



(51) International Patent Classification:

C12Q 1/6811 (2018.01) C12N 15/113 (2010.01)
C12Q 1/6869 (2018.01)

(21) International Application Number:

PCT/US2021/042733

(22) International Filing Date:

22 July 2021 (22.07.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/055,460 23 July 2020 (23.07.2020) US

(71) Applicant: **INTEGRATED DNA TECHNOLOGIES, INC.** [US/US]; 1710 Commercial Park, Coralville, Iowa 52241-2760 (US).

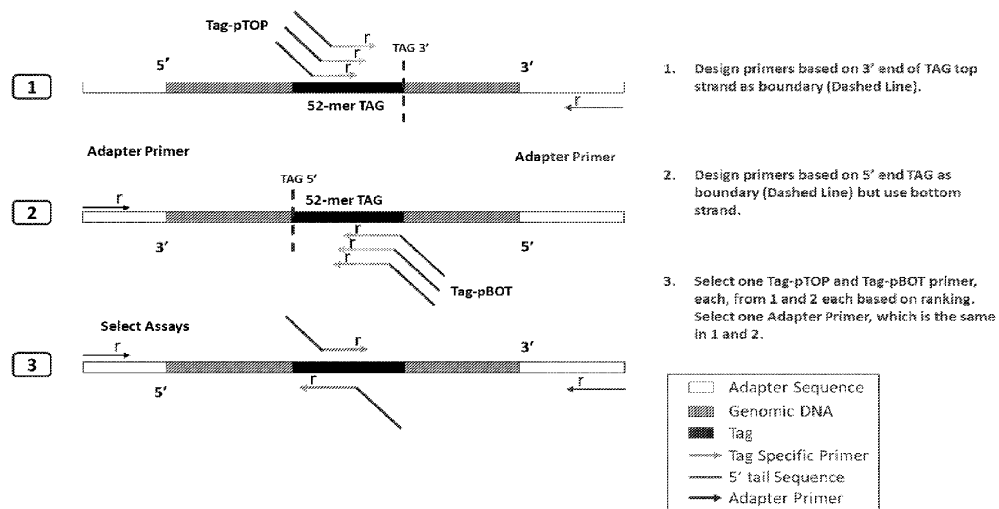
(72) Inventors: **MCNEILL, Matthew**; 135 N. Westminster Street, Iowa City, Iowa 52245 (US). **TURK, Rolf**; 415 Lee Street, Iowa City, Iowa 52246 (US). **RETTIG, Garrett**; 2432 Dempster Drive, Coralville, Iowa 52241 (US). **BLACK, Ellen**; 1416 Goldfinch Court NE, Swisher, Iowa 52338 (US). **SUN, Yongming**; 153 Paddington Court, San Ramon, California 94583 (US). **SAILOR, Chris**; 2502 Desoto Street SW, Cedar Rapids, Iowa 52404 (US). **WANG, Yu**; 8 Bridle Ridge Drive, North Grafton, Massachusetts 01536 (US). **GUNDERSON, Keith**; 1157 Oakes Drive, Iowa City, Iowa 52245 (US). **KINNEY, Kyle**; 260 Hawkeye Court, Apt. 102, Iowa City, Iowa 52246 (US).

(74) Agent: **BROWN, Bernard A., II** et al.; 790 N. Water Street, Suite 2500, Milwaukee, Wisconsin 53202 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,

(54) Title: METHODS FOR NOMINATION OF NUCLEASE ON-/OFF-TARGET EDITING LOCATIONS, DESIGNATED "CTL-SEQ" (CRISPR TAG LINEAR-SEQ)

FIG. 4



(57) Abstract: Methods for identifying and nominating on- and off-target CRISPR editing sites, particularly Cas9, with improved accuracy and sensitivity in which a tag is inserted a double-stranded breaks and the region comprising the tag and a universal adapter comprising a unique molecular index (UMI). The sequences to which the universal sequencing primer binds may be predesigned. A method for designing 52-base pair tag sequences based on (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 7m < 50 °C, and self-dimer 7m < 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

[Continued on next page]

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO,
NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:

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

(88) Date of publication of the international search report:

10 March 2022 (10.03.2022)

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.