Single-cell analysis of a mutant library generated using CRISPR-guided deaminase

Soyeong Jun[1, 4], Hyeonseob Lim[1, 4], Ji Hyun Lee[2,3,*], Duhee Bang[1,*]

[1]Department of Chemistry, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

[2]Department of Clinical Pharmacology and Therapeutics, College of Medicine, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Republic of Korea

[3]Department of Biomedical Science and Technology, Kyung Hee Medical Science Research Institute, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Republic of Korea

[4]These authors should be regarded as joint first authors.

*Correspondence should be addressed to D.B. (duheebang@yonsei.ac.kr) or J.H.L. (hyunihyuni@khu.ac.kr)

Abstract

CRISPR-based screening methods using single-cell RNA sequencing (scRNA-seq) technology enable comprehensive profiling of gene perturbations from knock-out mutations. However, evaluating substitution mutations using scRNA-seq is currently limited. We combined CRISPR RNA-guided deaminase and scRNA-seq technology to develop a platform for introducing mutations in multiple genes and simultaneously assessing the mutation-associated perturbations and signatures in a high-throughput manner. Using this platform, we generated a library consisting of 420 sgRNAs, performed sgRNA-tracking analysis, and assessed the effect size of the response to vemurafenib in the melanoma A375 cell line, which has been well-studied via GeCKO but not transcriptome analysis. A convenient and efficient workflow not possible with abundance-based assays enabled the characterization of surviving cells. Our platform permits discrimination of several hit-mutations within a large-scale library by integrating sgRNA hits and gene expression readout. We anticipate that our platform will enable high-throughput analyses of the mechanisms related to a variety of biological events.

Innovative biological research approaches based on the CRISPR/Cas9 system have been developed to facilitate investigations of the functional effects of genetic mutations[1, 2]. Although the functions of thousands of variants have been determined using functional screening approaches based on gene knock-out and activation, screening other genetic features, such as single nucleotide variants (SNVs) and structural variations, to examine related phenotypes remains challenging. Among these genetic features, SNVs are the most frequently observed, and they are associated with many diseases in humans[3]. For example, mutations related to drug resistance and cancer susceptibility significantly impact both the prognosis of therapy and disease progression. In interrogating the function of SNVs, oligo-mediated saturation editing based on homology-directed repair enables base-resolution mutagenesis[4], but this approach is limited to a single locus due to the difficulty of the preparation and size restriction of the oligos composed of various types of SNVs. An alternative approach based on a CRISPR RNA-guided deaminase[5] permits the specific and efficient mutagenesis of C to T without the need for a DNA template. Using this technology, disease-related missense mutations can be generated and analyzed. However, to date, this approach has been only used for multigene knock-out[6] and multiple point mutations within a single gene[7], and introducing and screening multiple point mutations in multiple genes remains challenging.

As the methods described above are based on population analyses and thus dependent on the number of clones, data regarding only significantly dominant clones are obtained when using RIGER[8] and other similar statistical algorithms[9]. However, recently developed techniques such as CROP-seq[10], PERTURB-seq[11] and CRISP-seq[12], which leverage single-cell RNA-seq, enable analysis of the various phenotypes and perturbations with each cell by integrating both the gene expression readout and CRISPR-based perturbations. For the perturbation read out, CROP-seq utilizes a vector that adds a poly(A) tail to the sgRNA transcript, whereas PERTURB-seq and CRISP-seq generate a “perturbation barcode” linked to the sgRNA. The use of these methods has thus far been limited to investigations of perturbations associated with knock-out mutations and transcriptional regulations[13] generated using a CRISPR library at the single-cell level, as it remains challenging to investigate perturbations for point mutations. However, a recent report demonstrated a spectrum of subclonal point mutations in the same tumor has implications for precision medicine and immune therapies[14]. This report emphasizes the need for an atlas of transcription profiles associated with SNVs.

To investigate the myriad of SNVs that affect biological function, a high-throughput, pooled screening method for SNVs which enables analysis of generated SNVs by tracking sgRNAs is required. Here, we demonstrate a novel method combining CRISPR RNA-guided deaminase and CROP-Seq technology that enables the introduction of SNVs in multiple genes and screening of the impact on function in addition to analyses of perturbations in single cells (**Fig. 1a**). To highlight the utility of our novel platform, we generated SNVs in each exon of three human genes (*MAP2K1*, *KRAS*, *NRAS*) associated with resistance to vemurafenib[15, 16], which is a cancer drug targeting the *BRAF* V600E mutation in melanoma patients. We then screened for point mutations that conferred resistance to vemurafenib, analyzed the perturbations in individual resistant clones, and assessed gene expression signatures for individual clones. Using this platform, we could classify resistant clones into two sub-types according to drug response. Moreover, we were able to identify sgRNAs more likely to be hit by combined analysis of candidate sgRNAs based on trackable genomic integrated sequences with transcriptome data. We anticipate that our platform can be extended to characterize individual responses related to biological stimulation in heterogeneous cells harboring single substitution mutations.

RESULTS

Overview of the system

To facilitate the generation and tracking of point mutations, lentivirus- and *piggyBac*-based delivery systems are needed for integration of single-copy sgRNAs in each cell and stable expression of the CRISPR RNA-guided deaminase, respectively. Because complete genes are too long for delivery by lentivirus or transposase in our system, we designed the system to utilize two vectors (**Fig. 1a**): lentiviral and *piggyBac* vectors encoding the sgRNA and CRISPR RNA-guided deaminase, respectively. For sgRNAs, we adopted the CROPseq-Guide-Puro plasmid[10] to enable simultaneous capture of the sgRNA with other mRNAs during the scRNA-seq preparation procedure and puromycin selection of sgRNA-bearing cells. For the CRISPR RNA-guided deaminase, we adopted the BE3 sequence–encoding Cas9 nickase fused with APOBEC and UGI[5]. We also constructed a cassette containing BE3 and a GFP reporter and blasticidine resistance marker, which was cloned into the *piggyBac* vector to provide high “cargo-carrying” capacity.

As BE3 can be used to introduce C to T mutations in the 4th to 8th positions from the PAM-distal end of the protospacer, only a limited variety of mutations can be introduced using the CRISPR RNA-guided deaminase (**Fig. 1b**). Thirteen of the 20 canonical amino acids can be converted to another amino acid by targeted deamination. The remaining seven amino acids can be converted only by either silent or nonsense mutations.

We utilized the above-described system to screen for mutations conferring resistance to vemurafenib to demonstrate that the system is suitable for large-scale screening of SNVs. Although mutations conferring resistance to vemurafenib have been thoroughly studied using approaches such as GeCKO[1], to date, there are no reports of systematic analyses of transcriptional changes induced by various SNVs. We therefore explored how such mutations affect the drug-response mechanism using our novel CRISPR RNA-guided deaminase system.

Characterization of mutations by all possible sgRNAs in the human genome

To investigate the feasibility of our system, we designed all possible sgRNAs (approximately 3.8 million) from all gene isoforms within the human genome (**Methods**, **Supplementary Table 1**). Then, we calculated the breadth of coverage for each gene isoform. These calculations indicated that mutations could be introduced in 41% of residues per gene isoform by targeting the “GG” and “AG” PAMs[17] **(Supplementary Fig. 1a)**. One sgRNA could induce multiple types of mutations depending on the number and position of the cytosine residues in the activity window. In most cases, silent and missense mutations can be produced alone or in combination with sgRNA (**Fig. 1c**). Many loci were covered by more than one sgRNA; therefore, more than one sgRNA can be selected from the candidates for a more robust statistical analysis (**Supplementary Fig. 1b**).

To illuminate the potential relevance of our system to address human cancer mutations, we examined the coverage of the mutations listed in the Cancer Gene Census (CGC)[3]. Our system covered 36,211 missense mutations and 3,491 nonsense mutations. This indicates that a large proportion of known cancer-related mutations can be generated and examined using our system. When considering only the possible types of missense mutations that can result from conversion of the 13 canonical amino acids (**Fig. 1b**), our system covered 56% of the mutations listed in the CGC (**Supplementary Fig. 1c**).

