

METHODS FOR NOMINATION OF NUCLEASE ON-/OFF-TARGET EDITING LOCATIONS, DESIGNATED “CTL-seq” (CRISPR Tag Linear-seq)

CROSS REFERENCE TO RELATED APPLICATIONS

- 5 This application claims priority to U.S. Provisional Patent Application No. 63/055,460, filed on July 23, 2020, which is incorporated by reference herein in its entirety.

REFERENCE TO SEQUENCE LISTING

- 10 This application is filed with a Computer Readable Form of a Sequence Listing in accordance with 37 C.F.R. § 1.821(c). The text file submitted by EFS, “013670-9056-WO01_sequence_listing_19-JUL-2021_ST25.txt,” contains 273 sequences, was created on July 19, 2021, has a file size of 153 Kbytes, and is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

- 15 Described herein are methods for identifying and nominating on- and off-target CRISPR editing sites with improved accuracy and sensitivity.

BACKGROUND

- 20 CRISPR (clustered regularly interspaced short palindromic repeats) has revolutionized genomics by permitting the simple introduction of changes to the genetic code. CRISPR systems, such as Cas9 and Cas12a proteins, are guided to their target by RNA oligonucleotide sequences bound by the Cas proteins (forming ribonucleoprotein protein; RNP), where the enzyme creates double stranded breaks (DSBs) in DNA sequences. Native cellular machinery repairs DSBs, generally using non-homologous end joining (NHEJ) or homology directed repair (HDR) molecular pathways. DNA repaired through NHEJ, which occurs at on- and off-target locations, often contains indels (insertions/deletions), which can lead to mutations and change the function of encoded genes. Thus, identifying these locations is critical to deconvoluting the impact of on- and off-target editing on biological phenotypes.

- 30 To date, no “gold standard” method exists to identify or nominate off-target editing locations for CRISPR or other nucleases. Many methods have been developed. These methods use a variety of strategies, including the detection of endogenous repair machinery assembled at DSBs (Discover-Seq [1]), the integration of a DNA tag sequence into the host cell genome (GUIDE-Seq; see U.S. Pat No. 9,822,407), iGUIDE [2, 3]), or by cutting DNA *in vitro* (BLISS [4], CIRCLE-Seq [5], SiteSeq [6]).

Cellular or cell based (sometimes referred to as *in vivo*) and biochemical (sometimes referred to as *in vitro*) off-target assay nomination systems each have their advantages. Proteins bound to the DNA and epigenetic marks modify the function of nuclease activity, suggesting that cellular or cell based methods may better identify actual editing targets [7]. However, biochemical methods have nominated sites not identified through cellular or cell based methods, suggesting biochemical methods may be more comprehensive [5, 6]. Nevertheless, these current tools tend to have imperfect sensitivity [5, 6] (see FIG. 1).

What is needed is a method for detecting and nominating on- and off-target CRISPR editing sites with improved accuracy and sensitivity.

SUMMARY

One embodiment described herein is a method for identifying and nominating on- and off-target CRISPR edited sites with improved accuracy and sensitivity, the process comprising the steps of: (a) co-delivering a guide sequence RNA (sgRNA) or a two-part CRISPR RNA:trans-activating crRNA (crRNA:tracrRNA) duplex, one or more tag sequences, and an RNA-guided endonuclease to cells; (b) incubating the cells for a period of time sufficient for double strand breaks to occur; (c) isolating genomic DNA from the cells, fragmenting the genomic DNA, and ligating the fragmented genomic DNA to a unique molecular index containing a universal adapter sequence; (d) amplifying the ligated DNA fragments using primers targeting the tag and universal adapter sequences to produce a first set of amplified sequences; (e) amplifying the first set of amplified sequences using universal sequencing primers targeting the tails of Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences; (f) sequencing the pooled sequences and obtaining sequencing data; and (g) identifying on-/off-target CRISPR editing loci. In one aspect, the universal sequencing primers target SP1 or SP2 sequence (SEQ ID NO: 7, 8) tails on the Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences. In another aspect, the universal sequencing primers target predesigned non-homologous sequence (SEQ ID NO: 269–273) tails on the Tag-pTOP or Tag-pBot primers to produce a second set of amplified sequences. In another aspect, the universal sequencing primers target predesigned 13-mer tails on the Tag-pTOP or Tag-pBot primers to produce a second set of amplified sequences. In another aspect, step (g) comprises executing on a processor: (i) aligning the sequence data to a reference genome; (ii) identifying on-/off-target CRISPR editing loci; and (iii) outputting the alignment, analysis, and results data as custom-formatted files, tables or graphics. In another aspect, the method further comprises a step following step (e) comprising: (e1) normalizing the second set of amplified sequences to produce concentration normalized libraries,

pooling the normalized libraries with other samples to produce pooled libraries; and continuing with steps (f)–(i). In another aspect, step (d) uses a suppression PCR method. In another aspect, the RNA-guided endonuclease comprises an endogenously-expressed Cas enzyme, a Cas expression vector, a Cas protein, or a Cas RNP complex. In another aspect, the RNA-guided endonuclease comprises an endogenously-expressed Cas9 enzyme, a Cas9 expression vector, a Cas9 protein, or a Cas9 RNP complex. In another aspect, the cells comprise human or mouse cells. In another aspect, the period of time is about 24 hours to about 96 hours. In another aspect, multiple tag sequences are co-delivered. In another aspect, the tag sequences comprise double-stranded deoxyribooligonucleotides (dsDNA) comprising 52-base pairs. In another aspect, the tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides. In another aspect, the tag sequences comprise a double stranded DNA comprising the complementary top and bottom strand pairs of SEQ ID NO: 1–2 or 7–268.

Other embodiments described herein are on- and off-target CRISPR editing sites identified or nominated using the methods described herein.

Another embodiment described herein is a method for designing 52-base pair tag sequences, the method comprising, executing on a processor: (a) randomly generating 13-nucleotide sequences with 40–90% GC content, max homopolymer length A:2, C:3, G:2, T:2, weighted homopolymer rate < 20, self-folding $T_m < 50$ °C, and self-dimer $T_m < 50$ °C; (b) removing sequences that perfectly align to a particular genome or that are homopolymers or GG or CC dinucleotide motifs and obtaining a set of 13-mers; (c) selecting a subset of the 13-mer sequences that contain one or less CC or GG dinucleotide motifs; (d) concatenating four of the of 13-mer subset sequences to form random 52-mer sequences; (e) aligning the random 52-mer sequences to a genome; (f) removing the random 52-mer sequences that have similarity to the genome to produce a subset of 52-mer sequences; and (h) outputting the subset of 52-mer sequences and generating the complementary strands to produce double stranded 52-base pair tag sequences. In one aspect, the genome is human or mouse. In another aspect, the 52-base pair tag sequences are non-complementary to the genome. In another aspect, the method further comprises designing primers for the 52-base pair tag sequences. In another aspect, the 52-base pair tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides of the 52-base pair tag sequences. In another aspect, the method further comprises synthesizing oligonucleotides comprising the 52-base pair tag sequences, the complement of the 52-base pair tag sequences, or primers for the 52-base pair tag sequences.

Other embodiments described herein are one or more 52-base pair tag sequences designed using the methods described herein. In one aspect, the 52-base pair tag sequence comprises a double stranded DNA comprising the top and bottom strand pairs of SEQ ID NO: 1–2 or 7–268.

5 Another embodiment described herein is a method for designing primers partially complementary to the 52-base pair tag sequences of claim 23 and an adapter primer, the method comprising, executing on a processor: (a) designing tag primers that are partially complementary to the top and bottom strands of tag sequences; and (b) designing an adapter primer that is partially complementary to the top strand of the adapter sequence; wherein: the tag primers
10 comprise a 5'-universal tail sequence; and the adapter primer comprises a sequence complementary to the tails of Tag-pTOP or Tag-pBOT primers. In one aspect, the 5'-universal tail sequence is complementary to an SP1 or SP2 sequence (SEQ ID NO: 7, 8), a locus specific segment, a ribonucleotide (rN) 6-nucleotides from the 3'-end, a 3'-end mismatch, a 3'-end block (3'-C₃ spacer), a predesigned non-homologous sequence (SEQ ID NO: 269–273), or a
15 predesigned 13-mer sequence. In another aspect, the primers partially complementary to top and bottom strands of the tag sequences comprise a tail sequence complementary to the SP1 sequence (SEQ ID NO: 7) and the adapter primer comprises a sequence complementary to the SP2 sequence (SEQ ID NO: 8) tail on the Tag-pTOP or Tag-pBOT primers; or the primers partially complementary to top and bottom strands of the tag sequences comprise a tail sequence
20 complementary to the SP2 sequence (SEQ ID NO: 8) and the adapter primer comprises a sequence complementary to the SP1 sequence (SEQ ID NO: 7) tail on the Tag-pTOP or Tag-pBOT primers. In another aspect, the amplification of a nucleic acid molecule with the primers that are complementary to the top and bottom strands of tag sequences and primers that are complementary to the top strand of the adapter sequence produces a PCR product that comprises
25 a portion of the tag sequence, a sgDNA sequence, and the adapter sequence. In another aspect, the method further comprises synthesizing oligonucleotides comprising the sequences of the forward and reverse tag primers and the adapter primer. In another aspect, the 52-base pair tag sequences and primers partially complementary to the 52-base pair tag sequences are designed and selected using an algorithm predicting whether the primers are likely to be partially
30 complementary and have a propensity to form primer-dimers.

Other embodiments described herein are one or more primers partially complementary to the 52-base pair tag sequences and one or more adapter primers designed using the methods described herein. In one aspect, the primers comprise the sequences of SEQ ID NO: 3, 4; and the adapter primer, wherein the adapter primer comprises the sequence of SEQ ID NO: 5.

Another embodiment described herein is the use of one or more double-stranded 52-base pair tag sequences for identifying on- and off-target CRISPR editing sites.

DESCRIPTION OF THE DRAWINGS

5 FIG. 1 shows fraction of reads shared by three biological replicates are shown in white sectors; whereas reads shared by two replicates, or present in a single replicate, are shown in black sectors. Table 1 shows GUIDE-seq [3] based nomination for 4 different gRNAs in triplicate in a 96-well format. gRNA complexes were generated by mixing equimolar amounts of Alt-R crRNA-XT and Alt-R tracrRNA. HEK293 cells stably expressing Cas9 were transfected with 10
10 μM gRNA and 0.5 μM dsODN GUIDE-seq tag using the Nucleofector™ system (Lonza). After 72 hrs, genomic DNA (gDNA) was isolated. Genomic DNA was fragmented, and adapters were ligated using the Lotus DNA library preparation kit (IDT). Libraries were generated by amplification from the inserted tag to the ligated adapters [3]. Libraries were then sequenced in paired-end fashion on an Illumina® platform.

15 FIG. 2 shows that GUIDE-Seq finds more off-target locations than can be validated through rhAmpSeq targeted amplification. Presented results are an aggregate of 331 GUIDE-Seq nominated sites when delivering gRNA sequences (internally named: AR, CTNNB1, EMX1, GRHPR, HPRT38087, HPRT38285, VEGFA) into HEK293 cells stably expressing WT Cas9. GUIDE-seq nominated off-targets assigned $\geq 0.1\%$ of the total reference genome aligned reads
20 for each guide were designed and targeted by one rhAmpSeq panel all reference genome aligned. In subsequent experiments, gRNAs were again delivered to the same cells, and editing was assayed with rhAmpSeq. Targets were called “edited” if the treated condition had observed indels $\geq$ the untreated control sample at $\geq 1\%$.

 FIG. 3 illustrates that GUIDE-Seq tag integration rate varies. The graph shows the
25 percentage of Tag integration (normalized to % Editing) for 118 unique Cas9 on/off-target sites that had InDel editing in rhAmpSeq panels targeting GUIDE-Seq nominated on/off-target loci for guide sequences targeting the RAG1, RAG2, and EMX1 genes. Each guide was co-delivered with the 34-base pair GUIDE-Seq, dsODN tag into HEK293 cells stably expressing Cas9 by nucleofection. DNA was extracted 72 hrs later, amplified by rhAmpSeq multiplex PCR,
30 sequenced on an Illumina® MiSeq, and analyzed through a custom pipeline. The normalized tag integration rate is calculated as the percentage of sequenced reads at each target containing the tag sequence divided by the total reads containing an allele divergent from the reference genome (indicating Cas9 editing).

FIG. 4 shows the design of rhAmpSeq primers against alien sequence tags. A cartoon diagram shows the steps of the design process using the rhAmpSeq design pipeline including design of forward primers against the top (1) and bottom (2) strands, discarding unneeded primers, and selecting tag-targeting primers that have 5'-overlapping, but not 3'-overlapping sequences, so that the top/bottom strand primer dimers would hairpin (3).

FIG. 5 shows an overview of the rhAmpSeq design pipeline used to construct the overlapping primer designs. In the pipeline, a known sequence is appended onto the 5'-end and 3'-end of each tag sequence, the inputs are quality-controlled and assays (shown in FIG. 4A) are designed against the top and bottom strand of each tag. Primers targeting each tag strand are paired such that at least 4-nucleotides 3' of the RNA nucleotide do not overlap between primers targeting the same tag, and primer pairs are ranked and selected. Hg38 and mm38 acronyms represent versions of the human and mouse genomes, respectively.

FIG. 6 illustrates hairpin formation if overlapping primers generate PCR amplicons. The diagram shows a representative target sequence and hairpin PCR product of undesired short amplicons from overlapping primer regions with complementary 5' primer tail ends at the 3'- and 5'-end of the PCR product.

FIG. 7 shows the number of target sites (black bars) with integration of the specified single tag (SEQ ID NO: 9–40) or pools of tags described in Table 5 (SEQ ID NO: 9–40, 45–268). The striped bar (CTLmax) shows the maximum number of target sites that theoretically can be found if a combination of the single tags (SEQ ID NO: 9–40) is used (23 sites out of a maximum of 32 sites). Pool A1 contains all the single tags (SEQ ID NO: 9–40). Pools B1–6 contain 16 different tags each (SEQ ID NO: 45–268). Pool C1 contains all tags tested (SEQ ID NO: 9–40, 45–268). Integration events were determined using an in-house data analysis tool.

