In vivo measurements of interindividual differences in DNA glycosylases and APE1 activities

Isaac A. Chaim,a,b Zachary D. Nagelb,a,b Jennifer J. Jordan,a,b Patrizia Mazzucatoa,b, Le P. Ngoa,b, and Leona D. Samsona,b,c,d,1

*Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; 2Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; 3Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and 4The David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Paul Modrich, Howard Hughes Medical Institute and Duke University Medical Center, Durham, NC, and approved October 20, 2017 (received for review July 6, 2017)

The integrity of our DNA is challenged with at least 100,000 lesions per cell on a daily basis. Failure to repair DNA damage efficiently can lead to cancer, immunodeficiency, and neurodegenerative disease. Base excision repair (BER) recognizes and repairs minimally helix-distorting DNA base lesions induced by both endogenous and exogenous DNA damaging agents. Levels of BER-initiating DNA glycosylases can vary between individuals, suggesting that quantifying and understanding interindividual differences in DNA repair capacity (DRC) may enable us to predict and prevent disease in a personalized manner. However, population studies of BER capacity have been limited because most methods used to measure BER activity are cumbersome, time consuming and, for the most part, only allow for the analysis of one DNA glycosylase at a time. We have developed a fluorescence-based multiplex flow-cytometric host cell reactivation assay wherein the activity of several enzymes [four BER-initiating DNA glycosylases and the downstream processing apurinic/apyrimidinic endonuclease 1 (APE1)] can be tested simultaneously, at single-cell resolution, in vivo. Taking advantage of the transcriptional properties of several DNA lesions, we have engineered specific fluorescent reporter plasmids for quantitative measurements of 8-oxoguanine DNA glycosylase, alkyladenine DNA glycosylase, MutY DNA glycosylase, uracil DNA glycosylase, and APE1 activity. We have used these reporters to measure differences in BER capacity across a panel of cell lines collected from healthy individuals, and to generate mathematical models that predict cellular sensitivity to methylmethane sulfonate, H2O2, and 5-FU from DRC. Moreover, we demonstrate the suitability of these reporters to measure differences in DRC in multiple pathways using primary lymphocytes from two individuals.

DNA repair | transcriptional mutagenesis | DNA glycosylase | apurinic/apyrimidinic endonuclease 1 | base excision repair

Our genomes are constantly challenged with damage that arises endogenously from products of cellular metabolic processes, as well as exogenously from environmental DNA damaging agents (1, 2). Failure to efficiently repair DNA damage can lead to cancer, degenerative disease, and premature aging (3–5). Thus, a variety of mechanisms to repair DNA damage have evolved that serve to delay the onset of such diseases.

The base excision repair (BER) pathway recognizes and repairs small, nonbulky DNA lesions, including alkylated, deaminated, and oxidized bases (5). BER involves five steps: (i) recognition and excision of a damaged base by a DNA glycosylase, forming an abasic site; (ii) incision of the sugar-phosphate DNA backbone at the abasic site; (iii) processing and removal of remaining sugar moieties; (iv) gap-filling by a DNA polymerase; and (v) sealing of the remaining nick by a DNA ligase (2).

BER deficiencies have been linked with several diseases. MutY DNA glycosylase (MUTYH) deficiency confers colorectal cancer predisposition (6), and uracil DNA glycosylase (UNG) deficiency leads to hyper IgM-syndrome (7). Furthermore, variations in the activities of several BER enzymes are associated with increased cancer risk and other diseases (8–10). However, the lack of high-throughput assays that can reliably measure interindividual differences in BER capacity have limited epidemiological studies linking BER capacity to disease. Moreover, BER requires DNA lesions induced by radiation and chemotherapy (3, 11), raising the possibility of personalized treatment strategies using BER capacity in tumor tissue to predict which therapies are most likely to be effective, and using BER capacity in healthy tissue to determine the dose individual patients can tolerate.

Although methods have been developed to measure BER activity for every step of the pathway (12), each has drawbacks that limit the potential for use in epidemiological studies and clinical translation. In vitro assays using cell lysates may not recapitulate physiological DNA repair conditions. With few exceptions (13, 14), the lack of multiplex assays for multiple DNA glycosylases and downstream BER steps has resulted in studies focusing on a single-repair step or a single-repair pathway. A recently described fluorescence-based multiplex flow-cytometric host cell reactivation (FM-HCR) overcomes these challenges, allowing for in vivo multiplexed DNA repair capacity (DRC) measurements for multiple DNA repair pathways (15). Here we report on the development of FM-HCR reporters for apurinic/apyrimidinic endonuclease 1 (APE1), and for four DNA glycosylase activities. We apply these reporters to measure interindividual variations in BER in a panel of 24 cell lines derived from healthy individuals, and build mathematical models that predict cellular sensitivity to several agents that induce DNA damage processed by the BER pathway. The models reveal the

Significance

The DNA in each cell is damaged thousands of times daily. Consequently, a battery of DNA repair pathways exist that allow repair of this damage. Failure to repair can lead to devastating diseases, including cancer and neurodegeneration. Each individual's DNA repair capacity (DRC) is inherently different. Being able to measure an individual's DRC could contribute to a personalized approach to prevent and treat disease. Here we present powerful tools for measuring in vivo base excision repair capacity for five distinct DNA lesions. We use these methods to predict the cellular responses to a variety of DNA damaging agents, and to monitor differences in DRC in primary human lymphocytes. Additionally, we unveil previously unknown transcriptional mutagenesis induced by DNA lesions.