Introduction and enrichment of resistant mutation MAP2K1 E203K

To assess the efficiency of our system, we investigated the rate of conversion of the C base targeted in a single locus. First, we introduced the *MAP2K1* p.E203K mutation (c.G607A) into human A375 melanoma cells carrying a homozygous *BRAF* V600E mutation (c.T1799A). The E203K point mutation results in the substitution of glutamic acid with lysine at codon 203 in exon 6 of *MAP2K1*, resulting in resistance to vemurafenib[15].

The *piggyBac* vector expressing BE3 was transfected into A375 cells to create a cell line that stably expresses BE3. Cells expressing the base editor were infected with a lentivirus expressing sgRNA, targeting codon 203 of *MAP2K1* to induce the E203K mutation. The cells were cultured for 10 days to ensure selection of cells harboring the sgRNA and to promote base editing. Then, frequency of base editing in terms of base and amino acid resolution was investigated by deep sequencing after 25 days of treatment with either vemurafenib or vehicle (dimethyl sulfoxide [DMSO]). At 10 days after puromycin selection, the substitution efficiency ranged from 4.24–5.65% in the “activity window” with a small indel frequency of 1.4% (**Fig. 1d**). After treatment with vemurafenib for 25 days, the substitution efficiency increased significantly, to 41.11–77.17%, whereas DMSO vehicle treatment yielded an efficiency of only 9.67–12.08%. These data suggest that continuous engineering occurred during culture and that the mutation confers resistance to vemurafenib. Therefore, cells harboring the E203K mutation survived over wild-type cells in the presence of vemurafenib, whereas DMSO treatment did not promote enrichment of the mutant clones as much as vemurafenib. Consistent with the base resolution analysis, codon-based analysis indicated that 77.19% of the viable cells were E203K clones (**Fig. 1e**). The same analysis of DMSO-treated cells revealed that 12.06% of the viable cells were E203K clones. These results confirmed that significant enrichment of the mutation occurred with vemurafenib treatment. As shown in a previous study[5], conversion of C residues can occur at a variety of positions, although conversion occurs most frequently within the activity window (**Fig. 1f**, **Supplementary Fig. 2a**). Furthermore, processive deamination tends to occur in same DNA strand (**Supplementary Fig. 2b**). Collectively, the results of our singleplex experiments suggest that the mutations artificially introduced via base editing functioned well and that mutants can be specifically enriched using drugs in a manner similar to naturally acquired mutations in patients. These data also demonstrate the possibility of using this system to measure the effect sizes of various mutations in terms of drug resistance.

Introduction and functional screening of multiple putative vemurafenib-resistance mutations using population analyses

To determine whether the use of our system can be extended to the analysis of multiple loci, we designed a sgRNA library for three genes (*MAP2K1*, *KRAS*, and *NRAS*) related to vemurafenib resistance in melanoma. We selected possible targets based on the criterion that spacers include cytosine residues in the activity window. A total of 420 sgRNAs were designed for all of the exons of *MAP2K1*, *KRAS*, *NRAS* (263, 80, and 77 sgRNAs, respectively, **Supplementary Table 2, Supplementary Fig. 3a**). We excluded the sgRNAs used in the previous singleplex experiments to avoid enrichment of the known E203K resistance clone in the pooled screen. The resulting sgRNA library covered 17.4% of the reported disease-related SNVs in the CGC (**Fig. 2a**). The designed library was then synthesized using a microarray and cloned into the CROP-guide-puro plasmid. A375 cells stably expressing BE3 were transduced in two independent replicates with the library and selected for 14 days using puromycin to ensure that most of the cells expressed sgRNA and were base edited. After selection, the cells were treated with either vehicle (DMSO) or vemurafenib for 28 days. Cells were sampled on the day treatment was started and 7, 14, 21, and 28 days after treatment (hereafter designated initial, D+7, D+14, D+21, and D+28). Before examining the transcriptome of single cells, the population of sgRNAs integrated into the genome was analyzed to determine which sgRNA was responsible for conferring resistance.

First, the distribution of sgRNAs for each condition was investigated. The sgRNA representation decreased over time, meaning that some sgRNAs were enriched over time (**Fig. 2b**, **Supplementary Fig. 3b**). An analysis using MAGeCK[9] indicated that sgRNA #176, which targets close to codon E203 in the *MAP2K1* gene, was enriched in both independent screens (**Fig. 2c**). It should be noted that sgRNAs targeting close to codon E203 (#176, #182) accounted for ~45% of all sgRNAs, and E203K mutant cells constituted over 70% of the total cell population in replicate 1, suggesting that the abundance of sgRNAs close to E203 is reflective of the frequency of E203K mutant cells (**Supplementary Fig. 3c, d**). Although we excluded the highly active sgRNA generating E203K, the data suggested that other sgRNAs in which the protospacer included the E203 residue in the extra position of the activity window robustly generated the E203K mutation. In sgRNA #176, the E203 residue is in the 13th to 15th position from the PAM-distal end of the protospacer, such that the E203K mutation is predominant compared with the mutation in the editing window due to the TC motif preferred by BE[5]. The E203K mutation was also detected in replicate 2, with an allele frequency of ~48% (**Supplementary Fig. 3d, e**). These results demonstrate that generating a library comprised of multiple mutants is possible with our novel system and that appropriate targets can be selected using our platform. However, further analyses using transcriptional data are warranted in order to identify other putative sgRNAs and elucidate the mechanism of enrichment.

Identifying cell subpopulations across different treatment periods using scRNA-seq

The scRNA-seq approach was employed to explore properties related to drug responses in individual cells. To assess the transcriptional changes occurring in each mutant, cells from each experimental condition (initial, DMSO[D+14], Vem[D+14], DMSO[D+28], Vem[D+28]) were individually harvested and subjected to Drop-seq[18]. The mRNAs and sgRNAs in each cell were captured and converted into a cDNA library for NGS. On average, we sequenced 82 million reads per sample (**Supplementary Table 3**). After filtering cells in replicates 1 and 2, 5707 and 7511 cells with more than 500 genes were identified, respectively, and 57.4 and 55.2% of the cells were assigned as sgRNAs, respectively. These results were comparable to data from a previous report[10].

We first compared the abundance of each sgRNA to the abundance determined from genomic DNA obtained from bulk cells. High correlations were observed between the sgRNA and genomic DNA data (**Fig. 3a**). In particular, enrichment of sgRNAs introducing the E203K mutation was observed in drug-treated samples from both analyses, indicating that the transcriptome data were reliable for analyzing the perturbations associated with the assigned sgRNAs.

Next, we performed a principal component analysis (PCA) and trajectory analysis to determine whether there is a characteristic cluster pattern according to period and treatment. The transcriptome of resistant cells at D+28 and D+14 was distinct from that of naïve cells in each replicate (**Supplementary Fig. 4a**). However, no common pattern was observed between replicates (**Supplementary Fig. 4b**). Therefore, we focused on D+28 cells to assess the mutational effects of sgRNAs in more detail.

D+28 cells were visualized using t-SNE and grouped by unbiased clustering (**Fig. 3b, c**). We hypothesized that the transcriptome pattern of cells that acquired resistance due to the E203K mutation would differ significantly from that of natural survivors. Thus, we expected that the cluster including the #176 sgRNA introducing E203K would be separated from other clusters. In the t-SNE visualization, we observed two and three major clusters from replicates 1 and 2, respectively. One cluster in both replicate screens (rep1-1, rep2-1) consisted primarily of #176 sgRNA, which targets close to E203 in *MAP2K1* (**Fig. 3d, Supplementary Fig. 4c**). In replicate 2, an additional cluster composed primarily of #56 sgRNA targeting close to Q61 of *KRAS* (rep2-2) was identified (**Fig. 3d, Supplementary Fig. 4d**). We determined that these clusters consisted of cells that had acquired resistance and therefore investigated them further to elucidate the resistance mechanism in more detail.