FIG. 8 shows the number of target sites (black bars) with integration of the specified single tag (SEQ ID NO: 9–40) or pools of tags described in Table 5 (SEQ ID NO: 9–40, 45–268). The striped bar (CTLmax) shows the maximum number of target sites that theoretically can be found if a combination of the single tags (SEQ ID NO: 9–40) is used (47 sites out of a maximum of 53 sites). Pool A1 contains all the single tags (SEQ ID NO: 9–40). Pools B1–6 contain 16 different tags each (SEQ ID NO: 45–268). Pool C1 contains all tags tested (SEQ ID NO: 9–40, 45–268). Integration events were determined using an in-house data analysis tool.

DETAILED DESCRIPTION

Described herein are methods for detecting and nominating on- and off-target CRISPR editing sites with improved accuracy and sensitivity. The intracellular context information is

maintained by building upon prior *in vivo* nomination methods. The sensitivity is expanded by co-delivering a set of unique, predefined sequence tags. In one aspect, the co-delivered set of predefined unique tags may range from 13–80 base pairs. In another aspect, the co-delivered set of predefined tags may be comprised of 13 base pair tag sequence tags, 26 base pair tag sequence tags, 39 base pair tag sequence tags, 52 base pair tag sequence tags, 65 base pair tag sequence tags, or 78 base pair tag sequence tags. In another aspect, the unique predefined tags are a set of 52-base pair tag sequence tags (the increased length of the sequence tags improves the ability to find good primer landing sites for rhPrimers). This limitation is believed to be mitigated by using a diversity of tag sequences that are distinct from human and mouse genomes. The specificity is improved by building upon Integrated DNA Technologies (IDT)'s rhAmp technology that uses RNAaseH2 (*Pyrococcus abyssi*) to unblock primers that have correctly annealed to their target; this yields lower rates of false priming. Specificity can be further enhanced by only nominating targets using reads that contain an expected tag sequence at the 5'-end. The incorporation of suppression PCR into this method permits ease of use. The prior *in vivo* methods (e.g., GUIDE-seq and iGUIDE) require parallel PCR reactions (2 pool amplification) to amplify by annealing to and extending from the top and bottom strand of the tags. Here, suppression PCR is used to allow both pools to be amplified simultaneously without causing problematic dimer sequences.

A GUIDE-Seq dsDNA tag was co-delivered with one guide RNA to HEK293 cells constitutively expressing Cas9 using nucleofection. See U.S. Pat. No. 9,822,407, which is incorporated by reference herein for such teachings. A total of four different guide RNAs were tested in this fashion. Ribonucleoprotein complexes (RNPs) between the expressed Cas9 and guide RNA form within the cells, introducing double stranded breaks. Repaired breaks can contain the co-delivered tags. After delivery, cells were incubated, and the resulting DNA was extracted. Target amplification was performed according to the GUIDE-Seq protocol and assayed with a modified version of the GUIDE-Seq analytical pipeline (github.com/aryeelab/guideseq). Nominated targets were compared between three biological replicates (unique guideRNA + Tag co-deliveries). Not all nominated targets were common to all biological replicates (commonly / total nominated targets: 7/31, 6/19, 2/4, 3/5 respectively; see Table 1). However, >90% of the total reads, attributed to any target, were attributed to common targets (on average; see FIG. 1).

Table 1. Identified off-target sites for four different gRNAs and relative level of editing at off-target sites compared to the on-target site

Location	C19orf84_BR1	C19orf84_BR2	C19orf84_BR3
chr19_51389306	100.00%	100.00%	100.00%

chr9_20224748	38.55%	16.43%	29.00%
chr4_28036434	16.33%	13.05%	14.36%
chr15_74256506	14.30%	18.18%	25.17%
chr2_171312919	11.40%	8.51%	7.93%
chr8_65742269	10.82%	1.17%	10.40%
chr13_96554656	8.70%	0.00%	0.00%
chr4_86807920	8.50%	9.21%	1.92%
chr3_124485356	6.57%	0.00%	0.00%
chr9_20330398	5.60%	0.00%	0.00%
chr11_71298123	5.12%	0.00%	0.00%
chr7_101729696	4.83%	0.00%	9.58%
chr19_10923882	3.67%	3.03%	0.00%
chr10_15548456	3.57%	15.38%	0.00%
chr12_117097457	2.80%	0.00%	2.60%
chr22_33493900	2.13%	0.00%	4.79%
chrX_149763439	2.13%	0.00%	3.83%
chr17_7435217	1.93%	0.00%	0.55%
chr12_26286721	1.74%	0.00%	5.06%
chr16_49704848	1.26%	5.01%	7.11%
chr12_51288216	1.06%	0.00%	0.00%
chr12_56010621	0.87%	0.00%	0.00%
chr13_29717148	0.48%	0.00%	0.00%
chr1_3088065	0.29%	0.00%	0.00%
chr15_73442915	0.19%	0.00%	0.55%
chr10_118045968	0.19%	0.00%	0.00%
chr14_102199972	0.00%	0.00%	0.68%
chr18_56334679	0.00%	0.00%	2.33%
chr21_36426137	0.00%	0.00%	2.19%
chr5_139002763	0.00%	0.00%	3.83%
chrX_58291642	0.00%	0.00%	3.83%
Location	C17orf99_BR1	C17orf99_BR2	C17orf99_BR3
chr17_78164110	100.00%	100.00%	100.00%
chr22_24471716	15.00%	13.24%	10.86%
chr10_101156881	6.22%	11.07%	9.79%
chr3_170476431	5.86%	3.97%	4.57%
chr17_17692965	4.94%	0.66%	8.62%
chr15_73400031	3.93%	4.63%	5.73%
chr19_15238775	0.00%	0.00%	2.56%
chr2_18362316	0.00%	0.00%	1.59%
chr2_171087784	0.00%	0.54%	0.84%
chr22_19959968	0.00%	1.26%	0.19%
chr22_32114104	0.00%	0.00%	4.06%
chr4_129034015	0.00%	0.00%	0.33%
chr5_61219030	0.00%	0.00%	0.33%

chr5_66209615	0.00%	0.00%	1.86%
chr7_69709389	0.00%	0.12%	2.75%
chr7_158662844	0.00%	1.44%	5.27%
chrX_9567397	0.00%	0.00%	0.23%
chr19_55657073	0.00%	0.66%	0.00%
chr22_43788032	0.00%	2.47%	0.00%
Location	C16orf90_BR1	C16orf90_BR2	C16orf90_BR3
chr16_3494817	100.00%	100.00%	100.00%
chr2_109189307	75.32%	4.27%	52.05%
chr22_24586001	45.45%	0.00%	0.00%
chr10_104736568	0.00%	0.00%	8.22%
Location	ATAD3C_BR1	ATAD3C_BR2	ATAD3C_BR3
chr1_1450685	100.00%	100.00%	100.00%
chr1_1503588	11.73%	10.07%	9.27%
chr1_1516015	2.47%	1.86%	5.14%
chr19_32167960	26.34%	0.93%	0.00%
chr2_111077960	0.00%	1.12%	0.00%

Additionally, nominated targets may not be replicable or detectable using orthogonal methods. Using the GUIDE-Seq method, the GUIDE-Seq DNA tag was co-delivered with each of 6 guides (each tag is delivered with one guide RNA) to HEK293 cells constitutively expressing Cas9 using nucleofection. rhAmpSeq multiplex amplicon panels were designed to amplify the nominated targets, and we quantified editing in biological replicates. Of the 331 targets nominated by GUIDE-Seq, only 41 (12%) could be verified with rhAmpSeq (see FIG. 2).

dsDNA tag sequences co-delivered with the guide RNAs into a stably expressing CRISPR cell line, which are used in the NHEJ repair, are incorporated at varying rates. Here, the GUIDE-Seq dsDNA tag was co-delivered with each of 6 guides into HEK293 cells constitutively expressing Cas9. In another aspect, the dsDNA tag sequences co-delivered with CRISPR RNP, which are used in the NHEJ repair, are incorporated at varying rates. Here, the GUIDE-Seq dsDNA tag was co-delivered with each of 6 guides into HEK293 cells constitutively expressing Cas9. rhAmpSeq panels were developed to amplify nominated targets, and in biological replicates, the rates of tag integration were analyzed using a custom analytical pipeline. These results demonstrate that tags are incorporated at 0–85% of edited genomic copies, varying by target (see FIG. 3). Without being bound by any theory, it is hypothesized that the rate varies by sequence context.

Described herein are methods to improve the signal to noise ratio by combining Integrated DNA Technology's rhAmpSeq™ technology, suppression PCR, and novel alien DNA sequence designs to nominate nuclease off-target editing locations within a host genome.

In this method, Cas9, a sgRNA or a two-part CRISPR RNA:trans-activating crRNA (crRNA:tracrRNA) duplex, and one or more double stranded DNA (dsDNA) tag sequences are delivered to cells. Co-delivering multiple tags permits improved tag integration at off-target sites (see below). The tag sequences have sequence content significantly different (i.e., alien) to the host genome. After nuclease introduced DSBs, NHEJ repair will insert the tag sequence(s) into the target site, forming known primer landing sites. After cells have time to repair the DSBs and possibly further divide (such as after 72 hr), genomic DNA is isolated, fragmented (e.g., Covaris® shearing, enzyme-based shearing, Tn5, etc.), ligated a unique molecular index (UMI)-containing universal adapter sequence to the fragmented DNA, and the un-ligated material is removed. Next, the DNA fragments are amplified by targeting primers to the tag and universal adapter sequences (Round 1 PCR). Using universal primers, a sample index (PCR2) is added, the amplified material is concentration normalized, pooled with other samples, and the pooled material is sequenced on an Illumina® (or similar) machine. The sequenced reads are aligned to a reference genome, and loci where large numbers of reads map may nominate on/off-target locations.

Alien sequences were designed by generating >1 M random 13-mer sequences with 40–90% GC content, max homopolymer length A:2, C:3, G:2, T:2, weighted homopolymer rate < 20, self-folding $T_m < 50$ °C, and self-dimer $T_m < 50$ °C. From the list of sequences, sequences that aligned perfectly against human (GRCh38.p2; hg38) or mouse (GRCh38.p4; mm38) reference genomes or had troubling motif sequences (homopolymers, most G-G or C-C dinucleotide motifs) were removed, resulting in 479 sequences.

To design the 52-base pair tag sequences described herein, 49 13-mer oligo sequences were selected that contain ≤ 1 C or G dinucleotide, and 10,000 unique combinations of four 13-mer sequences were generated. The length of each concatenated sequence (e.g., pasting four 13-mer sequences in a row using software) is 52-nucleotides. Next, each 52-nucleotide tag sequence was aligned against the human (GRCh38.p2) and mouse (GRChm38.p4) genomes using an internally modified version of *bwa*, called *bwa-psm*. Implementation of *bwa-psm* returns all possible secondary matches up to a defined threshold. A set of tag sequences (SEQ ID NO: 1–2) were designed that were intended to work as a group, that had no similarity to the human or mouse genomes (max seed size: 7, seed edit distance: 2, max edit distance: 21, max gap open: 2, max gap extension: 3, mismatch penalty: 1, gap open penalty: 1, gap extension penalty: 1).

Overlapping rhAmpSeq V1 primers (SEQ ID NO: 3–4) were designed complementary to the top and bottom strands of the tag and 5'-end of the adapter sequence (SEQ ID NO: 6) (FIG. 4). The tag-specific primers (SEQ ID NO: 3–4) contain a 5'-universal tail sequence matching the

SP1 and SP2 primer sequences (SEQ ID NO: 7–8), a locus specific segment, a ribonucleotide (rN) 6-nucleotides from the 3'-end, a 3'-end mismatch, and a 3'-end block (3'-C₃ spacer). The adapter-specific primer (SEQ ID NO: 5) targets the 5'-end of the 5'-P5 adapter sequence (SEQ ID NO: 6), and the adapter sequence contains unique molecular index (UMI) sequence (Table 2).

5 The primers were designed to target the plus and minus strands of the annealed tag such that, if these primers unexpectedly form a dimer, the formed product will hairpin, removing the oligo from the available reaction templates (e.g., suppression PCR). (FIG. 6A–B). Primer sequences targeting the tags were chosen based on a proprietary design algorithm designed and implemented by IDT (internal copy of the algorithm with a public-facing UI:
10 www.idtdna.com/site/account?ReturnURL=/site/order/designtool/index/RHAMPSEQ), which selects the most optimally performing primer pairs to amplify the intended template sequence. (FIG. 5). Primer sequences were assessed for non-specific binding to all other tag sequences and both human and mouse primary genome assemblies to verify they were unlikely to form off-target amplicons when combined with a universal adapter sequence and the presence of human
15 or mouse genomic DNA.

The primers were desired to work in pairs where one tag-specific primer (top or bottom strand) pairs with the adapter-specific primer (SEQ ID NO:5). This results in the amplification of a molecule that contains a portion of the tag, gDNA, and the adapter sequence when amplified using suppression PCR methods (FIG 4).

20

Table 2. Sequences Used for First Proof of Concept

Type	Name	Sequence (5'→3')	SEQ ID NO
Tag	9022179029169042579 04625907201907281	T*C*GTTCGTTCCGCTCTAACCGG CGAATCTACCGCGCATATCTACGC CGCA*A*T	SEQ ID NO: 1
Tag	9022179029169042579 04625907201907281_r ev	A*T*TGCGGCGTAGATATGCGCGG TAGATTCGCCGGTTAGAGCGGAAC GAAC*G*A	SEQ ID NO: 2
Tag Primers	pFWD.ID_Target1: 9022179029169042579 04625907201907281.12 7.150.1.SP1	acactctttccctacacgacgctc ttccgatctTCTACCGCGCATATC TACrGCCGCT/3SpC3/	SEQ ID NO: 3
Tag Primers	pFWD.ID_Target2: 9022179029169042579 04625907201907281.11 6.140.-1.SP1	acactctttccctacacgacgctc ttccgatctATATGCGCGGTAGAT TCGCrCGGTTT/3SpC3/	SEQ ID NO: 4
Adapter Primer	Adapter Primer	gtgactggagttcagacgtgtgct cttccgatctAATGATACGGCGAC CACCGAGATCTACArCAAGGC/3S pC3/ AATGATACGGCGACACCGAGATC	SEQ ID NO: 5
P5 Adapter	Example Sequence	TACACTAGATCGCENNWNWNACAC CTCTTTCCCTACACGACGCTCTTC CGATC*T	SEQ ID NO: 6
SP1	Sequencing Primer 1	acactctttccctacacgacgctc ttccgatct	SEQ ID NO: 7
SP2	Sequencing Primer 2	gtgactggagttcagacgtgtgct cttccgatct	SEQ ID NO: 8

“*” indicates a phosphorothioate linkage; “rN” indicates a ribonucleotide, where N is the nucleotide preceeded by the “r”; “/3SpC3/” indicates a 3'-C₃ spacer.