Author contributions: I.A.C. and L.D.S. designed research; I.A.C., Z.D.N., J.J.J., P.M., and L.P.N. performed research; I.A.C. contributed new reagents/analytic tools; I.A.C. and J.J.J. analyzed data; and I.A.C., Z.D.N., and L.D.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Published under the PNAS license.

1To whom correspondence should be addressed. Email: lsamson@mit.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/10.1073/pnas.1712032114/DCSupplemental.
relative contribution of multiple repair pathways to cellular sensitivity, and highlight the potential for FM-HCR assays to guide clinicians in selecting therapy for individual patients. Finally, we demonstrate the potential of our DRC reporters to inform epidemiological studies by measuring DRC in primary human T lymphocytes.

Results

Validation of FM-HCR Assays for Measuring DNA Glycosylase Activity. Building upon our previously described FM-HCR substrate that reports 8-oxoguanine DNA glycosylase (OGG1) activity (15), we took advantage of DNA-lesion–induced transcriptional mutagenesis to design FM-HCR substrates that report activity for UNG, alkyladenine DNA glycosylase (AAG, also known as MPG), and MUTYH.

We first generated plasmids bearing a nonsynonymous site-specific mutation that inactivates the chromophore for each of four fluorescent proteins, namely mOrange, BFP, GFP, and mPlum. We further modified these plasmids by positioning one of four different site-specific DNA lesions at each mutated site, each lesion being a substrate for repair by one of the four DNA glycosylases. The mOrange reporter contained 8-oxoguanine (8oxoG) to measure OGG1 activity; the BFP reporter contained a uracil (U) to measure UNG activity; the GFP reporter contained hypoxanthine (Hx) to measure AAG activity; and the mPlum reporter contained adenine opposite 8oxoG to measure MUTYH activity. Following transfection, transcriptional mutagenesis induced by the site-specific lesion leads to expression of fluorescent protein; detailed examples and the rationale for this approach are presented.

Fig. 1. Rationale for transcriptional mutagenesis-based fluorescent reporters. The base pairs shown correspond to sites that code for a key amino acid of the chromophores of the fluorescent proteins. The transcribed strand is always shown on top. The base shown in mRNA corresponds to the ribonucleotide incorporated by RNAPII across from the specific base in the plasmid. OG, 8-oxoguanine. (A) WT reporter codes for a fluorescent protein, incorporation of the base shown at that specific site is necessary for fluorescence to occur. (B) Nonfluorescent reporter variants. (C) Base misincorporation opposite DNA lesions. Only transcriptional mutagenesis events (or incorporation of U opposite A for the A:OG base pair), caused by the presence of the lesion (or incorrect base for the A:OG base pair) will result in the expression of a fluorescent variant of the reporter. Incorporation of any other ribonucleotides results in nonfluorescent proteins. (1) To generate substrates, WT reporters were mutated through QuikChange site-directed mutagenesis (Materials and Methods) to produce nonfluorescent variants. (2) Through primer extension (Materials and Methods), base lesion is incorporated on the transcribed strand at the mutated site of the nonfluorescent variant (with the exception of the A:OG base pair for which OG is incorporated on the nontranscribed strand of the WT variant). (3) Repair by the cellular DNA repair machinery leads to removal of the lesion (or incorrect base for the A:OG base pair) and the source of transcriptional mutagenesis events, so that only nonfluorescent variants are expressed. (4) For the A:OG base pair in particular, a new round of repair is needed to remove the OG lesion.
We reproduced our previous results measuring 8oxoG repair and plasmid combinations #5 and #6 in Table 1. Since Aag MEFs exhibited relatively low fluorescent reporter expression, whereas Ung−/− and Aag−/− MEFs exhibited ~8- and 12-fold higher fluorescence, respectively (Fig. 2 B and C).

**MUTYH.** During replication, DNA polymerases can incorporate A opposite 8oxoG, leading to production of a nonfluorescent protein. The ability of these two plasmids to report on U and Hx repair was tested in MEFS deficient in the main UNG (plasmid combinations #1 and #2 in Table 1), and in MEFS deficient in the only known Hx DNA glycosylase, AAG (plasmid combinations #3 and #4 in Table 1). As expected for both reporters, WT MEFS exhibited relatively low fluorescent reporter expression, whereas Ung−/− and Aag−/− MEFS exhibited ~8- and 12-fold higher fluorescence, respectively (Fig. 2 B and C).

**Fig. 2.** FM-HCR assays for measuring DNA glycosylase activity on alkylated, deaminated, oxidized, and misincorporated bases. Proof-of-concept BER capacity measurements comparing WT MEFS and MEFS deficient in DNA glycosylases known to repair the lesion of choice. (A) 8oxoG:C repair reporter in Ogg1+/− cells. (B) Uracl repair reporter in Ung−/− cells. (C) Hx repair reporter in Aag−/− cells. (D) A:8oxoG repair reporter in Mutyh−/− cells. Error bars represent the SD calculated from at least three biological replicates.

in Fig. 1. Validation of the ability of each construct to report on a specific DNA glycosylase activity is presented below.

**Ogg1.** We reproduced our previous results measuring 8oxoG repair in Ogg1-proficient and -deficient mouse embryonic fibroblasts (MEFS) (15). This plasmid (mOrange-A215C-8oxoG) reports on transcriptional misincorporation of A opposite the 8oxoG lesion. As expected, Ogg1+/− cells show approximately a 17-fold higher level of mOrange than WT MEFS, indicating a much lower 8oxoG repair capacity in Ogg1-deficient vs. WT cells, which efficiently remove the source of the transcriptional errors (Fig. 2A and plasmid combinations #5 and #6 in Table 1). Since this plasmid also reports on the activity of other DNA glycosylases that can remove 8oxoG (i.e., NEIL1 and NEIL2), we infer that these enzymes play a minor role in 8oxoG repair in MEFS. That the plasmid reports on 8oxoG repair by multiple DNA glycosylases is viewed as an advantage for ultimately measuring total DNA repair capacity for this lesion in epidemiological and clinical samples.