Investigation of cluster of resistance-acquired cells

We investigated whether transcriptional changes in the clusters composed primarily of #176 sgRNA targeting close to E203 in *MAP2K1* (rep1-1 and rep2-1) were related to vemurafenib resistance by attempting to identify marker genes. Among candidate up-regulated marker genes identified using the Wilcoxon rank sum test, the top 10 with lowest p-values were selected as “signature genes” (**Supplementary Table 4, Fig. 4a, Supplementary Figs. 5 and 6**). Gene ontology and pathway enrichment analyses[19, 20, 21, 22] of the signature genes showed that the clusters were enriched with gene sets related to immune responses such as antigen processing and presentation of peptides via MHC class II (**Fig. 4b, Supplementary Figs. 7 and 8**). This observation was similar to the results of a previous report[23]. *CD74*, *HLA-DRA*, *SLC26A2*, *HLA-DRB1*, *FOS*, and *HLA-DPA1* were commonly identified as signature genes in the clusters from both replicates (**Fig. 4c**). When we extended the criterion for marker genes to include all listed genes with a p-value <0.05, a total of 66 up-regulated genes and 163 down-regulated genes were identified as common (**Fig. 4c, Supplementary Figs. 9–11, Supplementary Tables 5 and 6**). This result indicates that the members of these clusters are similar and that their perturbations are reproducible. We assume that most members of these clusters are perturbed by the E203K mutation and associated with immune responses.

The rep2-2 cluster was composed primarily of #56 sgRNA, which targets close to Q61 of *KRAS*, whereas representation of #176 sgRNA was low (**Fig. 3c**). As indicated above, the top 10 genes with the lowest p-values were selected as signature genes, and then ontology and pathway enrichment analyses were performed. The signature genes of this cluster (i.e., rep2-2) differed completely from those of the rep1-1 and rep2-1 clusters (**Fig. 4c,**

Supplementary Fig. 11). Although their expression was not as significant compared with rep1-1 and rep2-1 (**Supplementary Fig. 12**), the signature genes were partially enriched in gene ontology (GO) terms and pathways associated with chemokine signaling (**Fig 4b**). It has been reported that activated CXC chemokine receptor (CXCR) signaling in melanoma cells contributes to vemurafenib resistance[24, 25]. Our results indicate that the transcriptional changes in the rep2-2 cluster are distinct from those of the rep1-1 and rep2-1 clusters. We hypothesize that the cells in the rep2-2 cluster composed primarily of #56 sgRNA survive via a mechanism different from that by which cells in the rep1-1 and rep2-1 clusters mainly composed of #176 sgRNA survive.

We next investigated whether other sgRNAs are common to rep1-1, rep2-1, and rep2-2. We identified nine, eight, and five sgRNAs with more than two cells in rep1-1, rep2-1, and rep2-2, respectively (**Supplementary Table 8**). Of these sgRNAs, two (#176, #182) that could introduce E203K in *MAP2K1* were common in these clusters, and the #56, #126, and #217 sgRNAs account partially or almost completely for the rep2-1 and rep2-2 clusters (**Supplementary Fig. 13**). Other sgRNAs were minor components (1.4% on average). Mutations that can be introduced by these sgRNAs (#56, #126, and #217) are thus potential candidates for conferring resistance to vemurafenib. We next examined whether these sgRNAs introduce cognate mutations.

Validation of genomic loci targeted by candidate sgRNAs via deep sequencing

We performed deep sequencing of genomic loci targeted by the candidate sgRNAs to verify the introduced mutations. Genomic loci of the #56, #126, and #217 sgRNAs were sequenced. An indel mutation (7.7%) was introduced in the target region of #56 sgRNA (**Supplementary Fig. 14a**). Because we used BE3, which employs Cas9 nickase, the indel mutation could be introduced into a small proportion of cells. There were two main types of in-frame indel mutations introduced (c.171insCTGTTGGATATTCTCGAC and c.176delCAG), and no substitutions were observed (**Supplementary Fig. 14b, c**). Enrichment in the indel mutations was observed compared to control cells treated with DMSO (0.03%). These mutations were not reported in previous studies, but region in which the mutations were introduced is next to the sequence encoding the Q61 residue of K-Ras, which is involved in constitutive activation of intrinsic GTPase activity[26]. Activated Ras is known to positively regulate the expression of various chemokines and ultimately promote tumorigenesis[27, 28], which is consistent with the present result demonstrating that chemokine signaling pathways and CXCR binding genes were enriched in clusters composed primarily of the #56 sgRNA. We concluded that cells with the #56 sgRNA that introduced indel mutations next to the *KRAS* Q61-encoding sequence conferred resistance to vemurafenib. In contrast, deep sequencing of the target regions of the #126 and #217 sgRNAs did not confirm introduction of cognate mutations (**Supplementary Fig. 15**), presumably because the cells with these sgRNAs were natural survivors.

In summary, we first observed enrichment of the #176 sgRNA in sgRNA abundance analyses and considered the #56, #126, and #217 sgRNAs as additional candidates as a result of transcriptome analyses. Of these selected sgRNAs, those targeting close to the sequences encoding E203 of the *MAP2K1* gene and Q61 of the *KRAS* gene (#176, #56) generated cognate mutations and induced different transcriptional changes.

Discussion

In the present study, we established a platform combining CRISPR RNA-guided deaminase and scRNA-seq technologies that enables the introduction of SNVs into multiple genes and facilitates their functional screening and measurement of perturbations in single cells. We demonstrated the introduction of SNVs into each exon of three genes and screened for SNVs conferring resistance to vemurafenib using population analysis. The results of population analyses indicated enrichment of the E203K SNV in *MAP2K1*, consistent with a previous study[15]. In addition, by employing scRNA-seq technology, we identified the perturbation as well as the signature of SNVs at the transcriptome level.

A recent study using targeted AID reported the introduction of multiple SNVs and assessed the effect of the introduced SNVs on the mechanism of imatinib resistance[7]. The introduced SNVs were confirmed and validated by amplification of targeted exons using genomic DNA. However, this technology was not expanded to include multiple genes or multiple exons. Our platform was expanded to include all exons in three genes, which illustrates that the confirmation procedure is not labor intensive. Unlike previous approaches that require amplification of individual targeted regions, we utilized a lentivirus-expressing sgRNA that can be tracked by amplification of the integrated sgRNA. Using this sgRNA approach, we could narrow the scope and amplify only the sgRNA-targeted region and thus confirm which mutation was introduced.

Improved engineering efficiency could provide for more-accurate assessment of the perturbations associated with individual sgRNAs. A knock-out–based study reported that approximately 70% of sgRNA-bearing cells can be edited using sgRNA[1]. In contrast, experiments using BE3 indicated that 5–20% of sgRNA-bearing cells can be edited using sgRNAs[5]. These data suggest the possibility of discordant transcriptomic changes resulting from use of the same sgRNA, which could create confusion in analysis. Optimization of the engineering efficiency can be achieved using BE4[29] or BE4max[30]. By achieving a higher efficiency in base editing (i.e., increasing the likelihood that sgRNA-bearing cells will be engineered), perturbations in a larger number of cells can be characterized.

We envisage that our approach could be optimized further by employing an alternative to BE3 protein. Recently developed BE variants[31] and other effector proteins[32] provide a broader range of targets, thus increasing the analytical coverage of diseases related to SNVs. In addition, broadening of the editing range by using eA3A-BE3 would permit narrowing to control for off-target effects[33]. We believe that more comprehensive analyses of mutations present could be achieved by optimizing protein selection. We used oligo-dT sequence–linked beads to capture polyadenylated mRNAs in the scRNA-seq experiment. By synthesizing target sequence–linked beads, targeted sequences could be captured directly without laborious PCR amplification of each targeted sequence. Alternatively, ligating target sequences to conventional beads could also be used[34]. Another application involves coupling with methods targeting individual cells using barcoding[35]. A target cell's cDNA can be amplified using this method in order to analyze the transcriptome, and the method is applicable to the study of cells with known mutations and validation of cells enriched with specific sgRNAs. Furthermore, although we demonstrated the effect of SNVs introduced into exons of three genes in the present study, further exploration of other exons in genes related to cancer[3] is expected to accelerate the identification of cancer-driving and drug-resistance mutations. We expect our platform to be extended to the examination of substitution mutations related to a variety of biological responses.

