



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

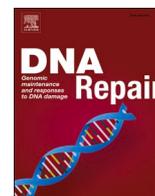
## **Stained DNA Dot Detection (SD<sup>3</sup>): An automated tool for quantifying fluorescent features along single stretched DNA molecules**

Downloaded from: <https://research.chalmers.se>, 2025-05-12 07:59 UTC

Citation for the original published paper (version of record):

Aning, O., Dvirnas, A., Nyblom, M. et al (2025). Stained DNA Dot Detection (SD<sup>3</sup>): An automated tool for quantifying fluorescent features along single stretched DNA molecules. *DNA Repair*, 149. <http://dx.doi.org/10.1016/j.dnarep.2025.103836>

N.B. When citing this work, cite the original published paper.



## Stained DNA Dot Detection (SD<sup>3</sup>): An automated tool for quantifying fluorescent features along single stretched DNA molecules

Obed A. Aning<sup>a</sup>, Albertas Dvirnas<sup>a,b</sup>, My Nyblom<sup>a</sup>, Jens Krog<sup>b</sup>, Johanna Carlson<sup>a</sup>, Pegah Johansson<sup>c,d</sup>, Tobias Ambjörnsson<sup>b</sup>, Fredrik Westerlund<sup>a,\*</sup>

<sup>a</sup> Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden

<sup>b</sup> Computational Science for Health and Environment, Centre for Environmental and Climate Science, Lund University, Lund, Sweden

<sup>c</sup> Region Västra Götaland, Department of Clinical Chemistry, Sahlgrenska University Hospital, Gothenburg, Sweden

<sup>d</sup> Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

### ARTICLE INFO

#### Keywords:

Automated image analysis  
Single molecule analysis  
DNA analysis  
Fluorescence Microscopy  
Fluorescent label localization  
DNA damage

### ABSTRACT

The main information in DNA is its four-letter sequence that builds up the genetic information and that is traditionally read using sequencing methodologies. DNA can, however, also carry other important information, such as epigenetic marks and DNA damage. This information has recently been visualized along single DNA molecules using fluorescent labels. Quantifying fluorescent labels along DNA is done by counting the number of “dots” per length of each DNA molecule on DNA stretched on a glass surface. So far, a major challenge has been the lack of standardized data analysis tools. Focusing on DNA damage, we here present a Matlab-based automated software, Stained DNA Dot Detection (SD<sup>3</sup>), which uses a robust method for finding DNA molecules and estimating the number of dots along each molecule. We have validated SD<sup>3</sup> by comparing the outcome to manual analysis using DNA extracted from cells exposed to H<sub>2</sub>O<sub>2</sub> as a model system. Our results show that SD<sup>3</sup> achieves high accuracy and reduced analysis time relative to manual counting. SD<sup>3</sup> allows the user to define specific parameters regarding the DNA molecule and the location of dots to include during analysis via a user-friendly interface. We foresee that our open-source software can have broad use in the analysis of single DNA molecules and their modifications in research and in diagnostics.

### 1. Introduction

The DNA of all cells contains several layers of information. The DNA sequence is typically analyzed using sequencing protocols [1,2]. The next layer of information is generally based on chemical modifications on the DNA, such as epigenetic marks or DNA damage. Such information has in recent years been visualized along single DNA molecule using fluorescent enzymatic labeling [3–7]. A major challenge has been the lack of automated and easy-to-use software to detect and quantify these fluorescent labels, commonly referred to as “dots”, along DNA. We here present a tool that detects any fluorescent label along stretched DNA. We use experiments based on DNA damage as a model system to show the useability of the developed automated software.

Cellular DNA is constantly challenged with various lesions resulting from endogenous (metabolic processes) or exogenous (chemicals, radiation) sources, with each cell in the human body estimated to experience about 10,000 DNA lesions per day [8,9]. Cells have evolved specialized

DNA damage response mechanisms to sense, monitor and repair these DNA lesions [10]. Single-strand damages and breaks are the most abundant form of DNA lesions and are repaired mainly by the base excision repair (BER) and nuclear excision repair (NER) pathways, where the damaged DNA is excised, leaving a gap that is subsequently filled by a DNA polymerase [11–13].