**UNG and AAG.** We developed a U repair reporter (BFP-A191G-U) that produces WT-BFP transcripts only when RNA polymerase II (RNAPII) incorporates an A opposite U present in the transcribed strand. Upon repair, U is replaced by C, leading to production of a nonfluorescent protein. We also developed a Hx repair reporter (GFP-C289T-Hx) that results in WT-GFP only upon incorporation of C opposite Hx. Upon repair, Hx is replaced by A, leading to production of a nonfluorescent protein. The ability of these two plasmids to report on U and Hx repair was tested in MEFS deficient in the main UNG (plasmid combinations #1 and #2 in Table 1), and in MEFS deficient in the only known Hx DNA glycosylase, AAG (plasmid combinations #3 and #4 in Table 1). As expected for both reporters, WT MEFS exhibited relatively low fluorescent reporter expression, whereas Ung−/− and Aag−/− MEFS exhibited ~8- and 12-fold higher fluorescence, respectively (Fig. 2 B and C).

**Simultaneous Measurement of Four DNA Glycosylase Activities.** A major advantage of FM-HCR assays is the ability to measure multiple repair activities simultaneously; however, multiplexing requires that any given reporter plasmid does not alter the ability of cells to repair damage in the other reporter plasmids and that the fluorescent signals produced by each of the reporters do not alter the ability to detect signals from the other reporters. To test whether interference occurs between the reporter plasmids upon cotransfection, we transfected a single BER reporter plasmid at a time (Fig. 3A and plasmid combinations #1 through #8 in Table 1) or cotransfected four BER reporter plasmids simultaneously (Fig. 3B and combinations #9 and #10 in Table 1). Nearly identical results were observed under both conditions, validating the ability of the FM-HCR assay to simultaneously assess the activity of at least four DNA glycosylases, thus decreasing the number of transfections needed to measure the four activities from eight to two. It should be noted that only live cells are considered for these analyses; this is achieved by incorporating the dead-stain TO-PRO-3 (Materials and Methods) as a sixth simultaneously measured fluorescent signal (besides the four DNA glycosylase reporters and the transfection efficiency control plasmid).
Development of FM-HCR Assays for Measuring APE1 Activity. DNA glycosylase activity determines the rate at which BER is initiated, but the ability to measure downstream steps of this pathway can also contribute to understanding the BER status of a cell, tissue, or individual. We engineered two different reporters for measuring APE1 activity, the major AP-endonuclease that cleaves DNA at abasic sites generated by DNA glycosylases. We built plasmids bearing a site-specific tetrahydrofuran (THF), an abasic site analog that is more chemically stable than a natural abasic site and which is efficiently recognized and nicked by APE1 (18). Substrates were developed using two complementary strategies, namely measuring the ability of THF to induce transcriptional mutagenesis, or measuring the ability of THF to block transcription (Fig. 4A). A transcriptional mutagenesis reporter analogous to the mOrange-A215C-8oxoG reporter described above (mOrange-A215C-THF) was developed, where RNAPII-mediated incorporation of A opposite THF in the transcribed strand leads to expression of a transcript encoding WT mOrange, and upon THF repair G replaces the abasic site and no fluorescent protein is expressed. A second reporter using a THF:C base pair (GFP-617-THF) was designed.

![Fig. 3. Comparison of single and multiplexed measurements of four BER DNA glycosylase substrates.](image)

**Fig. 3.** Comparison of single and multiplexed measurements of four BER DNA glycosylase substrates. (A) FM-HCR analysis with four different BER reporter plasmids transfected separately into repair-proficient (WT) cells or MEFs deficient for one DNA glycosylase known to repair each one of the tested lesions at a time. (B) FM-HCR analysis with four different BER reporter plasmids cotransfected into the same cell lines. An AmCyan plasmid was cotransfected as a transfection efficiency control for all of the experiments. %R.E. in mutant MEFs was normalized to %R.E. in WT cells. Error bars represent the SD calculated from at least three biological replicates.