One embodiment described herein is a method for identifying and identifying and nominating on- and off-target CRISPR editing sites with improved accuracy and sensitivity, the process comprising the steps of: (a) co-delivering a guide sequence RNA (sgRNA) or a two-part CRISPR RNA:trans-activating crRNA (crRNA:tracrRNA) duplex and one or more tag sequences to cells; (b) incubating the cells for a period of time; (c) isolating genomic DNA from the cells, fragmenting the genomic DNA, and ligating the fragmented genomic DNA to a unique molecular index containing a universal adapter sequence; (d) amplifying the ligated DNA fragments using primers targeting the tag and universal adapter sequences to produce a first set of amplified sequences; (e) amplifying the first set of amplified sequences using universal sequencing primers targeting the tails of Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences; (f) sequencing the pooled sequences and obtaining sequencing data; and (g) identifying on-/off-target CRISPR editing loci. In one embodiment, the universal sequencing primers target SP1 or SP2 sequence (SEQ ID NO: 7, 8) tails on the Tag-pTOP or Tag-pBOT

primers to produce a second set of amplified sequences. In another embodiment, the universal sequencing primers target predesigned non-homologous sequence (Table 6; SEQ ID NO: 269–273) tails on the Tag-pTOP or Tag-pBot to produce a second set of amplified sequences. In yet another embodiment, the universal primers target predesigned 13-mer tails on the Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences. In one embodiment, step (g) comprises executing on a processor: (i) aligning the sequence data to a reference genome; (ii) identifying on-/off-target CRISPR editing loci; and (iii) outputting the alignment, analysis, and results data as tables or graphics. In another embodiment, the method further comprises a step following step (e) comprising: (e1) normalizing the second set of amplified sequences to produce concentration normalized libraries, pooling the normalized libraries with other samples to produce pooled libraries; and continuing with steps (f)–(i). In one aspect, step (d) uses a suppression PCR method. In another aspect, the cells constitutively express a Cas enzyme, are co-delivered with a Cas expression vector, are co-delivered with a Cas protein, or are co-delivered with a Cas RNP complex. In another aspect, the cells constitutively express a Cas9 enzyme, are co-delivered with a Cas9 expression vector, are co-delivered with a Cas9 protein, or are co-delivered with a Cas9 RNP complex. In another aspect, the cells comprise human or mouse cells. In another aspect, the period of time is about 24 hours to about 96 hours. In another aspect, multiple tag sequences are co-delivered. In another aspect, the tag sequences comprise double-stranded deoxyribonucleotides (dsDNA) comprising 52-base pairs. In another aspect, the tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides. In another aspect, the tag sequences comprise a double stranded DNA comprising the top and bottom strand pairs of SEQ ID NO: 9–40 or 45–268.

Another embodiment described herein is on- and off-target CRISPR editing sites identified or nominated using the methods described herein.

Another embodiment described herein is a method for designing 52-base pair tag sequences, the method comprising, executing on a processor: (a) randomly generating 13-nucleotide sequences with 40–90% GC content, max homopolymer length A:2, C:3, G:2, T:2, weighted homopolymer rate < 20, self-folding $T_m < 50$ °C, and self-dimer $T_m < 50$ °C; (b) removing sequences that perfectly align to a particular genome or that are homopolymers or GG or CC dinucleotide motifs and obtaining a set of 13-mers; (c) selecting a subset of the 13-mer sequences that contain one or less CC or GG dinucleotide motifs; (d) concatenating four of the of 13-mer subset sequences to form random 52-mer sequences; (e) aligning the random 52-mer sequences to a genome; (f) removing the random 52-mer sequences that have similarity to the genome to

produce a subset of 52-mer sequences; and (h) outputting the subset of 52-mer sequences and generating the complementary strands to produce double stranded 52-base pair tag sequences. In one aspect, the genome is human or mouse. In one aspect, the 52-base pair tag sequences are not complementary to the genome. In another aspect, the method further comprises
5 designing primers for the 52-base pair tag sequences. In another aspect, the 52-base pair tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides of the 52-base pair tag sequences. In another aspect, the method further comprises synthesising oligonucleotides comprising the 52-base pair tag sequences, the complement of the 52-base pair tag sequences, or primers for the
10 52-base pair tag sequences.

Another embodiment described herein is one or more 52-base pair tag sequences designed using the methods described herein. In one aspect, the 52-base pair tag sequence comprises a double stranded DNA comprising the complementary top and bottom strand pairs of SEQ ID NO: 9–40 or 45–268.

15 Another embodiment described herein is a method for designing primers partially complementary to the 52-base pair tag sequences described herein and an adapter primer, the method comprising, executing on a processor: (a) designing tag primers that are partially complementary to the top and bottom strands of tag sequences; and (b) designing an adapter primer that is partially complementary to the top strand of the adapter sequence; wherein: the tag
20 primers comprise a 5'-universal tail sequence complementary to an SP1 or SP2 sequence (SEQ ID NO: 7, 8), a locus specific segment, a ribonucleotide (rN) 6-nucleotides from the 3'-end, a 3'-end mismatch, and a 3'-end block (3'-C₃ spacer); and the adapter primer comprises a sequence complementary to the SP1 or SP2 sequence (SEQ ID NO: 7, 8). In one aspect, the primers partially complementary to top and bottom strands of the tag sequences comprise a sequence
25 complementary to the SP1 sequence and the adapter primer comprises a sequence complementary to the SP2 sequence; or the primers partially complementary to top and bottom strands of the tag sequences comprise a sequence complementary to the SP2 sequence and the adapter primer comprises a sequence complementary to the SP1 sequence. In another aspect, amplification of a nucleic acid molecule with the primers that are complementary to the top and
30 bottom strands of tag sequences and primers that are complementary to the top strand of the adapter sequence produces a PCR product that comprises a portion of the tag sequence, a sgDNA sequence, and the adapter sequence. In another aspect, the method further comprises synthesising oligonucleotides comprising the sequences of the forward and reverse tag primers and the adapter primer.

In another embodiment described herein, the 52-base pair tag sequences and primers partially complementary to the 52-base pair tag sequences are designed and selected using an algorithm predicting whether the primers are likely to be partially complementary and have a propensity to form primer-dimers.

5 Another embodiment described herein is one or more primers partially complementary to the 52-base pair tag sequences and one or more adapter primers designed using the methods described herein. In one aspect, the primers partially complementary to the 52-base pair tag sequence comprise the sequences of SEQ ID NO: 3, 4; and the adapter primer comprises the sequence of SEQ ID NO:5.

10 Another embodiment described herein is the use of one or more double-stranded 52-base pair tag sequences for identifying on- and off-target CRISPR editing sites.

It will be apparent to one of ordinary skill in the relevant art that suitable modifications and adaptations to the compositions, formulations, methods, processes, and applications described herein can be made without departing from the scope of any embodiments or aspects thereof.

15 The compositions and methods provided are exemplary and are not intended to limit the scope of any of the specified embodiments. All the various embodiments, aspects, and options disclosed herein can be combined in any variations or iterations. The scope of the methods and processes described herein include all actual or potential combinations of embodiments, aspects, options, examples, and preferences herein described. The methods described herein may omit

20 any component or step, substitute any component or step disclosed herein, or include any component or step disclosed elsewhere herein. It should also be understood that embodiments may include and otherwise be implemented by a combination of various hardware, software, and electronic components. For example, various microprocessors and application specific integrated circuits ("ASICs") can be utilized, as can software of a variety of languages. Also, servers and

25 various computing devices can be used and can include one or more processing units, one or more computer-readable mediums, one or more input/output interfaces, and various connections (e.g., a system bus) connecting the components. Should the meaning of any terms in any of the patents or publications incorporated by reference conflict with the meaning of the terms used in this disclosure, the meanings of the terms or phrases in this disclosure are controlling.

30 Furthermore, the specification discloses and describes merely exemplary embodiments. All patents and publications cited herein are incorporated by reference herein for the specific teachings thereof.

Various embodiments and aspects of the inventions described herein are summarized by the following clauses:

- Clause 1. A method for identifying and nominating on- and off-target CRISPR edited sites with improved accuracy and sensitivity, the process comprising the steps of:
- (a) co-delivering a guide sequence RNA (sgRNA) or a two-part CRISPR RNA:trans-activating crRNA (crRNA:tracrRNA) duplex, one or more tag sequences, and an RNA-guided endonuclease to cells;
 - (b) incubating the cells for a period of time sufficient for double strand breaks to occur;
 - (c) isolating genomic DNA from the cells, fragmenting the genomic DNA, and ligating the fragmented genomic DNA to a unique molecular index containing a universal adapter sequence;
 - (d) amplifying the ligated DNA fragments using primers targeting the tag and universal adapter sequences to produce a first set of amplified sequences;
 - (e) amplifying the first set of amplified sequences using universal sequencing primers targeting the tails of Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences;
 - (f) sequencing the pooled sequences and obtaining sequencing data; and
 - (g) identifying on-/off-target CRISPR editing loci.
- Clause 2. The method of clause 1, wherein the universal sequencing primers target SP1 or SP2 sequence (SEQ ID NO: 7, 8) tails on the Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences.
- Clause 3. The method of clause 1 or 2, wherein the universal sequencing primers target predesigned non-homologous sequence (SEQ ID NO: 269–273) tails on the Tag-pTOP or Tag-pBot primers to produce a second set of amplified sequences.
- Clause 4. The method of any one of clauses 1–3, wherein the universal sequencing primers target predesigned 13-mer tails on the Tag-pTOP or Tag-pBot primers to produce a second set of amplified sequences.
- Clause 5. The method of any one of clauses 1–4, wherein step (g) comprises executing on a processor:
- Clause 6. aligning the sequence data to a reference genome;
- (a) (ii) identifying on-/off-target CRISPR editing loci; and
 - (b) (iii) outputting the alignment, analysis, and results data as custom-formatted files, tables or graphics.
- Clause 7. The method of any one of clauses 1–5, further comprising a step following step (e) comprising:

- (a) (e1) normalizing the second set of amplified sequences to produce concentration normalized libraries, pooling the normalized libraries with other samples to produce pooled libraries; and continuing with steps (f)–(i).
- Clause 8. The method of any one of clauses 1–6, wherein step (d) uses a suppression PCR method.
- Clause 9. The method of any one of clauses 1–7, wherein the RNA-guided endonuclease comprises an endogenously-expressed Cas enzyme, a Cas expression vector, a Cas protein, or a Cas RNP complex.
- Clause 10. The method of any one of clauses 1–8, wherein the RNA-guided endonuclease comprises an endogenously-expressed Cas9 enzyme, a Cas9 expression vector, a Cas9 protein, or a Cas9 RNP complex.
- Clause 11. The method of any one of clauses 1–9, wherein the cells comprise human or mouse cells.
- Clause 12. The method of any one of clauses 1–10, wherein the period of time is about 24 hours to about 96 hours.
- Clause 13. The method of any one of clauses 1–11, wherein multiple tag sequences are co-delivered.
- Clause 14. The method of any one of clauses 1–12, wherein the tag sequences comprise double-stranded deoxyribonucleotides (dsDNA) comprising 52-base pairs.
- Clause 15. The method of any one of clauses 1–13, wherein the tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides.
- Clause 16. The method of any one of clauses 1–14, wherein the tag sequences comprise a double stranded DNA comprising the complementary top and bottom strand pairs of SEQ ID NO: 1–2 or 7–268.
- Clause 17. On- and off-target CRISPR editing sites identified or nominated using the method of any one of clauses 1–15.
- Clause 18. A method for designing 52-base pair tag sequences, the method comprising, executing on a processor:
- (a) randomly generating 13-nucleotide sequences with 40–90% GC content, max homopolymer length A:2, C:3, G:2, T:2, weighted homopolymer rate < 20, self-folding $T_m < 50$ °C, and self-dimer $T_m < 50$ °C;
 - (b) removing sequences that perfectly align to a particular genome or that are homopolymers or GG or CC dinucleotide motifs and obtaining a set of 13-mers;

- (c) selecting a subset of the 13-mer sequences that contain one or less CC or GG dinucleotide motifs;
 - (d) concatenating four of the of 13-mer subset sequences to form random 52-mer sequences;
 - 5 (e) aligning the random 52-mer sequences to a genome;
 - (f) removing the random 52-mer sequences that have similarity to the genome to produce a subset of 52-mer sequences; and
 - (g) outputting the subset of 52-mer sequences and generating the complementary strands to produce double stranded 52-base pair tag sequences.
- 10 Clause 19. The method of clause 17, wherein the genome is human or mouse.
- Clause 20. The method of clause 17 or 18, wherein the 52-base pair tag sequences are-non complementary to the genome.
- Clause 21. The method of any one of clauses 17–19, further comprising designing primers for the 52-base pair tag sequences.
- 15 Clause 22. The method of any one of clauses 17–20, wherein the 52-base pair tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides of the 52-base pair tag sequences.
- Clause 23. The method of any one of clauses 17–21, further comprising synthesizing oligonucleotides comprising the 52-base pair tag sequences, the complement of the 52-
- 20 base pair tag sequences, or primers for the 52-base pair tag sequences.
- Clause 24. One or more 52-base pair tag sequences designed using the methods of clauses 17–22.
- Clause 25. The 52-base pair tag sequences of clause 23, wherein the 52-base pair tag sequence comprises a double stranded DNA comprising the top and bottom strand pairs of SEQ ID NO: 1–2 or 7–268.
- 25 Clause 26. A method for designing primers partially complementary to the 52-base pair tag sequences of clause 23 and an adapter primer, the method comprising, executing on a processor:
- (a) designing tag primers that are partially complementary to the top and bottom strands of tag sequences; and
 - 30 (b) designing an adapter primer that is partially complementary to the top strand of the adapter sequence;
 - (c) wherein:
 - (d) the tag primers comprise a 5'-universal tail sequence; and

- (e) the adapter primer comprises a sequence complementary to the tails of Tag-pTOP or Tag-pBOT primers.

