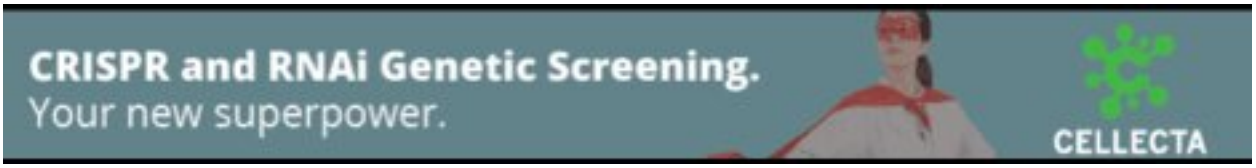
Global quantification exposes abundant low-level off-target activity by base editors

Ilana Buchumenski, Shalom Hillel Roth, Eli Kopel, et al.

Genome Res. published online October 19, 2021
Access the most recent version at doi:10.1101/gr.275770.121



Global quantification exposes abundant low-level off-target activity by base editors

Ilana Buchumenski[1], Shalom Hillel Roth[1], Eli Kopel[1], Efrat Katsman[1], Ariel Feiglin[1], Erez Y. Levanon[1,2*], Eli Eisenberg[3*]

Affiliations

[1] The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel

[2] The Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat Gan, 5290002, Israel

[3] Raymond and Beverly Sackler School of Physics and Astronomy and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, 69978, Israel

* Corresponding authors

Contact Info

Correspondence: elieis@post.tau.ac.il, erez.levanon@biu.ac.il

Running title: Low-level off-target activity by base editors

Keywords

Base-editors, off-targets, RNA editing

Abstract

Base editors are dedicated engineered deaminases that enable directed conversion of specific bases in the genome or transcriptome in a precise and efficient manner, and hold promise for correcting pathogenic mutations. A major concern limiting application of this powerful approach is the issue of off-target edits. Several recent studies have shown substantial off-target RNA activity induced by base editors and demonstrated that off-target mutations may be suppressed by improved deaminases versions or optimized guide RNAs. Here we describe a new class of off-target events that are invisible to the established methods for detection of genomic variations, and were thus far overlooked. We show that much of the off-target activity of the deaminases is nonspecific, seemingly stochastic, affecting a large number of sites throughout the genome or the transcriptome and accounting for the majority of off-target activity. We develop and employ a different, complementary, approach that is sensitive to the stochastic off-targets activity, and use it to quantify the abundant off-target RNA mutations due to current optimized deaminase editors. Engineered base editors enable directed manipulation of the genome or transcriptome at single-base resolution. We believe that implementation of this computational approach would facilitate design of more specific base editors. We provide a computational tool to quantify global off-target activity, which can be used to optimize future base editors.

Introduction

Base editors are dedicated engineered deaminases that enable directed conversion of specific bases in the genome or transcriptome in a precise and efficient manner and hold promise for correcting pathogenic mutations (Rees and Liu 2018; Komor et al. 2016). A major concern limiting application of this powerful approach is the issue of off-target edits(Gehrke et al. 2018; Zuo et al. 2019). Several recent studies(Grünewald et al. 2019a; Zhou et al. 2019; Rees et al. 2019) have shown substantial off-target RNA activity induced by base editors and demonstrated that off-target mutations may be suppressed

by improved deaminases versions or optimized guide RNAs. These studies have employed one of the established methods to find genomic variations reoccurring in multiple copies of the edited transcript(McKenna et al. 2010; Garrison and Marth 2012), which are well-suited for detection of genomic polymorphisms and reoccurring mutations, but are insensitive to weakly edited sites. Accordingly, these studies have mainly focused on specific off-target sites, where the guided deaminase binds efficiently and edits a substantial fraction of DNA/RNA molecules.

Endogenous deaminases (on which the engineered versions are based) are known to induce abundant low level RNA modifications (Rosenberg et al. 2011; Bazak et al. 2014a). In fact, most ADAR (also known as ADAR1) A-to-I mRNA editing activity occurs at sites edited up to a few percent level or even lower (Bazak et al. 2014a), and is therefore undetectable in a single sample. It is only natural to ask whether engineered deaminases too exhibit nonspecific, seemingly stochastic, off-target activity in addition to the previously studied specific off-target sites. As is the case with endogenous base-editors, the nonspecific activity may affect a large number of sites throughout the genome or the transcriptome, so that although these sites are edited at a low probability per site per molecule, nonspecific deamination events may outnumber the specific ones (Fig. 1A).