Methods

piggyBac-BE3-GFP: Construction and establishment of BE3-expressing cells

The *piggyBac*-BE3-GFP plasmid was constructed by insertion of the BE3-blasticidin fragment into plasmid PB-CA (Addgene #20960). The BE3-blasticidin fragment was obtained by assembly of the BE3 fragment amplified from the pCMV-BE3 vector (Addgene #73021) and the blasticidin fragment amplified from lentiCas9-Blast (Addgene #52962). To establish a line of BE3-expressing cells, we co-transfected A375 cells with transposase (pPbase, Sanger Institute, UK) and the *piggyBac*-BE3-GFP plasmid at a 1:4 molar ratio using Lipofectamine 3000 (Invitrogen). At 48 h after transfection, the cells were cultured in medium containing 5 μg/ml of blasticidin for 14 days to enrich for BE3-expressing cells.

sgRNA design

All sgRNAs were designed using an in-house program. We searched for all “GGs” and “AGs” as PAM sequences and considered sgRNA sequences in the form 5’-N_{20}-PAM-3’ as candidates in the CDS region of every gene isoform. The candidates were examined to determine whether at least one “C” was in the activity window (4th to 8th position from the PAM-distal end of the protospacer) and that a “C” was in the CDS region. All possible mutations in the activity window were calculated, and the associated amino acid changes were classified as either silent, nonsense, or missense by reference to the codon frame and strand of the gene. All information regarding region and frame were obtained from a GTF file of hg19 downloaded from the UCSC Table Browser.

Cell culture

A375 (ATCC) and HEK293T (ATCC) cells were maintained in RPMI medium (for A375 cells) or DMEM (for HEK293T cells) (Gibco) supplemented with 10% fetal bovine serum (Gibco) and penicillin/streptomycin (Gibco) in a humidified 5% CO_2 incubator at 37°C. Solutions of vemurafenib (Selleckchem) were prepared by dissolving the colorless powder in DMSO (BioReagent grade; Sigma-Aldrich). DMSO was used as a vehicle control. To analyze the response to vemurafenib, cells were treated with either DMSO alone or 2 μM vemurafenib in DMSO.

Generation of sgRNA library plasmid, virus, and infection

Designed sgRNA sequences were synthesized using programmable microarrays (CustomArray, USA). Sequences are listed in **Supplementary Table 2**. Oligos were cleaved from the microarray and PCR amplified using chip_fwd and chip_rev. PCR cycling was performed as follows: 95°C for 3 min, 25 cycles of 98°C for 20 s, 56°C for 15 s, 72°C for 30 s, and 72°C for 1 min. To ensure high-yield coverage of the PCR products, eight repeats of the PCR reaction were conducted. The second PCR was performed using the chip_2nd_fwd and chip_2nd_rev primers. The PCR cycling conditions were 95°C for 3 min, followed by 6 cycles of 98°C for 20 s, 58°C for 15 s, 72°C for 30 s, and 1 cycle of 72°C for 1 min. The primers used in these steps are listed in **Supplementary Table 9**. PCR products were purified using 2.0× (by volume) AMPure XP beads (Beckman Coulter). The purified amplicons were cloned into *Bsm*BI (NEB)-digested CROPseq-Guide-Puro plasmids (Addgene #86708) by Gibson assembly, as described previously[1].

To ensure high-yield coverage, four repeats of the Gibson reactions were performed. The products of the Gibson reactions were combined and electroporated into Endura cells (Lucigen) following the manufacturer's protocol. A 1000-fold dilution of the full transformation was spread on carbenicillin (50 μg/ml) LB agar plates to determine the library coverage, and the remainder of the culture was incubated overnight in 200 ml of LB medium. By counting the number of colonies on the plate, >300× library coverage was ensured. The plasmid library was then extracted using an EndoFree Plasmid Maxi kit (Qiagen). The lentivirus was then generated by co-transfecting the plasmid library, psPAX2 (Addgene #12260), and pMD2.G (Addgene #12259) into HEK293T cells using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol.

A375 cells were transduced with the lentivirus at a multiplicity of infection of 0.1–0.3 in each of two independent biological replicates. The cells were selected by culturing in medium containing either 1 μg/ml of puromycin for 14 days or DMSO or 2 μM vemurafenib for 28 days.

Deep sequencing and analysis of sgRNA-integrated regions

To determine the frequency of each sgRNA, sgRNA sequences integrated into the genome were PCR amplified using the primers listed in **Supplementary Table 9**. The resulting amplicons were adaptor-ligated using a SPARK kit (Enzymatics) and deep sequenced using NextSeq 500. Raw reads were quality trimmed via trimmomatic v0.33[36] using the following parameters: LEADING: 20, TRAILING: 20, SLIDINGWINDOW: 150:25, MINLEN: 36.

Reads containing each sgRNA spacer were counted, and these sgRNA read-counts were used as input for the MAGeCK software package to identify hits by comparing DMSO-treated and vemurafenib-treated cells. The p-value of every sgRNA was calculated based on the negative binomial model of read counts. Log_2(fold-change) was calculated as log 2 ratio of the normalized sgRNA count of vemurafenib-treated cells to that of DMSO-treated cells. Normalized sgRNA counts were calculated as reported previously[1].

Deep sequencing of genomic DNA samples and C to T substitution efficiency

To validate whether the genome was edited by specific sgRNAs, the targeted region was amplified from the genome and sequenced. The primers used in these amplifications are listed in **Supplementary Table 9**.

The PCR cycle conditions were as follows: 1 cycle at 95°C for 3 min, followed by 27 cycles of 98°C for 20 s, 56°C for 15 s, 72°C for 30 s, and 1 cycle of 72°C for 1 min. The resulting amplicons were adaptor-ligated and deep sequenced using NextSeq 500. If more than 50% of the bases that had a quality score lower than Q30, the sequenced reads were discarded, and base calls with a quality score below Q30 were converted to N. Ten-bp flanking sequences on both sides were used to identify the protospacer region for the targeted sgRNA. Protospacer reads that were not 20 bp in length were considered reads with indels. Base frequencies for each locus were calculated across the reads without indels. The plots are shown with the protein-coding strand, and if the protein-coding strand was the non-target strand of the sgRNA, G to A conversion efficiency was calculated (**Supplementary Fig. 2a, 3c, 3e, 15**).

Substitution efficiencies were calculated using the following equation:

$$\frac{\text{Reads with C to T substitution}}{\text{Reads covering protospacer region}} \times (1 - \text{fraction of reads with indels}).$$

scRNA-seq

A375 cells were harvested and divided into two pools on each sampling day. One pool was used for extraction of genomic DNA, and the other pool was methanol fixed for Drop-seq analysis. Methanol fixation was performed as described previously[37]. Briefly, cells were centrifuged at 300*g* for 5 min and resuspended in 80% methanol (BioReagent grade; Sigma-Aldrich) and 20% PBS (Gibco). The resuspended cells were incubated on ice for at least 15 min and then stored at −80°C. On the day of the Drop-seq experiment, cells were recovered by centrifugation at 2000*g* for 5 min, washed once with PBS-0.01% BSA, and resuspended in PBS-0.01% BSA to a concentration of 100 cells/µl. Finally, Drop-seq was performed as described previously[18].

Droplets were collected into 50-ml conical tubes over a 15 min time period. After which, the droplets were broken, and the RNAs on beads were subjected to reverse transcription.

