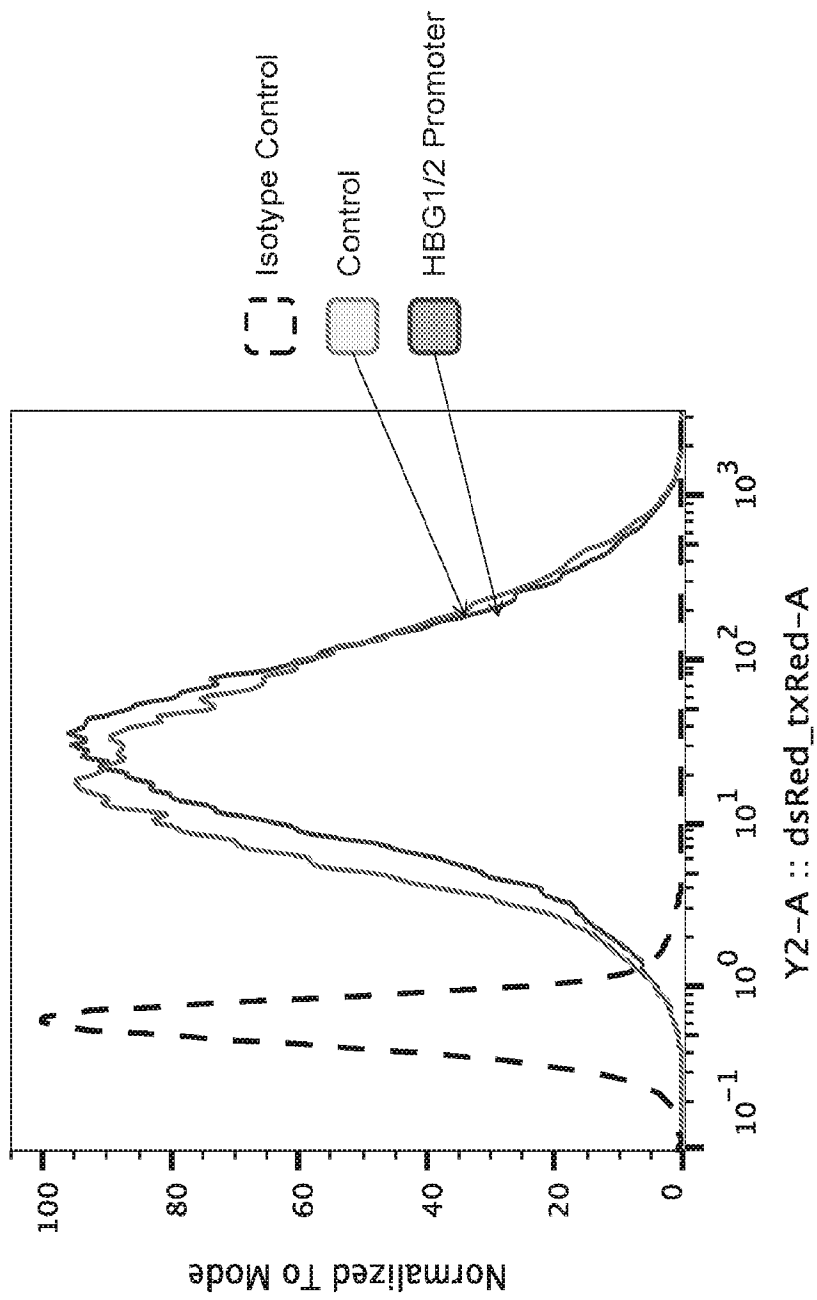

Isotype Control
Control
HBG1/2 Promoter
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Y2-A :: dsRed_tXRed-A

FIG. 43B

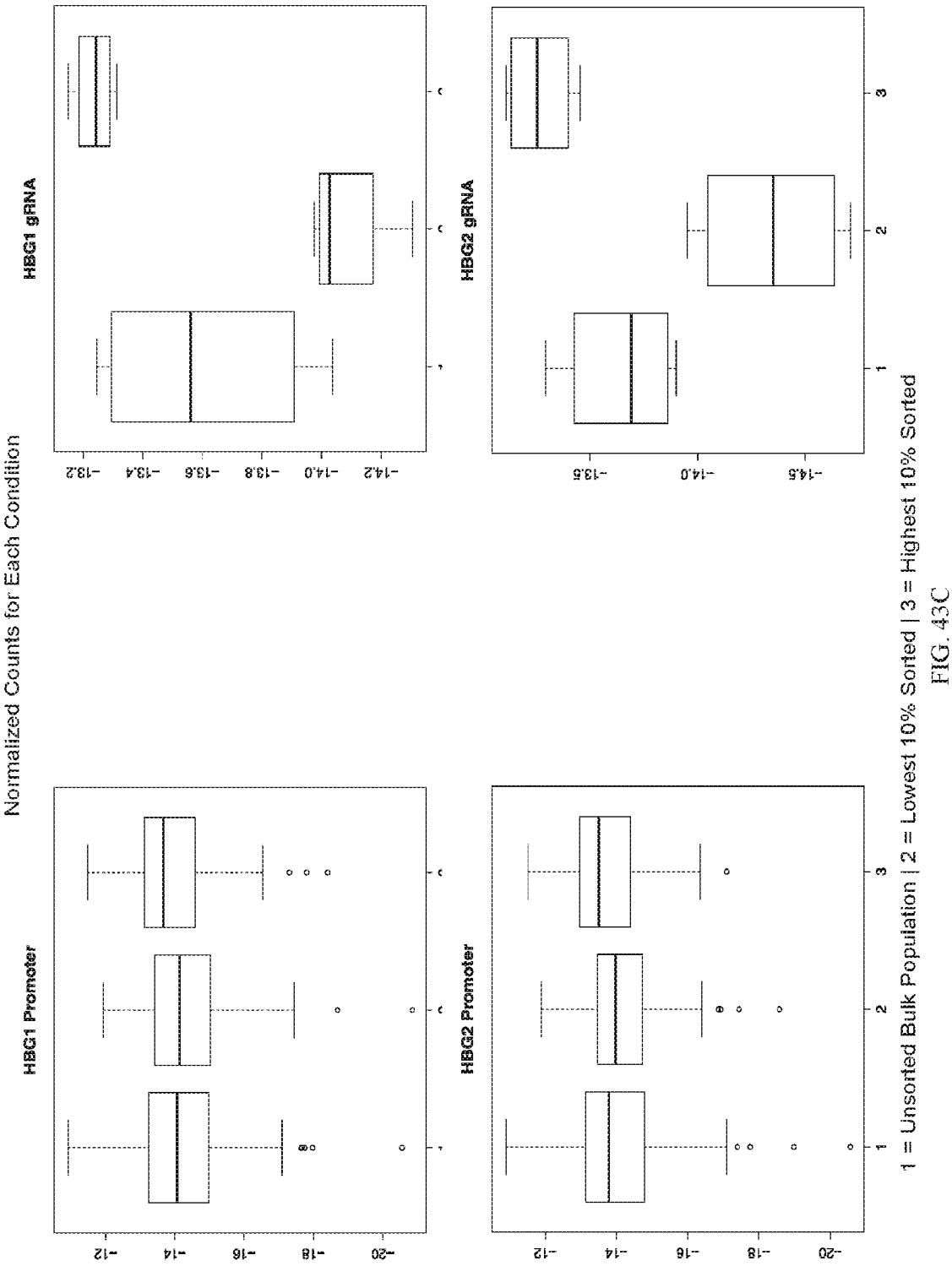

HBG1 Promoter
HBG1 gRNA
HBG2 Promoter
HBG2 gRNA
Normalized Counts for Each Condition
1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 43C

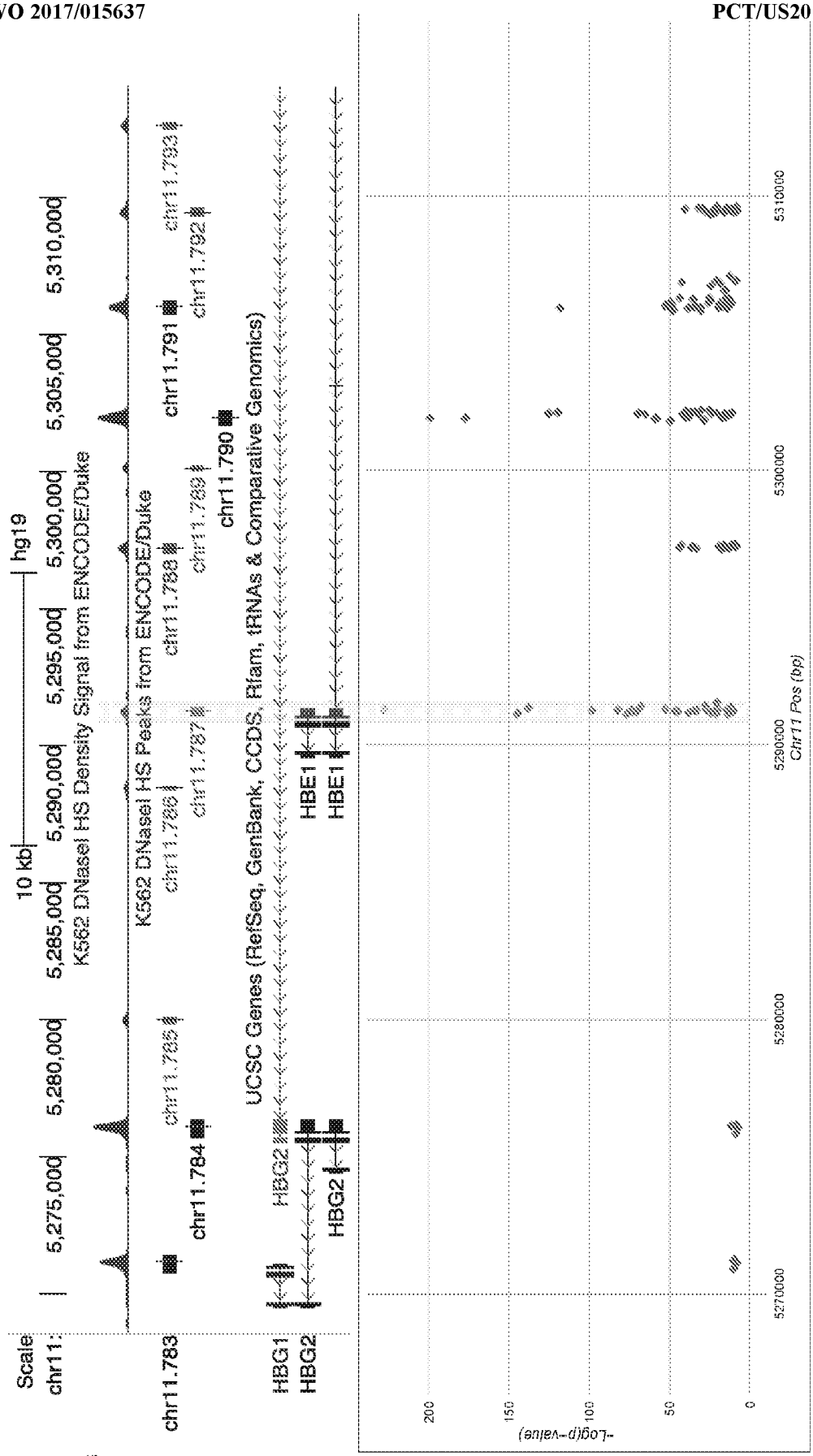

K562 DS
Scale
chr11:
5,275,000
5,280,000
5,285,000
10 kb
5,290,000
5,295,000
5,300,000
5,305,000
5,310,000
hg19
K562 DNaseI HS Density Signal from ENCODE/Duke
K562 DNaseI HS Peaks from ENCODE/Duke
chr11.783
chr11.786
chr11.784
chr11.787
chr11.788
chr11.789
chr11.790
chr11.791
chr11.792
chr11.793
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
HBG1
HBG2
HBE1
-Log(p-value)
200
150
100
50
0
5270000
5280000
5290000
5300000
5310000
Chr11 Pos (bp)

FIG. 44A

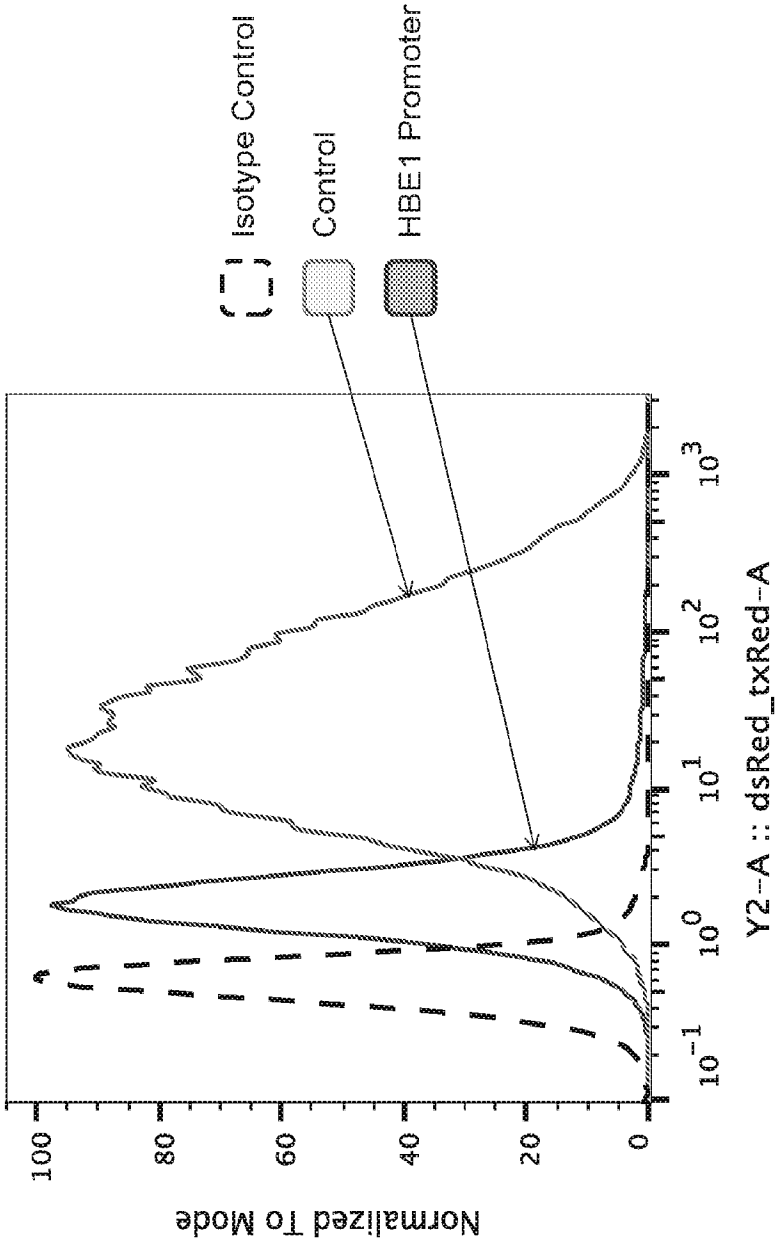

Isotype Control
Control
HBE1 Promoter
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Y2-A :: dsRed_txRed-A

FIG. 44B

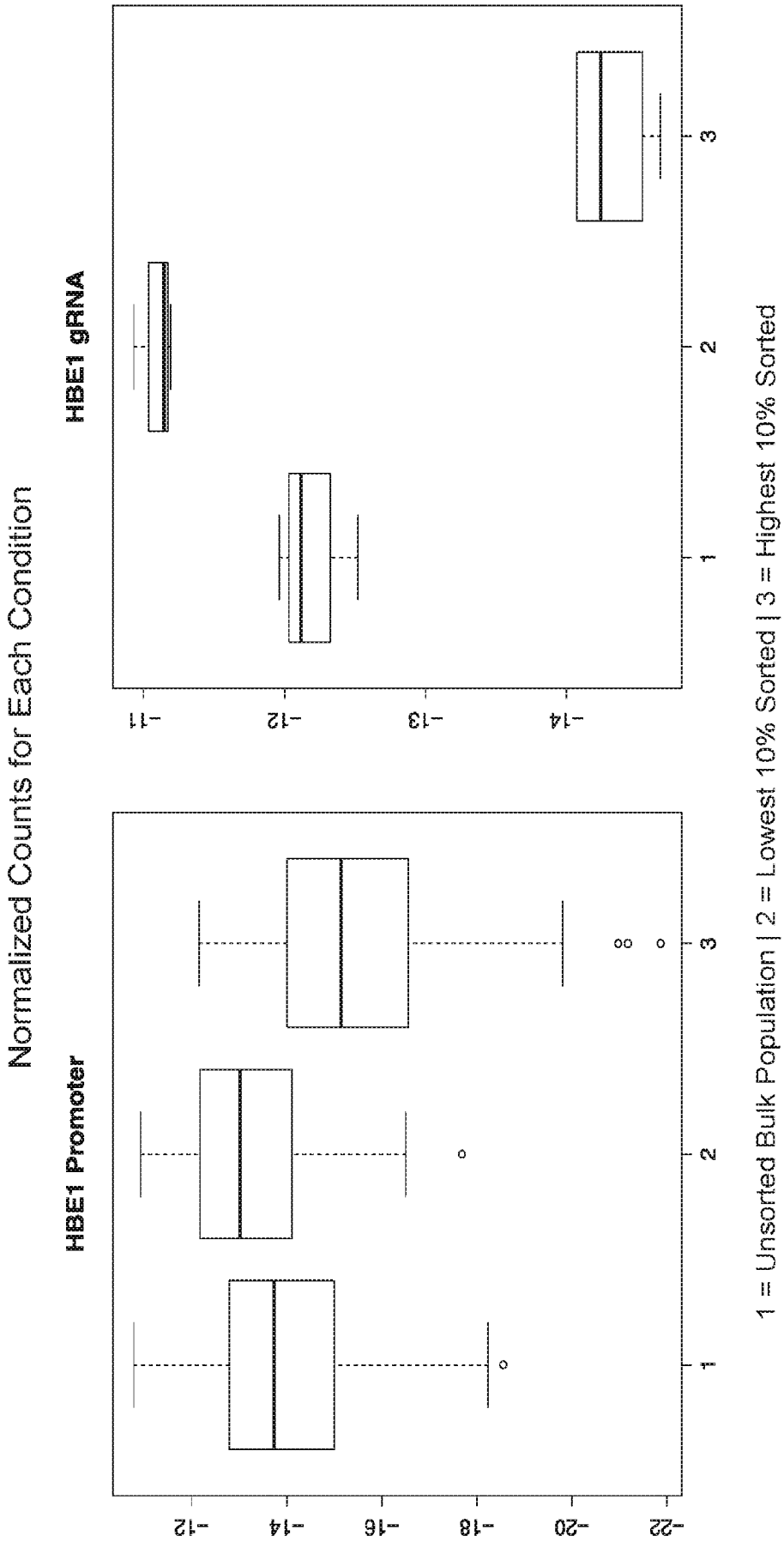

Normalized Counts for Each Condition
HBE1 Promoter
HBE1 gRNA
1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 44C

