Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases

Richard L Frock1–4, Jiazhi Hu1–4, Robin M Meyers1–3, Yu-Jui Ho1–3, Erina Kii1–3 & Frederick W Alt1–3

Although great progress has been made in the characterization of the off-target effects of engineered nucleases, sensitive and unbiased genome-wide methods for the detection of off-target cleavage events and potential collateral damage are still lacking. Here we describe a linear amplification–mediated modification of a previously published high-throughput, genome-wide, translocation sequencing (HTGTS) method that robustly detects DNA double-stranded breaks (DSBs) generated by engineered nucleases across the human genome based on their translocation to other endogenous or ectopic DSBs. HTGTS with different Cas9:sgRNA or TALEN nucleases revealed off-target hotspot numbers for given nucleases that ranged from a few or none to dozens or more, and extended the number of known off-targets for certain previously characterized nucleases more than tenfold. We also identified translocations between bona fide nuclease targets on homologous chromosomes, an undesired collateral effect that has not been described previously.

Finally, HTGTS confirmed that the Cas9D10A paired nickase approach suppresses off-target cleavage genome-wide.

Targeting endogenous loci in live cells with nucleases designed to generate DNA double-stranded breaks (DSBs) at specific endogenous sequences without the need for substrate integration has been very useful for introducing targeted mutations and holds great promise for targeted gene therapy in humans1–4. In this regard, the recently developed TALENs (transcription activator–like effector nucleases) and Cas9:sgRNA endonucleases (clustered, regularly interspaced, short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9): single guide RNA (sgRNA)) are particularly promising5–10. One continuing concern for employing TALENs and Cas9:sgRNAs for genome engineering, and for therapeutic human genome engineering in particular, is the potential for off-target DSB activity at nonconsensus sites within the genome for any given enzyme2. Current assays for such off-target nuclease activity involve cytotoxicity11, prediction-based modeling12–14, select screening12,15,16 and viral vector DSB traps17,18. Such assays have been valuable for testing approaches designed to minimize undesired DNA cleavage activities of these enzymes1–2.

TALENs are dimeric, site-specific nucleases with monomers consisting of an engineered DNA binding domain fused to a C-terminal FokI nuclease domain9,10. Specific TALEN activity requires the dimerization of the FokI domain from two TALEN subunits with each monomer providing half of the specific DNA recognition sequence2. The DNA-binding code for TALENs allows targeting of DSBs with 5′ overhangs at nearly any position across different genomes2,19,20. For Cas9:sgRNA endonucleases, the Cas9 nuclease forms a complex with an engineered sgRNA comprising a chimeric CRISPR RNA and trans-activating CRISPR RNA1. Cas9:sgRNA sequence-specificity relies on hybridization of a 20-nt targeting sequence on the 5′ end of the sgRNA to complementary DNA and recognition of an ‘NGG’ protospacer-adjacent motif (PAM) on the noncomplementary strand. Cas9:sgRNA complexes, which, again, can be designed to cleave a multitude of sites across the genome, generate blunt DSB ends 3 bp into the 20-nt target sequence proximal to the PAM1.

Chromosomal translocations can arise by fusion of the ends of two DNA DSBs lying on heterologous chromosomes or on separated regions of a homologous chromosome21,22. The HTGTS23 and translocation-capture sequencing approaches24 were developed to identify translocations of yeast I-SceI meganuclease-generated ‘bait’ DSBs at target sites introduced into the genome of mouse cells to other ‘prey’ cellular DSBs genome-wide. Correspondingly, these methods also identified various classes of endogenous DSBs in primary and transformed B-lymphocyte lineage cells23–27. HTGTS, which provides nucleotide-level resolution of junctions, further revealed I-SceI–generated DSBs at cryptic off-target sequences within the mouse genome23. Based on ability to detect off-target, I-SceI meganuclease sites across the mouse genome, we proposed HTGTS might be developed into a robust general method for determining off-target activity of engineered nucleases23. We now describe the development of an enhanced HTGTS approach and its application in human cells for identifying nuclease-generated, on-target and off-target DSBs and associated collateral chromosomal damage.