Aliquots of 2,000 beads were amplified in each tube using the following PCR steps: 95°C for 3 min, then four cycles of 98°C for 20 s, 65°C for 45 s, 72°C for 3 min, and 11 cycles of 98°C for 20 s, 67°C for 20 s, 72°C for 3 min, and 1 cycle of 72°C for 5 min. The amplified products were purified with 0.6× Ampure XP beads and fragmented, tagged, and amplified using a Nextera XT DNA Library Preparation kit (Illumina).

Preprocessing of single-cell transcriptome data

The pipeline was designed based on Drop-seq tools (ver.1.12) and CROP-seq software. Raw data were converted to bam files via FastqToSam in Picard. Cell barcodes and UMI for each mRNA were obtained from Read1. mRNA sequences obtained from Read2 were modified by FilterBAM, TrimStartingSequence, and PolyATrimmer in Drop-seq tools. The modified bam files were then converted to fastq files and aligned to the reference comprised of hg19 and BE genes and guide RNA sequences using STAR aligner. The aligned data were sorted via SortSam and merged with tags via MergeBamAlignment in Picard. Exon information was annotated using TagReadWithGeneExon in Drop-seq tools, and bead synthesis errors within a hamming distance of <4 were corrected via DetectBeadSynthesisErrors in Drop-seq tools. For assignment of sgRNAs per each cell, UMI counts of each sgRNA in the same cell were determined, and the most abundant sgRNA was selected. Finally, transcript data with more than 500 genes per cell were selected and converted to a digital gene expression matrix, and all matrices were merged and indexed by the cell barcode, condition, replicate, sgRNA, and gene. All program runs were managed using CROP-seq software[10].

Transcriptome analysis

Analyses were carried out based on the modules in Crop-seq software, Seurat[38], and Monocle 2[39].

Matrix modification. Merged matrices of digital gene expression were normalized per cell, pseudo count 1 was added, multiplied by 10,000, and the matrix was finally $\log_2$ transformed. After normalization, ribosomal and mitochondrial genes and pseudo references of sgRNAs and BE were filtered.

Dimensional reduction. First, PCA was performed using a PCA module in the sklearn package for global clustering. For characterization of D+28 samples, t-SNE was performed using Seurat software from the original matrix before normalization. Briefly, cells from the D+28 samples treated with vemurafenib or DMSO with the appropriate fraction of mitochondrial genes (<0.065) were selected, and gene expression counts were normalized by multiplying by 10,000. Highly variable genes were selected using the Seurat clustering algorithm, and the number of PCs was determined using PCElbowPlot and JackStaw. PCA was performed on selected variable genes using determined PCs.

Gene set enrichment analysis. The Wilcoxon rank-sum test of the FindMarkers function in Seurat was used to assess differences in gene expression. Marker genes with low p-values (Benjamini -Hochberg[40] corrected p<0.05) for each cluster were obtained. GO and KEGG pathway enrichment analyses were carried out on the signature genes via clusterProfiler[19] and Enrichr[20]. Multiple testing corrections using the Benjamini and Hochberg method with both p-value threshold and false discovery rate set to 0.05 were carried out.

Trajectory Analysis. Trajectory analysis was performed using the expression matrix obtained from Crop-seq software. Normalization and filtering were performed with default parameters, and ordered genes were obtained using the differentialGeneTest module by setting the fullModelFormulaStr option as a condition. Finally, DDRTree-based dimensional reduction and ordering were performed.

Reference

1. Shalem O*, et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).

2. Konermann S*, et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015).

3. Forbes SA*, et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* **45**, D777-D783 (2017).

4. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120-123 (2014).

5. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).

6. Kuscu C*, et al.* CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat Methods* **14**, 710-712 (2017).

7. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods* **13**, 1029-1035 (2016).

8. Luo B*, et al.* Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A* **105**, 20380-20385 (2008).

9. Li W*, et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).

10. Datlinger P*, et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297-301 (2017).

11. Dixit A*, et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866 e1817 (2016).

12. Jaitin DA*, et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896 e1815 (2016).

13. Adamson B*, et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821 (2016).

14. van Galen P*, et al.* Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281 e1224 (2019).

15. Trunzer K*, et al.* Pharmacodynamic effects and mechanisms of resistance to vemurafenib in patients with metastatic melanoma. *J Clin Oncol* **31**, 1767-1774 (2013).

16. Shi H*, et al.* Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov* **4**, 80-93 (2014).

17. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233-239 (2013).

18. Macosko EZ*, et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).

19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).

20. Chen EY*, et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

21. Ashburner M*, et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

22. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

23. Yu Y*, et al.* Bioinformatics analysis of gene expression alterations conferring drug resistance in tumor samples from melanoma patients with EGFR-activating BRAF mutations. *Oncol Lett* **15**, 635-641 (2018).

24. Vergani E*, et al.* Overcoming melanoma resistance to vemurafenib by targeting CCL2-induced miR-34a, miR-100 and miR-125b. *Oncotarget* **7**, 4428-4441 (2016).

25. Arozarena I, Wellbrock C. Overcoming resistance to BRAF inhibitors. *Ann Transl Med* **5**, 387 (2017).

26. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res* **72**, 2457-2467 (2012).

27. Stolze B, Reinhart S, Bulllinger L, Frohling S, Scholl C. Comparative analysis of KRAS codon 12, 13, 18, 61, and 117 mutations using human MCF10A isogenic cell lines. *Sci Rep* **5**, 8535 (2015).

28. Ancrile BB, O'Hayer KM, Counter CM. Oncogenic ras-induced expression of cytokines: a new target of anti-cancer therapeutics. *Mol Interv* **8**, 22-27 (2008).

29. Komor AC*, et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, eaao4774 (2017).

30. Koblan LW*, et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* **36**, 843-846 (2018).

31. Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376 (2017).

32. Gaudelli NM*, et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).

33. Gehrke JM*, et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol* **36**, 977-982 (2018).

34. Saikia M*, et al.* Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *Nat Methods* **16**, 59-62 (2019).

35. Ranu N, Villani AC, Hacohen N, Blainey PC. Targeting individual cells by barcode in pooled sequence libraries. *Nucleic Acids Res*, (2018).

36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

37. Alles J*, et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol* **15**, 44 (2017).

38. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).

39. Qiu X*, et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).

40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**, 289-300 (1995).

Data Availability

Raw sequencing data from this study are deposited in Short Read Archive (SRA) under project number PRJNA530348.

Acknowledgements

This work was supported by: (i) the Mid-career Researcher Program (NRF-2018R1A2A1A05079172), (ii) the Bio & Medical Technology Development Program (NRF-2016M3A9B6948494), (iii) the Bio & Medical Technology Development Program (NRF-2018M3A9H3024850), and (iv) Basic Science Research Program (NRF-2018R1A2B2001322) of the National Research Foundation of Korea, funded by the Ministry of Science, ICT & Planning.

Author contributions

S.J. and H.L. should be regarded as joint first authors.

D.B. and J.H.L. conceived, designed, managed, and supervised the study. S.J. and H.L. designed and performed the experiments, analyzed the data, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Figure 1 | Workflow of our platform and demonstration of method feasibility. (**a**) Schematic flow of our platform. (**b**) Types of introducible mutations. Thirteen of the canonical amino acids can be converted to another amino acids. (**c**) Fraction of each introducible type of mutation. (**d**) Targeted cytosine to thymine substitution efficiency in the "activity window" of the base editor. Solid line represents the median, with the upper and lower limits of each box corresponding to the 75% and 25% quantiles of the data. (**e**) Allele frequency of the E203K mutation before and 25 days after drug treatment. (**f**) Efficiency of cytosine substitution according to position in the protospacer of the sgRNA.