Accurate detection and quantification of DNA damage is vital for understanding repair mechanisms, assessing environmental and toxic exposures, enabling early diagnosis, and monitoring therapeutic responses [10,14,15]. In the past, various techniques have been developed to meet these needs, often relying on unwinding DNA using methods like the comet assay [16] or immunoassays that rely on antibodies that target specific DNA-adducts [15,17] like 8-oxo-G adduct formed from oxidative damage [18,19]. However, only a limited number of such specific antibodies are commercially available. To overcome these limitations, single-molecule approaches either fluorescence-based such as Repair Assisted DNA Damage (RADD) [6,7,20] or sequence-based such as

\* Corresponding author.

E-mail address: [fredrikw@chalmers.se](mailto:fredrikw@chalmers.se) (F. Westerlund).

<https://doi.org/10.1016/j.dnarep.2025.103836>

Received 11 October 2024; Received in revised form 23 March 2025; Accepted 4 April 2025

Available online 15 April 2025

1568-7864/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nick-seq [21] have emerged as powerful tools for sensitive detection and quantification of various types of DNA damage.

Nick-seq generates high-resolution genomic maps of chemical modifications to DNA using enzymatic or chemical methods combined with sequencing, making it ideal for high-throughput, genome-wide studies, such as profiling DNA damage across multiple samples or experimental conditions [21]. Despite its sensitivity, Nick-seq is limited in detecting a broad range of damage types and in concurrently identifying multiple damage types on the same DNA strand. RADD on the other hand, is an easy to implement fluorescence microscopy-based assay that can evaluate a wide range of DNA lesions, owing to the effectiveness of the repair cocktail in recognizing and repairing various types of DNA damage [6,7,20]. By using specific repair enzymes, the assay can be tailored to address specific damage types [6,22–24]. Interestingly, it has also been adapted for single-molecule DNA damage mapping on a genome-wide scale [25] and for simultaneous detection of multiple damage types on stretched DNA [26]. Additionally, it has been successfully applied to assess DNA damage in high-throughput genomic samples [27] and for genomic mapping through DNA sequencing [28].

In RADD, by using a set of repair enzymes, a processive polymerase (DNA polymerase I) and fluorescent nucleotides, each damage site is fluorescently labeled (Fig. 1A) [7]. After stretching the DNA on silanized coverslips, the damage sites can be visualized as fluorescent dots along the DNA backbone (Fig. 1B). Image analysis entails measuring the length of each DNA molecule and counting the number of fluorescent dots along the DNA backbone (Fig. 1C and D).

This assay has been previously used to detect and quantify DNA damage induced by ionizing radiation [29], ultraviolet radiation [22], reactive oxygen species (ROS) [30] as well as certain anticancer drugs, such as bleomycin [6] and etoposide [5]. These studies have demonstrated that the single molecule assay is a powerful tool to investigate the mechanisms and characteristics of DNA damage, as well as to assess

variations in patient cell response to DNA damaging agents.

In most studies that count fluorescent dots along DNA, the analysis was done in a non-automated fashion, often by counting dots manually, which has drawbacks of subjectivity and slow and labor-intensive analysis. Automated analysis software provides quantitative and reproducible measurements in a high throughput fashion which is required in order to analyze a large number of DNA molecules and allows better statistical analyses.