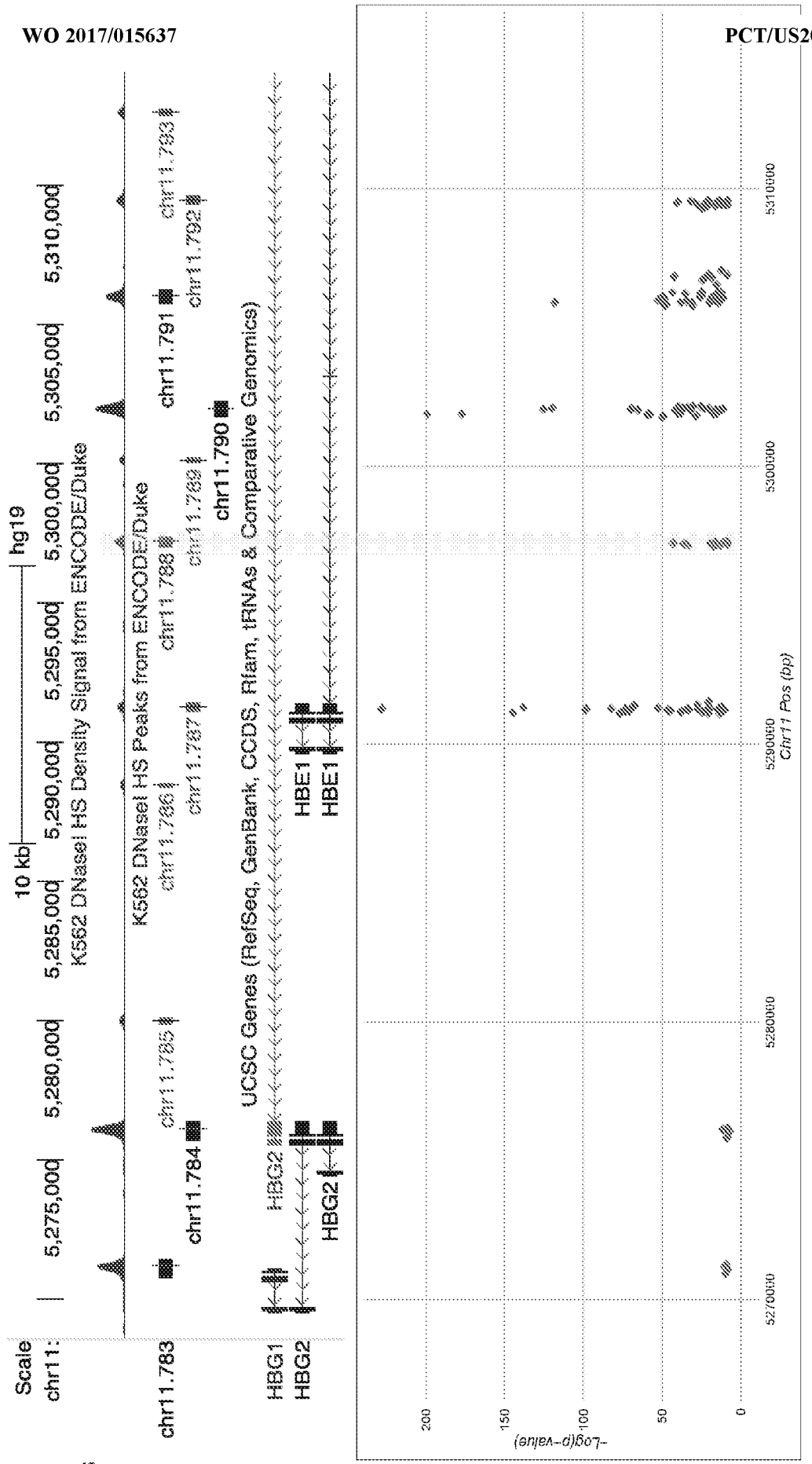

Scale
chr11:
10 kb
hg19
5,275,000
5,280,000
5,285,000
5,290,000
5,295,000
5,300,000
5,305,000
5,310,000
K562 DNaseI HS Density Signal from ENCODE/Duke
K562 DS
chr11.783
K562 DNaseI HS Peaks from ENCODE/Duke
chr11.786
chr11.788
chr11.791
chr11.793
chr11.785
chr11.784
chr11.787
chr11.789
chr11.792
chr11.790
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
HBG1
HBG2
HBG2
HBE1
HBG2
HBE1
-Log(p-value)
200
150
100
50
0
5270000
5280000
5290000
5300000
5310000
Chr11 Pos (bp)

FIG. 45A

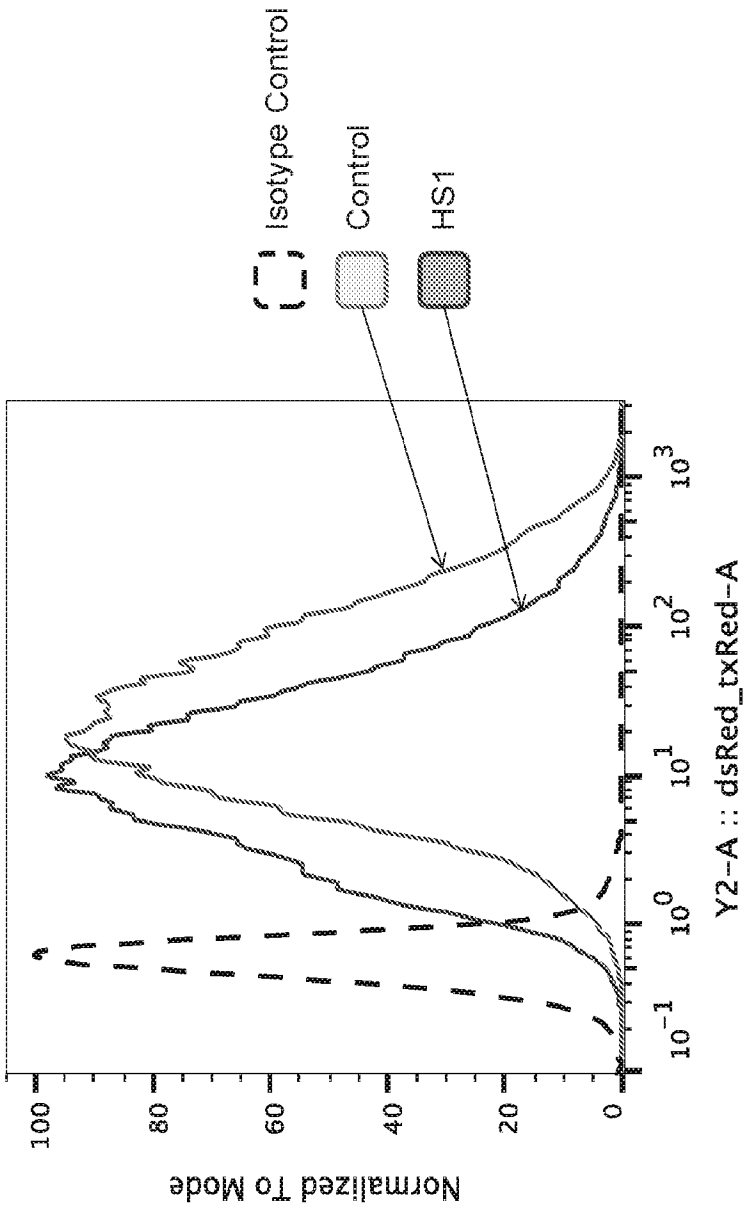

Isotype Control
Control
HS1
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Y2-A :: dsRed_txRed-A

FIG. 45B

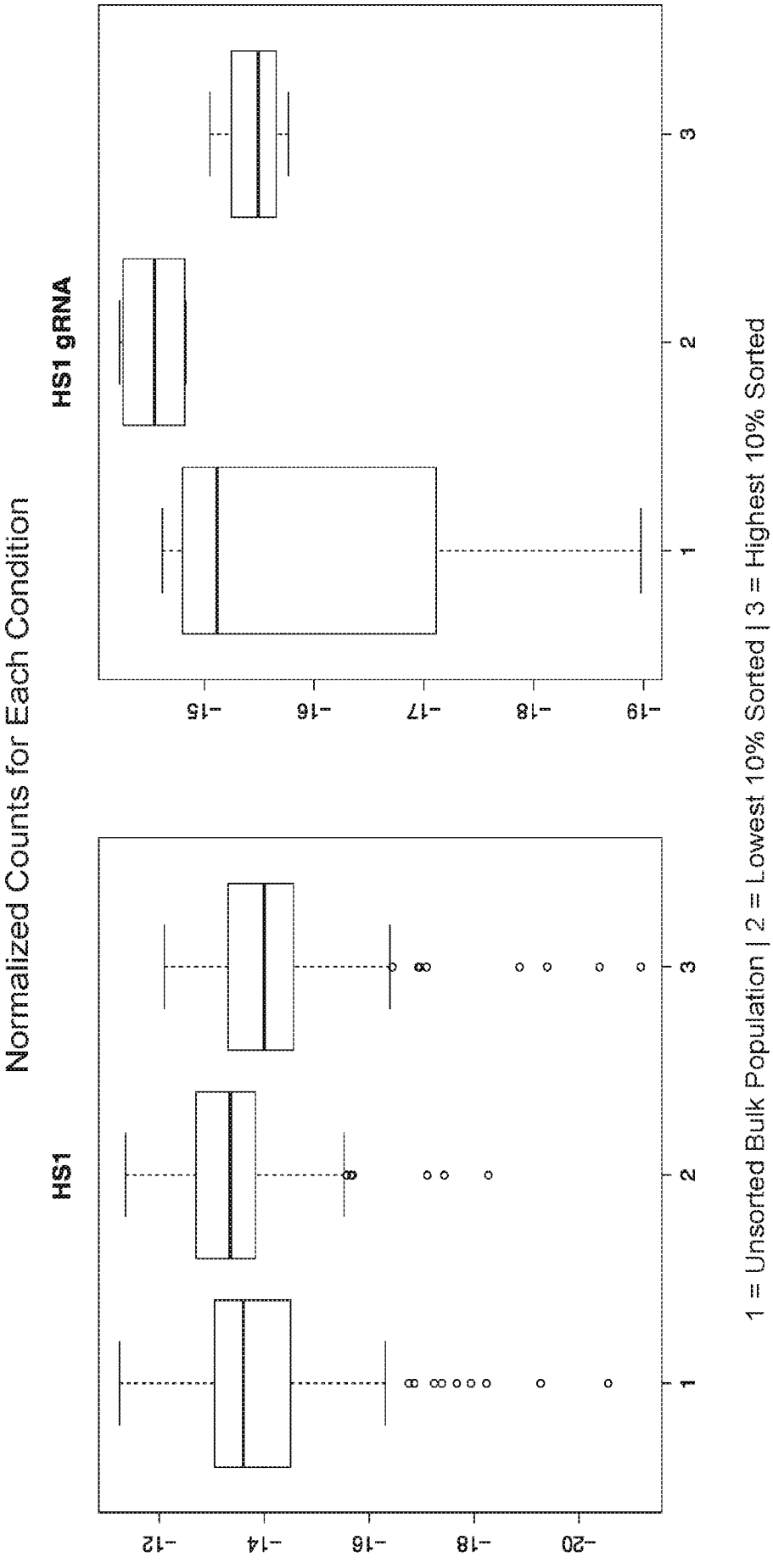

Normalized Counts for Each Condition
HS1
HS1 gRNA
1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 45C

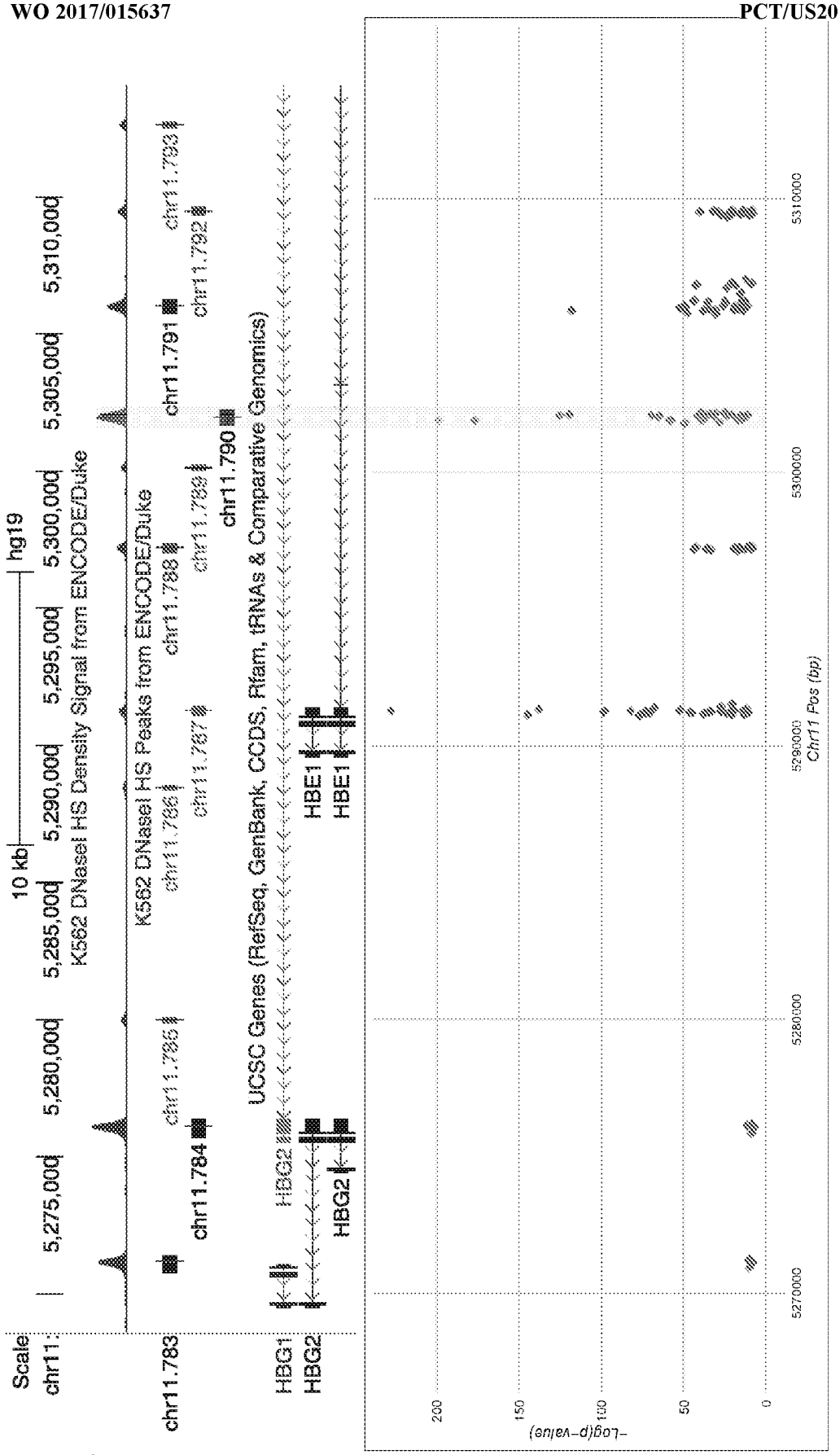

K562 DS
Scale
chr11:
5,275,000
5,280,000
5,285,000
5,290,000
5,295,000
5,300,000
5,305,000
5,310,000
10 kb
hg19
K562 DNaseI HS Density Signal from ENCODE/Duke
K562 DNaseI HS Peaks from ENCODE/Duke
chr11.783
chr11.785
chr11.784
chr11.786
chr11.788
chr11.787
chr11.789
chr11.790
chr11.791
chr11.792
chr11.793
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
HBG1
HBG2
HBG2
HBE1
HBE1
200
150
100
50
0
-Log(p-value)
5270000
5280000
5290000
5300000
5310000
Chr11 Pos (bp)

FIG. 46A

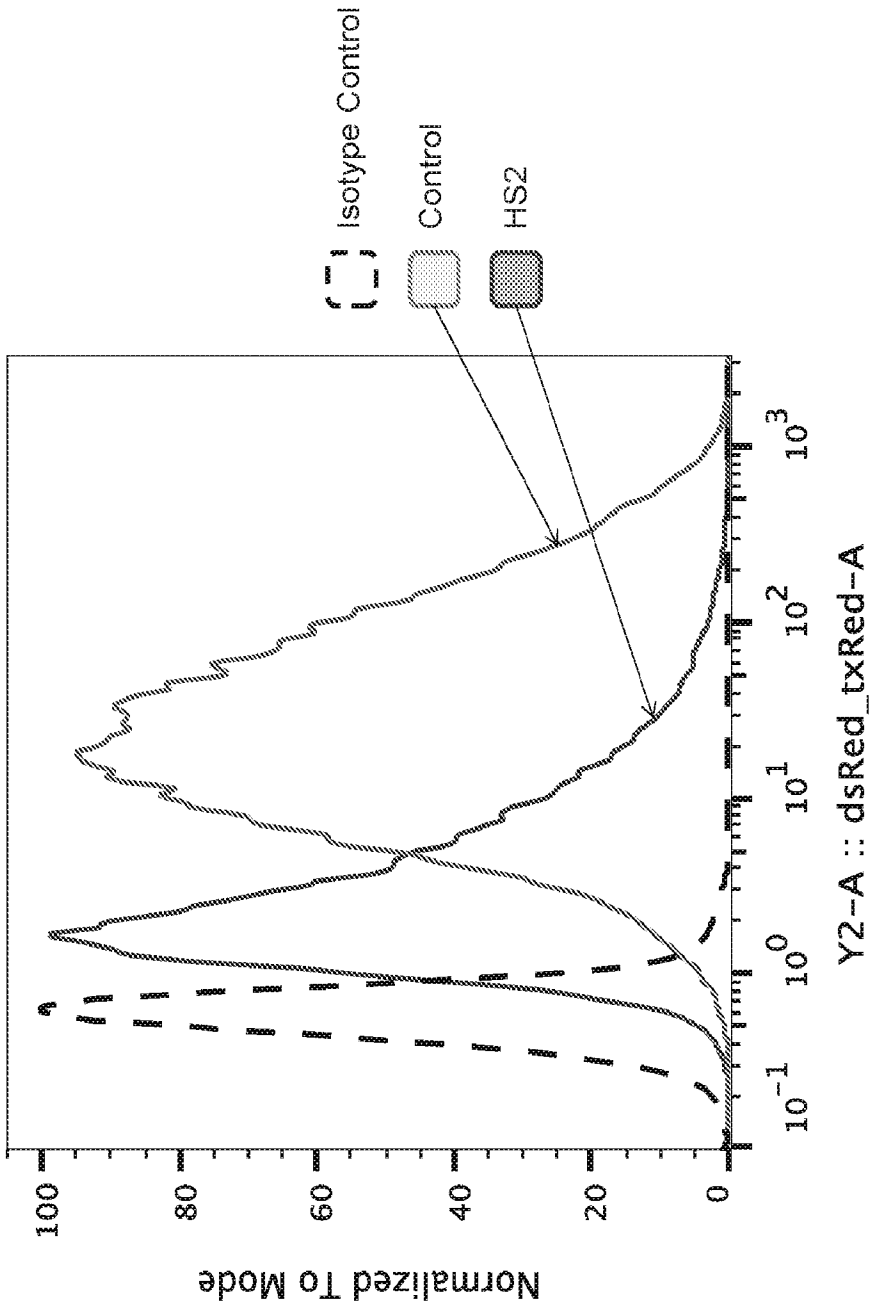

Isotype Control
Control
HS2
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Y2-A :: dsRed_txRed-A