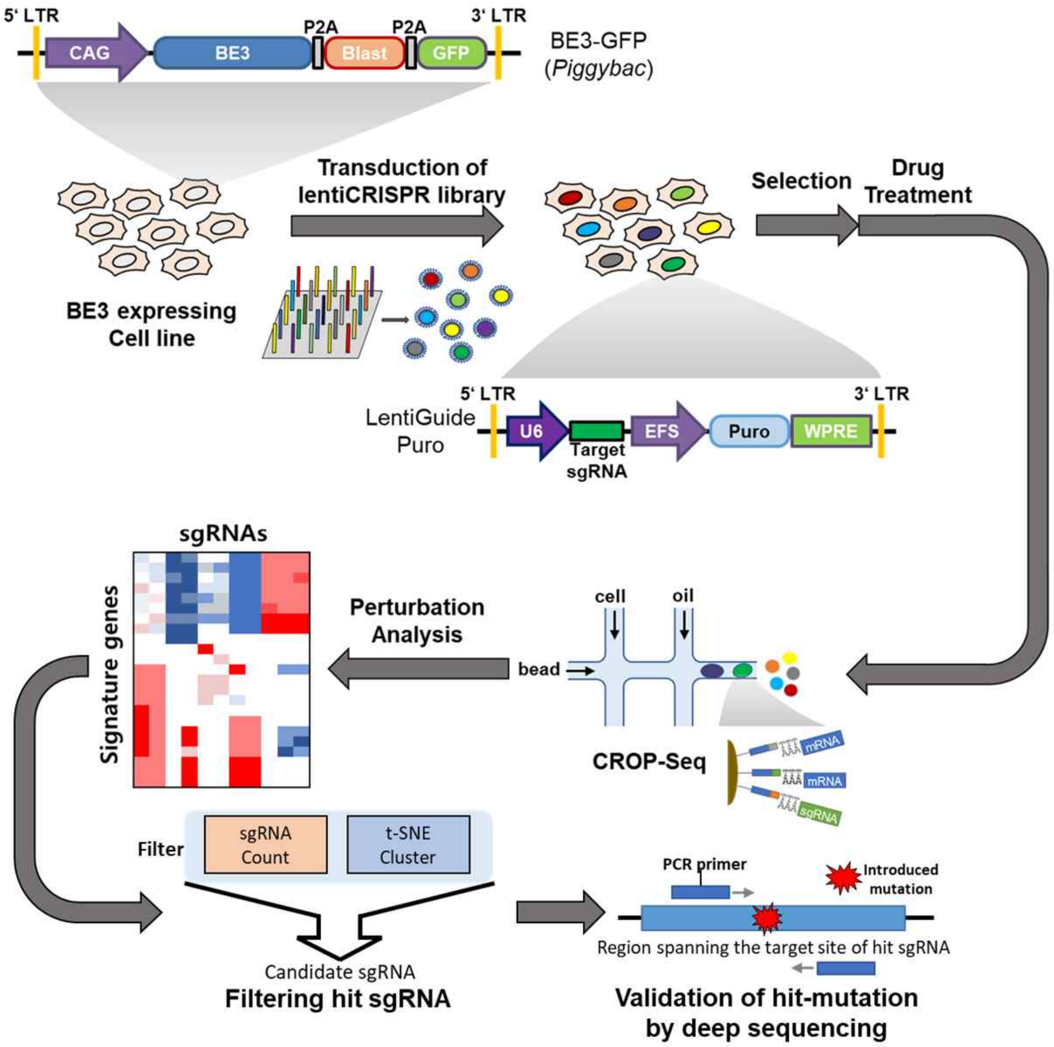
Figure 2 | Introduction and functional screening of multiple mutations using population analysis. (**a**) Introducible disease-related SNVs in a sgRNA library targeting *MAP2K1*, *KRAS*, and *NRAS*. Introducible silent, missense, nonsense, and CGC mutations are shown in red, green, blue, and yellow, respectively, according to residue position in the protein. Coverage of CGC mutations is indicated by a separate color (maximum coverage = 100%, indicated by dark yellow). (**b**) Box plot showing the distribution of reads from individual sgRNAs by time (initial, D+7, D+14, D+21, D+28) and condition (DMSO, vemurafenib [Vem]) in replicate 1. All p-values determined by Student's t-test, ****p < 0.0001. ns: not significant (p > 0.05). (**c**) Scatter plot showing enrichment of sgRNAs with their corresponding p-values. #176 sgRNAs targeting close to E203 of MAP2K1 are shown as red dots. P values less than 10^{-300} were arbitrarily fixed to 10^{-300} for easy comparison of the sgRNAs.

Figure 3 | Analysis of abundance and clusters in scRNA-seq data. (**a**) Concordance of sgRNA abundance from Drop-seq data and genomic data for each replication and condition. The #176 sgRNA, which potentially introduces a known mutation (E203K), is marked by a red circle, and the Pearson correlation coefficient is shown in the upper right corner of the plot. (**b**) t-SNE visualization of D+28 cells (left) and distribution of #176 sgRNA-containing cells (right) in replicate 1. (**c**) t-SNE visualization of D+28 cells (left), distribution of #176 and #56 sgRNA-containing cells (right) in replicate 2. (**d**) Distribution of sgRNAs for rep1-1, rep2-1, and rep2-2.

Figure 4 | Transcription profiles of signature genes from clusters composed primarily of #176 sgRNA. (**a**) Average expression of signature genes according to individual sgRNA. Each row represents one of the signature genes, and each column represents sgRNA-bearing cells. Vem(D+28) cells in other clusters and Vem(D+28) cells in rep1-1 are arranged from left to right. Color scale indicates standard deviation of gene expression from the mean expression value, with red indicating high expression and blue indicating low expression level. (**b**) Gene ontology and pathway enrichment analyses of marker genes. Partial list of the significantly enriched GO Molecular Function (top) and KEGG Pathways (bottom). (**c**) Venn diagram of signature genes from rep1-1, rep2-1, and rep2-2.

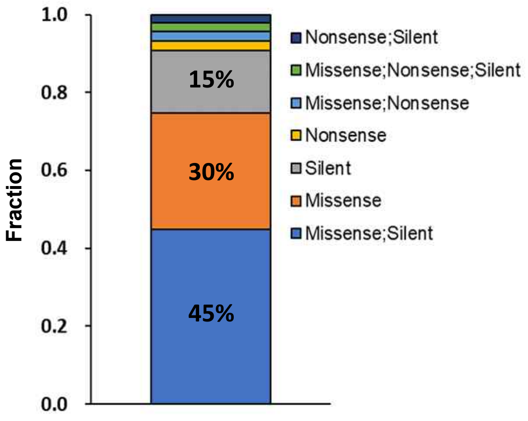
Figure 1

a



b

Wild Type	Introducible Mutation
A	T, V
C	Y
E	K
D	N
G	E, D, K, N, S, R
H	Y
M	I
L	F
P	S, L, F
S	F, L, N
R	Q, H, K, C, W
T	I, M
V	I, M