Clause 27. The method of clause 25, wherein the 5'-universal tail sequence is complementary to an SP1 or SP2 sequence (SEQ ID NO: 7, 8), a locus specific segment, a ribonucleotide (rN) 6-nucleotides from the 3'-end, a 3'-end mismatch, a 3'-end block (3'-C₃ spacer), a
5 predesigned non-homologous sequence (SEQ ID NO: 269–273), or a predesigned 13-mer sequence.

Clause 28. The method of clause 25 or 26, wherein the primers partially complementary to top and bottom strands of the tag sequences comprise a tail sequence complementary to the
10 SP1 sequence (SEQ ID NO: 7) and the adapter primer comprises a sequence complementary to the SP2 sequence (SEQ ID NO: 8) tail on the Tag-pTOP or Tag-pBOT primers; or the primers partially complementary to top and bottom strands of the tag sequences comprise a tail sequence complementary to the SP2 sequence (SEQ ID NO: 8) and the adapter primer comprises a sequence complementary to the SP1 sequence
15 (SEQ ID NO: 7) tail on the Tag-pTOP or Tag-pBOT primers.

Clause 29. The method of any one of clauses 25–27, wherein the amplification of a nucleic acid molecule with the primers that are complementary to the top and bottom strands of tag sequences and primers that are complementary to the top strand of the adapter sequence produces a PCR product that comprises a portion of the tag sequence, a sgDNA
20 sequence, and the adapter sequence.

Clause 30. The method of any one of clauses 25–28, further comprising synthesizing oligonucleotides comprising the sequences of the forward and reverse tag primers and the adapter primer.

Clause 31. The method of any one of clauses 17–21 and 25–29, wherein the 52-base pair tag sequences and primers partially complementary to the 52-base pair tag sequences are designed and selected using an algorithm predicting whether the primers are likely to be partially complementary and have a propensity to form primer-dimers.
25

Clause 32. One or more primers partially complementary to the 52-base pair tag sequences and one or more adapter primers designed using the method of clauses 22–25.

30 Clause 33. The primers of clause 32, wherein the primers comprise the sequences of SEQ ID NO: 3, 4; and the adapter primer, wherein the adapter primer comprises the sequence of SEQ ID NO: 5.

Clause 34. Use of one or more double-stranded 52-base pair tag sequences for identifying on- and off-target CRISPR editing sites.

REFERENCES

1. Wienert et al., "Unbiased detection of CRISPR off-targets *in vivo* using DISCOVER-seq," *Science* 364(6437): 286–289 (2019).
- 5 2. Nobles et al., "IGUIDE: An improved pipeline for analyzing CRISPR cleavage specificity," *Genome Biol.* 20(14): 4–9 (2019).
3. Tsai et al., "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases," *Nature Biotechnol.* 33(2): 187–197 (2015).
4. Yan et al., "BLISS is a versatile and quantitative method for genome-wide profiling of DNA
10 double-strand breaks," *Nature Commun.* 8: 15058 (2017).
5. Tsai et al., "CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR-Cas9 nuclease off-targets," *Nature Methods* 14(6): 607-614 (2017).
6. Cameron et al., "Mapping the genomic landscape of CRISPR-Cas9 cleavage," *Nature Methods* 14(6): 600-606 (2017).
- 15 7. Char and Moosburner, "Unraveling CRISPR-Cas9 genome engineering parameters via a library- on-library approach," *Nature Methods* 12(9): 823-826 (2015).
8. Rand et al., "Headloop suppression PCR and its application to selective amplification of methylated DNA sequences," *Nucleic Acids Res.* 33(14) :e127 (2005).

20 EXAMPLES

Example 1

This experiment demonstrates the increased efficiency in tag integration when using double-stranded DNA tags with a length of 52-base pairs and varying genetic sequence. The sequences used are shown in Tables 3–5. Double-stranded tags were generated by hybridization
25 of a top strand and a complementary bottom strand (Tables 3–4; SEQ ID NO: 9–40 or 45–268). Sixteen different tag designs were introduced separately into HEK293 cells constitutively expressing Cas9 together with a guideRNA which targets the EMX1 locus. Alternatively, either pools of 16 tags or one pool of 112 tags were introduced into HEK293 cells constitutively expressing Cas9 together with a guideRNA which targets the EMX1 locus. GuideRNAs were
30 electroporated at a concentration of 10 μ M, whereas the single Tag or pooled Tags were delivered at a final concentration of 0.5 μ M. Tag integration levels were determined by targeted amplification using rhAmpSeq primers (SEQ ID NO: 3–4), enriching for known on- and off-target sites of the EMX1 guideRNA. The rhAmpSeq pool for EMX1 consists of 32 sites, which represent empirically determined ON and OFF target loci. Amplified products were sequenced on an

Illumina® MiSeq, and tag integration levels were determined using custom software. This example shows that tag integration efficiency varies among single tag constructs individually with a range between 6 (CTL021) and 13 (CTL169, CTL079, CTL002) sites out of a maximum of 32 sites, and is therefore sequence dependent (Single Tags, FIG. 7). By taking the mathematical union of the single tag results, a hypothetical number of 23 sites was calculated (CTLmax, FIG. 7). The hypothesis that combining a pool of tags would increase the likelihood of tag integration was tested and was demonstrated (Pooled Tags, Table, FIG. 7). Pool A1 consists of the tags represented in the Single Tags (see Table 5) and demonstrated that 21 tag integration events were detected out of a maximum of 32 sites, which is higher than achieved with any of the single tags. Similarly, Pool B3 demonstrated integration of a tag at 21 sites out of a maximum of 32 sites. Again, variability between pools was shown (Pooled Tags, FIG. 7), indicating optimization of tag designs can potentially maximize tag integration.

Table 3. Sequences Used for Second Proof of Concept

Name	Sequence (5'→3')	SEQ ID NO
CTL085_TOP_tag	/5Phos/A*C*GAGCGGTAGTCACCTAGTCGTCGTACCAATTTCG ACGCACACTACTCGC*G*C	SEQ ID NO: 9
CTL085_BOT_tag	/5Phos/G*C*GCGAGTAGTGTGCGTCGAATTGGTACGACGACT AGGTGACTACCGCTC*G*T	SEQ ID NO: 10
CTL169_TOP_tag	/5Phos/T*A*GCGCGAGTAGTCGGACGAGCGGTTACCAATACG CCGCACCTTAATCCG*C*G	SEQ ID NO: 11
CTL169_BOT_tag	/5Phos/C*G*CGGATTAAGGTGCGGCGTATTGGTAACCGCTCG TCCGACTACTCGCGC*T*A	SEQ ID NO: 12
CTL137_TOP_tag	/5Phos/T*C*GCGACAGTAGTCGTTTCGGCTAGGTACCTATTAC CGCGTAGTTAGCGGC*G*T	SEQ ID NO: 13
CTL137_BOT_tag	/5Phos/A*C*GCCGCTAATACGCGGTAATAGGTACCTAGCCG AACGACTACTGTCGC*G*A	SEQ ID NO: 14
CTL042_TOP_tag	/5Phos/C*G*CGCTACTAGGTGCGTCGAATTGGTACCGATCCG CAATACACTACTCGC*G*C	SEQ ID NO: 15
CTL042_BOT_tag	/5Phos/G*C*GCGAGTAGTGTATTGCGGATCGGTACCAATTTCG ACGCACCTAGTAGCG*C*G	SEQ ID NO: 16
CTL051_TOP_tag	/5Phos/G*G*TAACGAGCGGTGCGTCGAATTGGTAACCGCTCG TCCGACCTTAATCGC*G*C	SEQ ID NO: 17
CTL051_BOT_tag	/5Phos/G*C*GCGATTAAGGTGCGACGAGCGGTTACCAATTTCG ACGCACCGCTCGTTA*C*C	SEQ ID NO: 18
CTL167_TOP_tag	/5Phos/T*T*CGGCGCTAGGTGCGGCGTATTGGTAACCGCTCG TCCGTTTCGGCGCTAG*G*T	SEQ ID NO: 19
CTL167_BOT_tag	/5Phos/A*C*CTAGCGCGAACGACGAGCGGTTACCAATACG CCGCACCTAGCGCCG*A*A	SEQ ID NO: 20
CTL026_TOP_tag	/5Phos/T*A*CGCGACTAGGTGCGCGATTAAGGTACCTATTAC CGCGCGACTATGTGC*G*C	SEQ ID NO: 21
CTL026_BOT_tag	/5Phos/G*C*GCACATAGTCGCGCGGTAATAGGTACCTTAATC GCGCACCTAGTCGCG*T*A	SEQ ID NO: 22
CTL068_TOP_tag	/5Phos/G*T*CGCGCAGTGTAGCGCGATTAAGGTACCTATTAC CGCGTCGCGACAGTA*G*T	SEQ ID NO: 23
CTL068_BOT_tag	/5Phos/A*C*TACTGTGCGACGCGGTAATAGGTACCTTAATC GCGCTACACTGCGCG*A*C	SEQ ID NO: 24

CTL138_TOP_tag	/5Phos/A*A*CCGTCGATCCGCGCGTAGTATGGTACCGATCCG CAATACTAGCGCGAC*A*A	SEQ ID NO: 25
CTL138_BOT_tag	/5Phos/T*T*GTCGCGCTAGTATTGCGGATCGGTACCATACTA CGCGCGGATCGACGG*T*T	SEQ ID NO: 26
CTL079_TOP_tag	/5Phos/T*C*GCTCGATTGGTTACGCGCACTACTTATGCGCTC GACTCGTTCGGCTAG*G*T	SEQ ID NO: 27
CTL079_BOT_tag	/5Phos/A*C*CTAGCCGAACGAGTCGAGCGCATAAGTAGTGCG CGTAACCAATCGAGC*G*A	SEQ ID NO: 28
CTL063_TOP_tag	/5Phos/A*C*TGCGAGCGTACTTGTGCGCGTAGTACCAATTTCG ACGCAACCGCTCGTC*C*G	SEQ ID NO: 29
CTL063_BOT_tag	/5Phos/C*G*GACGAGCGGTTGCGTCGAATTGGTACTAGCGCG ACAAGTACGCTCGCA*G*T	SEQ ID NO: 30
CTL168_TOP_tag	/5Phos/C*G*CATTAGTCGGTGC GGCGTATTGGTAACCGCTCG TCCGACGCGCTACCT*A*T	SEQ ID NO: 31
CTL168_BOT_tag	/5Phos/A*T*AGGTAGCGCGTCGGACGAGCGGTTACCAATACG CCGCACCGACTAATG*C*G	SEQ ID NO: 32
CTL021_TOP_tag	/5Phos/A*T*TGCGGATCGGTGCGTCGAATTGGTAACCGCTCG TCCGTACGCGCACTA*C*T	SEQ ID NO: 33
CTL021_BOT_tag	/5Phos/A*G*TAGTGC GCGTACGGACGAGCGGTTACCAATTTCG ACGCACCGATCCGCA*A*T	SEQ ID NO: 34
CTL151_TOP_tag	/5Phos/T*C*GGCGAGTAGTTGCGCGGTTATGGTACCATAACC GCGCAGTAGTACGCG*G*T	SEQ ID NO: 35
CTL151_BOT_tag	/5Phos/A*C*CGCGTACTACTGCGCGGTTATGGTACCATAACC GCGCAACTACTCGCC*G*A	SEQ ID NO: 36
CTL002_TOP_tag	/5Phos/A*C*TAGCGATCGGTACCTAGCGCCGAAACCTATTAC CGCGACCTAGCGTTG*C*G	SEQ ID NO: 37
CTL002_BOT_tag	/5Phos/C*G*CAACGCTAGGTCGCGGTAATAGGTTTCGGCGCT AGGTACCGATCGCTA*G*T	SEQ ID NO: 38
CTL134_TOP_tag	/5Phos/T*A*GCGCGTCAAGAGCGCGGTTATGGTTTCGGCGCT AGGTTAACAGCGCGT*C*G	SEQ ID NO: 39
CTL134_BOT_tag	/5Phos/C*G*ACGCGCTGTTAACCTAGCGCCGAAACCATTAACC GCGCTCTTGACGCGC*T*A	SEQ ID NO: 40
GuideSeq_TOP_tag	/5Phos/G*T*TTAATTGAGTTGTCATATGTTAATAACGGT*A* T	SEQ ID NO: 41
GuideSeq_BOT_tag	/5Phos/A*T*ACCGTTATTAACATATGACAACTCAATTAA*A* C	SEQ ID NO: 42
EMX1 protospacer	GAGTCCGAGCAGAAGAAGAA	SEQ ID NO: 43
AR protospacer	GTTGGAGCATCTGAGTCCAG	SEQ ID NO: 44

"/5Phos/" indicates a 5'-phosphate moiety; "*" indicates a phosphorothioate linkage.