FIG. 46B

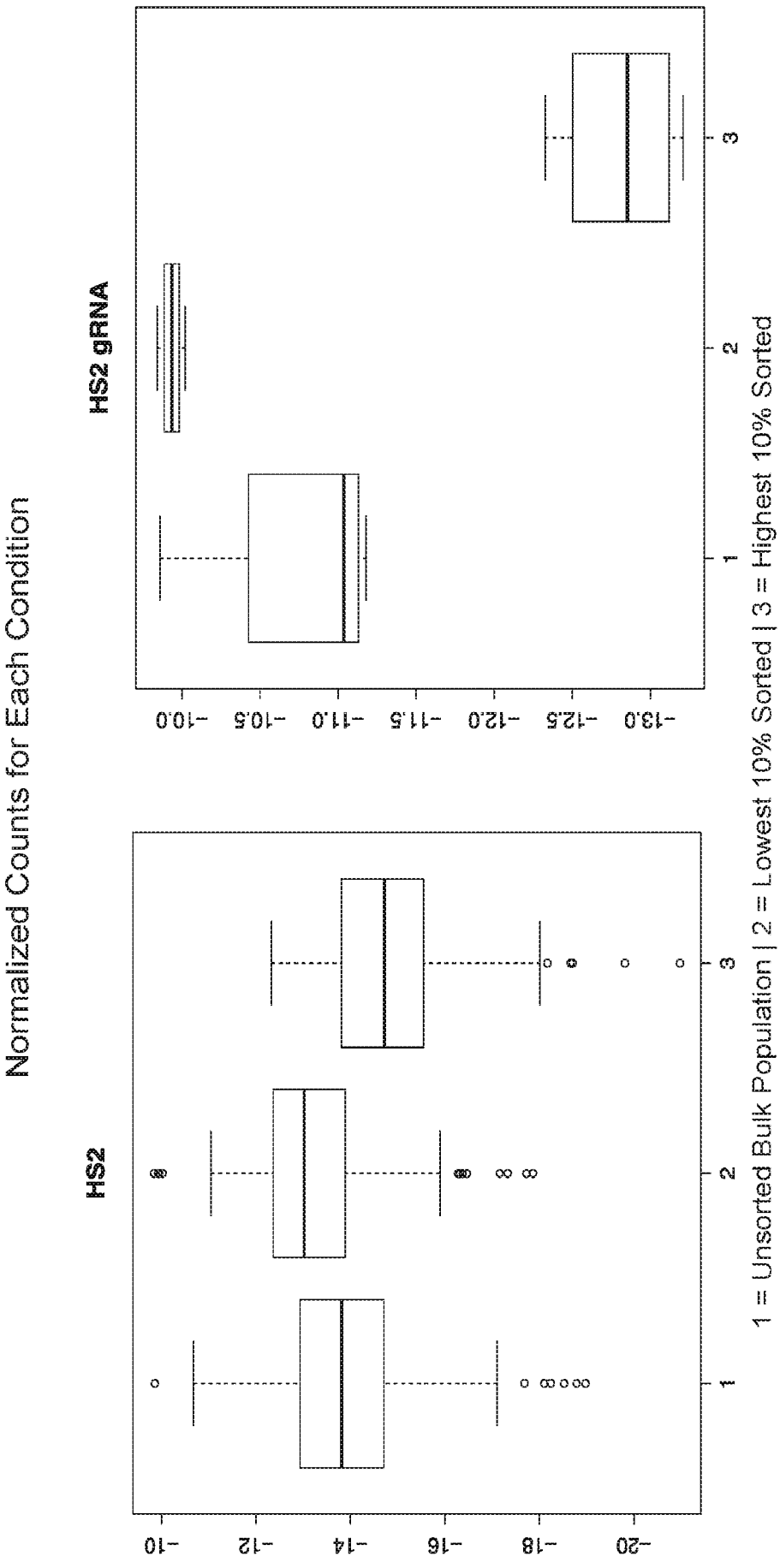

Normalized Counts for Each Condition
HS2
HS2 gRNA
1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 46C

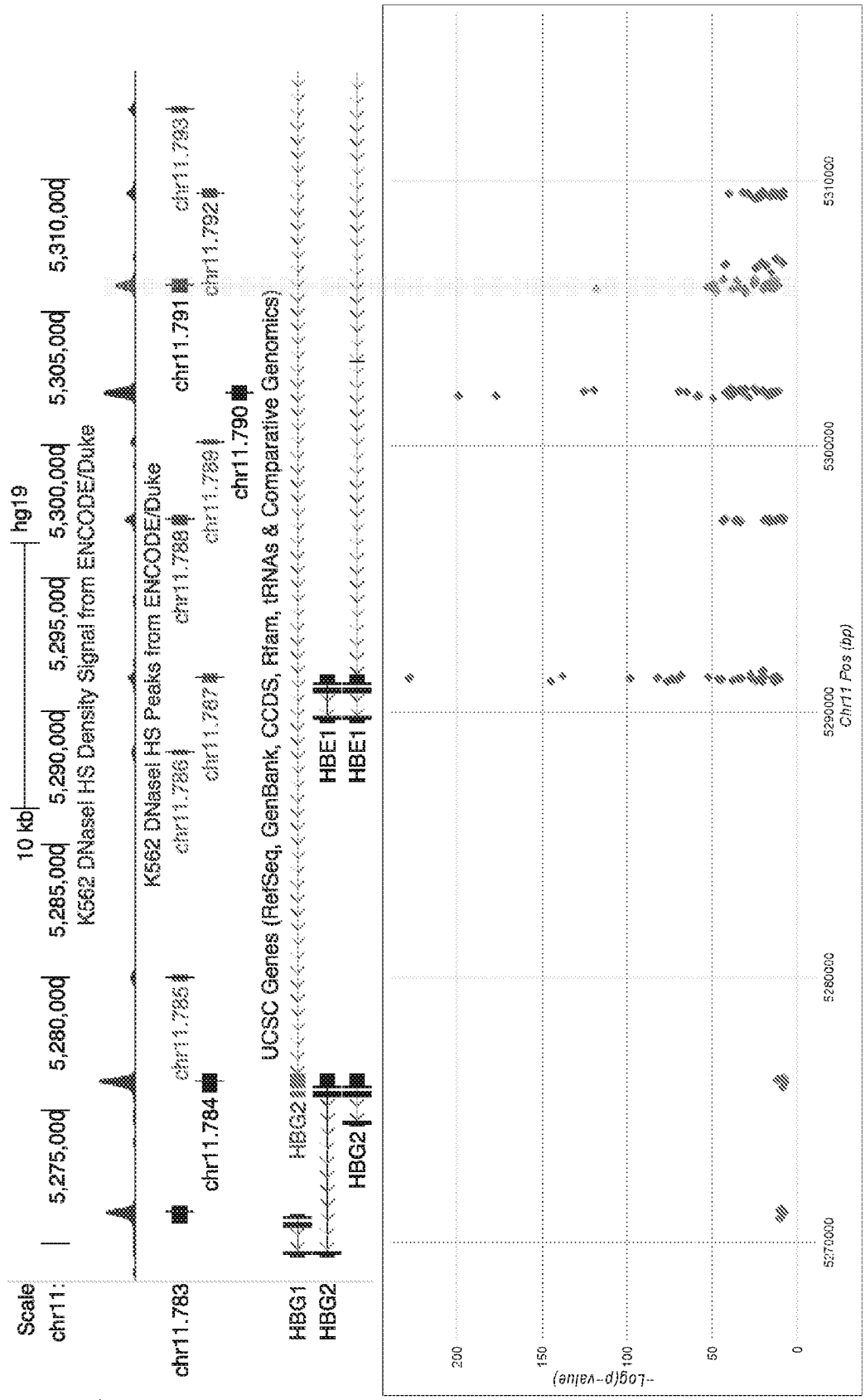

Scale
chr11:
5,275,000
5,280,000
5,285,000
10 kb
5,290,000
5,295,000
5,300,000
5,305,000
5,310,000
hg19
K562 DNaseI HS Density Signal from ENCODE/Duke
K562 DNaseI HS Peaks from ENCODE/Duke
chr11.783
chr11.786
chr11.784
chr11.787
chr11.788
chr11.789
chr11.790
chr11.791
chr11.792
chr11.793
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
HBG1
HBG2
HBG2
HBE1
HBE1
K562 DS
200
150
100
50
0
-Log(p-value)
5270000
5280000
5290000
5300000
5310000
Chr11 Pos (bp)

FIG. 47A

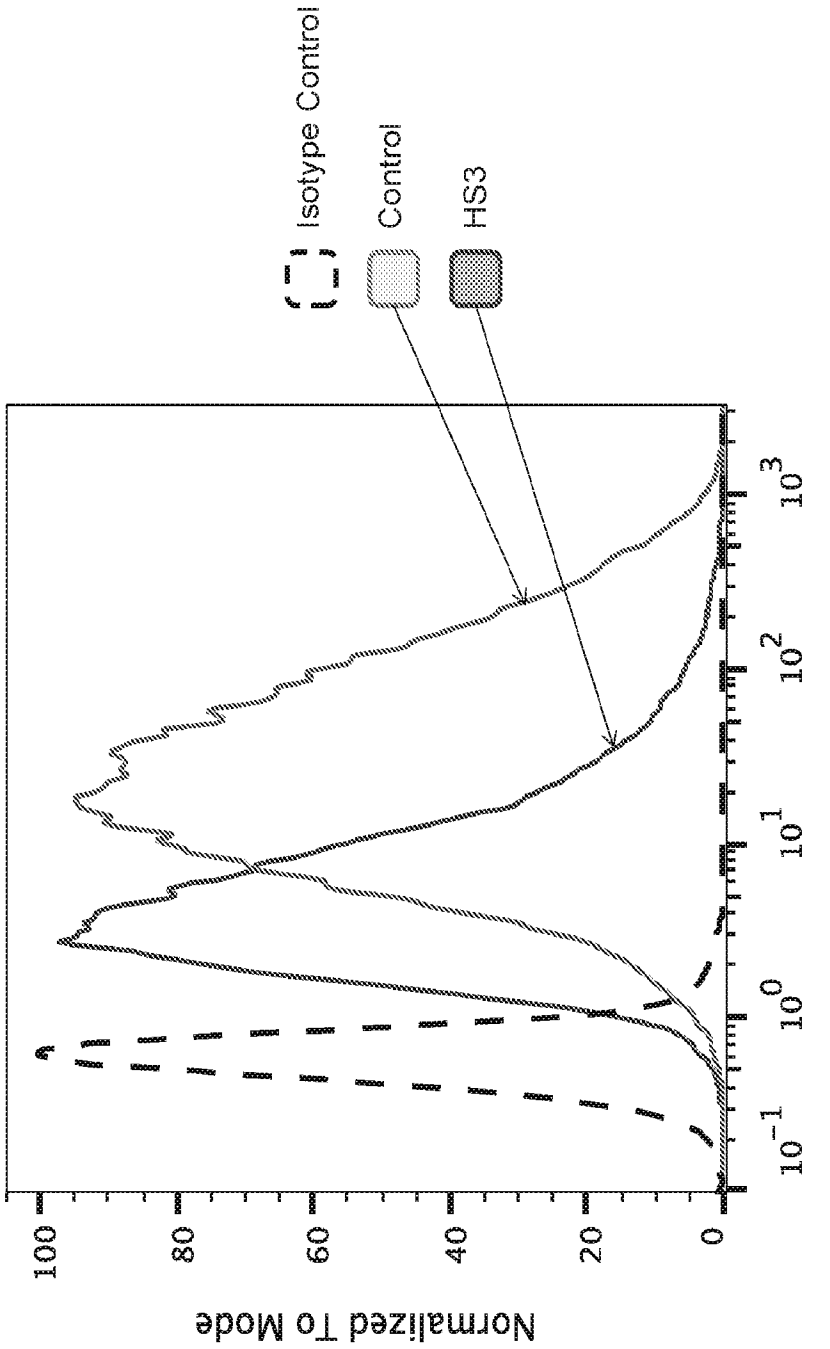

Isotype Control
Control
HS3
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Y2-A :: dsRed_txRed-A

FIG. 47B

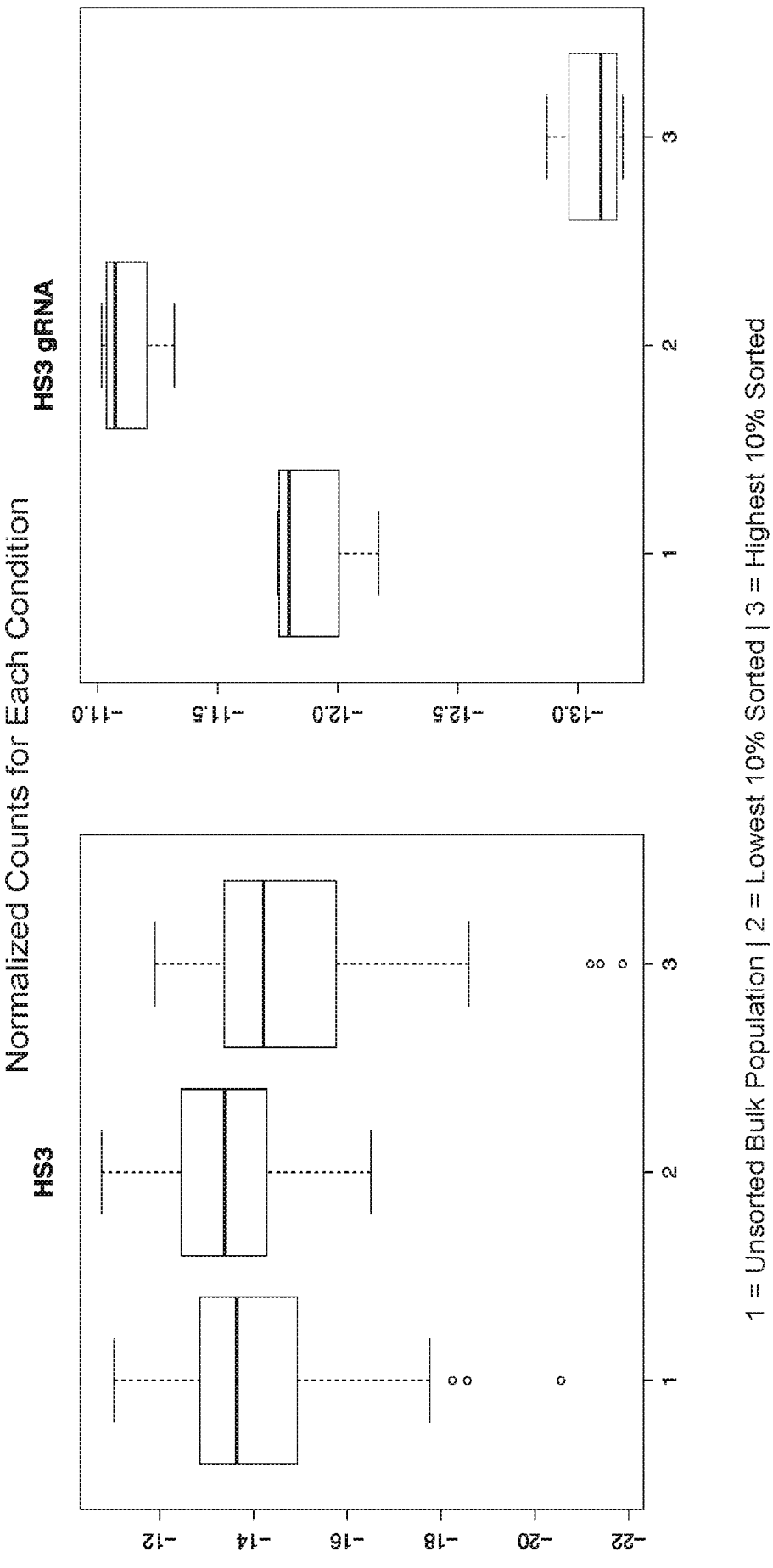

Normalized Counts for Each Condition
HS3
HS3 gRNA
1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 47C

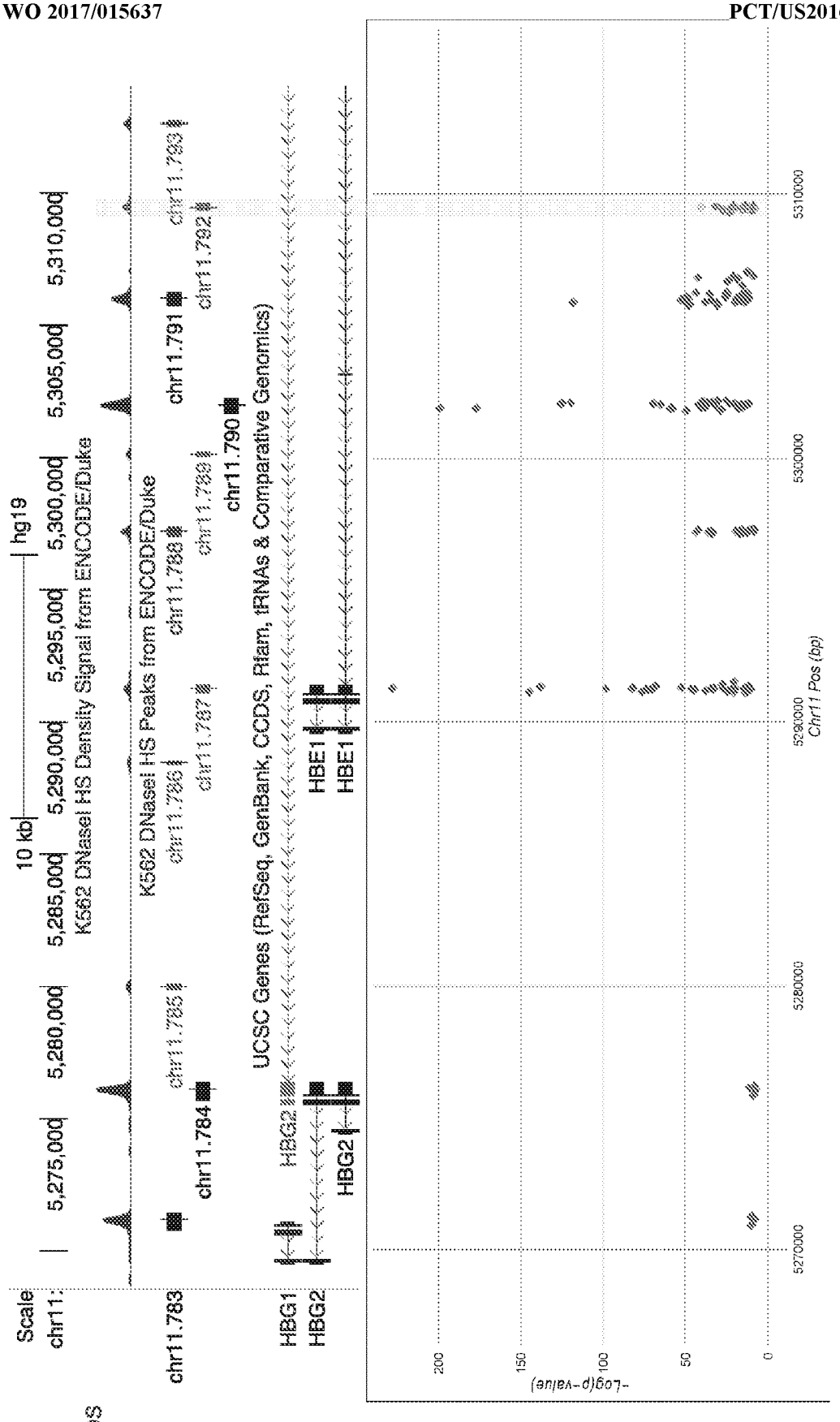

Scale
chr11:
5,275,000
5,280,000
5,285,000
10 kb
5,290,000
5,295,000
5,300,000
5,305,000
5,310,000
hg19
K562 DNaseI HS Density Signal from ENCODE/Duke
K562 DNaseI HS Peaks from ENCODE/Duke
chr11.783
chr11.785
chr11.786
chr11.788
chr11.791
chr11.793
chr11.784
chr11.787
chr11.789
chr11.792
chr11.790
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
HBG1
HBG2
HBG2
HBE1
HBG2
HBE1
K562 DS
200
150
100
50
0
-Log(p-value)
5270000
5280000
5290000
5300000
5310000
Chr11 Pos (bp)