RESULTS
HTGTS assay for Cas9-generated DSBs at the human RAG1 locus
To evaluate use of HTGTS for identifying on- and off-target custom nuclease activity in human cells, we first performed HTGTS using Cas9:sgRNA-generated DSBs as ‘bait’ to capture ‘prey’ sequences genome-wide in 293T cells culture for 48 h after transfection with Cas9:sgRNA. For these studies, we have now developed a modified HTGTS approach based on linear-amplification–mediated PCR (LAM-PCR)28 that is more robust, cost-efficient and rapid than our prior emulsion-PCR23 HTGTS (Fig. 1a, Supplementary Fig. 1 and
Figure 1 LAM-PCR HTGTS from Cas9:RAG1A on-target and off-target bait sites. (a) LAM-PCR HTGTS method overview. Major modifications of prior HTGTS method include enriching for translocations directly from sonicated DNA by LAM-PCR and ssDNA ligation using a bridge adapter. (b) Positions of Cas9:RAG1A–D sgRNA target sites within the RAG1 locus on chromosome (Chr) 11. Red arrow indicates the cloning primer used to capture the 3′ DSB ends from the RAG1A site (orange box); RAG1B uses the same primer but with longer bait. RAG1C and D use a separate primer strategy (Supplementary Fig. 2a). Cas9:RAG1 DSBs occur 3 bp proximal to the PAM within the sgRNA target sequence (red bar inside chr7 (N = 3 each; see Supplementary Table 1). (c) Circos plots1 of RAG1A HTGTS genome-wide prey junctions binned into 5-Mb regions (black bars) are plotted on a log scale with indicated custom ticks; frequency ranges are colored from white (0–10) to increasingly darker orange colors (>10) by factors of 10. Colored lines connect the RAG1 bait site to the prey hotspot; line color range from dark red to yellow indicates high to low hotspot enrichment, respectively (25,000 junctions; N = 6). Hotspots are identified as significantly focal sites (see online methods for more details), contain junctions to both chromosomal orientations, and are found in multiple library replicates; all hotspots shown are related to nuclease off-target activity. (d–f) RAG1A off-target (OT) bait HTGTS Circos plots using the (d) DAZAP OT1 on chr19, (e) 47.0 Mb OT2 on chr12 and (f) 103.6 Mb OT17 on chr7 (N = 3 each; see Supplementary Table 1). (g) Bar graph of average junction frequency per 10,000 unique junctions for each off-target site not located on bait chromosomes 7, 11, 12 and 19. Error bars, mean ± s.e.m. The junction scale displayed for the figure Circos plots is from 10–20,000.

Supplementary Table 1). For initial studies, we selected the human RAG1 gene, a proposed target for gene correction therapy29,30. To induce RAG1 DSBs, we generated four sgRNAs that each targeted a distinct sequence within a 317-bp region spanning the beginning of RAG1 exon 2; we refer to these four Cas9:sgRNA combinations as RAG1A, B, C and D (Fig. 1b and Supplementary Fig. 2a). We performed HTGTS from the 3′ DSB end (with respect to RAG1 transcriptional orientation) of a given Cas9:RAG1-generated DSB, cloning from the A or B site by a specific primer positioned, respectively, 152 bp and 194 bp centromeric to the sites or from the C or D sites with a second specific primer positioned, respectively, 106 bp and 227 bp centromeric to them (Fig. 1b and Supplementary Fig. 2a). For each HTGTS library, recovered junctions fused uniquely mapped coordinates corresponding to bait sequence and genome-wide prey sequences (Supplementary Figs. 2b–e and 3a–c) and were mainly direct (‘blunt’) or had short microhomologies (Supplementary Fig. 3d–g). On the bait-site side, junctions were enriched at or near the 3′ DSB bait-site end with enrichment decreasing along the bait sequence length, consistent with variable end resection before joining (Supplementary Fig. 3a–c; see below).

For each set of HTGTS libraries from a particular break-site or under particular conditions, we used modified Circos plots of the human genome organized into individual chromosomes to visualize overall junction patterns and key features31. In these plots, translocation hotspots, bioinformatically identified in an unbiased fashion as focally enriched HTGTS junction clusters, are indicated by lines that connect the bait-site to a given hotspot and range in color from dark red (highest junction enrichment) to yellow (lowest junction enrichment) (Fig. 1c and Supplementary Fig. 2b–e). We also denote HTGTS junction frequency within 5-Mb bins across all chromosomes on the Circos plots by black bars plotted on a log scale with custom axes (Fig. 1 legend). For each bait site analyzed in this study, at least three (and usually many more) separate HTGTS libraries were generated with individual libraries ranging in size from several thousand to 80,000 independent junctions (Supplementary Table 1 and detailed for each experimental figure). Independent HTGTS libraries for a given site or condition gave reproducible overall results and conclusions (Supplementary Fig. 2b–e and Supplementary Table 1; see below). Nonspecific HTGTS background was estimated as described23 and found to be low (Supplementary Table 2).

Genome-wide off-target activities of Cas9:RAG1 sgRNAs By convention, prey sequences are joined to bait DSBs as in ‘+’ (plus) orientation if they read from the junction in the p telomere to q telomere direction; correspondingly, junctions are in ‘−’ (minus) orientation if in the other direction (Supplementary Fig. 4a). Other than break-site junctions (see below), genome-wide Cas9:RAG1A–D HTGTS junctions occurred across the genome at similar frequencies in both orientations (Supplementary Fig. 1). Frequent endogenous DSBs at two loci can dominate translocation landscapes27,32, due to cellular...
heterogeneity in three-dimensional (3-D) genome organization. In this regard, we detected 33 highly significant, locally enriched prey junctional hotspots from RAG1A and two from RAG1B libraries (Fig. 1c, Supplementary Fig. 2b, c, and Supplementary Table 3). In contrast, no hotspots were detected for RAG1C and RAG1D libraries, which based on random library size normalizations would have been readily detectable if they occurred at the level of RAG1A or RAG1B off-target sites (Supplementary Fig. 2d and legend). The RAG1A or RAG1B hotspot junctions showed expected characteristics for involvement of off-target DSBs, peaking precisely at predicted off-target break sites based on their being highly related (with 2-7 nt mismatches) to the respective bona fide on-target sequences of these two enzymes (Supplementary Table 3 and Supplementary Fig. 4b). All junctions were fully consistent with DSB joining, displaying approximately equal numbers of (+) and (−) orientation joins that peaked at direct prey joins (no loss of nucleotides from predicted off-target break-site) and tailed off in both orientations into joins with up to 100 bp of resection (Supplementary Fig. 4c-e). Additional RAG1A studies at 24 h and 96 h after RAG1A transfection resulted in similar distributions of junctions genome-wide including RAG1A off-target sites (Supplementary Fig. 5). Also, RAG1B HTGTS libraries in a different human cell line (A549 lung carcinoma cells) revealed the same two off-target hotspots (Supplementary Fig. 6).