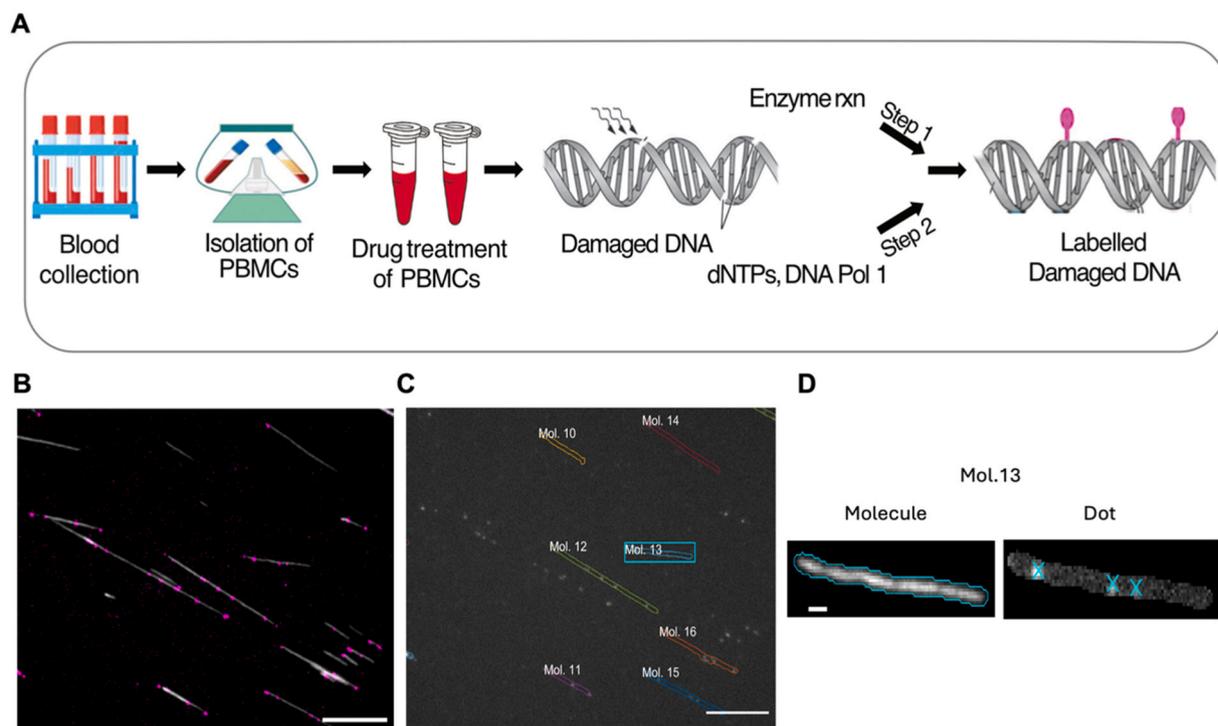
Here, we present Stained DNA Dot Detection (SD<sup>3</sup>), an automated software tool for identifying and quantifying fluorescent dots along single DNA molecules (available at 10.5281/zenodo.13889377). In SD<sup>3</sup>, we have implemented image processing algorithms for finding DNA molecules and measuring their lengths, as well as counting the number of fluorescent dots along the DNA backbone. We use DNA from cells treated with hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) to demonstrate the performance of SD<sup>3</sup> regarding accuracy and reproducibility. SD<sup>3</sup> has a user interface that gives the user flexibility to define molecule parameters to include in the analysis and generates a result output which gives detailed information on each single DNA molecule detected.

## 2. Materials and methods

### 2.1. DNA damage assay

#### 2.1.1. Blood sample collection

Excess blood (EDTA tubes) from individuals with normal differential blood count were collected from the Hematology Lab (Clinical Chemistry Department) at Sahlgrenska University Hospital in Gothenburg, Sweden. Peripheral mononuclear blood cells (PBMCs) were isolated from the blood by density gradient centrifugation using Lymphoprep (Axis-Shield PoC AS, Oslo, Norway) according to the manufacturer's instructions.