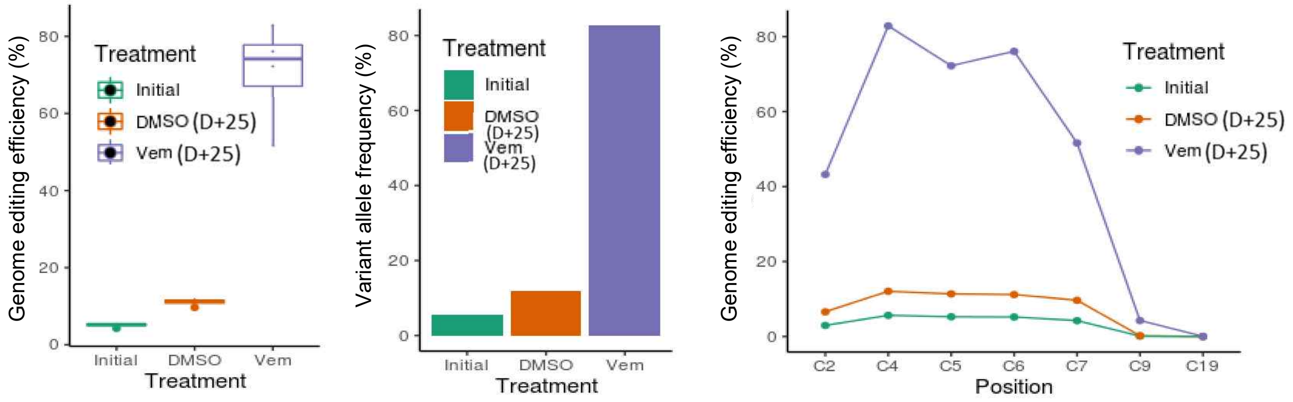


Figure 2

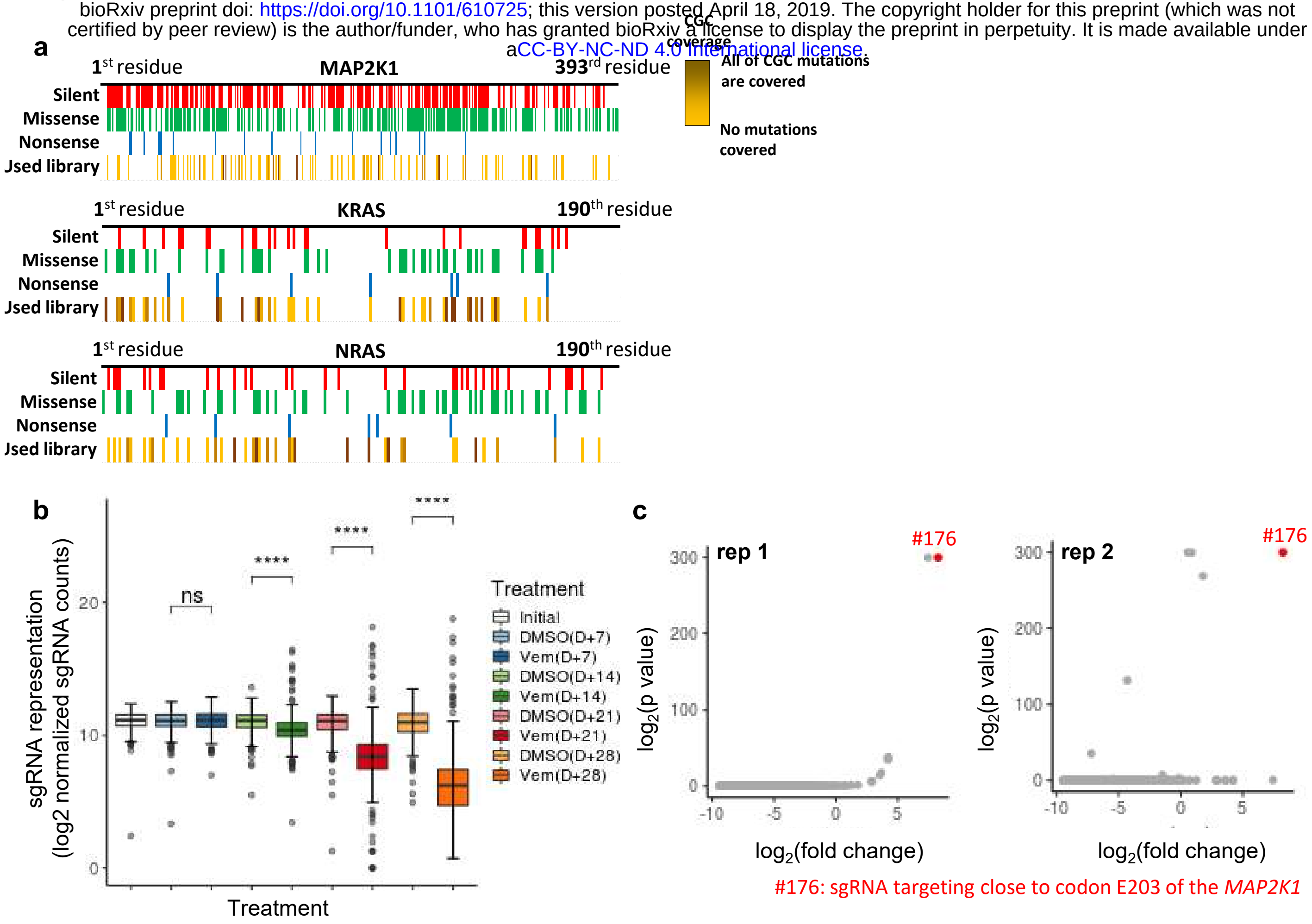

a
1st residue
MAP2K1
393rd residue
Silent
Missense
Nonsense
Jsed library
CGC coverage
All of CGC mutations are covered
No mutations covered
1st residue
KRAS
190th residue
Silent
Missense
Nonsense
Jsed library
1st residue
NRAS
190th residue
Silent
Missense
Nonsense
Jsed library
b
sgRNA representation (log2 normalized sgRNA counts)
0
10
20
ns
****
****
****
Treatment
Treatment
Initial
DMSO(D+7)
Vem(D+7)
DMSO(D+14)
Vem(D+14)
DMSO(D+21)
Vem(D+21)
DMSO(D+28)
Vem(D+28)
c
rep 1
#176
log2(p value)
0
100
200
300
-10
-5
0
5
log2(fold change)
rep 2
#176
log2(p value)
0
100
200
300
-10
-5
0
5
log2(fold change)
#176: sgRNA targeting close to codon E203 of the MAP2K1