To unequivocally confirm that identified HTGTS off-target sites represented DSBs and further test the ability of HTGTS to identify off-target nuclease-generated DSBs genome-wide, we performed HTGTS using, respectively, RAG1A high-level, off-target sites on chromosomes 12 or 19 or a RAG1A lower level, off-target site on chromosome 7 as bait (Fig. 1d-f). Strikingly, each bait produced HTGTS libraries with all characteristics expected for cloning from that specific off-target RAG1A bait DSB (Fig. 1d-f, see below). Moreover, all reproducibly captured the RAG1A on-target break-site as well as the vast majority of the off-target sites revealed by HTGTS from the on-target break-site (Fig. 1d-g). Indeed, the most highly enriched off-target translocation sites recovered from the bona fide RAG1A on-target HTGTS bait were similarly highly enriched when each of the three RAG1A off-target sites were used as HTGTS bait (Fig. 1g and Supplementary Table 3). The ability of these different HTGTS bait DSBs on different chromosomes to robustly identify essentially the same set of recurrent DSBs genome-wide is consistent with our prior findings on the influences of cellular heterogeneity in three-dimensional (3D) genome organization on the translocation landscape. Likewise, the major difference in frequency of translocations captured by these four different RAG1A HTGTS baits was the enriched recovery of prey DSB junctions that fell onto the same chromosome as the particular bait-site (Supplementary Table 3), as also predicted by our prior studies. Below, we present additional examples of such phenomena in the context of exploiting them to further facilitate HTGTS assays for nuclease off-target activity.

A common class of translocations for engineered nucleases

HTGTS junctions are highly enriched in regions immediately around the break-site owing to DSB rejoining following resection, as well as various types of break-site proximal translocations (deletions, inversions and excision circles) that are enhanced because of spatial proximity (Fig. 2a and Supplementary Fig. 7a). Engineered endonucleases usually have target sites on both homologous chromosomes in diploid cells, and primer sequences for detecting junctions from bait DSBs are usually present on both alleles (Supplementary Fig. 7b). Thus, most contributions from each of the homologous chromosome bait 3′ DSBs to break-site or genome-wide junctions cannot be distinguished. However, for the Cas9:RAG1A-D and RAG1A off-target bait break-sites (and others, see below), we find a high density of prey junctions at or very close to the break-site in the (+) orientation quadrant, of which many would correspond to inversive translocations that in cells would generate dicentric chromosomes (Fig. 2b,c and Supplementary Fig. 7c-e). In this regard, nucleotide sequences of such junctions confirm head-to-head inversive joins of the two break-sites, including perfect direct joins of the two Cas9 3′ DSB ends (Fig. 2d), with additional junctions in this inversion/dicentric quadrant extending several 100 bp or more upstream likely due to ‘prey’ sequence resection.
Figure 3  Cas9n paired nick DSB HTGTS libraries. (a) RAG1A is paired with nearby sgRNAs RAG1G, E and F, which target the opposing strand and generate DSBs with 5’ overhangs of 28, 36 and 51 bp in length, when used in combination with Cas9n. Red arrows indicate the orientation and position of the cloning primer used. Orange box indicates the sgRNA target sites with red bars indicating the position of Cas9 cleavage. (b–d) Circos plots (Fig. 1 legend) for HTGTS libraries of Cas9n (b) RAG1A/G (normalized to 20,000 junctions, N = 6), (c) RAG1A/E (N = 2), (d) and RAG1A/F (N = 2) paired nick using the RAG1A bait. See Supplementary Table 1 for junction frequencies. No hotspots relating to RAG1A off-target activity were identified.

(e) Prey junction enrichment within 500 bp of the RAG1A target site for Cas9n:RAG1A/F (N = 2). Red arrows indicate position of the bait primer, and the orange dashed line indicates the position of the opposing strand sgRNA target (RAG1F). Numbers in the lower left grid indicate total junctions for each region including the region between the paired offset target sites (middle). (f) Sequence indicating a head-to-head inversion of the RAG1A bait to the RAG1F ‘prey’ on the homologous chromosome, indicated as ‘1’ in f. PAM is boxed in red; the yellow and blue highlights indicate RAG1A and RAG1F target sites, respectively. Identical sequences between the two RAG1A target sites are underlined.

Cas9 paired nickase method suppresses off-target activity

One approach to reduce Cas9:sgRNA off-target activity is to use Cas9 D10A mutation (Cas9n), which renders the endonuclease into a nickase that generates DSBs from off-target activity. To test this approach by HTGTS, we paired the off-target-prone RAG1A sgRNA with nearby downstream sgRNA targets (RAG1G, E and F), which would result in 5’ overhang DSBs of 28 nt, 36 nt and 51 nt, respectively, when used with Cas9n (Fig. 3a and Supplementary Fig. 8a). Cas9n:RAG1A/G, A/E or A/F HTGTS libraries had similar genome-wide characteristics to standard Cas9:RAG1A HTGTS libraries except that they lacked hotspots (Fig. 3b–d); occasional junctions at RAG1A off-target sites were found upon inspection (Supplementary Table 3). Prey junctions around the break-site also revealed expected resection and recombination patterns (Fig. 3e and Supplementary Fig. 8b,c), including recurrent ‘denticentric’ (+) orientation junctions between break-sites on the two homologous chromosomes that encompassed the two offset nick sites (Fig. 3e,f and Supplementary Fig. 8b,c).