Example 2

This experiment demonstrates the increased efficiency in tag integration when using double-stranded DNA tags with a length of 52-base pairs and varying genetic sequence. The sequences used are shown in Tables 3–5. Double-stranded tags were generated by hybridization of a top strand and a complementary bottom strand (SEQ ID NO: 9–40 or 45–268). Sixteen different tag designs were introduced separately into HEK293 cells constitutively expressing Cas9 together with a guideRNA which targets the AR locus. Alternatively, either pools of 16 tags or one pool of 112 tags were introduced into HEK293 cells constitutively expressing Cas9 together with a guideRNA which targets the AR locus. GuideRNAs were electroporated at a concentration

of 10 μ M, whereas the single Tag or pooled Tags were delivered at a final concentration of 0.5 μ M. Tag integration levels were determined by targeted amplification using rhAmpSeq primers (SEQ ID NO: 3–4), enriching for known on- and off-target sites of the AR guideRNA. The rhAmpSeq pool for AR consists of 53 sites which represent empirically determined ON and OFF target loci. Amplified products were sequenced on an Illumina® MiSeq, and tag integration levels were determined using custom software. This example shows that tag integration efficiency varies among single tag constructs individually with a range between 35 (CTL085, CTL134) and 41 sites (CTL002) out of a maximum of 53 sites, and is therefore sequence dependent (Single Tags, Table 5, FIG. 8).

By taking the mathematical union of the single tag results, a hypothetical number of 47 sites was calculated (CTLmax, FIG. 8). The hypothesis that combining a pool of tags would increase the likelihood of tag integration was tested and was demonstrated (Pooled Tags, Table 5, FIG. 8). Pool B4 (see Table 5) demonstrated that 44 tag integration events were detected out of a maximum of 53 sites, which is higher than achieved with any of the single tags. Again, variability between pools was shown (Pooled Tags, Table 5, FIG. 8), indicating optimization of tag designs can potentially maximize tag integration.

Table 4. Tag Sequences

Name	Sequence (5'→3')	SEQ ID NO
CTL085_TOP_tag	/5Phos/A*C*GAGCGGTAGTCACCTAGTCGTCGTACCAATTCTGA CGCACACTACTCGC*G*C	SEQ ID NO: 45
CTL169_TOP_tag	/5Phos/T*A*GCGGAGTAGTCGGACGAGCGGTTACCAATACGC CGCACCTTAATCCG*C*G	SEQ ID NO: 46
CTL137_TOP_tag	/5Phos/T*C*GCGACAGTAGTCGTTCCGGCTAGGTACCTATTACC GCGTAGTTAGCGGC*G*T	SEQ ID NO: 47
CTL042_TOP_tag	/5Phos/C*G*CGCTACTAGGTGCGTCGAATTGGTACCGATCCGC AATACACTACTCGC*G*C	SEQ ID NO: 48
CTL051_TOP_tag	/5Phos/G*G*TAACGAGCGGTGCGTCGAATTGGTAACCGCTCGT CCGACCTTAATCGC*G*C	SEQ ID NO: 49
CTL167_TOP_tag	/5Phos/T*T*CGGCGCTAGGTGCGGCGTATTGGTAACCGCTCGT CCGTTGCGGCGTAG*G*T	SEQ ID NO: 50
CTL026_TOP_tag	/5Phos/T*A*GCGGACTAGGTGCGCGATTAAGGTACCTATTACC GCGGCGACTATGTGC*G*C	SEQ ID NO: 51
CTL068_TOP_tag	/5Phos/G*T*CGGCGAGTGTAGCGCGATTAAGGTACCTATTACC GCGTCGCGACAGTA*G*T	SEQ ID NO: 52
CTL138_TOP_tag	/5Phos/A*A*CCGTCGATCCGCGCGTAGTATGGTACCGATCCGC AATACTAGCGCGAC*A*A	SEQ ID NO: 53
CTL079_TOP_tag	/5Phos/T*C*GCTCGATTGGTTACGCGCACTACTTATGCGCTCG ACTCGTTCGGCTAG*G*T	SEQ ID NO: 54
CTL063_TOP_tag	/5Phos/A*C*TGCGAGCGTACTTGTGCGGCTAGTACCAATTCTGA CGCAACCGCTCGTC*G*G	SEQ ID NO: 55
CTL168_TOP_tag	/5Phos/C*G*CATTAGTCGGTGC GGCGTATTGGTAACCGCTCGT CCGACGCGCTACCT*A*T	SEQ ID NO: 56
CTL021_TOP_tag	/5Phos/A*T*TGCGGATCGGTGCGTCGAATTGGTAACCGCTCGT CCGTACGCGCACTA*C*T	SEQ ID NO: 57

CTL151_TOP_tag	/5Phos/T*C*GGCGAGTAGTTGCGCGGTTATGGTACCATAACCG CGCAGTAGTACGCG*G*T	SEQ ID NO: 58
CTL002_TOP_tag	/5Phos/A*C*TAGCGATCGGTACCTAGCGCCGAAACCTATTACC GCGACCTAGCGTTG*C*G	SEQ ID NO: 59
CTL134_TOP_tag	/5Phos/T*A*GCGCGTCAAGAGCGCGGTTATGGTTTCGGCGCTA GGTTAACAGCGCGT*C*G	SEQ ID NO: 60
CTL085_BOT_tag	/5Phos/G*C*GCGAGTAGTGTGCGTCGAATTGGTACGACGACTA GGTGACTACCGCTC*G*T	SEQ ID NO: 61
CTL169_BOT_tag	/5Phos/C*G*CGGATTAAGGTGCGGCGTATTGGTAACCGCTCGT CCGACTACTCGCGC*T*A	SEQ ID NO: 62
CTL137_BOT_tag	/5Phos/A*C*GCCGCTAACTACGCGGTAATAGGTACCTAGCCGA ACGACTACTGTCGC*G*A	SEQ ID NO: 63
CTL042_BOT_tag	/5Phos/G*C*GCGAGTAGTGTATTGCGGATCGGTACCAATTCTGA CGCACCTAGTAGCG*C*G	SEQ ID NO: 64
CTL051_BOT_tag	/5Phos/G*C*GCGATTAAGGTGCGACGAGCGGTTACCAATTCTGA CGCACCGCTCGTTA*C*C	SEQ ID NO: 65
CTL167_BOT_tag	/5Phos/A*C*CTAGCGCGAACGAGACGAGCGGTTACCAATACGC CGCACCTAGCGCGC*A*A	SEQ ID NO: 66
CTL026_BOT_tag	/5Phos/G*C*GCACATAGTCGCGCGGTAATAGGTACCTTAATCG CGCACCTAGTCGCG*T*A	SEQ ID NO: 67
CTL068_BOT_tag	/5Phos/A*C*TACTGTGCGGACGCGGTAATAGGTACCTTAATCG CGCTACACTGCGCG*A*C	SEQ ID NO: 68
CTL138_BOT_tag	/5Phos/T*T*GTCGCGCTAGTATTGCGGATCGGTACCATACTAC GCGCGGATCGACGG*T*T	SEQ ID NO: 69
CTL079_BOT_tag	/5Phos/A*C*CTAGCCGAACGAGTCGAGCGCATAAGTAGTGCGC GTAACCAATCGAGC*G*A	SEQ ID NO: 70
CTL063_BOT_tag	/5Phos/C*G*GACGAGCGGTTGCGTCGAATTGGTACTAGCGCGA CAAGTACGCTCGCA*G*T	SEQ ID NO: 71
CTL168_BOT_tag	/5Phos/A*T*AGGTAGCGCGTCGGACGAGCGGTTACCAATACGC CGCACCGACTAATG*C*G	SEQ ID NO: 72
CTL021_BOT_tag	/5Phos/A*G*TAGTGCGCGTACGGACGAGCGGTTACCAATTCTGA CGCACCGATCCGCA*A*T	SEQ ID NO: 73
CTL151_BOT_tag	/5Phos/A*C*CGCGTACTACTGCGCGGTTATGGTACCATAACCG CGCAACTACTCGCC*G*A	SEQ ID NO: 74
CTL002_BOT_tag	/5Phos/C*G*CAACGCTAGGTGCGGTAATAGGTTTCGGCGCTA GGTACCGATCGCTA*G*T	SEQ ID NO: 75
CTL134_BOT_tag	/5Phos/C*G*ACGCGCTGTTAACCTAGCGCCGAAACCTAACCG CGCTCTTGACGCGC*T*A	SEQ ID NO: 76
CTL161_TOP_tag	/5Phos/T*A*CACTGCGCGACACTGCGAGCGTACACCTTAATCG CGCTAGTTAGCGGC*G*T	SEQ ID NO: 77
CTL164_TOP_tag	/5Phos/A*A*CCGTCGAGTGCACCGCGTACTACTAATGTGGAAC CGCTACGCGCACTA*C*T	SEQ ID NO: 78
CTL030_TOP_tag	/5Phos/C*G*CGGACTAAGGTGCGCGAGTAGTGTTACGCGCACT ACTAATCTAGCCGC*G*A	SEQ ID NO: 79
CTL088_TOP_tag	/5Phos/A*C*TAGTGCGACGAACTACTCGCGCTAACCAATTCTGA CGCACCGATCGCTA*G*T	SEQ ID NO: 80
CTL148_TOP_tag	/5Phos/A*A*TGTCGAACCGCGCGGAGTAGTGTTACCATAACCG CGCACCTTAGTCCG*C*G	SEQ ID NO: 81
CTL152_TOP_tag	/5Phos/G*C*GTCGAATTGGTACCGCGGACTTATACCAATACGC CGCATAGGTAGCGC*G*T	SEQ ID NO: 82
CTL007_TOP_tag	/5Phos/A*C*CTAGTAGCGCGGCGTGAATTGGTACTAGCGCGA CAACGCGTAGTATG*G*T	SEQ ID NO: 83
CTL141_TOP_tag	/5Phos/A*C*CGCTCGTTACCGCGGATTAAGGTACGCCGCTAA CTACGGTACGGTCG*G*T	SEQ ID NO: 84
CTL064_TOP_tag	/5Phos/A*C*CGCCGACTTATCGTTGCGCTAGGTACCAATTCTGA CGCACTGCGAGCGT*A*C	SEQ ID NO: 85

CTL158_TOP_tag	/5Phos/A*C*CTTAATCCGCGACTGCGAGCGTACACCTATTACC GCGCGACGCGCTGT*T*A	SEQ ID NO: 86
CTL066_TOP_tag	/5Phos/A*C*GACGACTAGGTACCGCTCGTTACCTCTTGACGCG CTAACCAATTCGAC*G*C	SEQ ID NO: 87
CTL144_TOP_tag	/5Phos/A*C*CATACTACGCGCGGTTTCGACATTACCATAACCG CGCTAGTGCGAGCG*T*A	SEQ ID NO: 88
CTL107_TOP_tag	/5Phos/C*T*TGTACGGCGGTGCGGCGTATTGGTACCAATACGC CGCTCGTCGCACTA*G*T	SEQ ID NO: 89
CTL149_TOP_tag	/5Phos/G*T*ACGCTCGCAGTACCGCCGACTTATACCTTAATCG CGCACTAGCGCGAC*A*A	SEQ ID NO: 90
CTL008_TOP_tag	/5Phos/A*C*GACGACTAGGTTATGGTACGGCGTTAGCGCGAGT AGTACCTTAGTCCG*C*G	SEQ ID NO: 91
CTL099_TOP_tag	/5Phos/A*C*GAGCGGTAGTCATAGGTAGCGCGTTCTTGACGCG CTAACCGATCGCTA*G*T	SEQ ID NO: 92
CTL089_TOP_tag	/5Phos/A*C*CGATCCGCAATGCGTCGAATTGGTACCATAACCG CGCACCGCCGTACA*A*G	SEQ ID NO: 93
CTL081_TOP_tag	/5Phos/A*C*TAGTGCGACGAAGTACTGTCGCGAACCTATTACC GCGACCAATCGAGC*G*A	SEQ ID NO: 94
CTL075_TOP_tag	/5Phos/A*C*CGCGGTACAAGTCGCGACAGTAGTAACCGCTCGT CCGTTGCGGCGCTAG*G*T	SEQ ID NO: 95
CTL160_TOP_tag	/5Phos/T*C*GTCGCACTAGTCGCATTAGTCGGTAGTAGTACGC GGTATAGGTAGCGC*G*T	SEQ ID NO: 96
CTL133_TOP_tag	/5Phos/A*C*CAATTCGACGCTAGTTAGCGGCGTACACTACTCG CGCGCACTCGACGG*T*T	SEQ ID NO: 97
CTL076_TOP_tag	/5Phos/C*G*CGGTAATAGGTGCGGTAATAGGTACGAGCGGTA GTCACACTACTCGC*G*C	SEQ ID NO: 98
CTL024_TOP_tag	/5Phos/T*C*GGCGAGTAGTTTAGTGCGAGCGTAAGTAGTGCGC GTAACCAATCGAGC*G*A	SEQ ID NO: 99
CTL045_TOP_tag	/5Phos/G*T*CGCGCAGTGTAGCGCGGTTATGGTACCATAACCG CGCACTAGTGCGAC*G*A	SEQ ID NO: 100
CTL009_TOP_tag	/5Phos/T*A*TGCGCTCGACTGCGCGATTAAGGTAATGTGCAAC CGCAGTAGTACGCG*G*T	SEQ ID NO: 101
CTL055_TOP_tag	/5Phos/A*C*TAGCGCGACAACGACTATGTGCGCACCAATTCGA CGCTACGCGCACTA*C*T	SEQ ID NO: 102
CTL101_TOP_tag	/5Phos/A*A*CTACTCGCCGACTTGTACGGCGGTACCAATTCGA CGCAACTAATCCGC*G*C	SEQ ID NO: 103
CTL135_TOP_tag	/5Phos/C*G*CGGATTAAGGTCTGTACGGCGGTACCTAGCCGA ACGTACGCGCACTA*C*T	SEQ ID NO: 104
CTL155_TOP_tag	/5Phos/T*A*GCGCGTCAAGACTTGTACGGCGGTACCGATCCGC AATGCACTCGACGG*T*T	SEQ ID NO: 105
CTL122_TOP_tag	/5Phos/C*G*CATTAGTCGGTGC GGCGTATTGGTACGACGACTA GGTACCAATACGCC*G*C	SEQ ID NO: 106
CTL080_TOP_tag	/5Phos/A*C*CTAGTAGCGCGCGCGGTTATGGTACCGACTAAT GCGACTAGCGATCG*G*T	SEQ ID NO: 107
CTL126_TOP_tag	/5Phos/A*C*TACTCGCGCTAACCTAGTCGTCGTAATCTAGCCG CGATACGCTCGCAC*T*A	SEQ ID NO: 108
CTL098_TOP_tag	/5Phos/A*C*CGCCGCTATACGCGCGATTAAGGTGTACGCTCGC AGTCGCGGACTAAG*G*T	SEQ ID NO: 109
CTL038_TOP_tag	/5Phos/T*A*CGCGCACTACTAACCGTCGAGTGC GTACGCTCGC AGTACCGATCGCTA*G*T	SEQ ID NO: 110
CTL139_TOP_tag	/5Phos/G*T*CGCGCAGTGTATAACAGCGCGTCGTTAGTGCGCG AGAACGACGACTAG*G*T	SEQ ID NO: 111
CTL010_TOP_tag	/5Phos/G*C*GTCGAATTGGTCGCGTAGTATGGTACCGCCGCTA TACACCAATACGCC*G*C	SEQ ID NO: 112
CTL034_TOP_tag	/5Phos/T*A*CGCGCACTACTTACGCGACTAGGTACCGATCGCT AGTCGACGCGCTGT*T*A	SEQ ID NO: 113