FIG. 48A

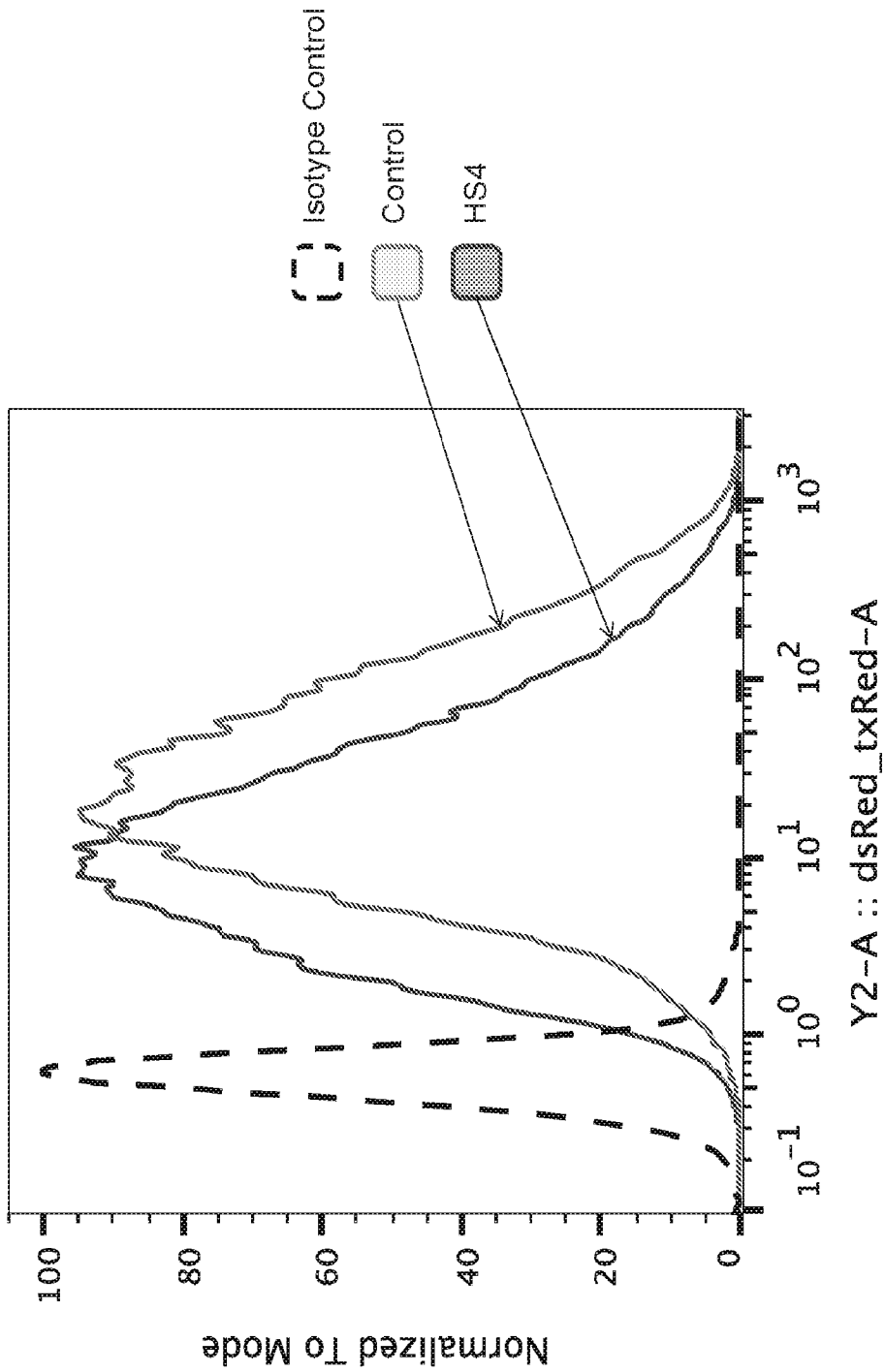

Isotype Control
Control
HS4
Normalized To Mode
100
80
60
40
20
0
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Y2-A :: dsRed_txRed-A

FIG. 48B

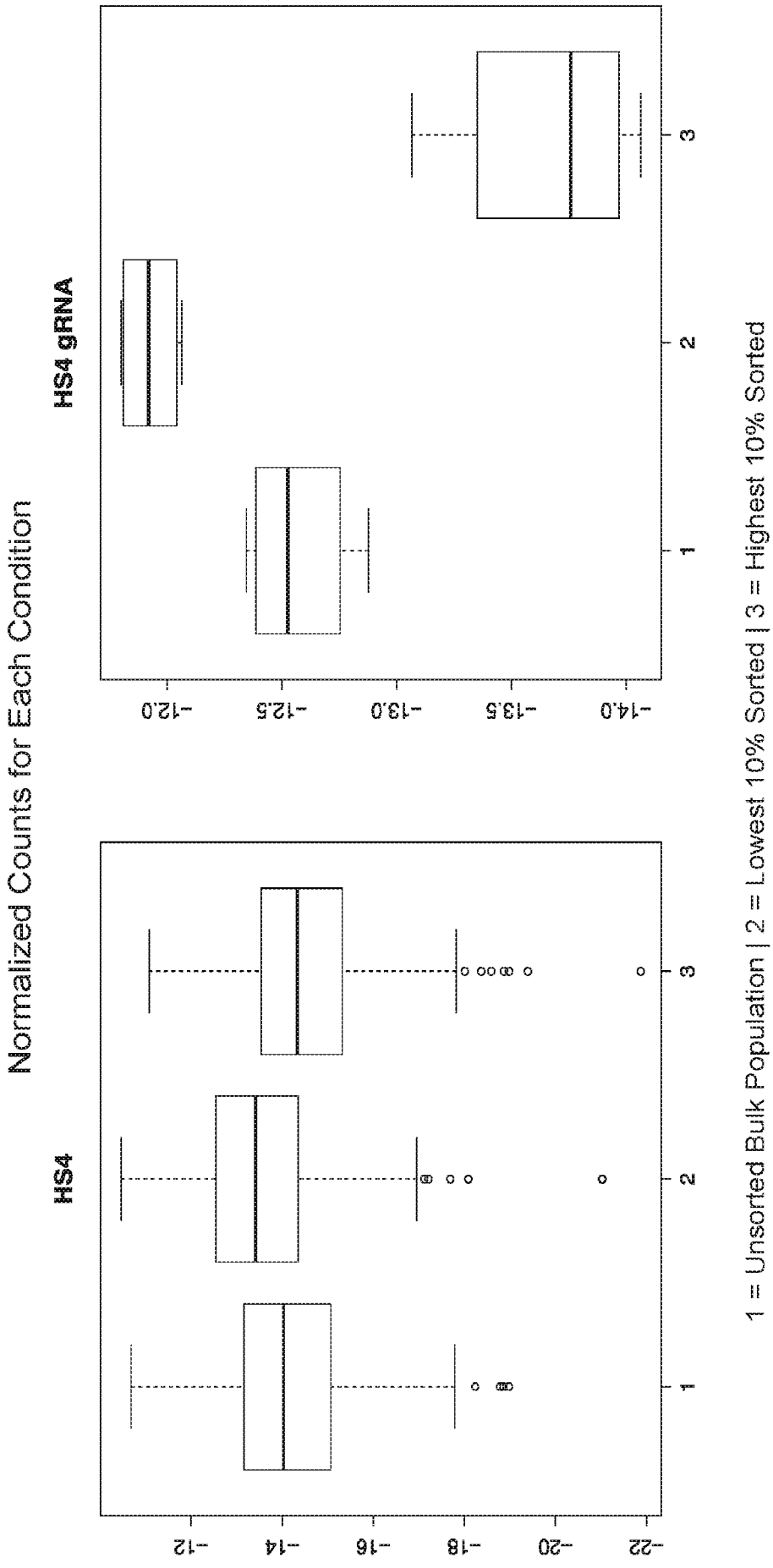

Normalized Counts for Each Condition
HS4
HS4 gRNA
-12
-14
-16
-18
-20
-22
-12.0
-12.5
-13.0
-13.5
-14.0
1
2
3
1 = Unsorted Bulk Population | 2 = Lowest 10% Sorted | 3 = Highest 10% Sorted

FIG. 48C

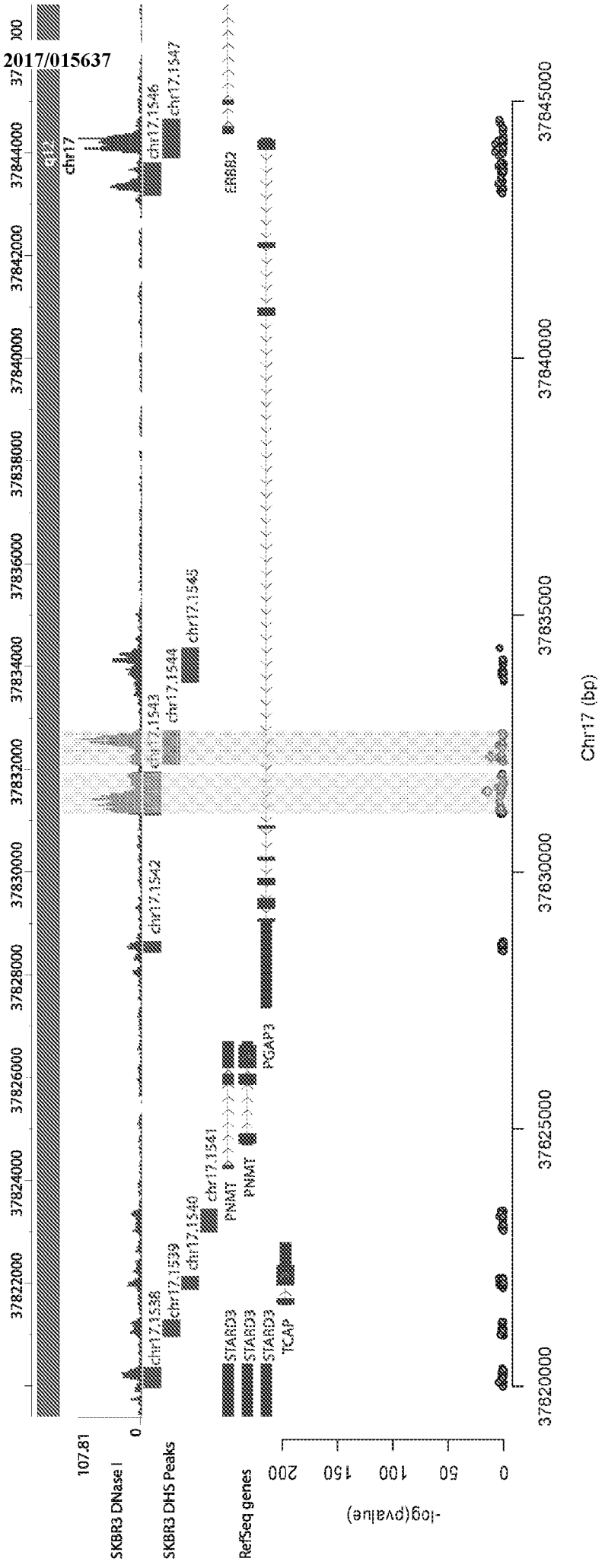

SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
Chr17 (bp)
chr17
ERBB2
PGAP3
PNMT
STARD3
TCAP

FIG. 49A

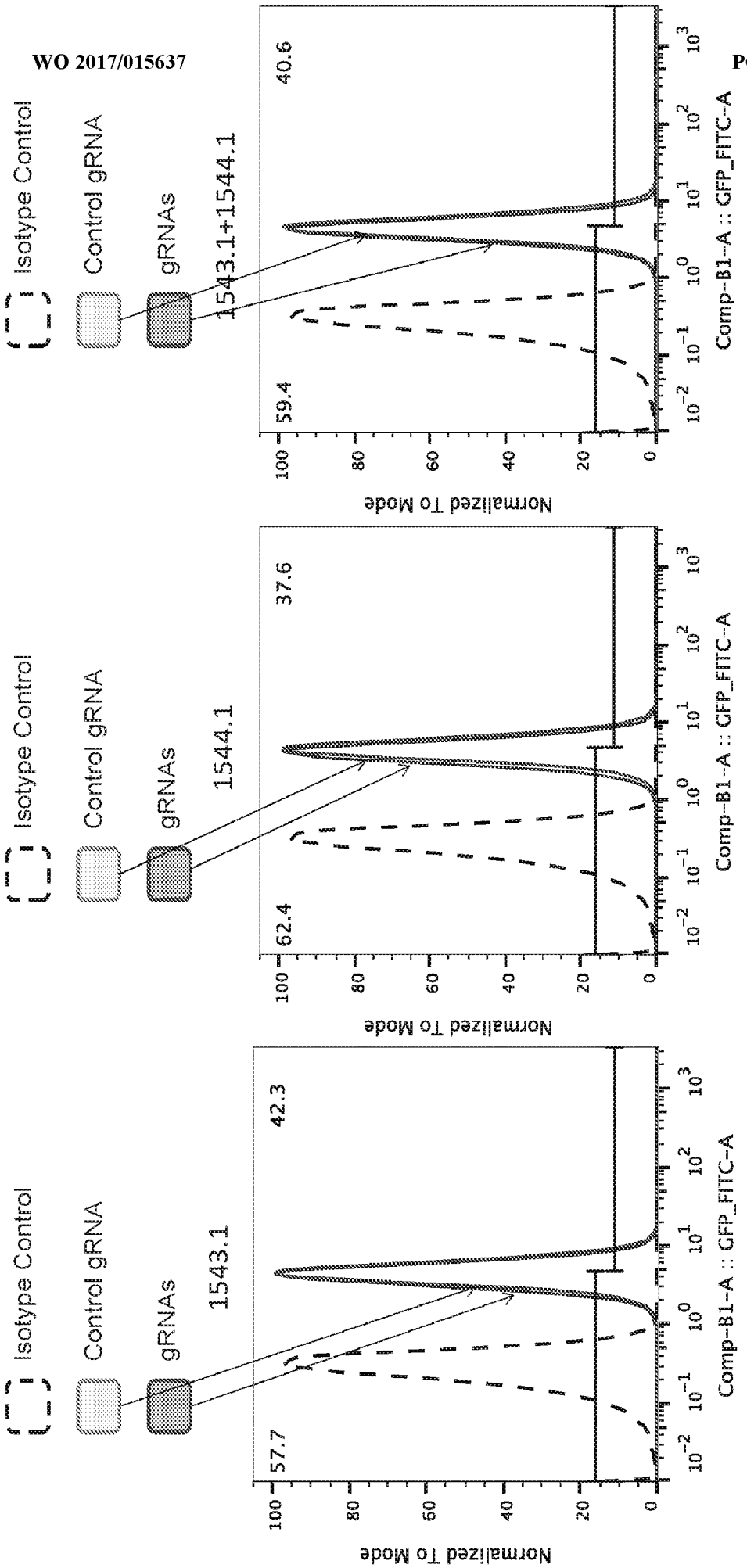

Isotype Control
Control gRNA
gRNAs
1543.1
42.3
57.7
1544.1
37.6
62.4
1543.1+1544.1
40.6
59.4
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 49B

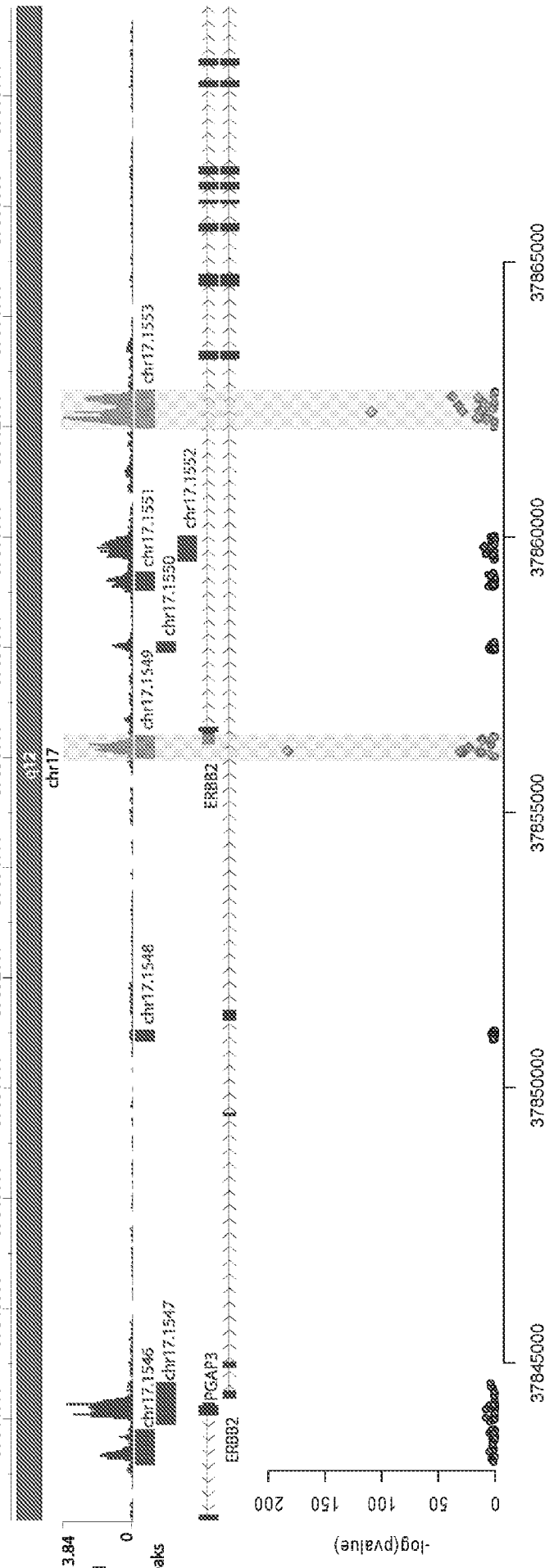

37844000
37846000
37848000
37850000
37852000
37854000
37856000
37858000
37860000
37862000
37864000
37866000
chr17
113.84
0
SKBR3 DNase I
SKBR3 DH5 Peaks
RefSeq genes
chr17.1546
chr17.1547
chr17.1548
chr17.1549
chr17.1550
chr17.1551
chr17.1552
chr17.1553
PGAP3
ERBB2
ERBB2
0
50
100
150
200
-log(pvalue)
37845000
37850000
37855000
37860000
37865000
Chr17 (bp)