TALEN-generated bait HTGTS libraries

To test ability of HTGTS to reveal TALEN off-target DSBs, we employed two previously described TALENs that, respectively, cleave the C-MYC gene on chromosome 8 or ATM gene on chromosome 11 (Supplementary Fig. 9a,b). The ATM and C-MYC TALEN bait HTGTS libraries showed similar patterns of break-site proximal junctions as those generated with Cas9n:sgRNAs, including readily detectable denticentric orientation joins between the TALEN break-sites on homologous chromosomes (Supplementary Fig. 9c–e). In addition, we detected a large number of off-target sites for both the ATM (522 off-target sites) and C-MYC (384 off-target sites) TALENs, of which all were lower frequency than the most robust Cas9:RAG1A off-targets (Supplementary Fig. 9a,b). Notably, many highly enriched TALEN off-targets were pseudo-palindromic sequences that corresponded to variants of the recognition site of a single TALEN monomer (Supplementary Tables 4 and 5 and Supplementary Fig. 9f,g; see Discussion). Both ATM and C-MYC TALEN bait libraries also reproducibly displayed a high enrichment of prey junctions along their respective break-site chromosomes (Supplementary Fig. 9a,b; see below).

Universal donor bait HTGTS assay for off-target detection

As illustrated by HTGTS from RAG1A off-target sites (Fig. 1), a fixed bait DSB from one nuclease should detect both on- and off-target DSBs of a second nuclease. To test the ability of this “universal donor bait” HTGTS approach to detect different types of potential nuclease-generated DSBs, we co-expressed RAG1B with I-SceI in 293T cells and used the RAG1B bait-site to capture I-SceI off-target sites (Fig. 4a–c). Indeed, beyond the two expected RAG1B off-target sites, we reproducibly identified nine I-SceI off-target sites with 2- to 4-nt mismatches from the consensus (Fig. 4a,b, Supplementary Fig. 10a and Supplementary Table 6). I-SceI off-target sites displayed expected characteristics of such prey DSBs (Supplementary Fig. 10b,c) and were confirmed by in vitro I-SceI digestion (Supplementary Fig. 10d).
We used the RAG1B universal donor bait HTGTS assay to evaluate previously described Cas9 EMX1 and VEGFA sgRNAs. For both EMX1 and VEGFA sgRNA targets, RAG1B bait HTGTS identified, respectively, the single and the four off-target sites previously documented by the established T7 endonuclease I (T7EI) cleavage assay, and also identified, respectively, an additional 12 and 34 novel off-target sites (Fig. 4d-f). Notably, all HTGTS-detected Cas9 EMX1 or VEGFA off-target sequences that we identified were related to the corresponding on-target sites and the majority were previously predicted but not confirmed (Supplementary Table 7 and Supplementary Fig. 11a). When tested by the T7EI assay, we also detected the two on-target break-sites and three previously described off-target sites; but only one of four tested off-target sites revealed by HTGTS (Supplementary Fig. 11b). Consistent with these findings, a prior T7EI assay study failed to detect 23 previously predicted Cas9 EMX1 or VEGFA sgRNA off-target sites that were clearly identified by our unbiased HTGTS assay. The RAG1B co-expression HTGTS assays also identified a large number of ATM and C-MYC TALEN off-target sites including all of the ~100 most dominant off-target sites detected by HTGTS using the individual TALEN break-sites as bait (Fig. 4g-i, Supplementary Fig. 9a,b and Supplementary Tables 4 and 5).

Low-level DSBs that occur widely across the genome can greatly influence translocation profiles of a given bait DSB. Specifically, treating cells with ionizing radiation to generate random ectopic DSBs ‘normalizes’ DSB frequency genome-wide, leading to diminution of dominant endogenous DSB hotspots and causing the length of a given break-site chromosome in cis to be a translocation hotspot region owing to a larger contribution of proximity effects.

To test ability of the RAG1B HTGTS assay to detect increased levels of widespread DSBs that do not qualify as hotspots individually, we introduced RAG1B into 293T cells for 24 h, then treated them with 7Gy of ionizing radiation (to introduce ~140 random DSBs per cell), further cultured for 24 h and performed HTGTS from the RAG1B bait-site. As predicted from prior mouse cell studies, ionizing radiation–treatment enhanced generation of HTGTS junctions that were greatly enriched across the entire RAG1B bait-site chromosome 11, with little or no increase on other chromosomes and a diminished recovery of break-site and recurrent off-target junctions (Fig. 4c, compare with Fig. 4a). Furthermore, similar ionizing radiation–treatment results were obtained with RAG1A break-site bait HTGTS libraries for the on-target site on chromosome 11 and also with two tested RAG1A off-target sites on chromosomes 12 and 19 (Supplementary Fig. 12). Thus, subsequent to induction of widespread ionizing radiation–generated DSBs genome-wide, each chromosome containing a particular on-target or off-target DSB hotspot became a hotspot region owing to a larger contribution of proximity effects.