CTL117_TOP_tag	/5Phos/A*C*GCCGCTAACTATAGTTAGCGGCGTACCAATTCTGA CGCAACTAATCCGC*G*C	SEQ ID NO: 114
CTL035_TOP_tag	/5Phos/C*G*CGGACTAAGGTTAGTTAGCGGCGTTACGCGCACT ACTACCGATCCGCA*A*T	SEQ ID NO: 115
CTL121_TOP_tag	/5Phos/A*C*GACGACTAGGTACCGCCGACTTATACGCCGCTAA CTAATAGGTAGCGC*G*T	SEQ ID NO: 116
CTL106_TOP_tag	/5Phos/C*G*GATCGACGGTTGCGCGAGTAGTGTAGTAGTACGC GGTTACACTGCGCG*A*C	SEQ ID NO: 117
CTL059_TOP_tag	/5Phos/A*T*TGCGGATCGGTACCGCCGACTTATACCGATCCGC AATTCGCTCGATTG*G*T	SEQ ID NO: 118
CTL157_TOP_tag	/5Phos/A*C*TGCGAGCGTACACTGCGAGCGTACACCTTAATCG CGCACCGCTCGTTA*C*C	SEQ ID NO: 119
CTL015_TOP_tag	/5Phos/A*C*TACTGTGCGGATCGTGCCTAGTTACGCTCGCA CTAATTGCGGATCG*G*T	SEQ ID NO: 120
CTL110_TOP_tag	/5Phos/G*G*TAACGAGCGGTTCTCGCGCACTAATTAGTGC GCG AGAACCATACTACG*C*G	SEQ ID NO: 121
CTL123_TOP_tag	/5Phos/A*C*TACTGCGCTAGCGGATTAAGGTACCTTAATCG CGCAACTACTCGCC*G*A	SEQ ID NO: 122
CTL014_TOP_tag	/5Phos/T*A*CGCGCACTACTCTGTACGGCGGTACCAATTCTGA CGCAACCGTCGAGT*G*C	SEQ ID NO: 123
CTL131_TOP_tag	/5Phos/A*A*CCGTCGATCCGATTGCGGATCGGTACCTTAATCG CGCACTAGTGCAC*G*A	SEQ ID NO: 124
CTL062_TOP_tag	/5Phos/A*G*TAGTGCGCGTATACACTGCGCGACACACTACTCG CGCACCTTAATCCG*C*G	SEQ ID NO: 125
CTL044_TOP_tag	/5Phos/A*C*GCCGTACCATACGCGGTAATAGGTAGTAGTGCGC GTATTCGGCGCTAG*G*T	SEQ ID NO: 126
CTL043_TOP_tag	/5Phos/T*A*GCGCGTCAAGAACCTAGCGTTGCGATAAGTCCGGC GGTAGTAGTACGCG*G*T	SEQ ID NO: 127
CTL118_TOP_tag	/5Phos/C*G*CATTAGTCGGTAATCTAGCCGCGAACCATAAACCG CGCACCGATCGCTA*G*T	SEQ ID NO: 128
CTL128_TOP_tag	/5Phos/T*A*TGGTACGGCGTGCGGCGTATTGGTACGCCGCTAA CTAATAAGTCGCGC*G*T	SEQ ID NO: 129
CTL067_TOP_tag	/5Phos/G*C*GCGGTTATGGTGCGGCGTATTGGTACGAGCGGTA GTCAACCGCTCGTC*C*G	SEQ ID NO: 130
CTL020_TOP_tag	/5Phos/C*G*ACTATGTGCGCACTACTCGCCGAACCATAAACCG CGCTATGCGCTCGA*C*T	SEQ ID NO: 131
CTL006_TOP_tag	/5Phos/T*A*GTTAGCGGCGTACCGCTCGTTACCACCTTAATCG CGCACCATACACTACG*C*G	SEQ ID NO: 132
CTL017_TOP_tag	/5Phos/C*G*CATTAGTCGGTAGTAGTGCGCGTAAACCGCTCGT CCGTTAGTGCGCGA*G*A	SEQ ID NO: 133
CTL057_TOP_tag	/5Phos/T*A*GCGCGAGTAGTACCGACTAATGCGTCTCGCGCAC TAAGACTACCGCTC*G*T	SEQ ID NO: 134
CTL078_TOP_tag	/5Phos/T*A*CGCTCGCACTATCGCTCGATTGGTACCGCCGCTA TACACCATAACCGC*G*C	SEQ ID NO: 135
CTL031_TOP_tag	/5Phos/A*C*CAATCGAGCGAAGTCGAGCGCATAACGCGCTACC TATACGCCGCTAAC*T*A	SEQ ID NO: 136
CTL136_TOP_tag	/5Phos/A*C*CTAATCCGCGACTGCGAGCGTACACCGACTAAT GCGACTACTGTGCGC*G*A	SEQ ID NO: 137
CTL165_TOP_tag	/5Phos/A*G*TAGTGCGCGTATCGCTCGATTGGTTCTTGACGCG CTAGTATAGCGGCG*G*T	SEQ ID NO: 138
CTL039_TOP_tag	/5Phos/T*C*GTCGCACTAGTCGGTACGGTCGGTGCGCACATAG TCGTATGGTACGCG*G*T	SEQ ID NO: 139
CTL036_TOP_tag	/5Phos/C*G*CGGATTAAGGTAGTCGAGCGCATAACCGCGTACT ACTACGACGACTAG*G*T	SEQ ID NO: 140
CTL048_TOP_tag	/5Phos/C*G*ACTATGTGCGCTACGCTCGCACTAACACTACTCG CGCACCTAGCGCCG*A*A	SEQ ID NO: 141

CTL053_TOP_tag	/5Phos/A*C*CGCCGACTTATTCTCGCGCACTAATCGTCGCACT AGTAACCGTCGATC*C*G	SEQ ID NO: 142
CTL072_TOP_tag	/5Phos/A*C*CTAGCGTTGCGACCGACTAATGCGGGTAACGAGC GGTTATGGTACGGC*G*T	SEQ ID NO: 143
CTL096_TOP_tag	/5Phos/C*G*CGCTACTAGGTCGCGGTAATAGGTACCTAGCGTT GCGACCTAGTCGCG*T*A	SEQ ID NO: 144
CTL150_TOP_tag	/5Phos/C*G*TTCGGCTAGGTACTACTCGCGCTACGCATTAGTC GGTTCGCGACAGTA*G*T	SEQ ID NO: 145
CTL084_TOP_tag	/5Phos/C*G*GACGAGCGGTTTCGCGGTAATAGGTACGACGACTA GGTTAGTTAGCGGC*G*T	SEQ ID NO: 146
CTL142_TOP_tag	/5Phos/T*A*CGCTCGCACTAATTGCGGATCGGTACCGACTAAT GCGACCGCGTACTA*C*T	SEQ ID NO: 147
CTL102_TOP_tag	/5Phos/A*C*CGACCGTACCGTATGGTACGGCGTTCTTGACGCG CTAACCTAGCGCCG*A*A	SEQ ID NO: 148
CTL154_TOP_tag	/5Phos/G*C*GCGGATTAGTTAACCGTCGAGTGCACACTACTCG CGCACTGCGAGCGT*A*C	SEQ ID NO: 149
CTL112_TOP_tag	/5Phos/A*C*CTTAATCCGCGACCGACTAATGCGTACGCGCACT ACTATAAGTCGGCG*G*T	SEQ ID NO: 150
CTL145_TOP_tag	/5Phos/A*C*CTTAATCCGCGCGCGGTTATGGTACCGACTAAT GCGAACCGCTCGTC*C*G	SEQ ID NO: 151
CTL060_TOP_tag	/5Phos/A*C*TGCGAGCGTACCTTGTACGGCGGTACCTAGTAGC GCGATAAGTCGGCG*G*T	SEQ ID NO: 152
CTL016_TOP_tag	/5Phos/T*T*CGGCGCTAGGTACCTTAGTCCGCGTTTCGGCGCTA GGTACCTAGCGTTG*C*G	SEQ ID NO: 153
CTL159_TOP_tag	/5Phos/A*C*CTAGTCGCGTACTTGTACGGCGGTACCTAGCCGA ACGAACCGTCGAGT*G*C	SEQ ID NO: 154
CTL056_TOP_tag	/5Phos/A*C*CATAACCGCGCTACACTGCGCGACACCAATACGC CGCTATGGTACGGC*G*T	SEQ ID NO: 155
CTL162_TOP_tag	/5Phos/A*C*ACTACTCGCGCTACGCGACTAGGTAATGTCGAAC CGCACGCCGCTAAC*T*A	SEQ ID NO: 156
CTL018_TOP_tag	/5Phos/A*C*CGACTAATGCGTAACAGCGCGTCGTTAGTGCGCG AGAACCTTAATCGC*G*C	SEQ ID NO: 157
CTL115_TOP_tag	/5Phos/A*C*GCCGTACCATAACCGACTAATGCGATAAGTCGGC GGTACCAATACGCC*G*C	SEQ ID NO: 158
CTL033_TOP_tag	/5Phos/G*T*ACGCTCGCAGTCGCGGTAATAGGTTTCGGCGAGTA GTTACCATAACCGC*G*C	SEQ ID NO: 159
CTL047_TOP_tag	/5Phos/C*G*GACGAGCGGTTGCGCGGTTATGGTACTAGTGCGA CGAGCGCACATAGT*C*G	SEQ ID NO: 160
CTL108_TOP_tag	/5Phos/A*C*TACTCGCGCTAGCGCGATTAAGGTACGCCGCTAA CTATCGCGGCTAGA*T*T	SEQ ID NO: 161
CTL041_TOP_tag	/5Phos/A*C*CAATTCGACGCACTAATCCGCGCACCAATTCGA CGCAGTAGTGCGCG*T*A	SEQ ID NO: 162
CTL061_TOP_tag	/5Phos/A*C*CGCCGCTATACACCTAGCGCCGAAGTACGCTCGC AGTGTATAGCGGCG*G*T	SEQ ID NO: 163
CTL166_TOP_tag	/5Phos/A*C*ACTACTCGCGCCGACGAGCGGTTACCAATACGC CGCTAGCGCGAGTA*G*T	SEQ ID NO: 164
CTL012_TOP_tag	/5Phos/T*C*GTCGCACTAGTACCTTAATCCGCGCGCAACGCTA GGTACACTACTCGC*G*C	SEQ ID NO: 165
CTL052_TOP_tag	/5Phos/C*G*CGCTACTAGGTACCGACTAATGCGCGCAACGCTA GGTAATGTGCAACC*G*C	SEQ ID NO: 166
CTL153_TOP_tag	/5Phos/A*C*GAGCGGTAGTCACTACTGTGCGACGCAACGCTA GGTTACACTGCGCG*A*C	SEQ ID NO: 167
CTL094_TOP_tag	/5Phos/A*C*CTAGTCGCGTACGCGTAGTATGGTACCGATCGCT AGTGGTAACGAGCG*G*T	SEQ ID NO: 168
CTL095_TOP_tag	/5Phos/G*C*GGTTCGACATTACCGACTAATGCGTATGCGCTCG ACTACCTAGCGTTG*C*G	SEQ ID NO: 169