**Fig. 1.** Schematic representation of the DNA damage detection protocol. (A) Two step reaction process where in step 1, damaged DNA extracted from peripheral blood mononuclear cells (PBMCs) undergoes enzymatic repair followed by step 2, which involves incorporation of fluorescent nucleotides at the damage sites using a DNA polymerase. (B) Image of labeled DNA stretched on a functionalized glass coverslip. DNA is stained with YOYO-1 (grey) and magenta dots are fluorescent nucleotides incorporated at damage sites. Scale bar = 10  $\mu$ m. (C) Output image provided by SD<sup>3</sup>. Scale bar = 10  $\mu$ m. (D) A zoomed in image of molecule 13: In the software, the DNA molecule is first identified (left) and thereafter the dots along the DNA are detected (right). SD<sup>3</sup> labels the DNA molecules detected and each dot along the DNA length is marked with x. Scale bar = 1  $\mu$ m.

### 2.1.2. Treatment of PBMCs with hydrogen peroxide

$0.5 \times 10^6$  PBMCs/sample were resuspended in RPMI 1640 in a total volume of 300  $\mu$ L. This was followed by 0.3 %  $H_2O_2$  treatment for 30 min at 37 °C.

### 2.1.3. Fluorescent labeling of DNA damage

$H_2O_2$ -treatment of PBMCs was followed by DNA extraction using GenElute-Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich) following the manufacturer's instructions. DNA concentrations were measured using a NanoDrop 1000 spectrophotometer. Shear-induced fragmentation of the DNA was minimized by using wide bore pipette tips throughout the procedure. The labeling of the damage sites was carried out as described previously [6,7,23,29]. 100 ng of DNA was incubated with 2.5 U of each of the repair enzymes, APE1, Endo III, Endo IV, Endo VIII, hAAG, Fpg and UDG (referred to as the enzyme cocktail) in  $1 \times$  CutSmart buffer (Bionordika Sweden) and incubated for 1 h at 37 °C. The *in vitro* DNA repair was followed by 1 h incubation at 20 °C with dNTPs (1  $\mu$ M of dATP, dGTP, dCTP, 0.25  $\mu$ M dTTP (Bionordika Sweden) and 0.25  $\mu$ M aminoallyl-dUTP-ATTO-647N (Jena Bioscience),  $1 \times$  NEBuffer 2 (Bionordika Sweden) and DNA polymerase I (1.25 U) (Promega). The reaction was terminated with 2.5  $\mu$ l of 0.25 M EDTA (Sigma-Aldrich).

### 2.1.4. Silanization of coverslips

$18 \times 18$  mm<sup>2</sup> glass coverslips (Thermo Fischer) were put in a coverslip rack which was carefully submerged in an acetone solution containing 1 % APTES (Sigma Aldrich) and 1 % ATMS (Sigma Aldrich)

[31,32]. The activated coverslips were rinsed with 2:1 v/v acetone: water solution and dried under a nitrogen gas flow right before DNA stretching and imaging.