---

### Table 1. Combinations of reporter plasmids and types of DNA damage used in each experiment

<table>
<thead>
<tr>
<th>Comb.</th>
<th>AmCyan</th>
<th>tagBFP</th>
<th>EGFP</th>
<th>mOrange</th>
<th>mPlum</th>
<th>Empty vector, ng</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>WT</td>
<td>WT</td>
<td>30</td>
<td></td>
<td></td>
<td>1,970</td>
</tr>
<tr>
<td>#2</td>
<td>WT</td>
<td>A191G-U</td>
<td>90</td>
<td></td>
<td></td>
<td>1,910</td>
</tr>
<tr>
<td>#3</td>
<td>WT</td>
<td>WT</td>
<td>25</td>
<td></td>
<td></td>
<td>1,975</td>
</tr>
<tr>
<td>#4</td>
<td>WT</td>
<td>C289T-Hx</td>
<td>25</td>
<td></td>
<td></td>
<td>1,975</td>
</tr>
<tr>
<td>#5</td>
<td>WT</td>
<td>WT</td>
<td>30</td>
<td></td>
<td></td>
<td>1,970</td>
</tr>
<tr>
<td>#6</td>
<td>WT</td>
<td>A215C-8oxoG</td>
<td>120</td>
<td></td>
<td></td>
<td>1,880</td>
</tr>
<tr>
<td>#7</td>
<td>WT</td>
<td>WT</td>
<td>90</td>
<td></td>
<td></td>
<td>1,910</td>
</tr>
<tr>
<td>#8</td>
<td>WT</td>
<td>T202WT-8oxoG(+)</td>
<td>135</td>
<td></td>
<td></td>
<td>1,865</td>
</tr>
<tr>
<td>#9</td>
<td>WT</td>
<td>WT</td>
<td>30</td>
<td>WT</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>#10</td>
<td>WT</td>
<td>A191G-U</td>
<td>90</td>
<td>C289T-Hx</td>
<td>25</td>
<td>135</td>
</tr>
<tr>
<td>#11</td>
<td>WT</td>
<td>WT</td>
<td>100</td>
<td>WT</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>#12</td>
<td>WT</td>
<td>A191G-U</td>
<td>300</td>
<td>C289T-Hx</td>
<td>50</td>
<td>300</td>
</tr>
<tr>
<td>#13</td>
<td>WT</td>
<td>WT</td>
<td>30</td>
<td></td>
<td></td>
<td>2,375</td>
</tr>
<tr>
<td>#14</td>
<td>WT</td>
<td>WT</td>
<td>30</td>
<td></td>
<td></td>
<td>2,375</td>
</tr>
<tr>
<td>#15</td>
<td>WT</td>
<td>WT</td>
<td>100</td>
<td>WT</td>
<td>50</td>
<td>617-THF</td>
</tr>
<tr>
<td>#16</td>
<td>WT</td>
<td>WT</td>
<td>100</td>
<td>617-THF</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>#17</td>
<td>WT</td>
<td>WT</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#18</td>
<td>WT</td>
<td>WT</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#19</td>
<td>WT</td>
<td>A215C-THF</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Headers correspond to the different plasmid reporters used. Nomenclature for plasmids can be found in Table 2. Comb., combination; ng, nanograms of plasmid used for each combination.
generated to measure APE1 activity based on the transcription-blocking properties of THF. To ensure that this substrate only reports a repair event and does not report a bypass of the THF lesion, GFP site C617 was mutated to the following three base combinations (C617A, C617G, and C617T) to test whether potential transcriptional bypass events placing A, G, or U opposite THF would be able to generate fluorescent protein events; none of them did, and thus only repair to the WT sequence in the transcribed strand (G), which leads to expression of WT transcripts (C) at position 617, could produce fluorescent protein.

We validated the mOrange-A215C-THF transcriptional mutagenesis reporter using isogenic mouse B cell lines that were APE1-proficient (Apel++) and APE1-deficient (ApelΔΔΔ) (19). Transfection of the mOrange-A215C-THF plasmid (plasmid combinations #17 and #18 in Table 1) into these cells resulted in no detectable fluorescent mOrange events for Apel++Δ cells and ~14% reporter expression for the deficient ApelΔΔΔ cells (Fig. 4B), indicating that WT cells efficiently remove the THF lesion before it can induce transcriptional mutagenesis detectable by FM-HCR. We also engineered a set of reporters to test whether transcriptional misincorporation of C, G, or U opposite THF would be able to generate fluorescent protein events when transfected into ApelΔΔΔ cells; they did not, suggesting that A is the major base misincorporated by RNASII opposite THF in vivo. This result also reiterates that in the case of the GFP-617-THF substrate, only removal of the transcription-blocking THF, and not misincorporation of C, can lead to GFP expression.

We validated the GFP-617-THF substrate, which reports APE1 activity based on the ability of THF to block transcription, using the same APE1-proficient (Apel++Δ) and APE1-deficient (ApelΔΔΔ) cell lines. As expected, following transfection with GFP-617-THF (Fig. 4C and plasmid combinations #15 and #16 in Table 1), repair-proficient Apel++Δ cells show approximately a fivefold higher level of fluorescent reporter expression as a result of higher repair capacity compared with the repair-deficient ApelΔΔΔ cells. For additional experiments, we chose to use the transcription-blocking GFP-617-THF substrate to report APE1 activity because, unlike for the transcriptional mutagenesis reporter, this substrate yielded a nonzero readout for repair-proficient cells (Fig. 4).

Analysis of BER Capacity in a Panel of 24 Cell Lines Derived from Apparently Healthy Individuals. Given that small variations in the activities of some BER enzymes have been correlated with disease risk (8–10), we sought to determine if our methods could reliably measure small interindividual differences in BER capacity. We used the four DNA glycosylase transcriptional mutagenesis-based reporters (for repair of Hx:T, 8oxoG:C, A:8xoG, and U:G), as well as the transcription blockage reporter for APE1 to measure the BER capacity across a panel of 24 human B lymphoblastoid cell lines derived from apparently healthy individuals of diverse ancestry (20) (plasmid combinations #11 and #12, and #15 and #16 in Table 1). A range of BER activity was observed across the 24 cell lines (Fig. S4 and detailed data in SI Appendix, Table S2). Moreover, each cell line displays a unique BER capacity fingerprint (Fig. S5B).