CTL105_TOP_tag	/5Phos/A*C*TGCGAGCGTACTCTCGCGCACTAAACGCCGCTAACTACGCGCTACTAG*G*T	SEQ ID NO: 170
CTL109_TOP_tag	/5Phos/C*G*GTACGGTCGGTAATCTAGCCGCGAACCTTAGTCCGCGACCGCCGTACA*A*G	SEQ ID NO: 171
CTL032_TOP_tag	/5Phos/T*C*GGCGAGTAGTTACGCGCTACCTATTCGCGGCTAGATTACGCCGCTAAC*T*A	SEQ ID NO: 172
CTL161_BOT_tag	/5Phos/A*C*GCCGCTAACTAGCGCGATTAAGGTGTACGCTCGCAGTGTGCGCAGTG*T*A	SEQ ID NO: 173
CTL164_BOT_tag	/5Phos/A*G*TAGTGCGCGTAGCGGTTTCGACATTAGTAGTACGCGGTGCACTCGACGG*T*T	SEQ ID NO: 174
CTL030_BOT_tag	/5Phos/T*C*GCGGCTAGATTAGTAGTGCGCGTAACACTACTCGCGCACCTTAGTCCG*C*G	SEQ ID NO: 175
CTL088_BOT_tag	/5Phos/A*C*TAGCGATCGGTGCGTCGAATTGGTTAGCGCGAGTAGTTCGTCGCACTA*G*T	SEQ ID NO: 176
CTL148_BOT_tag	/5Phos/C*G*CGGACTAAGGTGCGCGGTTATGGTACACTACTCGCGCGGTTTCGACA*T*T	SEQ ID NO: 177
CTL152_BOT_tag	/5Phos/A*C*GCGCTACCTATGCGGCGTATTGGTATAAGTCGGCGGTACCAATTCGAC*A*G	SEQ ID NO: 178
CTL007_BOT_tag	/5Phos/A*C*CATACTACGCGTTGTGCGCGTAGTACCAATTCGACGCCGCGCTACTAG*G*T	SEQ ID NO: 179
CTL141_BOT_tag	/5Phos/A*C*CGACCGTACCGTAGTTAGCGGCGTACCTTAATCGCGCGTAACGAGCG*G*T	SEQ ID NO: 180
CTL064_BOT_tag	/5Phos/G*T*ACGCTCGCAGTGCCTCGAATTGGTACCTAGCCGACGATAAGTCGGCG*G*T	SEQ ID NO: 181
CTL158_BOT_tag	/5Phos/T*A*ACAGCGCGTCGCGCGGTAATAGGTGTACGCTCGCAGTCGGGATTAAG*G*T	SEQ ID NO: 182
CTL066_BOT_tag	/5Phos/G*C*GTCGAATTGGTTAGCGCGTCAAGAGGTAACGAGCGGTACCTAGTCGTC*G*T	SEQ ID NO: 183
CTL144_BOT_tag	/5Phos/T*A*CGCTCGCACTAGCGCGGTTATGGTAATGTCGAACGCCGCGTAGTATG*G*T	SEQ ID NO: 184
CTL107_BOT_tag	/5Phos/A*C*TAGTGCGACGAGCGGCGTATTGGTACCAATACGCGCACCGCCGTACA*A*G	SEQ ID NO: 185
CTL149_BOT_tag	/5Phos/T*T*GTCGCGCTAGTGCGCGATTAAGGTATAAGTCGGCGGTACTGCGAGCGT*A*C	SEQ ID NO: 186
CTL008_BOT_tag	/5Phos/C*G*CGGACTAAGGTACTACTCGCGCTAACGCCGTACCATACCTAGTCGTC*G*T	SEQ ID NO: 187
CTL099_BOT_tag	/5Phos/A*C*TAGCGATCGGTTAGCGCGTCAAGAACGCGCTACCTATGACTACCGCTC*G*T	SEQ ID NO: 188
CTL089_BOT_tag	/5Phos/C*T*GTACGGCGGTGCGCGGTTATGGTACCAATTCGACGCATTGCGGATCG*G*T	SEQ ID NO: 189
CTL081_BOT_tag	/5Phos/T*C*GCTCGATTGGTCGCGGTAATAGGTTTCGCGACAGTAGTTCGTCGCACTA*G*T	SEQ ID NO: 190
CTL075_BOT_tag	/5Phos/A*C*CTAGCGCCGAACGGACGAGCGGTTACTACTGTCCGACTTGTACGGCG*G*T	SEQ ID NO: 191
CTL160_BOT_tag	/5Phos/A*C*GCGCTACCTATACCGCGTACTACTACCGACTAATGCGACTAGTGCGAC*G*A	SEQ ID NO: 192
CTL133_BOT_tag	/5Phos/A*A*CCGTCGAGTGCGCGCGAGTAGTGTACGCCGCTAACTAGCGTCGAATTG*G*T	SEQ ID NO: 193
CTL076_BOT_tag	/5Phos/G*C*GCGAGTAGTGTGACTACCGCTCGTACCTATTACCGGACCTATTACCG*C*G	SEQ ID NO: 194
CTL024_BOT_tag	/5Phos/T*C*GCTCGATTGGTTACGCGCACTACTTACGCTCGCACTAACTACTCGCC*G*A	SEQ ID NO: 195
CTL045_BOT_tag	/5Phos/T*C*GTCGCACTAGTGCGCGGTTATGGTACCATAACCGCGCTACACTGCGCG*A*C	SEQ ID NO: 196
CTL009_BOT_tag	/5Phos/A*C*CGGCTACTACTGCGGTTTCGACATTACCTTAATCGCGCAGTCGAGCGCA*T*A	SEQ ID NO: 197

CTL055_BOT_tag	/5Phos/A*G*TAGTGCGCGTAGCGTCGAATTGGTGCGCACATAGTCGTTGTGCGCGCTA*G*T	SEQ ID NO: 198
CTL101_BOT_tag	/5Phos/G*C*GCGGATTAGTTGCGTCGAATTGGTACCGCCGTAC AAGTCGGCGAGTAG*T*T	SEQ ID NO: 199
CTL135_BOT_tag	/5Phos/A*G*TAGTGCGCGTACGTTGCGCTAGGTACCGCCGTAC AAGACCTTAATCCG*C*G	SEQ ID NO: 200
CTL155_BOT_tag	/5Phos/A*A*CCGTCGAGTGCATTGCGGATCGGTACCGCCGTAC AAGTCTTGACGCGC*T*A	SEQ ID NO: 201
CTL122_BOT_tag	/5Phos/G*C*GGCGTATTGGTACCTAGTCGTCGTACCAATACGC CGCACCGACTAATG*C*G	SEQ ID NO: 202
CTL080_BOT_tag	/5Phos/A*C*CGATCGCTAGTCGCATTAGTCGGTACCATAACCG CGCCGCGCTACTAG*G*T	SEQ ID NO: 203
CTL126_BOT_tag	/5Phos/T*A*GTGCGAGCGTATCGCGCTAGATTACGACGACTA GGTAGCGCGAGTA*G*T	SEQ ID NO: 204
CTL098_BOT_tag	/5Phos/A*C*CTAGTCCGCGACTGCGAGCGTACACCTTAATCG CGCGTATAGCGGCG*G*T	SEQ ID NO: 205
CTL038_BOT_tag	/5Phos/A*C*TAGCGATCGGTACTGCGAGCGTACGCACTCGACG GTTAGTAGTGCGCG*T*A	SEQ ID NO: 206
CTL139_BOT_tag	/5Phos/A*C*CTAGTCGTCTTCTCGCGCACTAACGACGCGCTG TTATACACTGCGCG*A*C	SEQ ID NO: 207
CTL010_BOT_tag	/5Phos/G*C*GGCGTATTGGTGTATAGCGGCGGTACCATACTAC CGCAACCAATTCGAC*G*C	SEQ ID NO: 208
CTL034_BOT_tag	/5Phos/T*A*ACAGCGCGTCGACTAGCGATCGGTACCTAGTCGC GTAAGTAGTGCGCG*T*A	SEQ ID NO: 209
CTL117_BOT_tag	/5Phos/G*C*GCGGATTAGTTGCGTCGAATTGGTACGCCGCTAA CTATAGTTAGCGGC*G*T	SEQ ID NO: 210
CTL035_BOT_tag	/5Phos/A*T*TGCGGATCGGTAGTAGTGCGCGTAACGCCGCTAA CTAACCTTAGTCCG*C*G	SEQ ID NO: 211
CTL121_BOT_tag	/5Phos/A*C*GCGCTACCTATTAGTTAGCGGCGTATAAGTCGGC GGTACCTAGTCGTC*G*T	SEQ ID NO: 212
CTL106_BOT_tag	/5Phos/G*T*CGCGCAGTGTAACCGGTACTACTACACTACTCG CGCAACCGTCGATC*C*G	SEQ ID NO: 213
CTL059_BOT_tag	/5Phos/A*C*CAATCGAGCGAATTGCGGATCGGTATAAGTCGGC GGTACCGATCCGCA*A*T	SEQ ID NO: 214
CTL157_BOT_tag	/5Phos/G*G*TAACGAGCGGTGCGCGATTAAGGTGTACGCTCGC AGTGTACGCTCGCA*G*T	SEQ ID NO: 215
CTL015_BOT_tag	/5Phos/A*C*CGATCCGCAATTAGTGCGAGCGTAAGTAGTGCGA CGATCGCGACAGTA*G*T	SEQ ID NO: 216
CTL110_BOT_tag	/5Phos/C*G*CGTAGTATGGTTCTCGCGCACTAATTAGTGCGCG AGAACCCTCGTTA*C*C	SEQ ID NO: 217
CTL123_BOT_tag	/5Phos/T*C*GGCGAGTAGTTGCGCGATTAAGGTACCTTAATCG CGCTAGCGCGAGTA*G*T	SEQ ID NO: 218
CTL014_BOT_tag	/5Phos/G*C*ACTCGACGGTTGCGTCGAATTGGTACCGCCGTAC AAGAGTAGTGCGCG*T*A	SEQ ID NO: 219
CTL131_BOT_tag	/5Phos/T*C*GTCGCACTAGTGCGCGATTAAGGTACCGATCCGC AATCGGATCGACGG*T*T	SEQ ID NO: 220
CTL062_BOT_tag	/5Phos/C*G*CGGATTAAGGTGCGCGAGTAGTGTGTCGCGCAGT GTATACGCGCACTA*C*T	SEQ ID NO: 221
CTL044_BOT_tag	/5Phos/A*C*CTAGCGCGCAATACGCGCACTACTACCTATTACC GCGTATGGTACGGC*G*T	SEQ ID NO: 222
CTL043_BOT_tag	/5Phos/A*C*CGCGTACTACTACCGCGACTTATCGCAACGCTA GGTCTTGACGCGC*T*A	SEQ ID NO: 223
CTL118_BOT_tag	/5Phos/A*C*TAGCGATCGGTGCGCGTTATGGTTGCGCGCTAG ATTACCGACTAATG*C*G	SEQ ID NO: 224
CTL128_BOT_tag	/5Phos/A*C*CGCGGACTTATTAGTTAGCGGCGTACCAATACGC CGCACGCCGTACCA*T*A	SEQ ID NO: 225

CTL067_BOT_tag	/5Phos/C*G*GACGAGCGGTTGACTACCGCTCGTACCAATACGC CGCACCATAACCGC*G*C	SEQ ID NO: 226
CTL020_BOT_tag	/5Phos/A*G*TCGAGCGCATAGCGCGTTATGGTTCGGCGAGTA GTTGCGCACATAGT*C*G	SEQ ID NO: 227
CTL006_BOT_tag	/5Phos/C*G*CGTAGTATGGTGCGGATTAAGGTGGTAACGAGC GGTACGCCGCTAAC*T*A	SEQ ID NO: 228
CTL017_BOT_tag	/5Phos/T*C*TCGCGCACTAACGGACGAGCGGTTTACGCGCACT ACTACCGACTAATG*C*G	SEQ ID NO: 229
CTL057_BOT_tag	/5Phos/A*C*GAGCGGTAGTCTTAGTGCGCGAGACGCATTAGTC GGTACTACTCGCGC*T*A	SEQ ID NO: 230
CTL078_BOT_tag	/5Phos/G*C*GCGGTTATGGTGTATAGCGCGGTACCAATCGAG CGATAGTGCGAGCG*T*A	SEQ ID NO: 231
CTL031_BOT_tag	/5Phos/T*A*GTTAGCGGCGTATAGGTAGCGCGTTATGCGCTCG ACTTCGCTCGATTG*G*T	SEQ ID NO: 232
CTL136_BOT_tag	/5Phos/T*C*GCGACAGTAGTCGCATTAGTCGGTGTACGCTCGC AGTCGCGGATTAAG*G*T	SEQ ID NO: 233
CTL165_BOT_tag	/5Phos/A*C*CGCCGCTATACTAGCGCGTCAAGAACCAATCGAG CGATACGCGCACTA*C*T	SEQ ID NO: 234
CTL039_BOT_tag	/5Phos/A*C*GCCGTACCATACGACTATGTGCGCACCGACCGTA CCGACTAGTGCGAC*G*A	SEQ ID NO: 235
CTL036_BOT_tag	/5Phos/A*C*CTAGTCGTCGTAGTAGTACGCGGTTATGCGCTCG ACTACCTTAATCCG*C*G	SEQ ID NO: 236
CTL048_BOT_tag	/5Phos/T*T*CGCGCTAGGTGCGCGAGTAGTGTTAGTGCGAGC GTAGCGCACATAGT*C*G	SEQ ID NO: 237
CTL053_BOT_tag	/5Phos/C*G*GATCGACGGTTACTAGTGCGACGATTAGTGCGCG AGAATAAGTCGCGC*G*T	SEQ ID NO: 238
CTL072_BOT_tag	/5Phos/A*C*GCCGTACCATAACCGCTCGTTACCCGCATTAGTC GGTCGCAACGCTAG*G*T	SEQ ID NO: 239
CTL096_BOT_tag	/5Phos/T*A*CGCGACTAGGTGCGAACGCTAGGTACCTATTACC GCGACCTAGTAGCG*C*G	SEQ ID NO: 240
CTL150_BOT_tag	/5Phos/A*C*TACTGTGCGGAACCGACTAATGCGTAGCGCGAGT AGTACCTAGCCGAA*C*G	SEQ ID NO: 241
CTL084_BOT_tag	/5Phos/A*C*GCCGCTAATAACCTAGTCGTGCTACCTATTACC GCGAACCGCTCGTC*C*G	SEQ ID NO: 242
CTL142_BOT_tag	/5Phos/A*G*TAGTACGCGGTGCGATTAGTCGGTACCGATCCGC AATTAGTGCGAGCG*T*A	SEQ ID NO: 243
CTL102_BOT_tag	/5Phos/T*T*CGGCGCTAGGTTAGCGCGTCAAGAACGCCGTACC ATACGGTACGGTCG*G*T	SEQ ID NO: 244
CTL154_BOT_tag	/5Phos/G*T*ACGCTCGCAGTGCGCGAGTAGTGTGCACTCGACG GTTAACTAATCCGC*G*C	SEQ ID NO: 245
CTL112_BOT_tag	/5Phos/A*C*CGCCGACTTATAGTAGTGCGCGTACGCATTAGTC GGTCGCGGATTAAG*G*T	SEQ ID NO: 246
CTL145_BOT_tag	/5Phos/C*G*GACGAGCGGTTGCGATTAGTCGGTACCATAACCG CGCCGCGGATTAAG*G*T	SEQ ID NO: 247
CTL060_BOT_tag	/5Phos/A*C*CGCCGACTTATCGCGCTACTAGGTACCGCCGTAC AAGGTACGCTCGCA*G*T	SEQ ID NO: 248
CTL016_BOT_tag	/5Phos/C*G*CAACGCTAGGTACCTAGCGCCGAACGCGGACTAA GGTACCTAGCGCCG*A*A	SEQ ID NO: 249
CTL159_BOT_tag	/5Phos/G*C*ACTCGACGGTTCGTTTCGGCTAGGTACCGCCGTAC AAGTACGCGACTAG*G*T	SEQ ID NO: 250
CTL056_BOT_tag	/5Phos/A*C*GCCGTACCATAGCGGCGTATTGGTGTGCGCGAGT GTAGCGCGGTTATG*G*T	SEQ ID NO: 251
CTL162_BOT_tag	/5Phos/T*A*GTTAGCGGCGTTCGGTTCGACATTACCTAGTCGC GTAGCGCGAGTAGT*G*T	SEQ ID NO: 252
CTL018_BOT_tag	/5Phos/G*C*GCGATTAAGGTTCTCGCGCACTAACGACGCGCTG TTACGCATTAGTCG*G*T	SEQ ID NO: 253