FIG. 50A

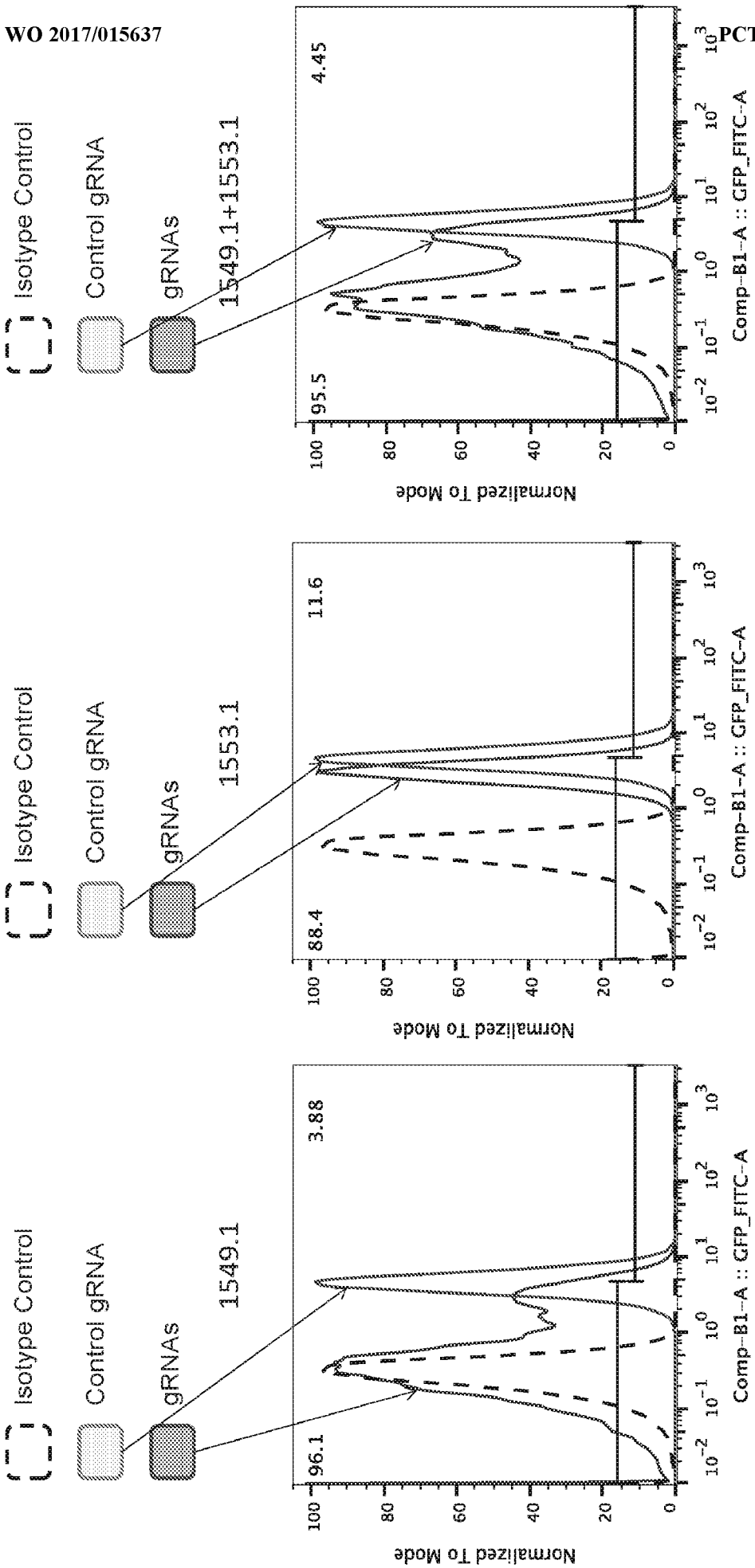

Isotype Control
Control gRNA
gRNAs
1549.1
96.1
3.88
1553.1
88.4
11.6
1549.1+1553.1
95.5
4.45
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 50B

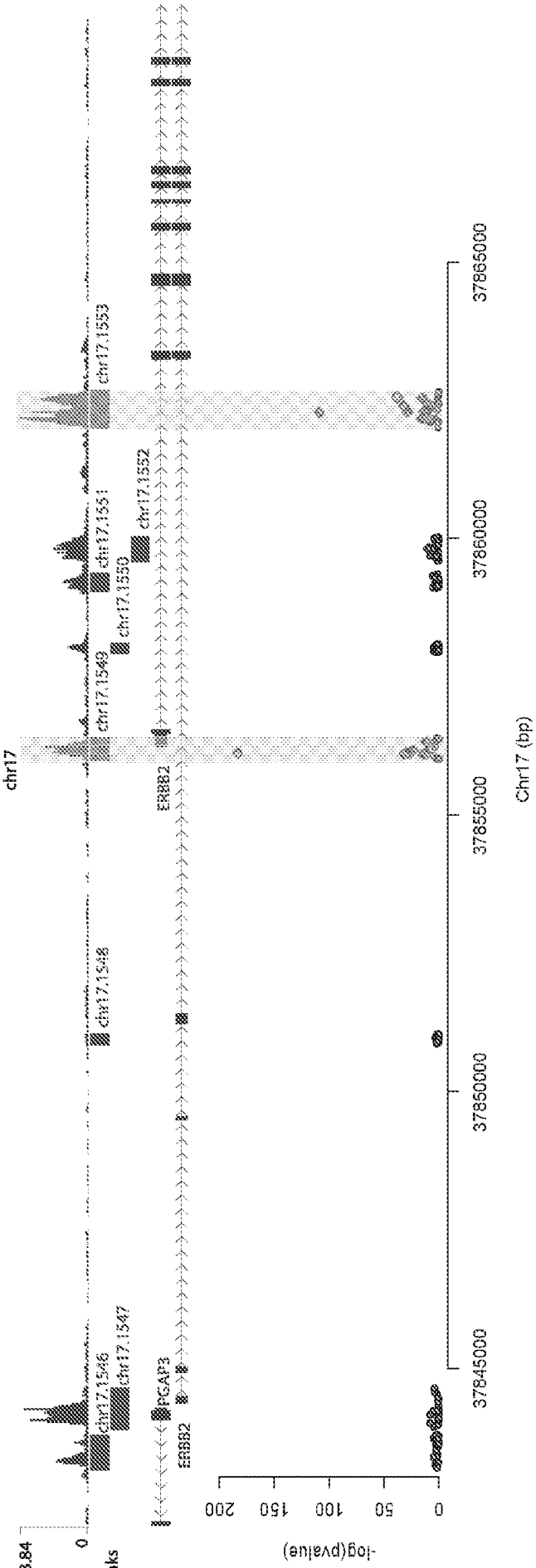

37844000
37846000
37848000
37850000
37852000
37854000
37856000
37858000
37860000
37862000
37864000
37866000
37868000
chr17
113.84
0
SKBR3 DNase I
SKBR3 DH5 Peaks
RefSeq genes
chr17.1546
chr17.1547
chr17.1548
chr17.1549
chr17.1550
chr17.1551
chr17.1552
chr17.1553
PGAP3
ERBB2
ERBB2
200
150
100
50
0
-log(pvalue)
37845000
37850000
37855000
37860000
37865000
Chr17 (bp)

FIG. 51A

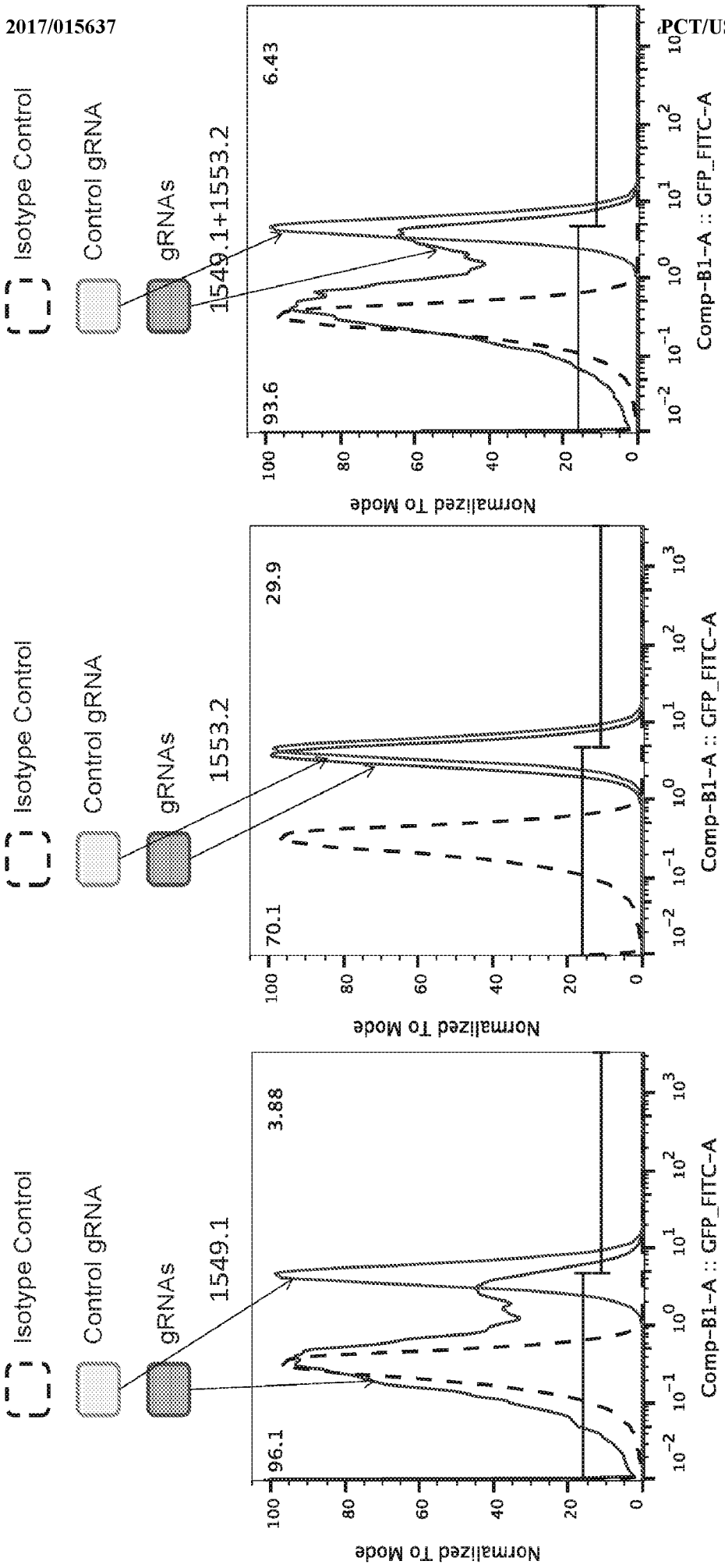

Isotype Control
Control gRNA
gRNAs
1549.1
96.1
3.88
1553.2
70.1
29.9
1549.1+1553.2
93.6
6.43
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 51B

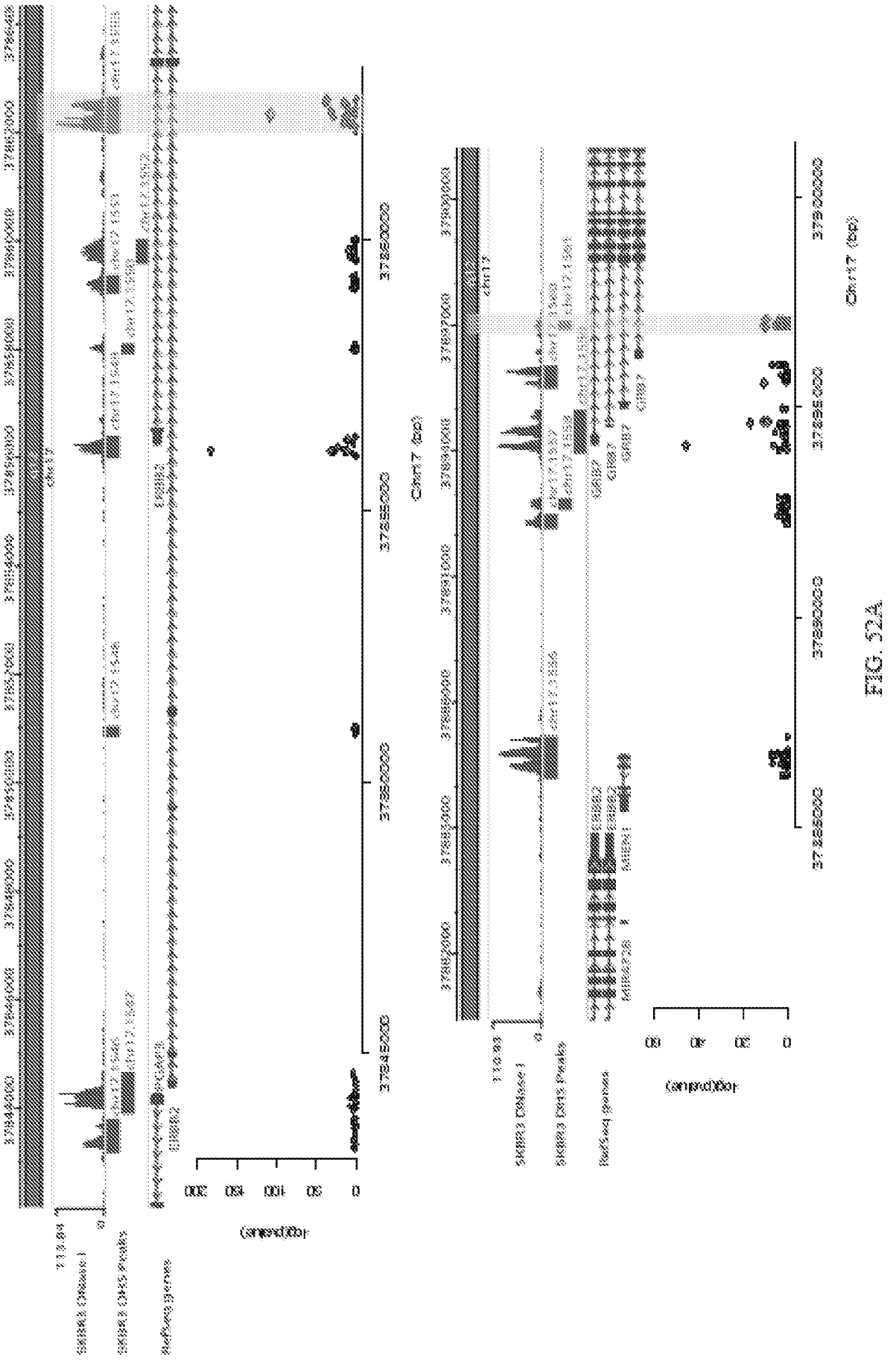


FIG. 52A

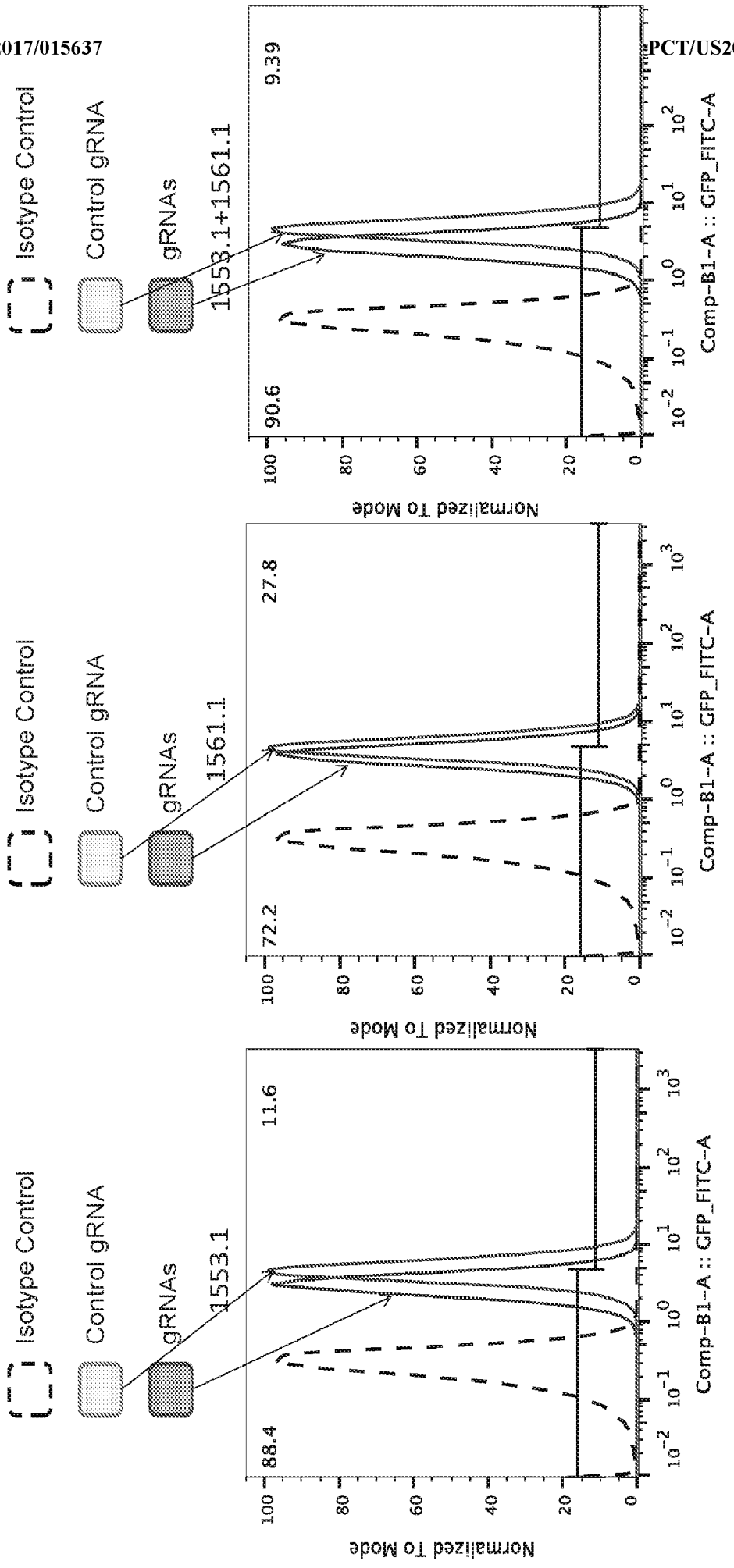

Isotype Control
Control gRNA
gRNAs
1553.1
88.4
11.6
1561.1
72.2
27.8
1553.1+1561.1
90.6
9.39
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 52B