Global quantification methods have been devised to study endogenous deaminases activity, including nonspecific one (Bazak et al. 2014b; Roth et al. 2019). Inspired by this approach, we develop here a method to quantify the global variation rate induced by engineered base-editors, including low-level variations that cannot be individually resolved. We apply it to various state-of-the-art base editors, either DNA or RNA editors, engineered ADAR and APOBEC or other deaminases including ones that are fused to CAS proteins or to other guiding systems (Table S1)(Abudayyeh et al. 2019; Vallecillo-Viejo et al. 2018; Rees et al. 2019; Grünewald et al. 2019b; Yu et al. 2020; Vogel et al. 2018; Katrekar et al. 2019; Zhou et al. 2019; Grünewald et al. 2019a; Zuo et al. 2019; Lee et al. 2020; Doman et al. 2020).

Results

Most RNA mutations due to off-target base editing are non-specific

Genomic variations of interest typically occur in a sizable fraction of chromosome copies. For example, heterozygous polymorphisms are present in half of the molecules, and cancer-related somatic mutations of interest are those that have been selected and appear in multiple clonal copies. Methods for calling variations and mutations based on DNA-sequencing data(McKenna et al. 2010; Garrison and Marth 2012) use this fact to filter out sequencing and alignment technical noise, and analyze multiple alignment of many DNA reads to the reference genome looking for reoccurring mismatches. At typical read coverage, these approaches are insensitive to low-level variations that affect only a small fraction of the transcript copies and are thus almost indistinguishable from the technical noise level(Martincorena et al. 2018).

Most of the endogenous editing activity occurs in such low-level editing sites that are “invisible” to standard variant calling methods. Looking at several examples of engineered base editors we observe a similar phenomenon. In addition to specific off-target sites where the guided deaminase binds efficiently and edits a sizable fraction of DNA/RNA molecules, most off-target activity is nonspecific, seemingly stochastic, affecting a large number of sites at a low editing level. While these sites are edited at a low probability per site per molecule, nonspecific deamination events outnumber the specific ones. Recent optimization efforts have considerably lowered the volume of nonspecific off-target activity, but many state-of-the-art base editors exhibit sizable amounts of RNA off-targets in weakly edited sites (Fig. 1B and Supplemental fig. S1).

Technical errors in the sequencing process are still a major source for deviations between the original biological information and the output sequencing reads(Alioto et al. 2015; Zaranek et al. 2010). Therefore, a single read supporting an isolated mismatch cannot be reliably attributed to deaminase activity, and may result from sequencing errors. However, endogenous deaminases often edit multiple sites in the same molecule, resulting in clusters of mismatches of the same type that may be identifiable

even at the single read level (Carmi et al. 2011). Scanning RNA-seq data representing dozens of recently-developed base editors(Abudayyeh et al. 2019; Vallecillo-Viejo et al. 2018; Vogel et al. 2018; Katrekar et al. 2019; Grünewald et al. 2019b; Yu et al. 2020; Zhou et al. 2019; Grünewald et al. 2019a; Rees et al. 2019) (Supplementary table 1), we found in some of the samples up to a million off-target sites (Supplemental fig. S2A) in excess of the endogenous A-to-I editing signal observed in control samples(Porath et al. 2014). Only a small fraction of these were previously identified for the same samples (Supplemental fig. S2b), indicating that the classical SNV detection schemes may overlook a large fraction of sites. Note that for some editors no excess of hyper-editing sites is observed. However, the hyper-editing analysis is not sensitive enough and the clusters of sites it finds may be just the tip of the iceberg, attesting for a much wider non-clustered editing activity that is overlooked by standard variant-calling methods. Therefore, a different approach is required to explore non-specific off-target activity.

In the following, we apply global quantification methods for the editing activity, which are complementary to the above-mentioned variant calling approaches, to reveal the full scope of off-target editing activity.

Editing activity is globally enhanced following introduction of base-editors

To quantify the total off-target activity, we follow an approach developed previously for studying global endogenous RNA editing that takes into account the loads of editing activity occurring at low levels(Roth et al. 2019). Briefly, we apply a strict alignment approach and look at all mismatches to the reference genome, not trying to determine whether each one of them can reliably be attributed to deaminase activity. We then compare the editing index, the average mismatch level weighted by expression level, between samples (Methods). Clearly, this approach includes contributions from sequencing and other errors. However, the excess signal seen in samples expressing the deaminases over the baseline (control) signal attests for global editing activity.

Applying this approach, we find a statistically significant excess of 35 out of the 37 active enzymes analyzed, indicating widespread off-target RNA editing for current best

optimized A-to-I (Fig. 2A) and C-to-U (Fig. 2B) DNA and RNA base editors. Even for adenosine base editors, where endogenous A-to-I activity is known to be widespread(Roth et al. 2019), the effect of the engineered deaminase is clearly noticeable. This excess is seen genome-wide (Supplemental fig. S3), as well as in coding sequence regions. The sites harboring the mismatches exhibit a clear sequence motif, supporting their being targeted by the base editors (Supplemental fig. S4). Samples sequenced 36h post-treatment show a 2-3 fold higher level of induced mutations compared to ones sequenced 72h post treatment (Fig. 2C). Of the enzymes screened in this work, adenosine base editors seem to be noisier, showing up to 0.1% off-target editing in some of the optimized deaminases (i.e., one in a thousand adenosines is deaminated into an inosine). As expected, there is only a little overlap between the sites contributing to the global A-to-I signal and those detected by previous methods which are designed to locate specific editing sites. The excess signal comes mostly from weakly edited sites which are usually invisible to standard SNV detection tools (Supplemental fig. S1). Studies B and D provides matched data for RNA base editors with and without a nuclear localization signal (NLS). Enzymes containing NLS exhibit reduced non-specific off-target levels compared with matched enzymes containing a nuclear export signal (NES) (Supplemental fig. S5) as they have lower chances to meet other mRNAs.