Regression Models Based on BER Capacity Can Predict Sensitivity to DNA Damaging Agents. We used mathematical modeling to test the potential for BER activity measurements to predict sensitivity to three DNA damaging agents known to form DNA lesions processed by BER. Sensitivity to 0.4 mM methylmethane sulfonate (MMS; previously measured) (21), 25 μM H2O2, and 100 μM 5-Fluourouracil (5-FU) were measured for this cell panel. The 24 cell lines showed a broad and continuous range of sensitivities, specifically, a 9.5-, 4.7-, and 2.7-fold range were observed for MMS, H2O2, and 5-FU, respectively (SI Appendix, Fig. S2; see data in SI Appendix, Tables S4 and S5). While H2O2 and 5-FU sensitivity were modestly correlated ($R = 0.47, P = 0.02$), MMS was not significantly correlated to either H2O2 or 5-FU (SI Appendix, Table S6), indicating that sensitivity to one agent does not necessarily predict sensitivity to other agents (22).

Multiple linear regression (MLR) seeks to find the best model describing the linear relationship between independent variables (z-scored FM-HCR reporter data) (Fig. 5B) and a dependent variable (sensitivity to DNA damaging agents reported as percent control growth). For this analysis, the BER plasmid reporter measurements reported here were combined with our previously published FM-HCR data for five other DNA repair pathways in the same cell lines (15). Specifically, we included DRC measured for nucleotide excision repair (NER), MMR, direct reversal by MGMT, nonhomologous end-joining (NHEJ), and homologous recombination (HR) (SI Appendix, Table S3).
A variety of MLR models were built for each agent (SI Appendix, Table S7). The models yielding the best correlation fit following leave-one-out cross-validation (LOOCV, $R^2_{CV}$) and the lowest root mean square error (RMSE) were further considered. Inclusion of all of the variables in the model results in overfitting of the data and a low predictive power; as such, $R^2_{CV}$ is prioritized over model $R^2$ to minimize overfitting of the data. The following models for predictions of percent control growth (\%C.G.) were selected:

$$%C.G_{MMS} = -13.2 \times 8\text{oxoG} : C + 8.7 \times U : G - 8.3 \times \text{THF} + 8.0 \times \text{HR} + 5.5 \times \text{NHEJ} + 44.2;$$

[1]

$$%C.G_{SFU} = -4.6 \times 8\text{oxoG} : C + 3.8 \times U : G - 3.1 \times \text{THF} - 1.9 \times \text{NHEJ} + 1.8 \times \text{MGMT} - 1.1 \times A:8\text{oxoG} + 19.3;$$

[2]

$$%C.G_{H2O2} = -14.5 \times 8\text{oxoG} : C + 14.1 \times U : G - 10.5 \times \text{THF} - 6.0 \times \text{HR} + 4.8 \times \text{MGMT} + 4.3 \times \text{NER} + 52.2.$$  

[3]

Special attention should be given to contributions by transcriptional mutagenesis-based reporters (Hx:T, U:G, 8oxoG:C, A:8oxoG, and MGMT) for which a negative sign indicates that increased repair activity for that particular lesion correlates with increased resistance. As shown in Eq. 4 (below) the $\beta$ slopes (here, the numerical values that accompany each term) represent the percentage point change in \%C.G. that is associated with a 1 SD increase in repair capacity in the “$x$” relevant pathway; for example a 1 SD increase in 8oxoG:C repair activity is associated with a 13.2% point increase in \%C.G. for cells treated with MMS. MLR provides a confidence interval for each of the $\beta$ slopes; those that were significantly different from zero with at least 95% confidence have been underlined in the models and are prioritized in the discussion of the implications of these models below.

**Discussion**

The ability to simultaneously measure the in vivo repair activity of multiple BER substrates extends our previously published FM-HCR substrate toolbox to include nearly all of the major DNA repair pathways (15), namely BER (initiated by AAG, MUTYH, OGG1, and UNG, and abasic site processing by APE1), plus HR, MGMT, MMR, NER, and NHEJ.

**Functional DRC assays, such as the FM-HCR presented here,** hold the advantage that multiple gene regulatory levels—such as epigenetic, transcriptional, translational, and posttranslational—are integrated into the readout of the fluorescent reporter (12). Moreover, assays performed in vivo report the activity of the repair machinery in its physiological context. Therefore, the involvement of redundant repair pathways is also integrated into the final repair score. For example, APE1 (23) and NEIL1 (24) have been shown to stimulate OGG1’s activity by increasing its turnover. Therefore, an in vitro assay that doesn’t recapitulate physiological conditions for OGG1, NEIL1, and APE1 could result in misrepresentation of...
of AP site-induced disruption of transcription is certainly not negligible, given that ~20,000 abasic sites are formed on average in a mammalian cell every day (41). Unrepaired abasic sites as well as Hx, U, and 8oxoG may thus give rise to mutations during DNA replication, and may also result in truncated or mutated mRNA transcripts that produce faulty proteins, having a direct impact on cellular physiology (36).

**Interindividual Variation in BER Capacity.** Most analyses of interindividual variation in BER have been performed in vitro with lysates of PBMCs isolated from the blood of healthy donors. This complicates a direct comparison with the fold-range of activities observed using in vivo FM-HCR methods in immortalized human B lymphoblastoid cell lines. Nevertheless, with the exception of AAG, for which a 3.3- to 10-fold activity range has been reported in PBMCs (42–44) (compared with the 1.7-fold described here), the degree of interindividual variation reported here agrees with previously observed ranges for OGG1 in PBMCs [2- to 2.9-fold (45, 46), compared with 4.3-fold] and UNG in colon, stomach, and liver [5.5-, 3.2-, and 3.1-fold (47), respectively, compared with 5.3-fold].