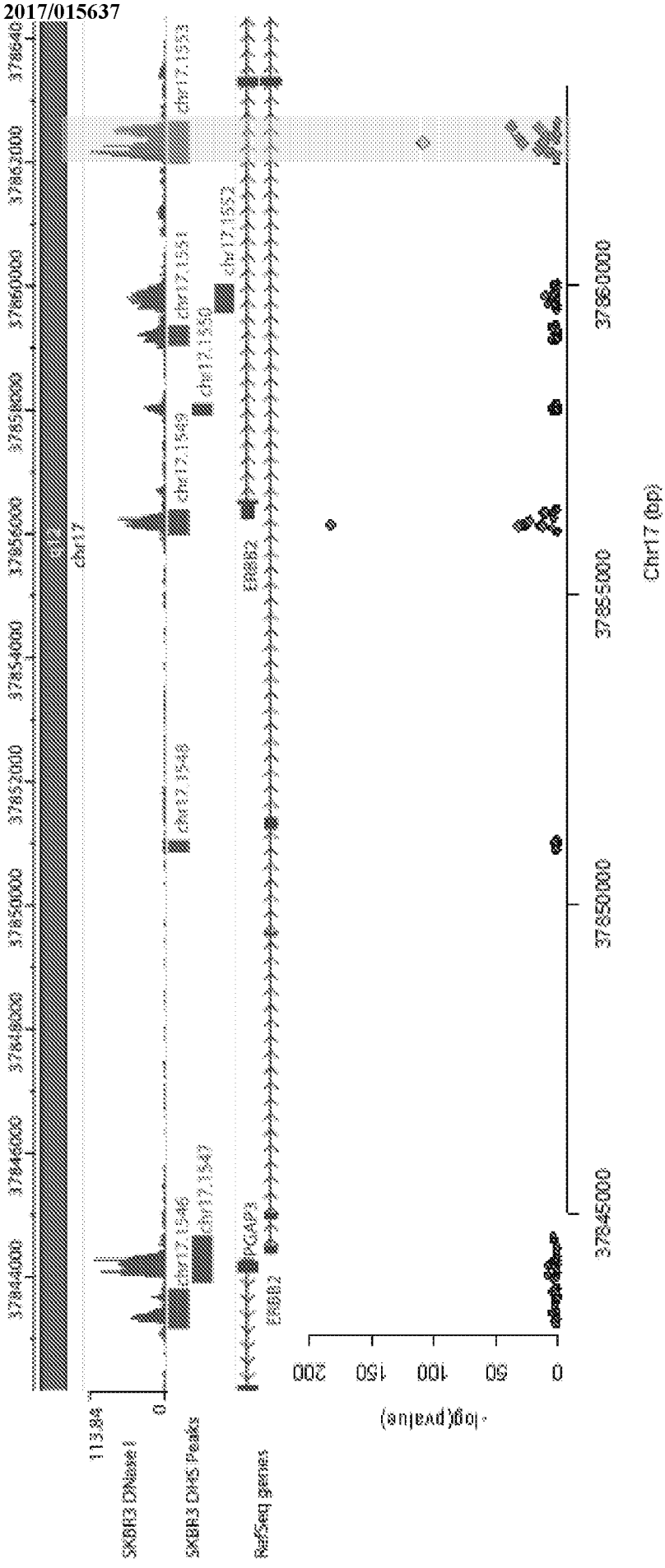

37844000
37846000
37848000
37850000
37852000
37854000
37856000
37858000
37860000
37862000
37864000
Chr17
Chr17 (bp)
37845000
37850000
37855000
37860000
SKBR3 DNase I
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
0
50
100
150
200
ERBB2
PGAP3
chr17:1546
chr17:1547
chr17:1548
chr17:1549
chr17:1550
chr17:1551
chr17:1552
chr17:1553

FIG. 33A

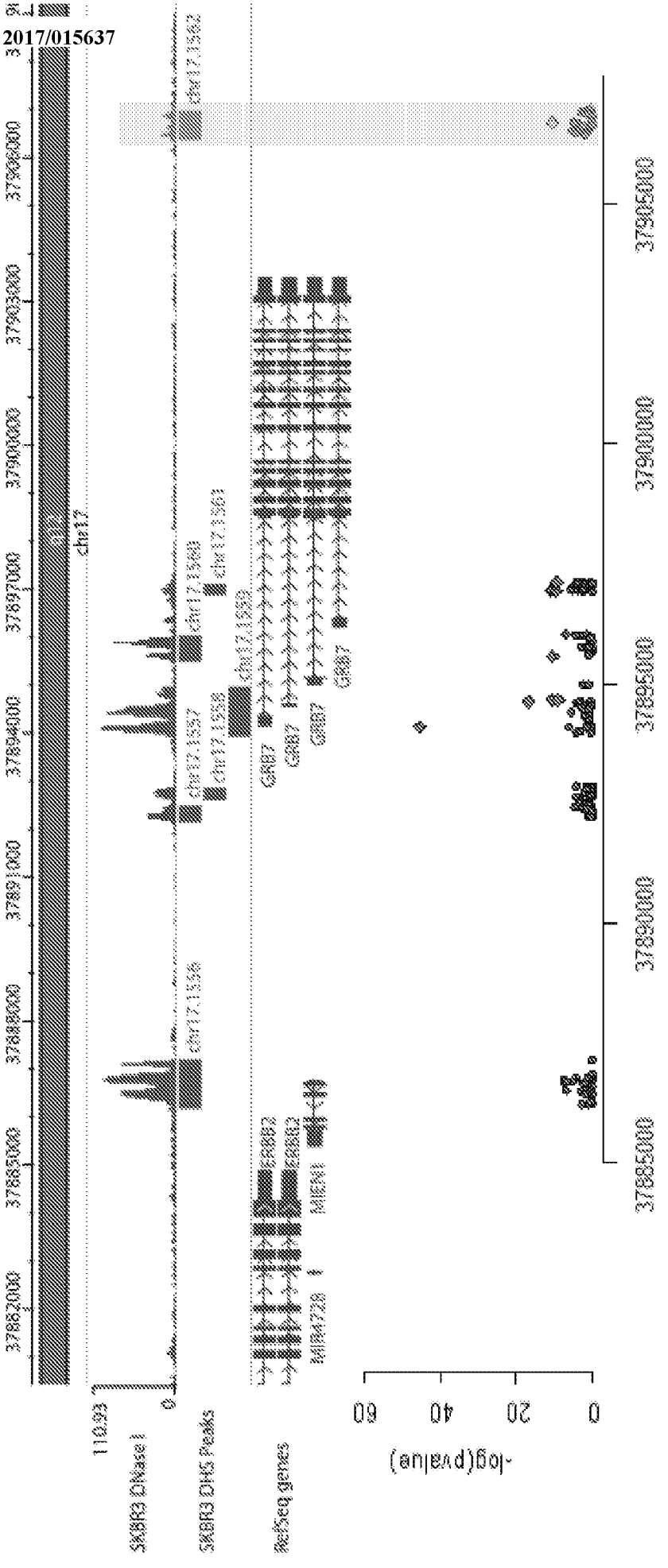

chr17
37862000
37865000
37868000
37871000
37874000
37877000
37880000
37883000
37886000
37889000
37892000
37895000
37898000
37901000
37904000
37907000
37910000
SKBR3 DNase I
110.93
0
SKBR3 DH5 Peaks
RefSeq genes
MIR4728
MIEN1
ERBB2
GRB7
-log(pvalue)
0
20
40
60
Chr17 (bp)

FIG. 53B

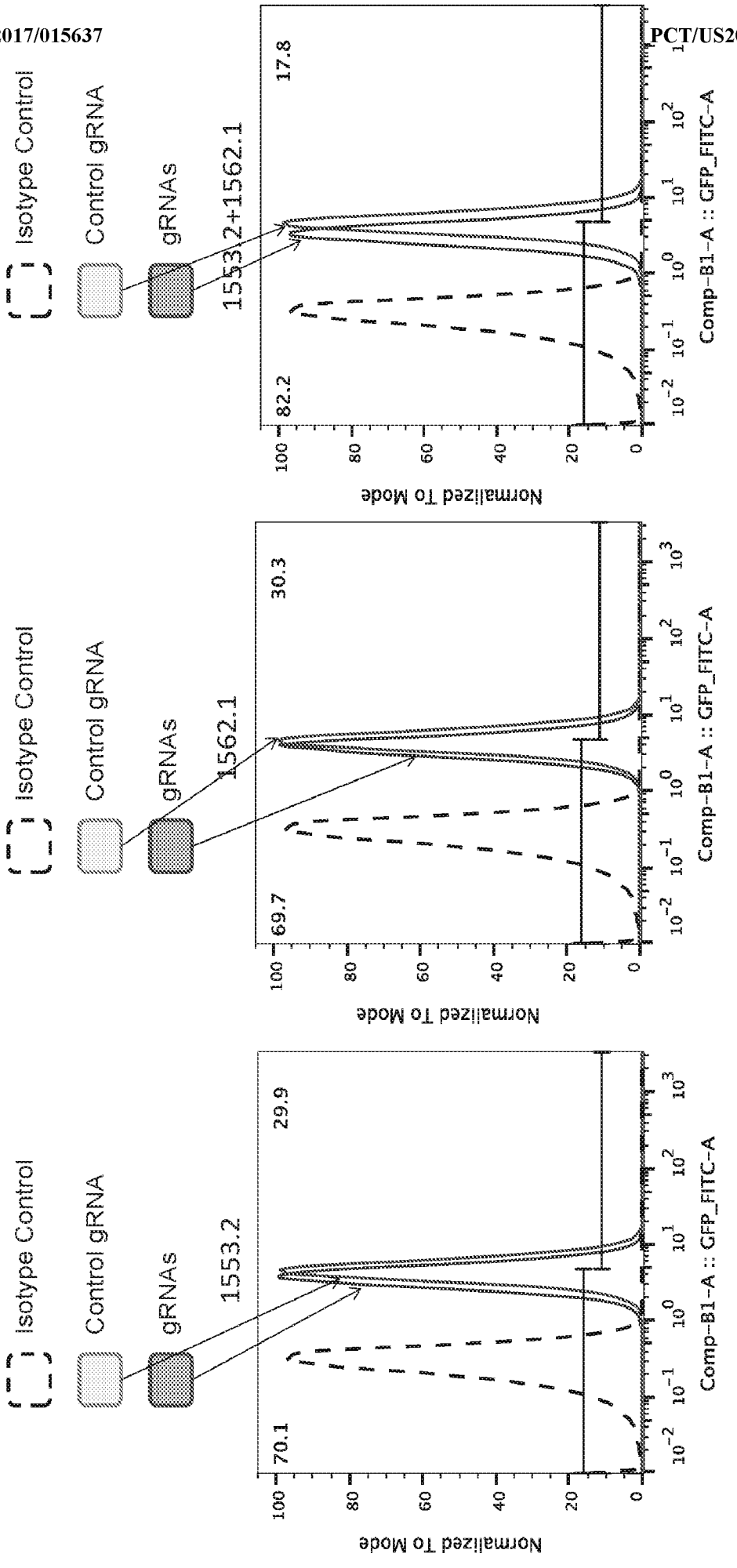

Isotype Control
Control gRNA
gRNAs
1553.2
70.1
29.9
1562.1
69.7
30.3
1553.2+1562.1
82.2
17.8
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 53C

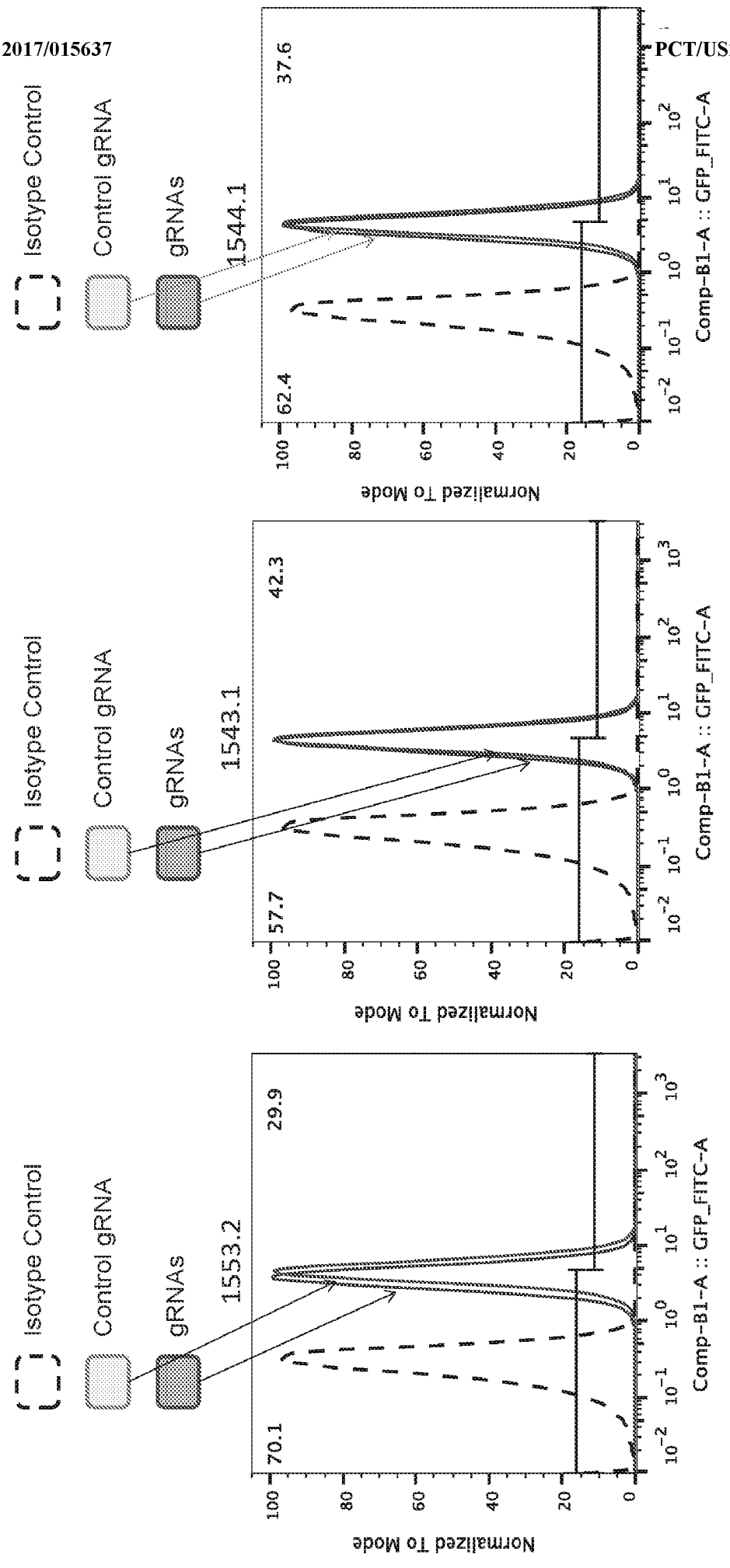

Isotype Control
Control gRNA
gRNAs
1544.1
62.4
37.6
1543.1
57.7
42.3
1553.2
70.1
29.9
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 54A

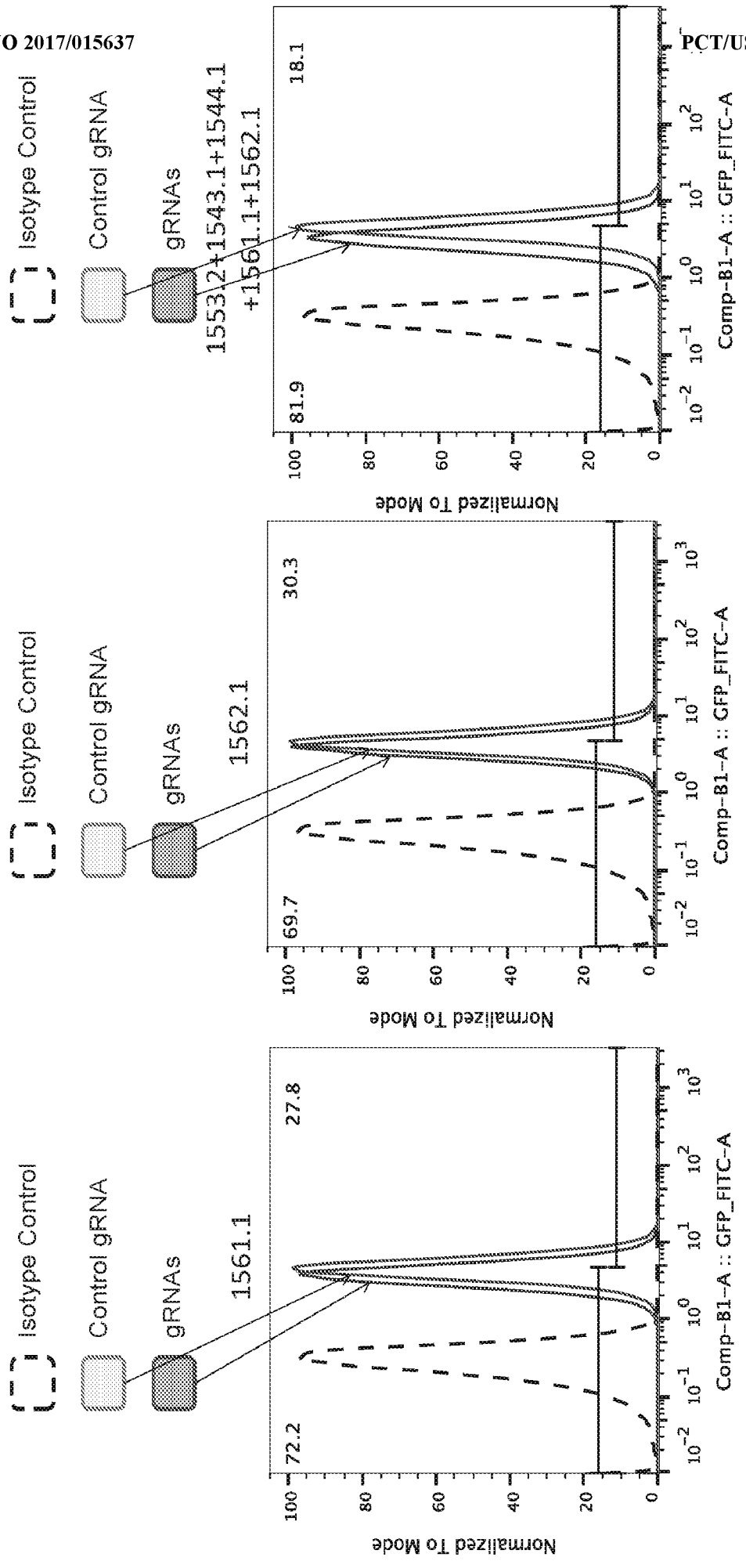

Isotype Control
Control gRNA
gRNAs
1561.1
27.8
72.2
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
Isotype Control
Control gRNA
gRNAs
1562.1
30.3
69.7
Normalized To Mode
Comp-B1-A :: GFP_FITC-A
Isotype Control
Control gRNA
gRNAs
1553.2+1543.1+1544.1
+1561.1+1562.1
18.1
81.9
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 54B