Fig. 4 Identification of off-target sites from other nucleases using the universal Cas9:RAG1B bait. (a,b) Circos plots of RAG1B only (N = 6) (a) and RAG1B + I-SceI (N = 3) (b) HTGTS libraries. (c) Circos plot of ionizing radiation (IR)-treated Cas9:RAG1B (N = 3) libraries. Note the enrichment on the break-site chromosome. (d,e) Circos plots of RAG1B + Cas9:EMX1 (N = 3) (d) or Cas9:VEGFA (N = 3) (e) HTGTS libraries. Red asterisks indicate previously identified sites and blue asterisks indicate the on-target site (blue asterisks)—are shown for RAG1B bait. (f) Plot of on-target and off-target frequencies per 10,000 unique junctions from Cas9:EMX1 and Cas9:VEGFA (N = 3 each) using Cas9:RAG1B as the bait. (g,h) Circos plots of RAG1B + (g) ATM or (h) c-MYC TALEN HTGTS libraries. Note that only the top 10% of identified recurrent sites—including the on-target site (blue asterisks)—are shown for TALEN off-target sites. (i) Average junction frequencies for the top 50 recurrent ATM and c-MYC TALEN sites are listed. Blue colored lines for all Circos plots in this figure indicate RAG1B off-target sites (dark blue to light blue = high to low enrichment), whereas red to yellow colored lines (dark red to yellow = high to low enrichment) indicate the on and off-target sites of the co-expressed nuclease. (a–e,g,h) all Circos plots shown are normalized to 25,000 junctions and presented as similarly described in Figure 1. junction scale for all Circos plots: 20–20,000. (f,i) Error bars, mean±s.e.m. See Figure 1 legend for further description of Circos plots.
**Figure 5** Nuclease titration to measure on-target/background efficiency. (a–c) Circos plots (see Fig. 1 legend) of fixed concentration RAG1B bait + titrated ATM TALEN prey HTGTS libraries at 1 µg (a), 3 µg (b), 10 µg (c) of each ATM TALEN plasmid; total junctions for each library are shown (N = 1 each). See Supplementary Figure 15 for additional replicates and concentrations. Dark blue to light blue colored lines indicate RAG1B off-target sites, dark red to yellow colored lines designate ATM TALEN recurrent sites. (d) Graph of multiple RAG1B bait ATM TALEN titration HTGTS libraries plotting the ATM on-target frequency, the sum of the top 5 ATM TALEN off-target sites and the average of junction frequencies along the break-site chromosome (100 regions minus bait break-site region and known ATM off-target sites: 174 sites on chromosome 11; 880-bp mean size/site) ranging from 1 µg to 100 µg each (1 µg N = 2; 3–10 µg N = 3; 20–100 µg N = 2). Libraries were normalized to RAG1B OT1 junction enrichment. (e) The fold enrichment of the ATM TALEN on-target and the sum of the top five ATM TALEN off-target sites and the average of junction frequencies along the break-site chromosome. Overall, HTGTS not only revealed apparently widespread, but very low-level, DSB-inducing activities of some nucleases. Further studies will be needed to determine whether this type of activity is generated by a large number of very infrequently cleaved off-target sites (i.e., requiring some degree of sequence recognition), an even more random DSB-generating activity, or a combination of the two. Whatever the case, this latter application of the HTGTS assay also should be useful for testing nonspecific DSB activity of chemotherapeutic and other agents.

LAM-PCR–based HTGTS is a versatile assay that goes beyond simply detecting nuclease off-target sites by also revealing collateral damage in the form of recurrent translocations between on-target DSBs and off-target DSBs, as well as translocations between different off-target DSBs. Although not an ‘off-target’ event, HTGTS also revealed that a major translocation hotspot for on-target Cas9:sgRNA, Cas9-paired nickases, and TALEN-induced DSBs is the corresponding on-target DSB on the homologous chromosome that would lead to dicentric chromosome formation. Although probably being controlled substantially by cellular checkpoint responses in normal cells, dicentrics have the potential to generate additional genomic instability through breakage-fusion-bridge cycles. Many such dicentric junctions likely originate from translocations between break-sites on the two homologs that could be eliminated by engineered nucleases that recognize sequence nucleotide polymorphisms; but some conceivably could also result from stalled replication-fork-mediated template switching. Finally, our HTGTS findings indicate that widespread, low-level nuclease-generated DSBs can make each chromosome in a cell a marked hotspot region for translocations of on-target and/or off-target sites within it. Overall, HTGTS not only reveals all of these complex patterns of collateral damaged generated in the context of certain nucleases, but also provides an approach to estimate their relative frequency.

**Titration of engineered nuclease genome-wide activities**

Increasing levels of ATM TALEN over a tenfold range revealed additional lower-level, off-target sites and suggested an apparent increase in widespread, low-level DSBs (Supplementary Fig. 14). Assaying a single nuclease over increasing levels is not optimal for titrating on-target versus the various types of potential off-target activities, as both bait and prey breaks are influenced. Therefore, we employed the RAG1B bait assay to determine whether HTGTS could help assess optimal ATM TALEN nuclease levels for on-target versus off-target DSB activities. Frequencies of recovered ATM TALEN on-target versus the top five ATM TALEN off-target sites over the ATM TALEN widespread, low-level DSB activity is indicated for each transfected ATM TALEN plasmid concentration assayed.