Adenosine base editors target preferentially the endogenous targets of the ADAR enzymes. The vast majority of endogenous A-to-I editing occurs within the million *Alu* copies in the genome(Levanon et al. 2004; Athanasiadis et al. 2004; Blow et al. 2004; Kim et al. 2004). Consistently, the index calculated over *Alu* sequences is considerably elevated for all adenosine base editors (Supplemental fig. S3), and in some cases is as high as 10%, namely 1/10 of all *Alu* adenosines are deaminated. In contrast, editing of well-covered endogenous recoding sites, mostly edited by ADARB1 (also known as ADAR2) (Tan et al. 2017), is not generally elevated (Supplemental fig. S6).

The deamination rates observed (excess of index over then control base-line level) vary considerably across the different active base-editors, and ranges between 0.004% (study G, BE3(hA3AY130F)–site 3) and 3% (study G, BE3) in the coding regions of the human

genome (Fig. 2). Note that these rates are based on RNA collected 72h post-treatment. The deamination rate is actually 2-3 fold higher 36h post-treatment (Fig. 2C). Study E provides the expression level of the enzymes. The index values observed are not correlated to the expression level. In addition, we did not find a significant correlation between on-target and off-target for any of the Studies for which on-target data is available (A,C,D,E,F,G).

In terms of the total transcriptome mutation, the observed range of rates (0.004% to 3%) is equivalent to a range of 675 up to 658,000 heterozygous genomic mutations in the coding sequence alone (Fig. 2). (Fig. 2). The non-synonymous to synonymous ratio, the prevalence of non-conservative amino-acid substitutions, the distribution over genes with different expression levels, the level in essential genes and oncogenes, and the occurrence of known harmful mutations (Supplemental fig. S7) are all consistent with these non-specific off-target editing events being spread randomly and uniformly over the coding sequence of all genes.

Endogenous genetic information flow is not perfect and introduces errors at all levels. The sources of information are more tightly controlled than the end-point products. Thus, the genomic information itself is most tightly regulated, with replication error rates as low as $\sim 10^{-10}$ mutations per base pair per cell division(Drake et al. 1998), while protein production, the end point of this process, can tolerate error rates as high as 10^{-4} (Mordret et al. 2019). Due to the transient nature of RNA, transcriptional fidelity is much lower than replication but still much higher than translation, with error rates estimated to be $4 \cdot 10^{-6}$ (Gout et al. 2017, 2013) in eukaryotic cells. The evolutionary pressure to optimize eukaryotic transcription fidelity, in comparison with the bacterial one which is an order of magnitude higher(Traverse and Ochman 2016), suggests that higher rates of RNA substitutions are detrimental. Accordingly, the off-target activity identified here, which even for the best active base-editor analyzed here is an order of magnitude higher than the endogenous eukaryotic error rate, could be potentially harmful.

Low-level non-specific base editing affects only a small fraction of the copies of each protein. For the low end of the deamination rates (0.004%), a typical 1000bp-long coding mRNA sequence harboring ~250 cytosines will show an off-target deamination in only 1% of its copies. For highly expressed genes, where multiple copies of the transcript exist in each cell, the effect would probably be minimal. However, thousands of genes are expressed at a level of 1-2 transcripts per cell or less(Mele et al. 2015). Thus, each cell would have dozens of genes for which the only transcript in the cell is mutated. Furthermore, low level of RNA mutations may have harmful impact even if most copies of the transcript are not affected. Furthermore, 3.7% of A–to-G and 4.7% of C–to-T edits in coding regions are expected to create neoantigens (Methods) that can provoke the immune system to attack self-tissue. In addition, accumulation of mis-folded proteins may lead to aggregation, a key contributor to neurodegenerative diseases.

Further support for the potential damaging effect of low-level off-target RNA editing comes from the observation that the index over A-to-G harmful mutation sites is somewhat lower than the overall index (Supplemental fig. S7D (blue)). This depletion may indicate that weak off-target editing of these critical adenosines by the endogenous A-to-I editor is deleterious, and therefore selected against. A similar effect was recently reported for mammalian housekeeping genes and viruses that adapted to avoid editing by endogenous APOBEC enzymes(Chen and MacCarthy 2017).