To our knowledge, only a single functional study of human MUTYH-associated variants has been reported (48), but an extensive study on the functional variation of MUTYH across healthy individuals, like the one performed here (showing a 7.2-fold range), had not taken place so far. The low interindividual variation in AAG and APE1 activity across all 24 cell lines is particularly striking. These phenomena may reflect a limitation for these reporters to distinguish among cell lines in which the pathway is uniformly active or inactive throughout all samples. AAG activities for all cell lines appear to be relatively low [high percent fluorescent reporter expression (%R.E.) in Fig. 5A, similar to the %R.E. by Aag<sup>+/−</sup> MEFs in Fig. 2C], potentially indicating that B lymphoblastoid cells in general might not have particularly high AAG activity. In contrast, APE1 activity for all cell lines is relatively high (high %R.E. in Fig. 5A, similar to the %R.E. by Ape1<sup>+/+Δ</sup> in Fig. 4C). Considering that deficiencies in APE1 are associated with disease (49, 50), and Ape1<sup>+/−</sup> mice are embryonic-lethal, it is not particularly striking that “healthy” cells stay within a small range of variation. However, a 4.9-fold range has been reported in a panel of 100 human PBMCs for an in vitro APE1 assay (51).

**Mathematical Modeling.** Chemo- and radiotherapy are widely used to treat cancer, and both their effectiveness and the severity of side effects depends, at least in part, upon DRC (21, 52–54). The ability to determine ahead of time whether a particular treatment would be effective could improve therapy and avoid putting patients through unnecessary exposures. Recently, we utilized such an approach to build predictive models relating alkylating agent sensitivity to DRC for five repair pathways (21). Here, we build upon these methods by incorporating BER capacity measurements to evaluate the contribution of 10 different functional DRC scores toward predicting sensitivity of B lymphoblastoid cells to three DNA damaging agents known to induce damage that is repaired, in part, by the BER machinery.

---

**Fig. 6.** Interindividual variability in DRC of primary T lymphocytes. FM-HCR with 10 DRC substrates was assessed in PHA-stimulated PBMCs isolated from fresh whole blood from two different donors (empty bars correspond to donor “A” and filled bars to donor “B”). Error bars show SD of four replicates. Two-tailed unpaired t test *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001. No results are shown for MGMT as none of the samples met the 30-fluorescent-event threshold described.
The fact that the strongest contributor in all three MLR models (Eqs. 1–3) is repair of 8oxoG:C supports the idea that sensitivity for these three agents is driven, in part, by an oxidative stress response and in consequence, partially, by oxidative DNA damage. Importantly, the association between 8oxoG repair activity and resistance to MMS is consistent with the observation that HCT116 cells overexpressing OGG1 acquire resistance to MMS-induced cell death (55), as well as with the detection of fluorescence from a reactive oxygen species reporter gene, sulfredoxin-1 (SRXN1), in SRXN1-GFP mouse embryonic stem cells upon MMS treatment (56). Moreover, MMS indirectly induces an increase in reactive oxygen species levels in yeast (57). Oxidative damage to cytosine can lead to DNA lesions that are substrates for the same glycosylases that excise U (58); as with 5-FU, the seemingly counterintuitive correlation between U excision activity and sensitivity to H2O2 could reflect BER imbalance. Furthermore, increased U excision activity contributes toward sensitivity (rather than resistance) for all three agents and might reflect the role that UNG and SMUG1 play in the excision of not only U but a variety of base lesions, such as 5-hydroxyuracil, isodouracil acid, alloxan, and 5-hydroxymethyluracil (59, 60), all of which are formed, like 8oxoG, in mammalian cells following oxidative challenge (61).

Although it remains unclear why BER imbalance is associated with elevated activity for some glycosylases and not others, we note that OGG1, for which higher activity does not appear to lead to imbalance in BER, is a bifunctional DNA glycosylase. In contrast, the best-characterized example of BER imbalance (AAG) (3), as well as the present case, wherein elevated uracil DNA glycosylase activity appears to be toxic to cells upon exposure to agents that produce UNG substrates, both involve multifunctional DNA glycosylases. Thus, the differing demands placed on downstream repair factors by the two DNA glycosylase mechanisms could potentially explain the opposite correlations between sensitivity to 5-FU, H2O2, and OGG1 vs. UNG activity.

Overall, incorporation of BER measurements into the MLR model for MMS sensitivity results in a substantial improvement in prediction capacity over our previous modeling, which lacked BER components, specifically regarding the correlation fit following LOOCV (R2 CV) (21). This model results in a correlation fit of 0.49 for MMS sensitivity, compared with the previously reported R2 CV of 0.07 (21), highlighting the role of BER components in predicting sensitivity to MMS. We acknowledge the limitations of the models, namely that they are based on measurements on B lymphoblastoid samples, which might not represent the effects of DRC on DNA damage-based chemotherapy in tumors, and that the R2 CV could still be substantially improved. As such, further validation of our DRC reporters and mathematical modeling, as well as incorporation of other non-DRC variables associated with therapeutic response to DNA damaging agents, will need to take place in primary “healthy” and tumors tissues.

Interindividual Variation in DRC Can Be Measured in Fresh Blood Samples. Because blood can be accessed easily in a minimally invasive manner, it is commonly used as a surrogate for assessing the biological properties of other tissues. For example, a direct correlation between 8oxoG repair capacity in blood and lung tissue was previously established through in vitro assays (62). Furthermore, numerous epidemiological studies have demonstrated an association between cancer and DRC in lymphocytes (12, 63), highlighting the increased potential that the use of in vivo DRC measurements in isolated blood can have on assessing an individual’s overall DRC and, more specifically, as a biomarker for disease risk. Even though it is still to be determined if these kinds of measurements can help inform an individual’s predisposition to disease or their potential tolerance to treatment with DNA damaging agents (12), we not only demonstrate that our DRC reporters can be utilized in primary blood samples but also that they can detect differences between individuals.