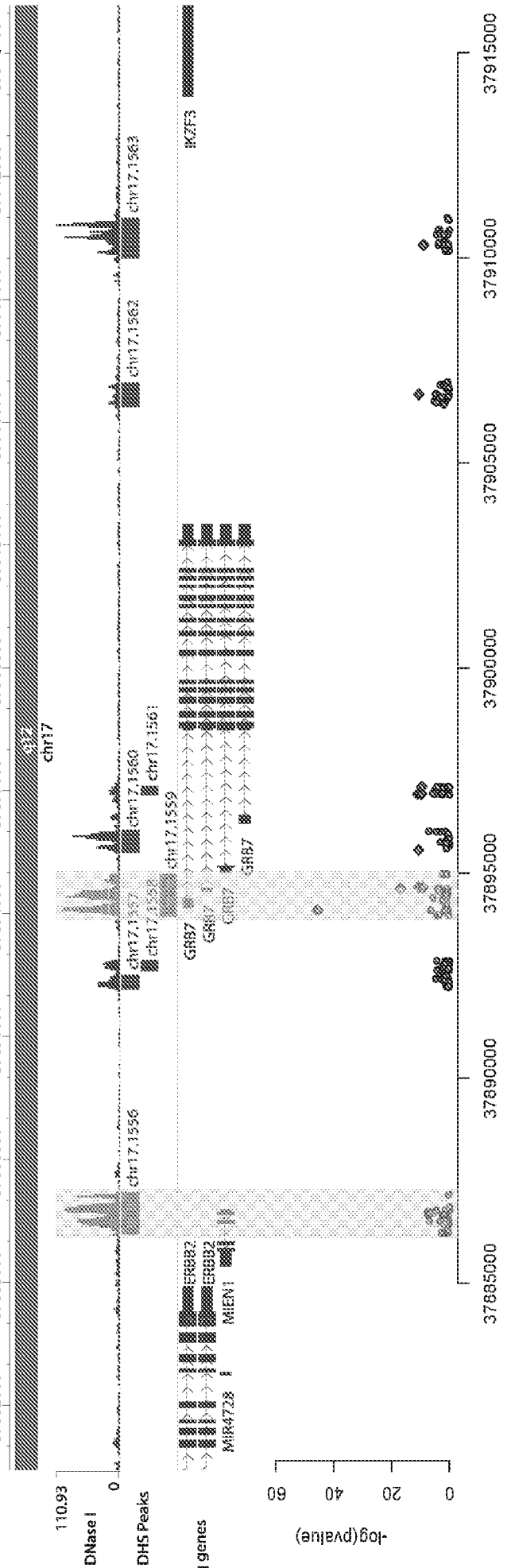

chr17
37882000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915000
Chr17 (bp)
SKBR3 DNase I
110.93
0
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
60
40
20
0
chr17:1556
chr17:1557
chr17:1558
chr17:1559
chr17:1560
chr17:1561
chr17:1562
chr17:1563
GRB7
ERBB2
MIEN1
MIR4728
IKZF3

FIG. 55A

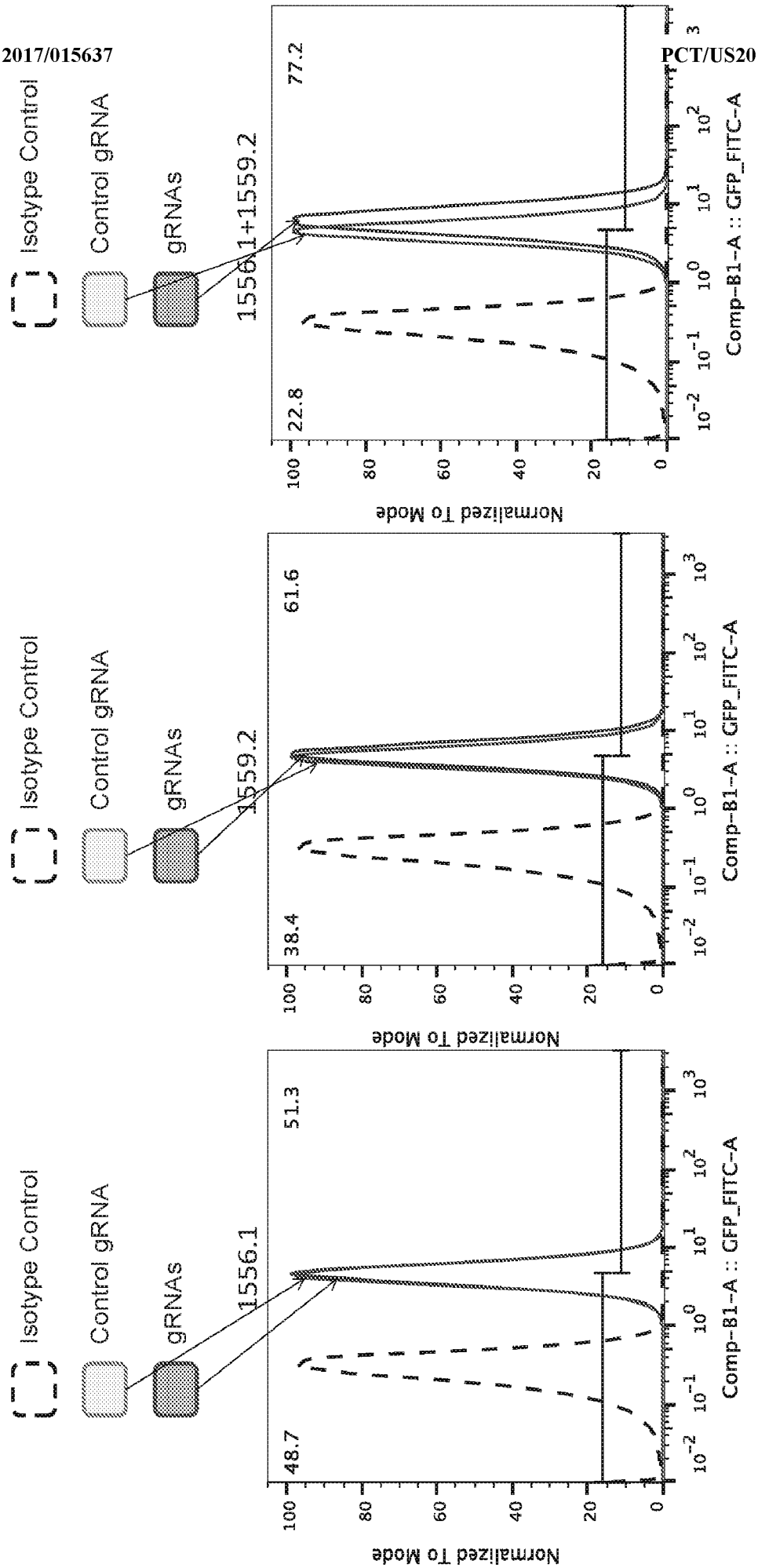

Isotype Control
Control gRNA
gRNAs
1556.1
48.7
51.3
1559.2
38.4
61.6
1556.1+1559.2
22.8
77.2
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 55B

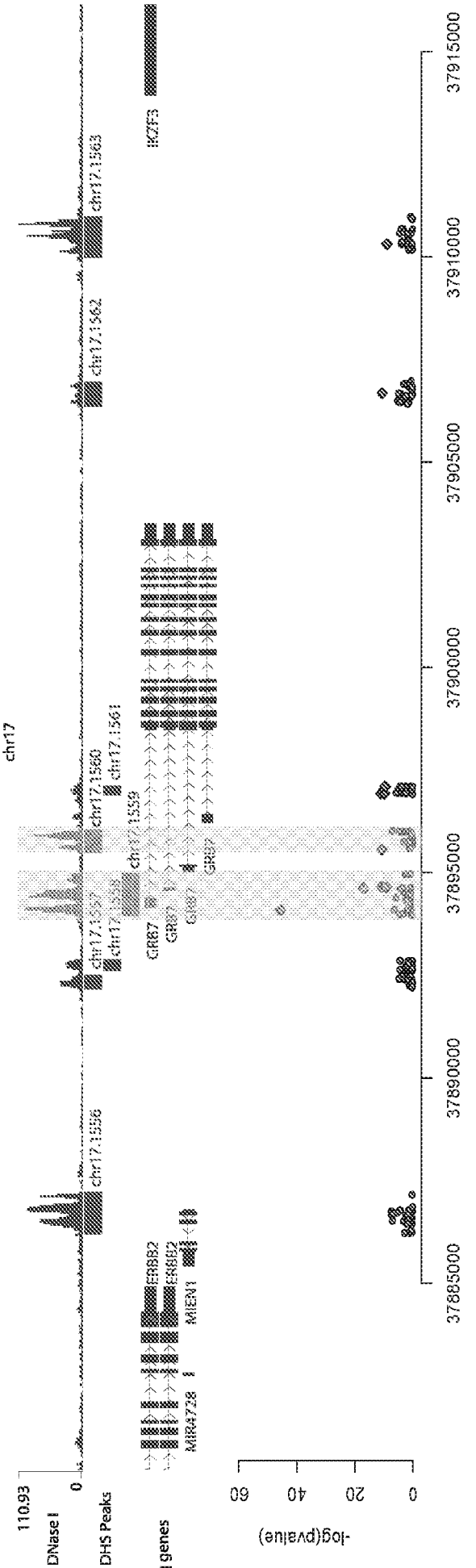

chr17
37882000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915000
chr17:1556
chr17:1557
chr17:1558
chr17:1559
chr17:1560
chr17:1561
chr17:1562
chr17:1563
ERBB2
ERBB2
MIEN1
MIR4728
GRB7
GRB7
GRB7
GRB7
HZF3
SK8R3 DNase I
110.93
0
SK8R3 DHS Peaks
RefSeq genes
-log(pvalue)
60
40
20
0
37885000
37890000
37895000
37900000
37905000
37910000
37915000
Chr17 (bp)

FIG. 56A

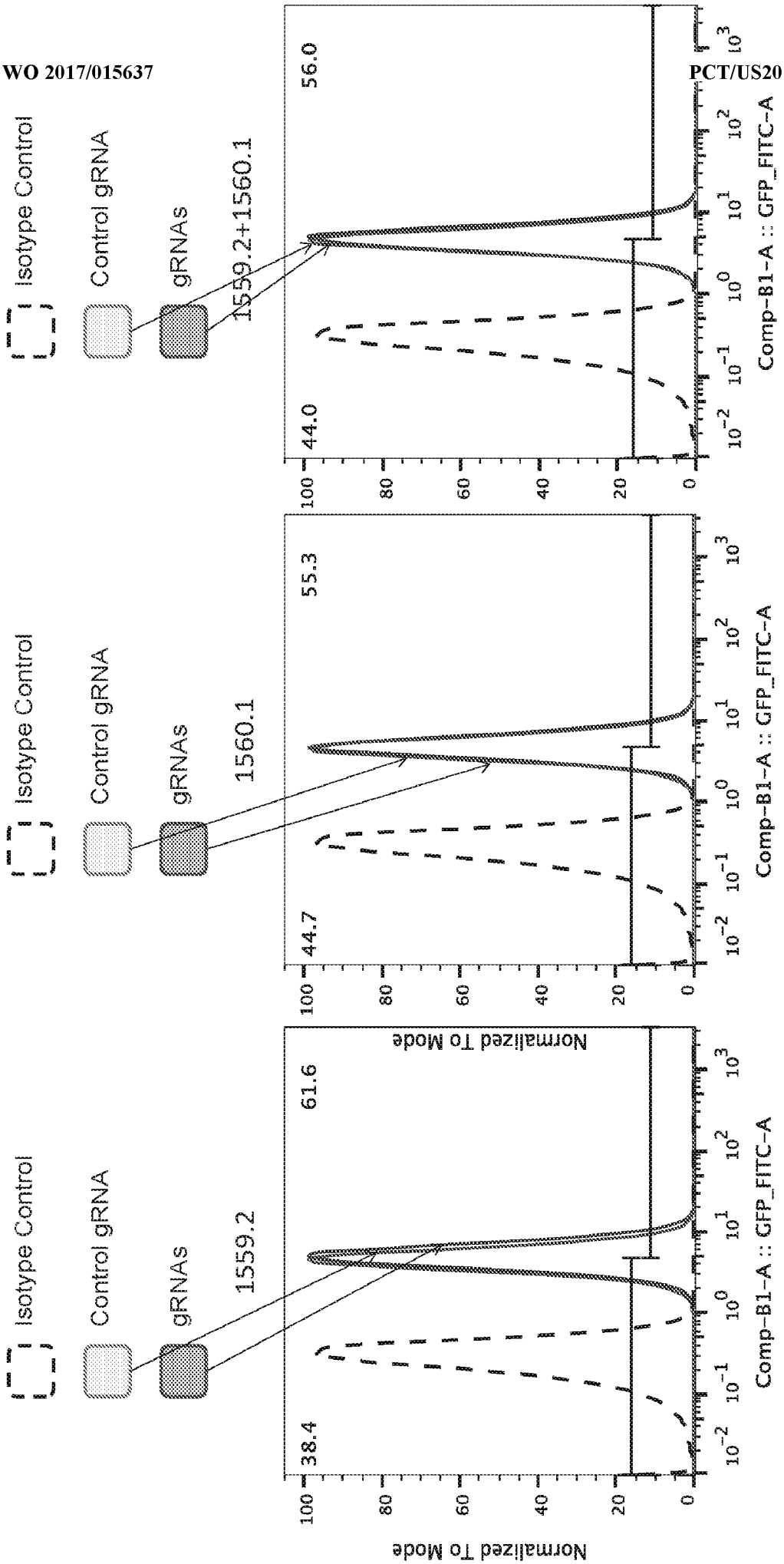

Isotype Control
Control gRNA
gRNAs
1559.2
1560.1
1559.2+1560.1
38.4
61.6
44.7
55.3
44.0
56.0
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 56B

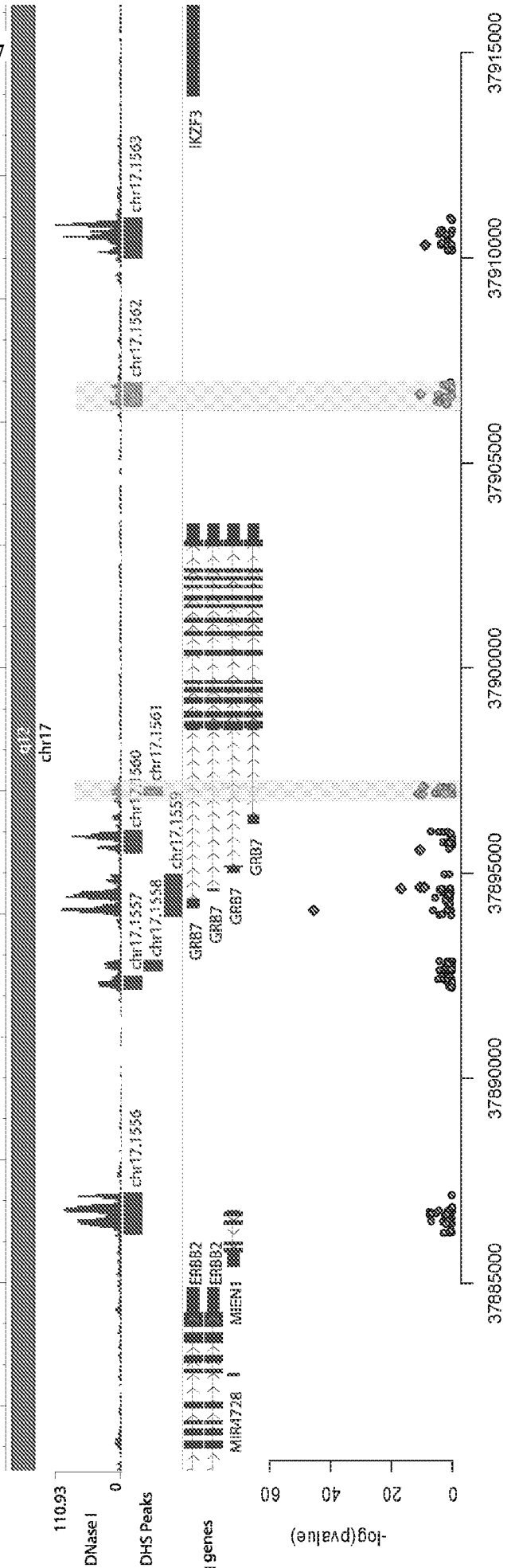

chr17
37882000
37885000
37888000
37891000
37894000
37897000
37900000
37903000
37906000
37909000
37912000
37915000
Chr17 (bp)
SKBR3 DNase I
110.93
0
SKBR3 DHS Peaks
RefSeq genes
-log(pvalue)
0
20
40
60
chr17:1556
chr17:1557
chr17:1558
chr17:1559
chr17:1560
chr17:1561
chr17:1562
chr17:1563
GRB7
ERBB2
MIEN1
MIR4728
IKZF3

FIG. 57A

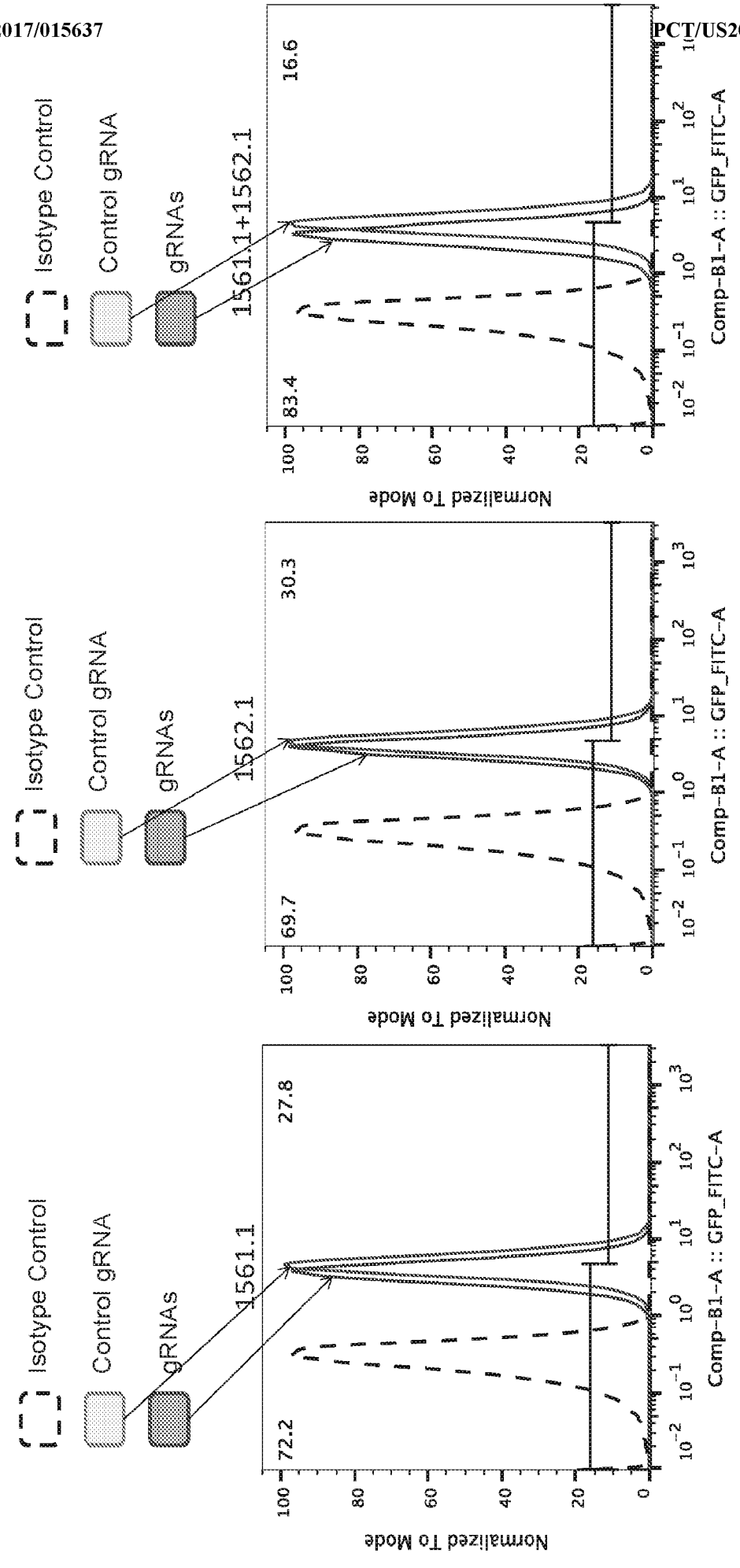

Isotype Control
Control gRNA
gRNAs
1561.1
72.2
27.8
1562.1
69.7
30.3
1561.1+1562.1
83.4
16.6
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 57B