Finally, we have looked for signs of deleterious effect at the cellular level, and used the GSEA tool and the hallmark gene-set(Liberzon et al. 2015) to look for over-expression of the apoptosis related gene-group (Methods). Significant over-expression is observed in 28 of the 33 enzymes analyzed (Table S11). For example, CDKN1A (also known as p21) is a key DNA damage–inducible protein whose transcriptional induction can occur dependent on TP53 and is considered a archetype of the cell response to genotoxic damage. We find significantly elevated expression of *cdkn1a* for most enzymes analyzed (Supplemental fig. S8 and Table S11). These results further support the need to control the level of off-target activity and to understand its cellular effects.

Off-target DNA editing activity following introduction of base-editors

The same approach may be used to quantify off-target activity not only in the transcriptome but also in the genome. We have analyzed four recent studies for which DNA-seq data was available(Lee et al. 2020; Doman et al. 2020; Zuo et al. 2019; Yu et al. 2020), and found detectable off-target DNA-editing activity for nine of the 22 enzymes studied (Studies D,K,I,J; Fig. 3A). Removing putative polymorphisms in the untreated samples (see Methods) improves the sensitivity, and reveals a significant signal for DNA off-target activity in ten of the 15 enzymes (Fig. 3B).

The absolute level of the excess index, representing off-target DNA mutations, is 2.2-$4.6\cdot10^{-5}$ (Fig. 3), lower than the one observed in RNA. However, it is still orders of magnitude higher than the natural DNA polymerase error rate of ~10^{-10}, and the impact of these heritable mutations is much more severe. To put it into a physiological context, one may compare the base-editors' off target activity to ionizing radiation, a different heavily-studied source of mutations. The mutation load due to (low rate) radiation is estimated to be $7.3\pm0.8\cdot10^{-6}$ mutations/bp/cell/Sievert(Russell and Kelly 1982b, 1982a). Note that our detection limit for mutation rate is roughly 10^{-5}, equivalent to ~1 Sievert, is still much higher than the accepted ionizing radiation safety limit. Thus, the mutation load detected in some of the editors calls for a more accurate sequencing and quantification methods to assess the risk due to off-target DNA editing, even if it is too weak a signal to be detected using current standard sequencing protocols and our approach.

Discussion

We present here a new method to quantify non-specific off-target activity. Off-target RNA mutations are found to be abundant even for current optimized deaminase editors, and some of these editors result in abundant DNA mutations as well. We provide a computational tool (https://github.com/a2iEditing/BEIndexer) to quantify global off-target activity, which can be used to optimize future base editors.

Note that this approach does not replace currently-used methods which are designed to identify specific off-target sites, but is presented as a complementary approach, focusing on a different manifestation of off-target activity. This method reveals varying levels of appreciable off-target activity induced by state-of-the-art base editors. It is not straight-forward to compare the base editors using these data, gathered under varying experimental conditions. However, these results emphasize the need for further optimization of base-editors with respect to their non-specific off-target activity, genomic and transcriptomic. The approach presented here may be utilized for these optimization efforts. In addition, cut-off values for tolerable RNA and DNA mutation rate in the context of genetic therapy should be established.

Methods

Sequencing Data and Alignment

To analyze off-target editing of RNA by recently developed base editors, we downloaded RNA-seq data from seven different studies(Abudayyeh et al. 2019; Vallecillo-Viejo et al. 2018; Grünewald et al. 2019b; Vogel et al. 2018; Katrekar et al. 2019; Zhou et al. 2019; Grünewald et al. 2019a; Yu et al. 2020) (Supplementary table 1). In total, we analyzed 306 RNA-seq samples, representing 48 Adenosine base editors, 29 Cytidine base editors and controls. Details are provided in Supplementary table 1. RNA-seq reads were aligned to the human (hg38) genome using STAR2.6.0(Dobin et al. 2013) (default parameters), keeping only uniquely-aligned reads.

To analyze genomic off-target editing by recently developed base editors, we downloaded DNA-seq data from four recently published studies(Zuo et al. 2019; Lee et al. 2020; Yu et al. 2020; Doman et al. 2020) (Supplementary table 1). In total we analyzed 493 DNA-seq samples, representing three base-editors (BE3, BE4 and ABE7.10) and controls. DNA-seq reads were uniquely aligned to the mouse (mm10) or human (hg38) genomes using STAR2.6.0(Dobin et al. 2013) (default parameters, except of alignIntronMin= 2, scoreDelOpen= -10000, scoreInsOpen= -10000 in order to avoid spliced alignments).

It is worth noting that, studies E and L have applied single cell sequencing to the DNA samples, and show low alignment levels (<40%).

Genomes (hg38 and mm10) and gene annotations (RefSeq data) were downloaded from the UCSC Genome Browser(Haeussler et al. 2019) (Table ncbiRefSeq).