Overall, the methods presented here represent powerful tools for measuring cellular BER capacity in mammalian cells. Additionally, we have unveiled previously unknown in vivo transcriptional mutagenesis and blockage events that have important consequences in protein homeostasis and cell physiology following DNA damage. Finally, we demonstrate that differences in DRC between individuals can be measured in lymphocytes isolated from whole-blood samples.

Materials and Methods

Plasmids. As described previously (15), AmCyan, EGFP, mOrange, mPlum, and tagBFP reporter genes were subcloned into the pmaxCloning Vector (Lonza) between the KpnI and SacI restriction sites in the multiple cloning site. Plasmids were amplified in E. coli DH5α (Invitrogen) and purified using Qiagen endotoxin-free Maxi and Giga kits.

Generating Plasmids Containing Site-Specific DNA Damage Reporting via Transcriptional Mutagenesis or Transcriptional Blockage. Nonfluorescent variants of the reporter plasmids containing a point mutation at a site coding for their respective chromophores were identified and generated via standard QuikChange site-directed mutagenesis (Agilent Technologies) (Table 2).

To produce single-stranded plasmid DNA (ssDNA), previously described methods were followed (15) with minor modifications. Reporter plasmids were nicked with either Nb.BtsI or Nt.BspQI (New England Biolabs), depending on the lesion-containing strand (Table 2). The nicked strand was then digested with ExoIII (New England Biolabs), and the remaining ssDNA was purified using a 1% agarose gel. The circular ssDNA was extracted from the agarose gel using a Gel Extraction Kit (Qiagen). Fifteen picomoles of the respective phosphorylated lesion-containing oligonucleotide (Table 2) were combined with 3.2 µg of the complementary single-stranded plasmid in 1× Pfu polymerase AD buffer (Agilent Biotechnologies) in a final volume of 46 µl (4:1 molar ratio). The mixture was heated to 85 °C in a thermal cycler for 6 min, and then allowed to anneal by cooling to 40 °C at 1 °C per minute. To extend the primer, five units of Pfu polymerase AD (Agilent) and 0.4 µM dNTP were added. The incubation parameters used for each ssDNA-plasmid/oligonucleotide combination are shown in Table 2. Following extension, the reaction mixture was cleaned up with a Qiagen PCR Purification kit column, eluted in EB buffer and subsequently combined with 1× Ligase Buffer (New England Biolabs), 0.4 µM dNTP, 1 µM ATP, 10 ng/µl BSA, 1.5 units T4 DNA polymerase, and 80 units T4 DNA Ligase (both New England Biolabs), and incubated for an additional hour at 16 °C to yield closed circular plasmid. Finally, the product was purified from a 1% agarose gel using a Qiagen gel extraction kit.

The presence of desired base lesions in the plasmid reporters was confirmed by analytical digests (SI Appendix, Figs. 53–55). The names for the modified plasmids are listed in Table 2.

Cell Culture. The panel of 24 B lymphoblastoid cell lines was obtained from the Coriell repository. This panel consists of a representative subset of a larger 450 cell line panel specifically chosen as representative samples of United States residents with ancestry from all major regions of the world (20). These and all other cell lines plus their respective culture conditions are detailed in SI Appendix, Table S1.

MUTYH Knockdown Cell Lines. MUTYH was knocked down in HCT116 + Chromosome 3 cells, as previously described (64). shRNAs expressed from a lentiviral plasmid (pGIPZ) were purchased from Open Biosystems to target MUTYH transcript (v1: RHR54340-98904053 and v2: RHR54340-99140608). Knockdown cells were compared with cells expressing a nontargeting shRNA (RHR54346). Virus was generated in 293T cells using packaging plasmids (pSPAX2 and pMD2.G; Addgene). Cells were infected with virus and stable clones were selected using Puromycin.

MUTYH Nuclear Overexpression. Nuclear MUTYH isoform 4 cDNA was cloned into a retroviral plasmid (pBABE, C-terminal Flag-tagged). HCT116 + Chromosome 3 cells overexpressing MUTYH were compared with the same cells expressing empty vector. Virus was generated in 293T cells, and stable clones were selected using Puromycin.

In Vitro MUTYH DNA Glycosylase Assays. Cells were sonicated in MUTYH DNA glycosylase dilution buffer (30 mM Tris pH 7.5, 1 mM EDTA, 1 mM β-mercaptoethanol, 50 mM NaCl, and 30% glycerol) with protease inhibitor mixture. Protein concentration was measured using micro BCA Kit (Pierce). DNA glycosylase assays were performed as previously published (65). A double-stranded
oligonucleotide containing a [32P]labeled strand (5′-TTGGGGAATGAGTCAGGCCCACTG-3′) and a nonlabeled strand (5′-GGTGCCTGGCCCTGAAGGTGTGGCCTG-3′) was incubated with an amount of extract determined to be in the linear range for activity at 37 °C for 60 min (base pair serving as target substrate for MUTYH is underlined). The resulting AP sites were cleaved by incubation with 0.1 M NaOH at 75 °C for 15 min, the mixture heat denatured and subjected to 7 M urea–10% polyacrylamide gel electrophoresis. Phosphorimaging was used to visualize and quantitate MUTYH DNA glycosylase activity.

DNA Repair Assay Transfections. 