**DISCUSSION**

Robust and accessible methods to test for off-target DSB-inducing activities of engineered nucleases are important as this class of enzymes continues to be groomed for human therapeutic purposes. We demonstrate that LAM-PCR–based HTGTS employing Cas9:sgRNAs, Cas9-paired nickases or TALEN bait DSBs provides a robust assay for identifying endogenous cellular DSB targets of these enzymes genome-wide in human cells. Thus, LAM-PCR–based HTGTS readily revealed off-target sites for a variety of different engineered nucleases. With respect to sensitivity, the off-target sites that we reproducibly detected in these studies included numerous sites predicted for previously tested nucleases but that had failed to be documented by existing methods, as well as a large number of off-target sites that were not predicted, but which were highly specific for each individually tested nuclease. Beyond off-target activities, we also found that HTGTS revealed apparently widespread, but very low-level, DSB-inducing activities of some nucleases. Further studies will be needed to determine whether this type of activity is generated by a large number of very infrequently cleaved off-target sites (i.e., requiring some degree of sequence recognition), an even more random DSB-generating activity, or a combination of the two. Whatever the case, this latter application of the HTGTS assay also should be useful for testing nonspecific DSB activity of chemotherapeutic and other agents.
Consistent with cellular heterogeneity in 3D genome organization allowing dominant DSB across the genome to drive recurrent translocations to each other^19,31, we identified the same large set of Cas9:RAG1A off-target DSBs in HTGTS assays that employed as bait, respectively, either the RAG1A on-target DSB site or three different RAG1A off-target DSB sites (each on a different chromosome). Based on this finding, we further improved the HTGTS assay by using the RAG1B DSB as a universal donor bait to identify on-target, off-target and low-level, widespread DSB activities of co-expressed nucleases. Indeed, this approach identified the known EMX1 and VEGFA sgRNA on-target and off-target sites, as well as many additional off-target sites. Thus, the modification of the HTGTS assay should facilitate rapid evaluation of on-target, off-target and low-level, widespread, DSB-generating activities of candidate nucleases from fixed-bait DSB sites without the need for generating and optimizing bait-site primers. Indeed, this approach can be used to identify engineered nuclease endogenous target sequences genome-wide even in cells that lack a known ‘on-target’ site for the nuclease tested as we showed for I-SceI.

The frequency of off-target sites for the four RAG1 Cas9:sgRNAs tested varied considerably, with two showing no detectable off-target activity. If desired, HTGTS could be scaled-up for even greater sensitivity and sensitivity also could be enhanced by performing HTGTS from target sites on each individual chromosome to increase identification of off-target sites on given chromosomes due to 3D proximity effects^27,33. HTGTS confirmed that off-target activity of the RAG1A sgRNA was dramatically suppressed genome-wide by the Cas9 D10 nickase approach^35,36, but also revealed that this approach does not suppress translocations involving DSBs on both bait-site chromosomes. Although two tested TALENs had numerous off-target sites, some. Although two tested TALENs had numerous off-target sites, suppression translocations involving DSBs on both bait-site chromosom -

## METHODS

Methods and any associated references are available in the online version of the paper.

### Accession codes.

The list of nuclease off-target sites is provided as Supplementary Data Set 1. HTGTS data are available from Gene Expression Omnibus GSE57283.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

## ACKNOWLEDGMENTS

This work is supported by the National Institutes of Health Grants P01CA109901 and P01AI076210. R.L.F was supported by the National Institutes of Health NRSA T32CA070803 and is supported by the National Health Institutes of Health NRSA T32AI007512. J.H. is supported by Robertson Foundation/Cancer Research Institute Irvington Fellowship. F.W.A. is an investigator of the Howard Hughes Medical Institute. A549 cells were provided by David Weinstock (Dana-Farber Cancer Institute).

### AUTHOR CONTRIBUTIONS

R.L.F., J.H. and F.W.A. designed the research; R.L.F., J.H. and E.K. performed the research; R.L.F., J.H., R.M.M., Y.-J.H. and E.K. analyzed the data; R.L.F. and F.W.A. wrote the paper.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.
34. Gostissa, M. et al. IgH class switching exploits a general property of two DNA breaks to be joined in cis over long chromosomal distances. Proc. Natl. Acad. Sci. USA 111, 2644–2649 (2014).
35. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838 (2013).
36. Ran, F.A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389 (2013).
37. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. 30, 460–465 (2012).
38. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233–239 (2013).
39. Asaithamby, A. & Chen, D.J. Cellular responses to DNA double-strand breaks after low-dose gamma-irradiation. Nucleic Acids Res. 37, 3912–3923 (2009).
40. Franco, S. et al. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. Mol. Cell 21, 201–214 (2006).
41. Ramiro, A.R. et al. Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature 440, 105–109 (2006).
42. Hastings, P.J., Lupski, J.R., Rosenberg, S.M. & Ira, G. Mechanisms of change in gene copy number. Nat. Rev. Genet. 10, 551–564 (2009).
43. Guilinger, J.P. et al. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. Nat. Methods 11, 429–435 (2014).
ONLINE METHODS
Plasmid DNA construction. Cas9 gRNAs targeting the RAG1 locus were cloned into pX330 or pX335 (Addgene plasmids 42230 and 42335, respectively) as described.15 The following gRNA targeting sequences (PAM) were used for Cas9 targeting: RAG1A: GCCTCTTTCCACCCACCTT (GGG), RAG1B: GACCTTATTCCATTTTGGAT (GGG), RAG1C: GCCAATATAATATA (AGG), RAG1D: GACCTTAAGTTTCGAGTTGGA (AGG), RAG1E: GCCATGCTTGGCTGAGACCT (AGG), RAG1F: GTACCGTGAGAACATGAAA (AGG), RAG1G: AAAAGGAGCCGGCATCTG GCTG (AGG). Guide RNAs for EMX1 were: GAGTCCGAGCTTTGGAAG (AGG), VEGFA: GGGTGGGGGGGAGTTTGCTCC (AGG), and VEGF: GGGTGGGGGGGAGTTTGCTCC (AGG). The EcoRI/XhoI 1-cleaved I-SceI vector from the pMX-I-SceI vector was cloned into the EcoRI/XhoI- cleaved pHR′-IRES-eGFP vector to generate pHR′-I-Sce1-IRES-eGFP.