Hyper-editing

We used the hyper-editing algorithm as previously described(Porath et al. 2014) to identify heavily edited RNA-seq reads which the aligner fails to align to the genome. Many of the off-targets identified by this tool occur within coding regions. Our computation tool for detecting hyper-edited reads is available at https://github.com/hagitpt/Hyper-editing

Off-target index calculation

To measure the total off-target activity we followed the previously developed approach for calculating the *Alu* editing index(Roth et al. 2019). This measure is robust, and takes into account low-level variations that cannot be individually determined. The details of the present calculation are identical to those specified for the *Alu* editing index, except that (i) we did not look at *Alu* elements alone, but also at the coding regions, or the whole genome; and (ii) we calculated the A-to-G index or the C-to-U index, depending of the base editor analyzed. Most of our results deal with the coding region, for which we assumed all reads were expressed from the annotated coding strand (see Roth et al(Roth et al. 2019) for analysis of the accuracy of this approach). Briefly, we calculated the weighted average over millions of genomic cytosines (9,700,565 C locations in coding regions for C-to-T) and adenosines (8,754,152 A locations in coding regions for A-to-G), while the weights are the total number of reads in these sites. To estimate the noise level, we measured the abundance of A-to-T substitution, which is the substitution type (other than A-to-G and C-to-T) with the highest noise level in most studies. Genomic sites overlapping common single nucleotide polymorphisms (SNPs) (human:dbSNP150; mouse:dbSNP142) were excluded.

We have made the computation tool for calculating the base editing index available at https://github.com/a2iEditing/BEIndexer. Importantly, the editing index is sensitive to reads' length(Roth et al. 2019). Longer reads may be mapped even if they include multiple editing events, and thus lead to higher values of the index. The different studies examined here have used reads of varying lengths, and therefore the index values should not be compared across studies. We decided not to trim the reads of all studies to a uniform length, since some of them are very short (35bp).

In addition, we calculate the index for sites or regions of special interest: (1) Housekeeping exons(Eisenberg and Levanon 2013) which are essential for the existence of a cell. (2) Oncogenes(Tate et al. 2019) with known mutations that have been causally implicated in cancer. (3) Reported pathogenic clinical variations with known human phenotype (ClinVar(Landrum et al. 2014)).

We annotated the potential substitutions in the coding regions using ANNOVAR(Wang et al. 2010). Amino-acids substitutions were classified to non-conservative if the substitution results in an amino acid of a different group, using the following classification: electrically charged side chain (R, H, K, D and E), polar uncharged (S, T, N and Q), hydrophobic (A, I, L, M, F, W, Y and V), and three amino-acids with special side chain cases – Cysteine (C), Glycine (G) and Proline (P). These mismatches can lead to changes in the protein structure and function. Amino-acids substitutions within the same group are termed conservative.

For DNA off-target editing, we took an additional step to improve our sensitivity, and discarded mismatch sites that were detected in at least in half of a set of untreated samples (Cas9 for study H and parental samples for study I) from all other samples. These sites are likely to be enriched in genomic polymorphisms that are common to all samples, and are not due to the base editors' activity.

Conserved recoding sites editing

Editing levels at the endogenous mammalian conserved RNA editing sites(Pinto et al. 2014) were calculated using REDItoolsKnown.py script that is part of REDItools(Picardi

and Pesole 2013) package. We calculated the editing level only if site coverage exceeds 10 reads. The weighted average of all these sites was used to calculate the conserved recoding index for each case.

Gene expression analysis

Gene expression levels of all genes were calculated by using the Salmon(Patro et al. 2017) tool.

Neoantigen simulation

Neoantigen creation by off-target editing was simulated as follows. All 57,204 coding mRNA sequences were downloaded from the UCSC Table Browser (table "UCSC RefSeq (refGene)", filtered by "*name does match NM_**" and "*chrom doesn't match alt* fix* random* chrUn**"). These sequences represent the transcriptome (including splice-variants), and include 58,472,242 adenosines and 50,119,924 cytosines. A Python script was used to perform 100,000 A-to-G or C-to-T substitutions at randomly chosen locations. The edited mRNA sequences were translated to proteins, and the resulting peptides were analyzed using netMHCpan(Lundegaard et al. 2008) to identify potentially immunogenic peptides. Binding was evaluated for 9-mer sequences with the human HLA-A02:01 MHC-I allele. For amino-acid substitutions, peptide sequences ranging eight amino acids up- and down-stream of the edited amino acid were used for binding prediction. For stop-loss edits, the sequence up to the next stop codon (or the end of the UTR if no stop codon occurred) was examined. "Strong binders" were determined using default netMHCpan parameters and cutoffs, with the "Binding Affinity" option selected. We then excluded strong binding 9-mers were that exist in the naturally occurring human proteome (downloaded from UniProtKB(Bateman 2019) database using the following query: "organism:"*Homo sapiens* (Human) [9606]" AND proteome:up000005640"). The resulting strong-binders 9-mers were designated as neoantigens.