Electroporation. Cells growing in suspension (3 × 10^6 cells in 100 μL complete medium) were combined with a reporter plasmid mixture (Table 1) and electroporated using a 96-well Bio-Rad Mxcell gene pulser, with an exponential waveform at 260 V and 950 μF. Following electroporation, 100 μL complete medium was added to each well in the electroporation plate and gently mixed. The electroporation mix was transferred to a 24-well cell culture plate, each well prefilled with 1.3 mL complete medium and incubated at 37 °C and 5% CO_2 for 12 h. The cells were then centrifuged for 5 min at 300 × g, resuspended in 250 μL of complete medium containing TO-PRO-3 dead stain, and transferred to a 96-well plate for flow cytometry.

Lipofection. For lipofection, 150,000 adherent cells were plated in six-well plates 1 d before transfection to achieve 50–80% confluency on the day of transfection. Transfection with Lipofectamine LTX (Life Technologies) was performed according to the manufacturer's instructions. Briefly, 2.5 μg of total plasmid DNA (Table 1) mixed with 2.5 μL plus reagent and Opti-MEM, was further mixed with 6.2 μL lipofectamine LTX in Opti-MEM, and incubated at room temperature for 5 min before layering 200 μL of the transfection reaction on top of the cells. Lipofection conditions were scaled down for 12- and 24-well plates when necessary. Transfected cells were incubated for 18h at 37 °C and 5% CO_2, then trypsinized and resuspended in a total of 250 μL of complete medium containing TO-PRO-3 dead stain and transferred to 75-mm Falcon tubes with Cell Strainer Caps (Fisher Scientific).

Flow Cytometry. Flow cytometric parameters and calculation of %R.E. have been previously described (15, 34) and are detailed in SI Appendix, Supplementary Materials and Methods.

Cell Sensitivity Assay. 

Chemicals. Hydrogen Peroxide (H_2O_2) (Sigma) was prepared at working concentrations in complete media. Next, 5-FU (Sigma) was prepared as a 100 mM stock in DMSO, and diluted to working concentrations in complete media before addition to cells.

XTT assay. For XTT assay, 25,000 lymphoblastoid cells were seeded in 100 μL in 96-well tissue culture plates and incubated with drugs (added in a volume of 50 μL complete media to a final concentration of 25 μM H_2O_2 and 100 μM 5-FU). Each condition, including controls, was set up in three replicates in 24- or 96-well plates when necessary. Transfected cells were incubated for 18h at 37 °C and 5% CO_2, then trypsinized and resuspended in a total of 250 μL of complete medium containing TO-PRO-3 dead stain and transferred to 75-mm Falcon tubes with Cell Strainer Caps (Fisher Scientific).

MLR Models. Z-scored DRC values for each of 10 FM-HCR reporters used to analyze DRC in the 24 Correll lymphoblastoid cell lines, served as the (independent) predictor variables (five BER reporters developed here (SI Appendix, Table S2); five previously published reporters for other pathways (15) (SI Appendix, Table S3)). Sensitivity to three DNA damaging agents for the same 24 cell lines is reported in SI Appendix, Table S4 and expressed in terms of %C.G. MLR models were developed as described previously (21) and as detailed below, and take the following form:

\[ Y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + b, \]

where Y represents the predicted response variable (sensitivity) reported as %C.G.; x_n are the z-scored DNA repair capacity predictor variables for the 10 substrates reported as %R.E.; \beta_i are the substrate slopes along the corresponding dimensions, and b is a constant that represents the y-axis intercept. Z-scores for relative DNA repair capacity among the 24 cell lines were generated for each pathway as follows.

\[ Z_i = \frac{x_i - \bar{x}_i}{\sigma_i}, \]

where Z_i is the z-score for a DNA repair pathway in cell line i in pathway j, x_i is the DRC for a given pathway j in cell line i, \bar{x}_i is the mean value of the DNA repair capacity in pathway j over the 24 lymphoblastoid cell lines (i = 1–24), and \sigma_i is the SD of the DNA repair capacity in pathway j. MGMT scores were transformed to log10 for before z-scoring.

MLR models were generated by running MATLAB's regress function. \( \beta \) Scores (slopes) for each variable as well as the model fit to the data (R^2) were calculated for each generated model. LOOCV was used to assess the prediction power of each model. The correlation coefficient (R^2CV) between the observed cellular sensitivities and the predicted ones following LOOCV as well as their RMSE were calculated for each model. Initial models were generated by including all variables (model 0). To identify the optimal models (Eq. 1–3), variable selection was carried out by sequentially removing predictor variables with the lowest contribution to the model (lowest \( \beta \) slope). In cases where two or more variables were tied for the smallest slope (\( \beta \)), each of the possible models were tested wherein one of the tied variables was removed. The combination of variables that yielded the smallest RMSE and largest R^2CV was selected as the best model for each DNA damaging agent (SI Appendix, Table S7).

PBMCS. Fresh whole blood from two different donors, collected in sodium heparin Vacutainer collection tubes, was purchased from Research Blood Components. PBMCS were isolated using standard Ficol-Paque (GE) density gradient centrifugation. PBMCS freezing and thawing procedures as well as PHA stimulation of T lymphocytes and transfection conditions are described in detail in SI Appendix, Supplementary Materials and Methods.

ACKNOWLEDGMENTS. We thank Dr. Kefei Yu (Michigan State University) for the apurinic/apyrimidinic endonuclease 1-deficient cells, and Dr. Samuel Wilson (National Institute on Environmental Health Sciences) for the uracil DNA glycosylase-deficient cells. This work was supported by a NIH Directors Pioneer Award DPI-ES022576, and NIH Grants R01-CA075576, R01-CA55042, R01-CA49261, R01-E002872, and P30-ES02109. L.D.S. is American Cancer Society Research Professor Emeritus.

Chaim et al.