Cell lines and transfection. 293T and A549 cells were maintained at 37 °C, 5% CO2 and cultured in DMEM with glutamine supplemented with 10% FCS and 0.5% penicillin/streptomycin (Invitrogen). 293T translocation libraries were prepared by CaPO4 co-transfection in 10-cm dishes of either 20 µg pX330 or 20 µg each pX335 gRNA combinations with 5 µg pCMX-eGFP followed by FACs analysis of GFP and DNA isolation 48 h after transfection. A549 cells (obtained from the American Type Culture Collection (ATCC)) and the CM130 program. A549 cells were cultured for 48 h before DNA isolation. For Cas9/I-SceI co-expression studies, pX330 gRNA vectors were co-transfected with pHR′-I-SceI followed by FACs analysis 48 h after transfection. TALEN pairs (previously generated by FLASH assembly37) corresponding to ATM and MYC targeting (Addgene plasmids 36805/36806 and 36713/36714, respectively)37 were co-transfected into 293T using 20 µg each—otherwise indicated—with 5 µg pCMX-eGFP and cultured for 48 h before FACs and DNA isolation. In some experiments 293T cells were gamma irradiated with 7 Gy 24 h after transfection.

HTGTS. General overview of the original emulsion-PCR method is described elsewhere.23 Junction cloning involved sonication (700 bp–1.5-kb fragment target size), end-repair (T4 DNA polymerase, polynucleotide kinase and Klenow Fragment DNA polymerase; New England BioLabs (NEB)), A-tailing, adaptor ligation, blocking digest to remove germline sequence, and locus-specific nested priming coupled with step-out adaptor priming to both enrich for captured junctions and suppress adaptor-ender fragments (Supplementary Table 8). Libraries included Illumina paired-end sequence for MiSeq along with the inclusion of extra nucleotides of variable length to enhance diversity of reads from the same locus. Illumina paired-end, sequence-specific primer tails (15, 17 sequence added to nested and adaptor primers, respectively) were used for emulsion PCR II followed by PCR III primers, P5 and P7, recognizing 15 and 17 sequences, respectively, to reconstruct the requisite P515 and P717 sequences necessary for MiSeq sequencing. Primers for HTGTS are listed (Supplementary Table 8). Taq polymerase (Qiagen) was used for all PCR translocation cloning steps. RAGA-D translocation cloning used EcoNI (NEB) to minimize amplification of germline fragments. RAG1 A/B HTGTS cloning conditions included the following: Biotin-PCR: 94 °C 120 s; 94 °C 20 s; 66 °C 60 s; 72 °C 60 s; 20 cycles; 72 °C 600 s. Emulsion PCR II: 94 °C 120 s; 94 °C 20 s; 66 °C 30 s; 72 °C 60 s; 30 cycles; 72 °C 600 s. PCR III: 94 °C 120 s; 94 °C 20 s; 64 °C 60 s; 72 °C 60 s; 12 cycles; 72 °C 600 s. RAG1 C/D HTGTS PCR conditions were identical to RAG1 A/B cloning but instead used 60 °C annealing for all steps.

HTGTS libraries were also generated using a modified protocol, which involved linear amplification-mediated (LAM)-PCR43 and bridge adaptor ligation44 to bypass end-repair and A-tailing steps followed by nested Illumina sequence-tailed PCR, blocking digest and a step-out tagging PCR to suppress germline sequence and to incorporate Illumina sequence tags for sequencing. This method is typically performed over 2 d with very little hands-on time the first day. Together with other HTGTS components (transfection, genomic DNA isolation, sequencing, filtering and analysis), the entire process is complete in less than 1 week. Briefly, sonicated DNA was subjected to LAM-PCR using Taq polymerase and the single biotinylated primer for 50 cycles. More Taq polymerase was added to the reaction mixture and proceeded for an additional 50 cycles. The 50 cycle PCR for all sites tested consisted of the following conditions: 94 °C 180 s; 94 °C 30 s; 58 °C 30 s; 72 °C 90 s; 50 cycles; 72 °C 600 s. Biotinylated DNA fragments were bound to MyOne C1 streptavidin beads (Invitrogen) before overnight on-bead ligation with bridge adapters (Supplementary Table 8) in the presence of 15% PEG-8000 (Sigma) and 1 mM hexamethane cobalt chloride (Sigma). Ligation conditions were as follows: 25 °C, 60 min; 22 °C, 120 min; 16 °C, 8–12 h. Adaptor-ligated products were subjected to nested-locus PCR for 15 cycles with primer tails corresponding to Illumina I5 and 17 sequences, digested with EcoNI to block germline sequence accumulation, and a final PCR for another 10–15 cycles with P5 and P7 primers to fully reconstruct MiSeq sequence tags (Supplementary Table 6).