CTL115_BOT_tag	/5Phos/G*C*GGCGTATTGGTACCGCCGACTTATCGCATTAGTC GGTTATGGTACGGC*G*T	SEQ ID NO: 254
CTL033_BOT_tag	/5Phos/G*C*GCGGTTATGGTAACTACTCGCCGAACCTATTACC GCGACTGCGAGCGT*A*C	SEQ ID NO: 255
CTL047_BOT_tag	/5Phos/C*G*ACTATGTGCGCTCGTCGCACTAGTACCATAACCG CGCAACCGCTCGTC*C*G	SEQ ID NO: 256
CTL108_BOT_tag	/5Phos/A*A*TCTAGCCGCGATAGTTAGCGGCGTACCTTAATCG CGCTAGCGCGAGTA*G*T	SEQ ID NO: 257
CTL041_BOT_tag	/5Phos/T*A*CGCGCACTACTGCGTCGAATTGGTGCGCGGATTA GTTGCGTCGAATTG*G*T	SEQ ID NO: 258
CTL061_BOT_tag	/5Phos/A*C*CGCCGCTATACACTGCGAGCGTACTTCGGCGCTA GGTGTATAGCGGCG*G*T	SEQ ID NO: 259
CTL166_BOT_tag	/5Phos/A*C*TACTCGCGCTAGCGGCGTATTGGTAACCGCTCGT CCGGCGCGAGTAGT*G*T	SEQ ID NO: 260
CTL012_BOT_tag	/5Phos/G*C*GCGAGTAGTGTACCTAGCGTTGCGCGCGGATTAA GGTACTAGTGCGAC*G*A	SEQ ID NO: 261
CTL052_BOT_tag	/5Phos/G*C*GGTTCGACATTACCTAGCGTTGCGCGCATTAGTC GGTACCTAGTAGCG*C*G	SEQ ID NO: 262
CTL153_BOT_tag	/5Phos/G*T*CGCGCAGTGTAACCTAGCGTTGCGTCGCGACAGT AGTGACTACCGCTC*G*T	SEQ ID NO: 263
CTL094_BOT_tag	/5Phos/A*C*CGCTCGTTACCACTAGCGATCGGTACCATACTAC GCGTACGCGACTAG*G*T	SEQ ID NO: 264
CTL095_BOT_tag	/5Phos/C*G*CAACGCTAGGTAGTCGAGCGCATACGCATTAGTC GGTAATGTGCAACC*G*C	SEQ ID NO: 265
CTL105_BOT_tag	/5Phos/A*C*CTAGTAGCGCGTAGTTAGCGGCGTTTAGTGCGCG AGAGTACGCTCGCA*G*T	SEQ ID NO: 266
CTL109_BOT_tag	/5Phos/C*T*TGTACGGCGGTCGCGGACTAAGGTTGCGGCTAG ATTACCGACCGTAC*C*G	SEQ ID NO: 267
CTL032_BOT_tag	/5Phos/T*A*GTTAGCGGCGTAATCTAGCCGCGAATAGGTAGCG CGTAACTACTCGCC*G*A	SEQ ID NO: 268

“/5Phos/” indicates a 5'-phosphate moiety; “*” indicates a phosphorothioate linkage.

Table 5. Pools of Tag Sequences

	Pools							
	Pool A1	Pool B1	Pool B2	Pool B3	Pool B4	Pool B5	Pool B6	Pool C1
Tags Present in Pools	CTL085	CTL161	CTL089	CTL098	CTL062	CTL048	CTL018	Pool A1
	CTL169	CTL164	CTL081	CTL038	CTL044	CTL053	CTL115	Pool B1
	CTL137	CTL030	CTL075	CTL139	CTL043	CTL072	CTL033	Pool B2
	CTL042	CTL088	CTL160	CTL010	CTL118	CTL096	CTL047	Pool B3
	CTL051	CTL148	CTL133	CTL034	CTL128	CTL150	CTL108	Pool B4
	CTL167	CTL152	CTL076	CTL117	CTL067	CTL084	CTL041	Pool B5
	CTL026	CTL007	CTL024	CTL035	CTL020	CTL142	CTL061	Pool B6
	CTL068	CTL141	CTL045	CTL121	CTL006	CTL102	CTL166	
	CTL138	CTL064	CTL009	CTL106	CTL017	CTL154	CTL012	
	CTL079	CTL158	CTL055	CTL059	CTL057	CTL112	CTL052	
	CTL063	CTL066	CTL101	CTL157	CTL078	CTL145	CTL153	
	CTL168	CTL144	CTL135	CTL015	CTL031	CTL060	CTL094	
	CTL021	CTL107	CTL155	CTL110	CTL136	CTL016	CTL095	
	CTL151	CTL149	CTL122	CTL123	CTL165	CTL159	CTL105	
	CTL002	CTL008	CTL080	CTL014	CTL039	CTL056	CTL109	
	CTL134	CTL099	CTL126	CTL131	CTL036	CTL162	CTL032	

Table 6. Non-homologous tails

Name	Sequence (5'→3')	SEQ ID NO:
H1	ACGCGACTATACGCGCAATATGGT	SEQ ID NO: 269
H2	CTAGCGATACTACGCGATACGAGAT	SEQ ID NO: 270
H3	CATAGCGGTATTACGCGAGATTACGA	SEQ ID NO: 271
H4	CGCGAGTACGTACGATTACCG	SEQ ID NO: 272
H5	ACGCGCGACTATACGCGCCTC	SEQ ID NO: 273

CLAIMS

What is claimed:

1. A method for identifying and nominating on- and off-target CRISPR edited sites with improved accuracy and sensitivity, the process comprising the steps of:
 - (a) co-delivering a guide sequence RNA (sgRNA) or a two-part CRISPR RNA:trans-activating crRNA (crRNA:tracrRNA) duplex, one or more tag sequences, and an RNA-guided endonuclease to cells;
 - (b) incubating the cells for a period of time sufficient for double strand breaks to occur;
 - (c) isolating genomic DNA from the cells, fragmenting the genomic DNA, and ligating the fragmented genomic DNA to a unique molecular index containing a universal adapter sequence;
 - (d) amplifying the ligated DNA fragments using primers targeting the tag and universal adapter sequences to produce a first set of amplified sequences;
 - (e) amplifying the first set of amplified sequences using universal sequencing primers targeting the tails of Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences;
 - (f) sequencing the pooled sequences and obtaining sequencing data; and
 - (g) identifying on-/off-target CRISPR editing loci.
2. The method of claim 1, wherein the universal sequencing primers target SP1 or SP2 sequence (SEQ ID NO: 7, 8) tails on the Tag-pTOP or Tag-pBOT primers to produce a second set of amplified sequences.
3. The method of claim 1 or 2, wherein the universal sequencing primers target predesigned non-homologous sequence (SEQ ID NO: 269–273) tails on the Tag-pTOP or Tag-pBot primers to produce a second set of amplified sequences.
4. The method of any one of claims 1–3, wherein the universal sequencing primers target predesigned 13-mer tails on the Tag-pTOP or Tag-pBot primers to produce a second set of amplified sequences.
5. The method of any one of claims 1–4, wherein step (g) comprises executing on a processor:

- (i) aligning the sequence data to a reference genome;
 - (ii) identifying on-/off-target CRISPR editing loci; and
 - (iii) outputting the alignment, analysis, and results data as custom-formatted files, tables or graphics.
6. The method of any one of claims 1–5, further comprising a step following step (e) comprising:
- (e1) normalizing the second set of amplified sequences to produce concentration normalized libraries, pooling the normalized libraries with other samples to produce pooled libraries; and continuing with steps (f)–(i).
7. The method of any one of claims 1–6, wherein step (d) uses a suppression PCR method.
8. The method of any one of claims 1–7, wherein the RNA-guided endonuclease comprises an endogenously-expressed Cas enzyme, a Cas expression vector, a Cas protein, or a Cas RNP complex.
9. The method of any one of claims 1–8, wherein the RNA-guided endonuclease comprises an endogenously-expressed Cas9 enzyme, a Cas9 expression vector, a Cas9 protein, or a Cas9 RNP complex.
10. The method of any one of claims 1–9, wherein the cells comprise human or mouse cells.
11. The method of any one of claims 1–10, wherein the period of time is about 24 hours to about 96 hours.
12. The method of any one of claims 1–11, wherein multiple tag sequences are co-delivered.
13. The method of any one of claims 1–12, wherein the tag sequences comprise double-stranded deoxyribooligonucleotides (dsDNA) comprising 52-base pairs.
14. The method of any one of claims 1–13, wherein the tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides.

15. The method of any one of claims 1–14, wherein the tag sequences comprise a double stranded DNA comprising the complementary top and bottom strand pairs of SEQ ID NO: 1–2 or 7–268.
16. On- and off-target CRISPR editing sites identified or nominated using the method of any one of claims 1–15.
17. A method for designing 52-base pair tag sequences, the method comprising, executing on a processor:
 - (a) randomly generating 13-nucleotide sequences with 40–90% GC content, max homopolymer length A:2, C:3, G:2, T:2, weighted homopolymer rate < 20, self-folding $T_m < 50$ °C, and self-dimer $T_m < 50$ °C;
 - (b) removing sequences that perfectly align to a particular genome or that are homopolymers or GG or CC dinucleotide motifs and obtaining a set of 13-mers;
 - (c) selecting a subset of the 13-mer sequences that contain one or less CC or GG dinucleotide motifs;
 - (d) concatenating four of the of 13-mer subset sequences to form random 52-mer sequences;
 - (e) aligning the random 52-mer sequences to a genome;
 - (f) removing the random 52-mer sequences that have similarity to the genome to produce a subset of 52-mer sequences; and
 - (h) outputting the subset of 52-mer sequences and generating the complementary strands to produce double stranded 52-base pair tag sequences.
18. The method of claim 17, wherein the genome is human or mouse.
19. The method of claim 17 or 18, wherein the 52-base pair tag sequences are-non complementary to the genome.
20. The method of any one of claims 17–19, further comprising designing primers for the 52-base pair tag sequences.

21. The method of any one of claims 17–20, wherein the 52-base pair tag sequences comprise a 5'-terminal phosphate, and phosphorothioate linkages between the 1st and 2nd, 2nd and 3rd, 50th and 51st, and 51st and 52nd nucleotides of the 52-base pair tag sequences.
22. The method of any one of claims 17–21, further comprising synthesizing oligonucleotides comprising the 52-base pair tag sequences, the complement of the 52-base pair tag sequences, or primers for the 52-base pair tag sequences.
23. One or more 52-base pair tag sequences designed using the methods of claims 17–22.
24. The 52-base pair tag sequences of claim 23, wherein the 52-base pair tag sequence comprises a double stranded DNA comprising the top and bottom strand pairs of SEQ ID NO: 1–2 or 7–268.
25. A method for designing primers partially complementary to the 52-base pair tag sequences of claim 23 and an adapter primer, the method comprising, executing on a processor:
 - (a) designing tag primers that are partially complementary to the top and bottom strands of tag sequences; and
 - (b) designing an adapter primer that is partially complementary to the top strand of the adapter sequence;wherein:
the tag primers comprise a 5'-universal tail sequence; and
the adapter primer comprises a sequence complementary to the tails of Tag-pTOP or Tag-pBOT primers.
26. The method of claim 25, wherein the 5'-universal tail sequence is complementary to an SP1 or SP2 sequence (SEQ ID NO: 7, 8), a locus specific segment, a ribonucleotide (rN) 6-nucleotides from the 3'-end, a 3'-end mismatch, a 3'-end block (3'-C₃ spacer), a predesigned non-homologous sequence (SEQ ID NO: 269–273), or a predesigned 13-mer sequence.
27. The method of claim 25 or 26, wherein the primers partially complementary to top and bottom strands of the tag sequences comprise a tail sequence complementary to the SP1

sequence (SEQ ID NO: 7) and the adapter primer comprises a sequence complementary to the SP2 sequence (SEQ ID NO: 8) tail on the Tag-pTOP or Tag-pBOT primers; or the primers partially complementary to top and bottom strands of the tag sequences comprise a tail sequence complementary to the SP2 sequence (SEQ ID NO: 8) and the adapter primer comprises a sequence complementary to the SP1 sequence (SEQ ID NO: 7) tail on the Tag-pTOP or Tag-pBOT primers.

28. The method of any one of claims 25–27, wherein the amplification of a nucleic acid molecule with the primers that are complementary to the top and bottom strands of tag sequences and primers that are complementary to the top strand of the adapter sequence produces a PCR product that comprises a portion of the tag sequence, a sgDNA sequence, and the adapter sequence.
29. The method of any one of claims 25–28, further comprising synthesizing oligonucleotides comprising the sequences of the forward and reverse tag primers and the adapter primer.
30. The method of any one of claims 17–21 and 25–29, wherein the 52-base pair tag sequences and primers partially complementary to the 52-base pair tag sequences are designed and selected using an algorithm predicting whether the primers are likely to be partially complementary and have a propensity to form primer-dimers.
31. One or more primers partially complementary to the 52-base pair tag sequences and one or more adapter primers designed using the method of claims 22–25.
32. The primers of claims 31, wherein the primers comprise the sequences of SEQ ID NO: 3, 4; and the adapter primer, wherein the adapter primer comprises the sequence of SEQ ID NO: 5.
33. Use of one or more double-stranded 52-base pair tag sequences for identifying on- and off-target CRISPR editing sites.