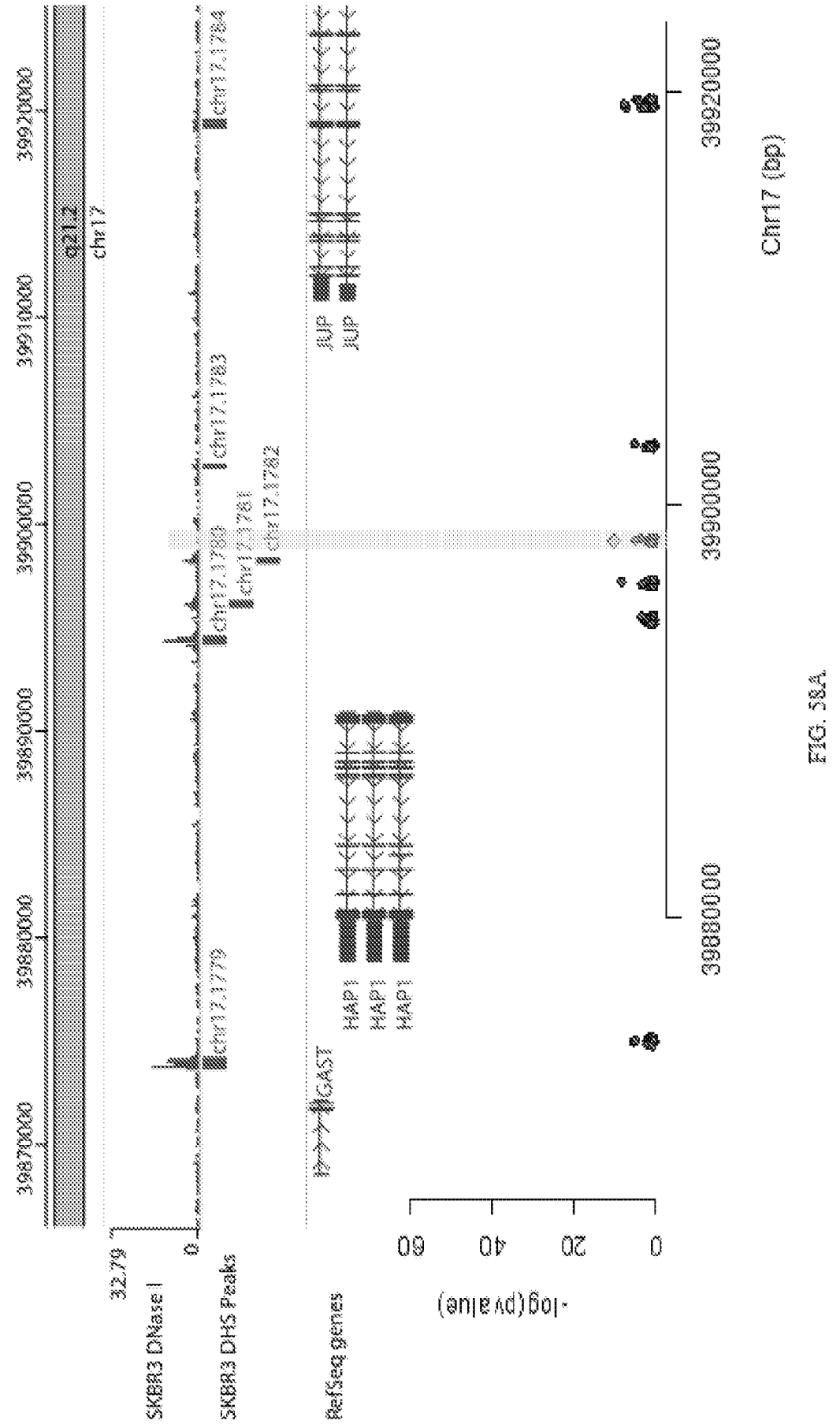

39870000
39880000
39890000
39900000
39910000
39920000
chr17
q21.2
32.79
0
SKBR3 DNase I
SKBR3 DHS Peaks
chr17.1779
chr17.1780
chr17.1781
chr17.1782
chr17.1783
chr17.1784
RefSeq genes
GAST
HAP1
HAP1
HAP1
JUP
JUP
-log(pvalue)
0
20
40
60
39880000
39900000
39920000
Chr17 (bp)

FIG. 58A

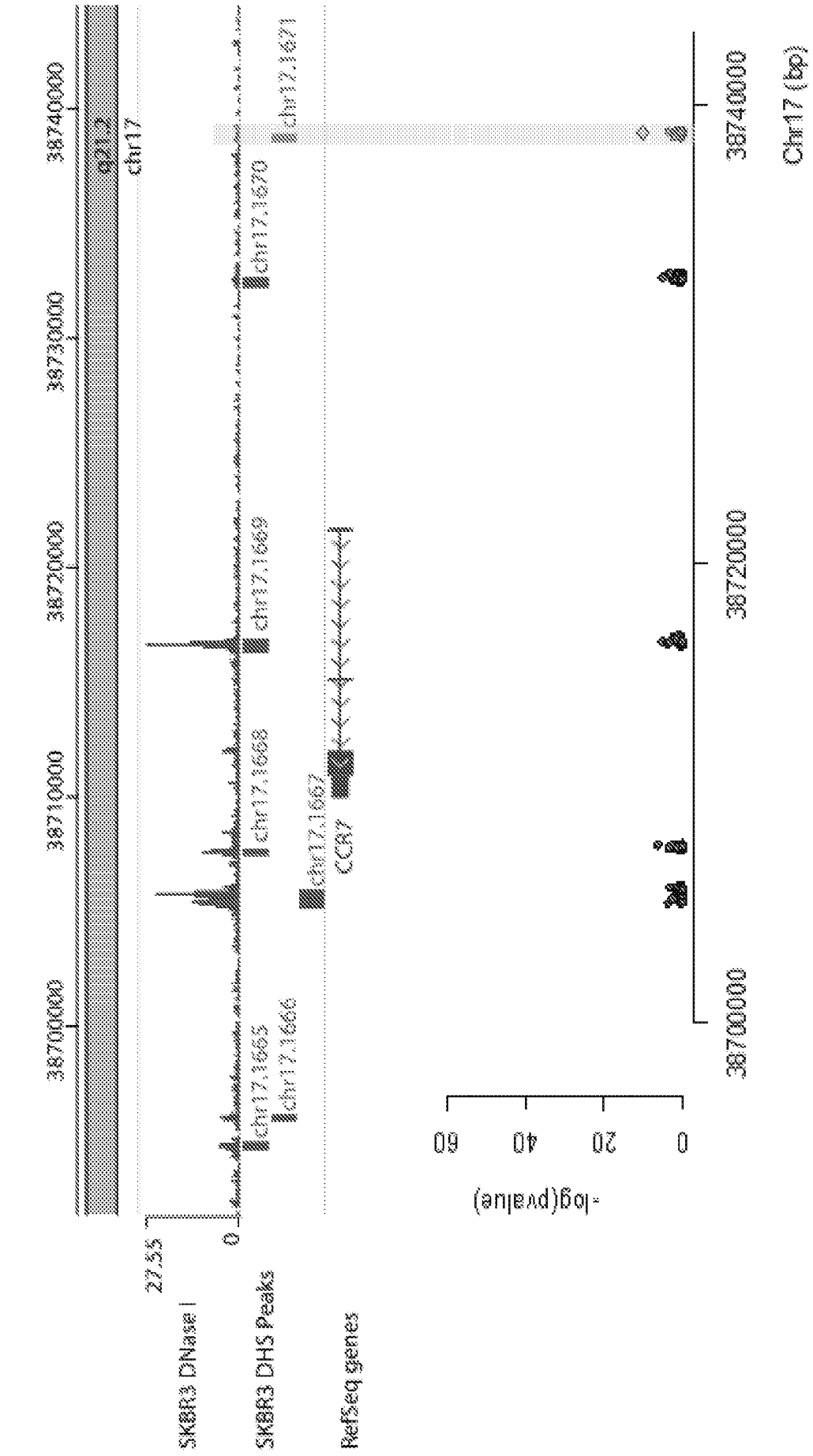

chr17
q21.3
38700000
38710000
38720000
38730000
38740000
Chr17 (bp)
-log(pvalue)
0
20
40
60
SKBR3 DNase I
27.55
0
SKBR3 DHS Peaks
RefSeq genes
chr17.1665
chr17.1666
chr17.1667
chr17.1668
chr17.1669
chr17.1670
chr17.1671
CCR7

FIG. 18B

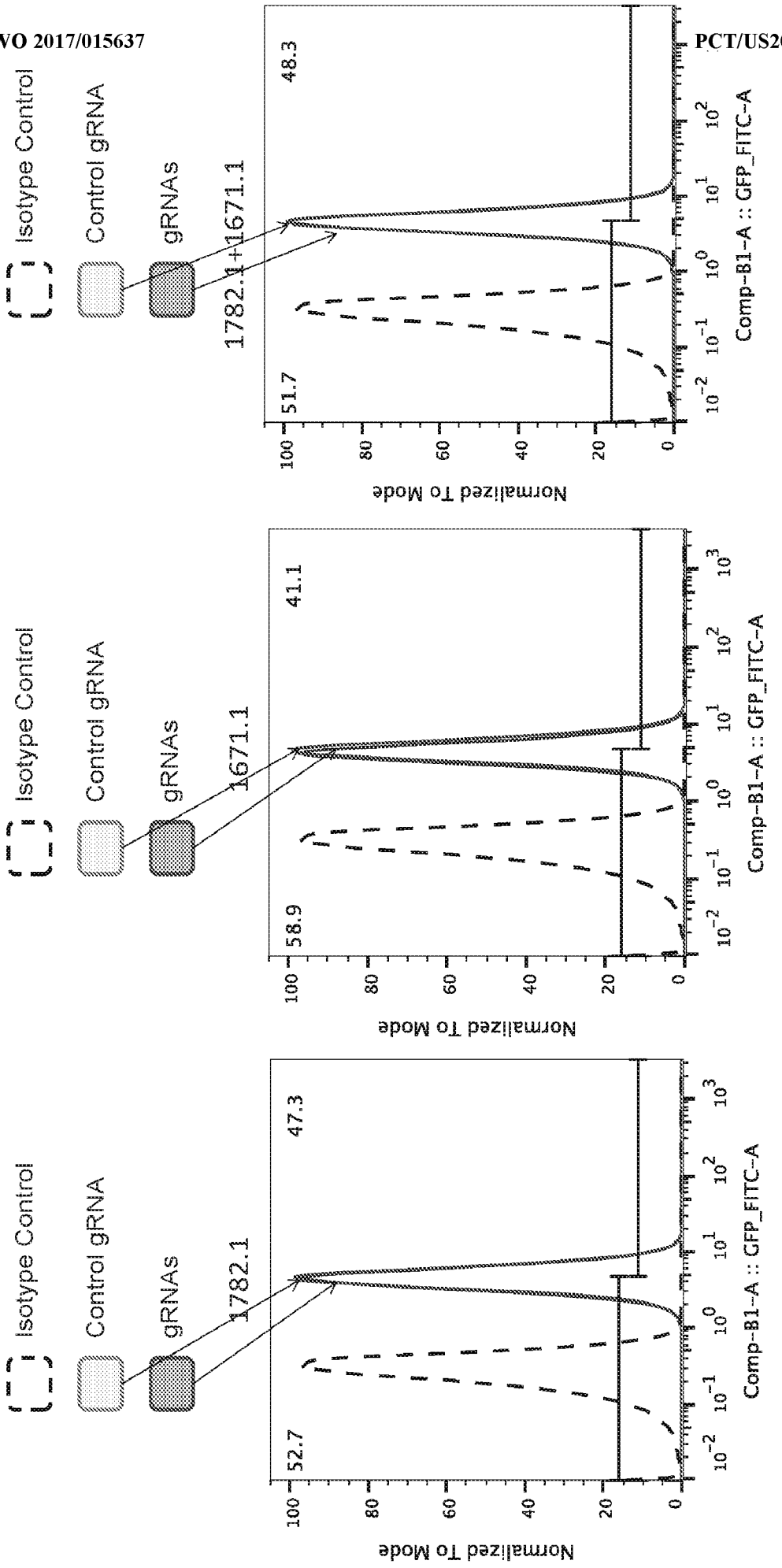

Isotype Control
Control gRNA
gRNAs
1782.1
52.7
47.3
1671.1
58.9
41.1
1782.1+1671.1
51.7
48.3
Normalized To Mode
Comp-B1-A :: GFP_FITC-A

FIG. 58C

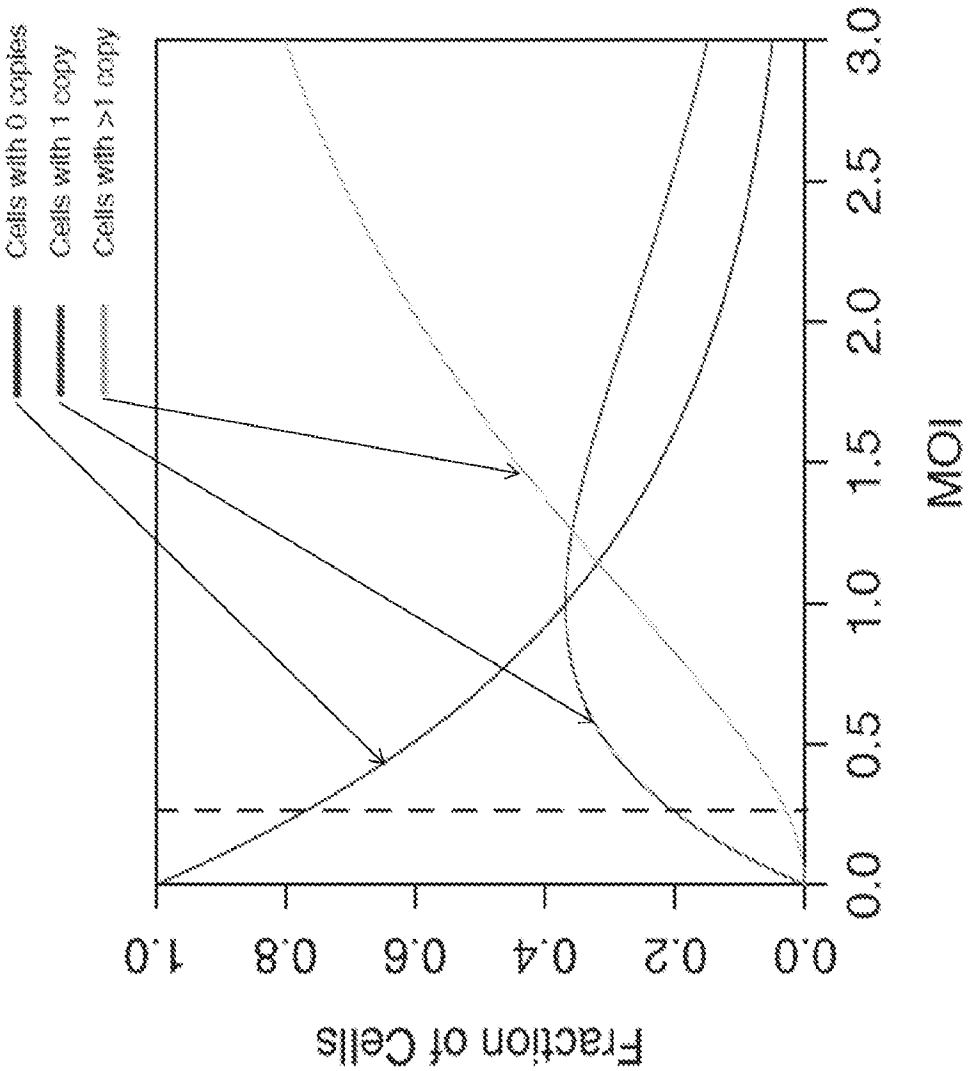

Cells with 0 copies
Cells with 1 copy
Cells with >1 copy
Fraction of Cells
0.0
0.2
0.4
0.6
0.8
1.0
MOI
0.0
0.5
1.0
1.5
2.0
2.5
3.0

FIG. 59

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/043756

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61K 48/00; C07K 19/00; C12N 5/22; C12N 15/63; C12N 15/65; C12N 15/867 (2016.01)
CPC - A61K 48/00; A61K 48/005; C07K 2319/71; C07K 2319/80; C12N 5/06; C12N 9/22 (2016.11)
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)
IPC - A61K 48/00; C07K 19/00; C12N 5/22; C12N 15/63; C12N 15/65; C12N 15/867; C12N 15/90
CPC - A61K 48/00; A61K 48/005; C07K 2319/71; C07K 2319/80; C12N 5/06; C12N 9/22; C12N 15/85; C12N 15/867; C12N 2015/859

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/366; 435/320.1; 435/325; 435/462; 536/23.2 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Orbit, Google Patents, Google Scholar
Search terms used: crispr dcas9 fusion gma library (activator or repressor) high-throughput screening regulatory elements her2 B-globin "red fluorescent protein"

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ZHOU et al. "High-Throughput Screening of a CRISPR/Cas9 Library for Functional Genomics in Human Cells," Nature, 22 May 2014 (22.05.2014), Vol. 509, Pgs. 487-91. entire document	1-31, 33-52
Y	WO 2014/197748 A2 (DUKE UNIVERSITY) 11 December 2014 (11.12.2014) entire document	1-31, 33-52
Y	US 2014/0093913 A1 (SANGAMO BIOSCIENCES, INC.) 03 April 2014 (03.04.2014) entire document	26, 27
Y	WO 2014/172470 A2 (WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH) 23 October 2014 (23.10.2014) entire document	28
A	WO 2014/204723 A1 (THE BROAD INSTITUTE INC. et al) 24 December 2014 (24.12.2014) entire document	1-52
A	HILTON et al. "Epigenome Editing by a CRISPR-Cas9-Based Acetyltransferase Activates Genes from Promoters and Enhancers," Nature Biotechnology, 01 May 2015 (01.05.2015), Vol. 33, Pgs. 510-7. entire document	1-52
A	US 2015/0044772 A1 (SAGE LABS, INC.) 12 February 2015 (12.02.2015) entire document	1-52

☐ 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	Date of mailing of the international search report
09 November 2016	05 DEC 2016
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/043756

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☒ furnished subsequent to the international filing date for the purposes of international search only:

      ☒ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☒ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NOs:2 and 3 were searched.