Gene-set enrichment analysis

Gene expression levels were computed (using Salmon) for all genes and analyzed for enrichment of the apoptosis-related gene-set using GSEA software and version 4.0.2 (10,000 permutations) and the hallmark gene-set collection(Liberzon et al. 2015).

Data analysis

The statistical analysis was calculated using R v.3.5.1(Core Development Team 2018). All tests conducted were two-sided, and the significant difference was considered as p.value < 0.05.

Software availability

Editing index software is available at GitHub (https://github.com/a2iEditing/BEIndexer) and as Supplemental Code.

Competing Interest Statement

E.E. is a consultant to Korro Bio, a company that develops RNA editors. E.Y.L is a consultant to ADARx.

Acknowledgments

We thank Ayal Hendel, Shay Ben-Aroya and the Levanon lab members for fruitful discussions.

This work was supported by the Israel Science Foundation [grant numbers 2673/17, 1945/18 to EE and 2039/20, 231/21 to EYL], as well as an International Collaboration Grant from the Jacki and Bruce Barron Cancer Research Scholars' Program, a partnership of the Israel Cancer Research Fund and City of Hope, as supported by The Harvey L. Miller Family Foundation [grant number 205467 to EYL].

Authors' Contributions

IB performed most of the bioinformatics data analyses. SHR designed and wrote the software. EK has initiated the bioinformatics work. AF have contributed to specific data analyses. EYL and EE conceived the study, designed the analyses, and wrote the paper. All authors read and approved the final manuscript.

References

Abudayyeh OO, Gootenberg JS, Franklin B, Koob J, Kellner MJ, Ladha A, Joung J, Kirchgatterer P, Cox DBT, Zhang F. 2019. A cytosine deaminase for programmable single-base RNA editing. *Science (80-)* **365**: 382–386. http://www.ncbi.nlm.nih.gov/pubmed/31296651.

Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, Heisler LE, Beck TA, Simpson JT, Tonon L, et al. 2015. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat Commun* **6**: 10001. http://www.ncbi.nlm.nih.gov/pubmed/26647970.

Athanasiadis A, Rich A, Maas S. 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* **2**: e391. http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0020391 (Accessed May 8, 2015).

Bateman A. 2019. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res* **47**: D506--D515.

Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, Eisenberg E, et al. 2014a. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res* **24**: 365–76. http://genome.cshlp.org/content/24/3/365 (Accessed October 24, 2015).

Bazak L, Levanon EY, Eisenberg E. 2014b. Genome-wide analysis of Alu editability. *Nucleic Acids Res* **42**: 6876–84.

Blow M, Futreal AP, Wooster R, Stratton MR. 2004. A survey of RNA editing in human brain. *Genome Res* **14**: 2379–2387. http://www.ncbi.nlm.nih.gov/pubmed/15545495.

Carmi S, Borukhov I, Levanon EY. 2011. Identification of widespread ultra-edited human RNAs. *PLoS Genet* **7**: e1002317. http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002317 (Accessed November 12, 2015).

Chen J, MacCarthy T. 2017. The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retrotransposon and virus sequences compared to host genes. *PLoS Comput Biol* **13**: e1005471. http://www.ncbi.nlm.nih.gov/pubmed/28362825.

Core Development Team R. 2018. A Language and Environment for Statistical Computing. *R Found Stat Comput*.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. http://bioinformatics.oxfordjournals.org/content/29/1/15 (Accessed July 13, 2014).

Doman JL, Raguram A, Newby GA, Liu DR. 2020. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat Biotechnol* **38**: 620–628. https://doi.org/10.1038/s41587-020-0414-6.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. *Genetics* **148**: 1667–1686.

Eisenberg E, Levanon EY. 2013. Human housekeeping genes, revisited. *Trends Genet* **29**: 569–574. http://www.ncbi.nlm.nih.gov/pubmed/23810203.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. http://arxiv.org/abs/1207.3907.

Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, Pinello L, Joung JK. 2018. An apobec3a-cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol* **36**: 977. http://www.ncbi.nlm.nih.gov/pubmed/30059493.

Gout J-F, Li W, Fritsch C, Li A, Haroon S, Singh L, Hua D, Fazelinia H, Smith Z, Seeholzer S, et al. 2017. The landscape of transcription errors in eukaryotic cells. *Sci Adv* **3**: e1701484. http://www.ncbi.nlm.nih.gov/pubmed/29062891.

Gout J-F, Thomas WK, Smith Z, Okamoto K, Lynch M. 2013. Large-scale detection of in vivo transcription errors. *Proc Natl Acad Sci U S A* **110**: 18584–18589. http://www.pnas.org/cgi/doi/10.1073/pnas.1309843110.

Grünewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, Joung JK. 2019a. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**: 433–437. http://www.nature.com/articles/s41586-019-1161-z.

Grünewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, Joung JK. 2019b. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol* **37**: 1041–1048. http://www.nature.com/articles/s41587-019-0236-6.

Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. 2019. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* **47**: D853--D858. http://www.ncbi.nlm.nih.gov/pubmed/30407534.

Katrekar D, Chen G, Meluzzi D, Ganesh A, Worlikar A, Shih Y-R, Varghese S, Mali P. 2019. In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat Methods*. http://www.ncbi.nlm.nih.gov/pubmed/30737497.

Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A. 2004. Widespread RNA editing of embedded Alu elements in the human transcriptome. *Genome Res* **14**: 1719–1725. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15342557.

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**: 420–424. http://www.nature.com/articles/nature17946.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**.

Lee HK, Smith HE, Liu C, Willi M, Hennighausen L. 2020. Cytosine base editor 4 but not adenine base editor generates off-target mutations in mouse embryos. *Commun Biol* **3**: 1–6.

Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, et al. 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* **22**: 1001–5. http://dx.doi.org/10.1038/nbt996 (Accessed June 18, 2015).

Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**: 417–425. https://pubmed.ncbi.nlm.nih.gov/26771021/ (Accessed September 14, 2021).

Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res* **36**: W509--12. http://www.ncbi.nlm.nih.gov/pubmed/18463140.

Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, et al. 2018. Somatic mutant clones colonize the human esophagus with age. *Science (80-)* **362**: 911–917.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303.

Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, et al. 2015. The human transcriptome across tissues and individuals. *Science (80-)* **348**: 660–665. http://www.ncbi.nlm.nih.gov/pubmed/25954002.

Mordret E, Dahan O, Asraf O, Rak R, Yehonadav A, Barnabas GD, Cox J, Geiger T, Lindner AB, Pilpel Y. 2019. Systematic Detection of Amino Acid Substitutions in Proteomes Reveals Mechanistic Basis of Ribosome Errors and Selection for Translation Fidelity. *Mol Cell* **75**: 427--441.e5. http://www.ncbi.nlm.nih.gov/pubmed/31353208.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**: 417–419. http://www.nature.com/articles/nmeth.4197 (Accessed July 3, 2019).

Picardi E, Pesole G. 2013. REDItools: high-throughput RNA editing detection made easy. *Bioinformatics* **29**: 1813–4.

Pinto Y, Cohen HY, Levanon EY. 2014. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biol* **15**: R5. http://genomebiology.com/2013/15/1/R5 (Accessed October 24, 2015).

Porath HT, Carmi S, Levanon EY. 2014. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat Commun* **5**: 4726. http://www.nature.com/ncomms/2014/140827/ncomms5726/full/ncomms5726.html?WT.ec_id=NCOMMS-20140903 (Accessed February 17, 2015).

Rees HA, Liu DR. 2018. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* **19**: 770–788.

Rees HA, Wilson C, Doman JL, Liu DR. 2019. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci Adv* **5**: eaax5717. https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aax5717.

Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S, Papavasiliou FN. 2011. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA editing targets in transcript 3 UTRs. *Nat Struct Mol Biol* **18**: 230.

Roth SH, Levanon EY, Eisenberg E. 2019. Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat Methods* **16**: 1131–1138. http://www.ncbi.nlm.nih.gov/pubmed/31636457.

Russell WL, Kelly EM. 1982a. Mutation frequencies in male mice and the estimation of genetic hazards of radiation in men. *Proc Natl Acad Sci U S A* **79**: 542–544. http://www.ncbi.nlm.nih.gov/pubmed/6952206.

Russell WL, Kelly EM. 1982b. Specific-locus mutation frequencies in mouse stem-cell spermatogonia at very low radiation dose rates. *Proc Natl Acad Sci U S A* **79**: 539–541. http://www.ncbi.nlm.nih.gov/pubmed/6952205.

Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, Ariyoshi K, et al. 2017. Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**: 249–254.

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, et al. 2019. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**: D941--D947.

Traverse CC, Ochman H. 2016. Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proc Natl Acad Sci* **113**: 3311–3316.

Vallecillo-Viejo IC, Liscovitch-Brauer N, Montiel-Gonzalez MF, Eisenberg E, Rosenthal JJC. 2018. Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol* **15**: 104–114. http://www.ncbi.nlm.nih.gov/pubmed/29099293.

Vogel P, Moschref M, Li Q, Merkle T, Selvasaravanan KD, Li JB, Stafforst T. 2018. Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat Methods* **15**: 535–538. http://www.nature.com/articles/s41592-018-0017-z.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**: e164. http://www.ncbi.nlm.nih.gov/pubmed/20601685.

Yu Y, Leete TC, Born DA, Young L, Barrera LA, Lee SJ, Rees HA, Ciaramella G, Gaudelli NM. 2020. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat Commun* **11**: 1–10. https://doi.org/10.1038/s41467-020-15887-5.