Sequence analysis and hotspot identification. MiSeq reads were de-multiplexed and adaptor sequence trimmed using the fastx-mulx tool from ea-utils (https://code.google.com/p/ea-utils/) and the SeqPrep utility (https://github. com/jstjohn/SeqPrep), respectively. Reads were mapped to the hg19 reference genome using Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml) with the top 50 alignments reported that had an alignment score above 50, representing a perfect 25-nt local alignment. On average, 94% of de-multiplexed reads per library harbored a bait sequence alignment with <10% of these reads containing an alignable prey junction. We used a best-path searching algorithm to select the optimal sequence of alignments that describe the read’s composition, typically finding the bait and prey alignments. This approach was inspired by the YAHA read aligner and breakpoint detector45. Aligned reads were filtered on the following conditions: (i) reads must include both a bait alignment and a prey alignment and (ii) the bait alignment cannot extend more than 10 nt beyond the targeted site. For vector controls and offset nicking with multiple sites, the distal targeted site was used. We compared discarded alignments to the selected prey alignment; if any of the discarded alignments surpassed both a coverage and score threshold with respect to the prey alignment, the read was filtered owing to low mapping quality. To remove possible mispriming events and other artifacts, the bait alignment must extend 10 nt past the primer. We removed potential duplicates by comparing the coordinates of the end of the bait alignment and the start of the prey alignment across all reads. A read will be marked as a duplicate if it has a bait alignment offset within 2 nt and a prey alignment offset within 2 nt of another read’s bait and prey alignments. Post-filter stringency was applied to remove junctions with gaps larger than 30 nt and bait sequences shorter than 50 nt. Reads with prey alignments to telomere repeat sequences were also removed. Genome mixing experiments were similarly filtered as described above but with a combined hg19/mm9 reference.

Identification of enriched regions was performed using the MACs2 software,46 designed for ChIP-seq peak calling, which gave similar results to the previously described method.23 Junctions associated with MACs-defined peaks (FDR-adjusted P-value enrichment threshold of 10−9) were extracted for further analysis. Hotspots were defined as having significant focal enrichment and present in more than one biological replicate library. Hotspots proximal to the break-site (~100 kb) were excluded from analysis. Off-target sites were defined as hotspots that contained genomic sequence differing from the on-target sequence by less than or equal to one-half the targeted sequence length.

Code availability. Programs/scripts used for this manuscript are listed above and described elsewhere. Details of additional parameters not described above are available upon request.

1-Scel off-target in vitro digestion and T7EI cleavage assay. 1-Scel off-target sites were amplified from 293T DNA using phusion polymerase (NEB) and standard PCR conditions. Purified amplicons were digested with 1-Scel for 1 h at 37 °C. T7EI assays were performed as described previously using phusion polymerase and standard PCR conditions followed by ethanol precipitation before denaturation, reannealing, T7EI digestion of amplicons and agarose gel electrophoresis. Primers for each site are listed (Supplementary Table 8). EMX1 and VEGFA on- and off-target site primers are described elsewhere.

doi:10.1038/nbt.3101
To derive the fraction cleaved, we measured germline (amplicon) band intensities (quantified by ImageJ) and used the following formula:

\[ \text{Fraction Cleaved} = 1 - \left( \frac{I_{C+} - I_{N+}}{I_{C-} - I_{N-}} \right) \]

where \( I_{C+} \) is the intensity of T7EI-digested, nuclease-expressed sample, \( I_{N+} \) is the intensity of mock digested (no T7EI) nuclease-expressed sample, \( I_{C-} \) is the intensity of T7EI-digested, nuclease-deficient sample and \( I_{N-} \) is the intensity of the mock-digested, nuclease-deficient sample.

**Western analysis.** 293T cell protein lysates were separated by SDS-PAGE. Antibodies to detect FLAG-Cas9 (FLAG; Sigma F7425) and tubulin (loading control; Sigma T5168) were used at 1:2,000 and 1:5,000, respectively.

**Nucleic acid multiple sequence alignment.** Sequence logos for Cas9:gRNA and I-SceI off-target sites were generated by using the weblogo interface at weblogo.berkeley.edu.

**Statistical analysis.** Where appropriate, data were expressed as mean ± s.e.m. Two-way ANOVA and Tukey post-tests were done to compare the same individual off-target site frequencies from two different libraries. \( P < 0.05 \) was considered significant.

44. Zhou, Z.X. et al. Mapping genomic hotspots of DNA damage by a single-strand-DNA-compatible and strand-specific ChIP-seq method. *Genome Res.* **23**, 705–715 (2013).

45. Faust, G.G. & Hall, I.M. YAHA: fast and flexible long-read alignment with optimal breakpoint detection. *Bioinformatics* **28**, 2417–2424 (2012).

46. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).