Figure 3

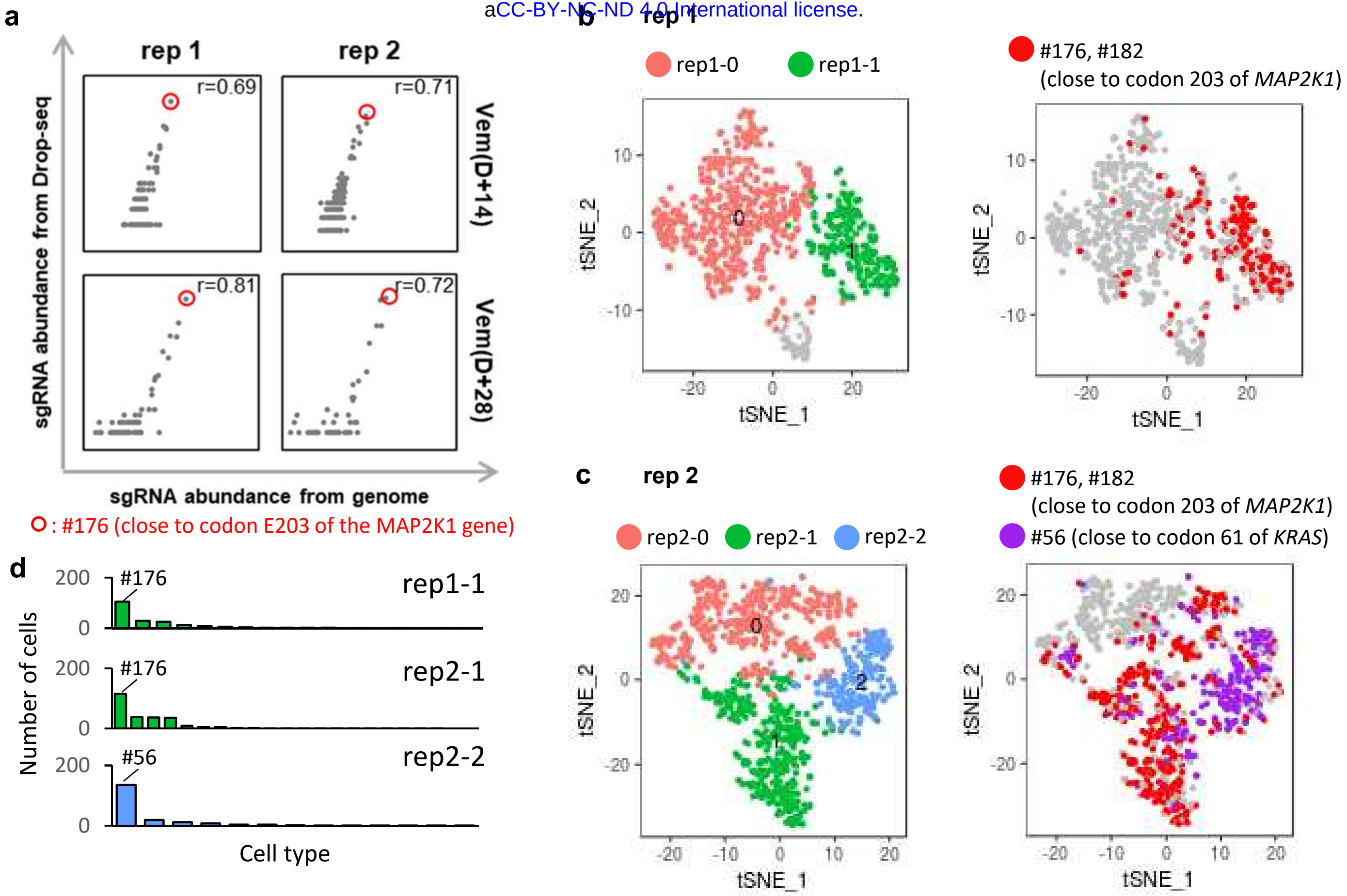
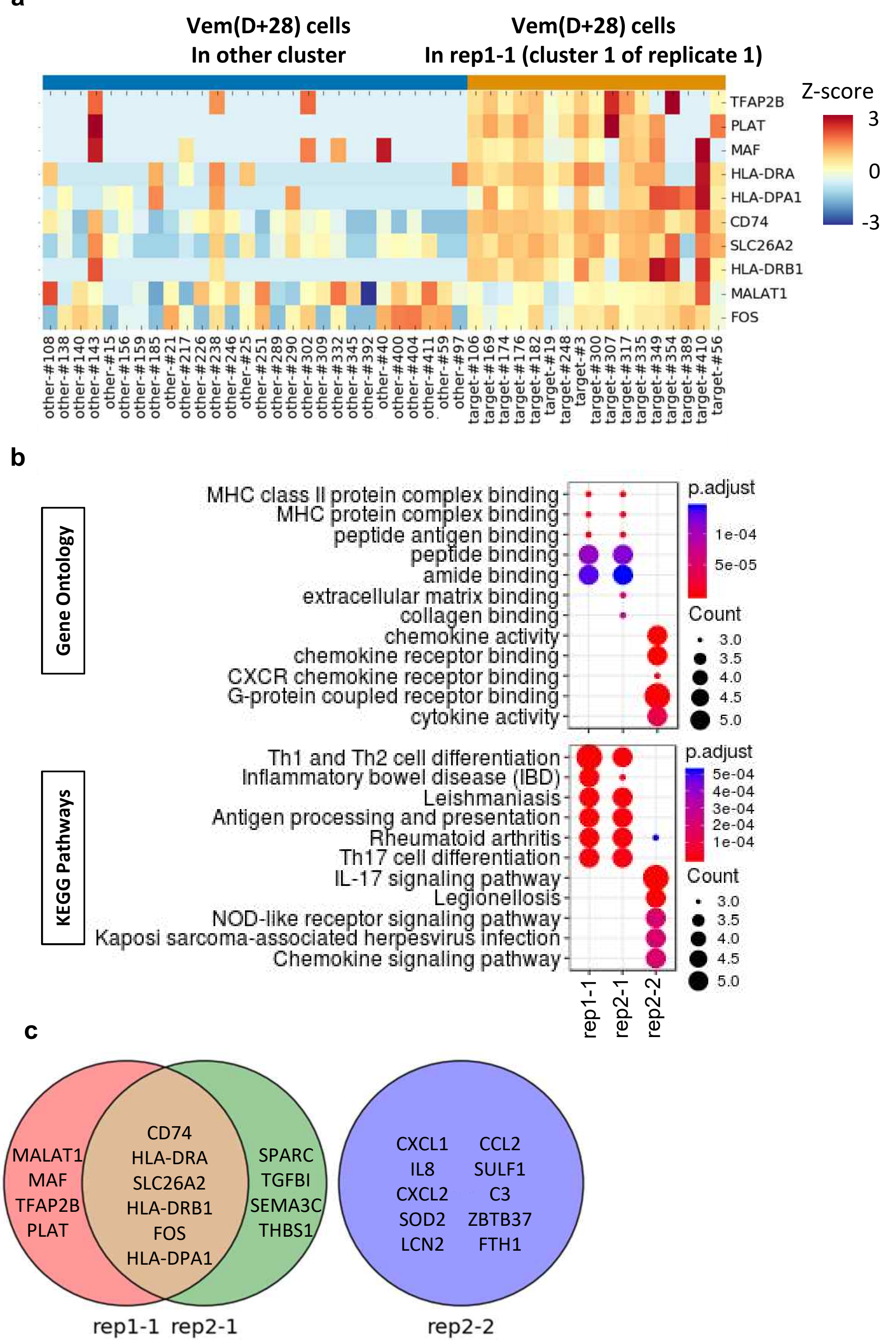

a
rep 1
rep 2
r=0.69
r=0.71
r=0.81
r=0.72
Vem(D+14)
Vem(D+28)
sgRNA abundance from Drop-seq
sgRNA abundance from genome
○ : #176 (close to codon E203 of the MAP2K1 gene)
b
rep 1
rep1-0
rep1-1
#176, #182
(close to codon 203 of MAP2K1)
tSNE_2
tSNE_1
10
0
-10
-20
20
c
rep 2
rep2-0
rep2-1
rep2-2
#56 (close to codon 61 of KRAS)
d
200
0
#176
#56
rep1-1
rep2-1
rep2-2
Number of cells
Cell type

Figure 4


a
Vem(D+28) cells
In other cluster
Vem(D+28) cells
In rep1-1 (cluster 1 of replicate 1)
Z-score
3
0
-3
TFAP2B
PLAT
MAF
HLA-DRA
HLA-DPA1
CD74
SLC26A2
HLA-DRB1
MALAT1
FOS
other-#108
other-#138
other-#140
other-#143
other-#15
other-#156
other-#159
other-#185
other-#21
other-#217
other-#226
other-#238
other-#246
other-#25
other-#251
other-#289
other-#290
other-#302
other-#309
other-#332
other-#345
other-#392
other-#40
other-#400
other-#404
other-#411
other-#59
other-#97
target-#106
target-#169
target-#174
target-#176
target-#182
target-#19
target-#248
target-#3
target-#300
target-#307
target-#317
target-#335
target-#349
target-#354
target-#389
target-#410
target-#56
b
Gene Ontology
MHC class II protein complex binding
MHC protein complex binding
peptide antigen binding
peptide binding
amide binding
extracellular matrix binding
collagen binding
chemokine activity
chemokine receptor binding
CXCR chemokine receptor binding
G-protein coupled receptor binding
cytokine activity
p.adjust
1e-04
5e-05
Count
3.0
3.5
4.0
4.5
5.0
KEGG Pathways
Th1 and Th2 cell differentiation
Inflammatory bowel disease (IBD)
Leishmaniasis
Antigen processing and presentation
Rheumatoid arthritis
Th17 cell differentiation
IL-17 signaling pathway
Legionellosis
NOD-like receptor signaling pathway
Kaposi sarcoma-associated herpesvirus infection
Chemokine signaling pathway
p.adjust
5e-04
4e-04
3e-04
2e-04
1e-04
Count
3.0
3.5
4.0
4.5
5.0
rep1-1
rep2-1
rep2-2
c
MALAT1
MAF
TFAP2B
PLAT
CD74
HLA-DRA
SLC26A2
HLA-DRB1
FOS
HLA-DPA1
SPARC
TGFBI
SEMA3C
THBS1
CXCL1
IL8
CXCL2
SOD2
LCN2
CCL2
SULF1
C3
ZBTB37
FTH1
rep1-1 rep2-1
rep2-2