Zaranek AW, Levanon EY, Zecharia T, Clegg T, Church GM. 2010. A survey of genomic traces reveals a common sequencing error, RNA editing, and DNA editing ed. D. Schübeler. *PLoS Genet* **6**: 8. http://dx.plos.org/10.1371/journal.pgen.1000954.

Zhou C, Sun Y, Yan R, Liu Y, Zuo E, Gu C, Han L, Wei Y, Hu X, Zeng R, et al. 2019. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**: 275–278. http://www.nature.com/articles/s41586-019-1314-0.

Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, Yuan L, Steinmetz LM, Li Y, Yang H. 2019. Cytosine

base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science (80-)* **364**: 289–292. http://www.ncbi.nlm.nih.gov/pubmed/30819928.

Figure legends

Fig. 1: Most RNA mutations due to off-target base editing are non-specific (A) Engineered guided deaminases may target efficiently some off-target locations (marked with arrows). These strongly edited sites result in a DNA-RNA mismatch seen in a large fraction of RNA molecules, resembling the mismatch profile observed at genomic polymorphism and clonally selected somatic mutation sites. In parallel, nonspecific off-target base editing activity affects multiple additional sites but a small fraction of RNA molecules per site. However, due to their large number, the total number of non-specific events could surpass that of specific editing sites, as illustrated in the Figure. Hyper-edited reads, where multiple mismatches are seen in the same read (right panel, top read), occasionally appear. They provide a strong indication of off-target activity, but account for a small minority of mismatches. **(B)** Relative contribution of editing sites to the editing index, by their observed editing rates, for two engineered base-editors. Most off-target activity occurs at weakly-edited sites in agreement with the scenario depicted in panel A. Pie area is proportional to the number of detected off-target events. Four base editors from study H are shown (two Cytidine deaminases: BE3, BE3 (hA3AY130F)–site 3; two Adenine base-editors: ABE7.10 and ABE7.10 (F148A) site 1), see fig. S1 for similar data for all enzymes.

Fig. 2: Editing activity is globally enhanced following introduction of base-editors. The editing index is a global measure of editing activity, quantifying the fraction (per cent) of the RNA nucleotides exhibiting a DNA-RNA mismatch (i.e. A-to-G index of 1 means that 1% of the RNA nucleotides mapped to a genomic adenine are guanosines). **(A)** For Adenine base-editors (Studies A, B, C, D, F, G, H), the A-to-G index (blue circles) per-sample over the coding sequence (See fig. S3 for whole-genome calculation) is presented. In almost all cases tested, the index is significantly elevated for base editors compared to the control samples (two-sided *t*-test for log(index); * $p<0.05$, ** $p<0.01$; *** $p<0.001$). Significance was not assessed for study A (did not include untreated controls) and studies B and G (only one replicate per condition). The two cases in study F that do not show a significant difference exhibit weak on-target activity as well. In order to appreciate the significance of the high index values obtained, the index values are translated into an equivalent number of heterozygous mutations (Methods), right axis. Note that the index cannot be directly compared between samples of different reads' lengths (Methods). **(B)** Same as A, C-to-U index (red) for Cytidine deaminase samples (Studies A, D, E, H, I). **(C)** Samples sequenced 36h post-treatment show a 2-3 fold higher level of induced mutations compared to ones sequenced 72h post treatment (as are all samples in panels A,B). The data per sample is available in Supplemental Table S2. Exact p-values are presented in Supplemental Table S3.

Fig. 3: Off-target DNA editing activity following introduction of base-editors. (A) The editing index is a global measure of editing activity, quantifying the fraction (per cent) of the DNA nucleotides exhibiting a mismatch with respect to the reference genome. Common polymorphic sites are excluded. Four studies are analyzed (E, L, J, K), examine Adenine base-editors and Cytidine deaminases. For each enzyme, the relevant indices (A-to-G, blue, and C-to-T, red) are

calculated over the coding region. The relevant index (A-to-G and C-to-T for Adenine and Cytidine deaminases, respectively) are compared with control samples (Study E: NC/None, Study L: nCas9, Study J: Cre, Study K: Control). A significant increase is observed for nine enzymes. **(B)** In order to suppress baseline contributions to the index due to genomic polymorphisms unrelated to the base-editor, we repeated the calculation excluding mismatch sites appearing in at least half on the untreated samples not used for the statistical tests (Study E: Cas9, Study L: parent, Study J: Cas9, Study K: parents). Ten of the 15 examined editors exhibit a significant increase in the index, indicating off-target DNA editing. In study E (mRNA delivery), there were no samples suitable for excluding shared SNVs, and therefore filtering was not applied. Two-sided *t*-test for log(index) (* $p<0.05$, ** $p<0.01$; *** $p<0.001$). The data per sample is available in Supplemental TableS4 and exact p-values are presented in Supplemental Table S5.