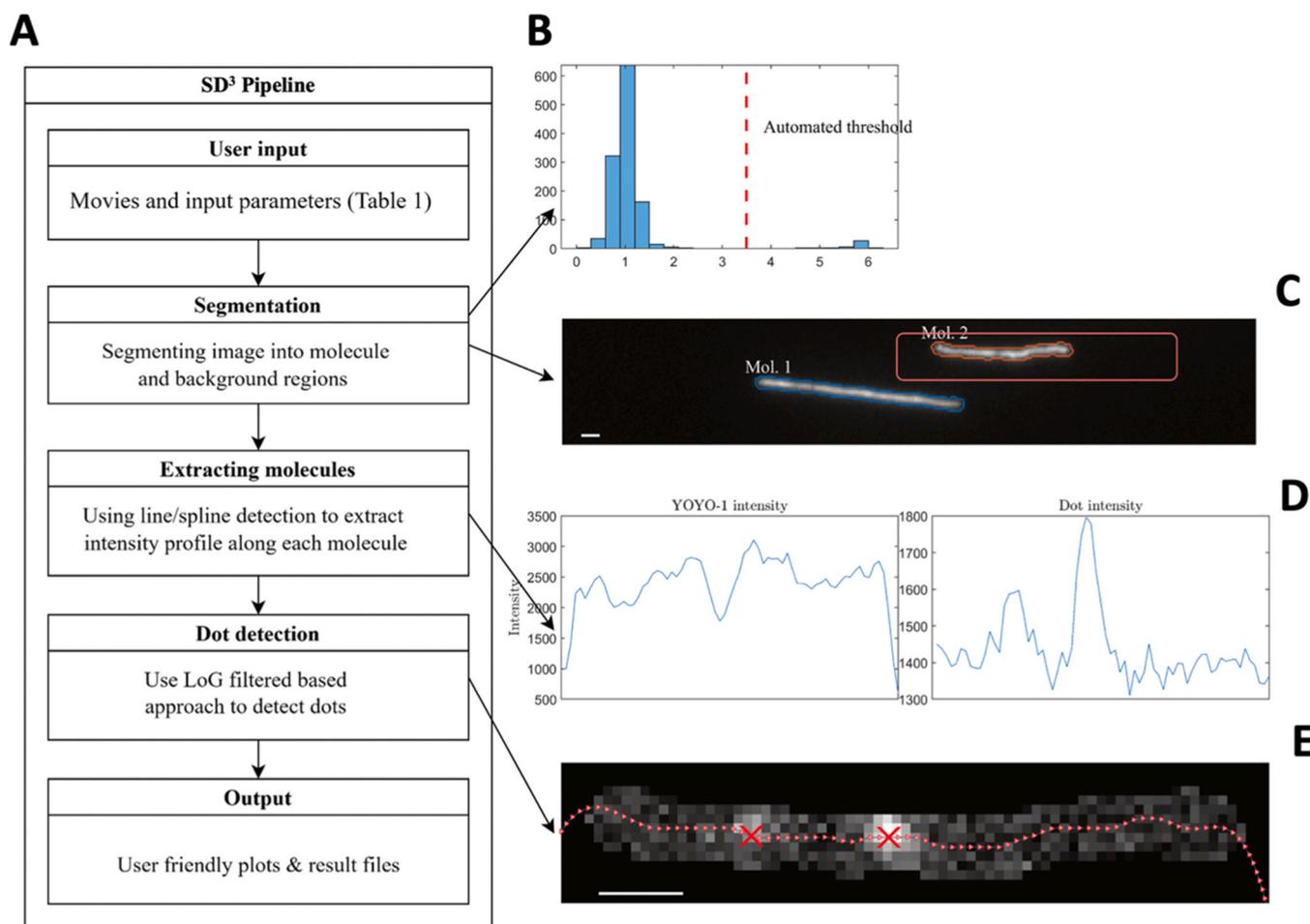
### 2.1.5. DNA staining and imaging

Fluorescently labeled DNA (7  $\mu$ L/sample) was diluted in 0.5X TBE and stained with 320 nM YOYO-1 (Invitrogen) in a total volume of 50  $\mu$ L. To prevent photobleaching, 2 %  $\beta$ -mercaptoethanol (BME, Sigma-Aldrich) was added prior to image acquisition. 3.2  $\mu$ L of the stained DNA sample was put at the interface of a silanized coverslip and a clean microscopy slide (VWR). The extended DNA molecules were imaged with a fluorescence microscope (Zeiss Observer.Z1) using an Andor iXON Ultra EMCCD camera equipped with a Colibri 7 LED illumination system. Band-pass excitation (475/40 and 640/30) and emission filters (530/50 and 690/50) were used for YOYO-1 and ATTO-647N, respectively. Each imaging output produces two images: one for the DNA molecule (YOYO-1) and one for the fluorescent dots (ATTO-647N).

## 2.2. Data analysis

### 2.2.1. Analysis overview

Analysis of data was done using the Stained DNA Dot Detection (SD<sup>3</sup>) software or, for comparison, using ImageJ straight line selection tool and visual detection and manual counting of fluorescent dots. The analysis pipeline for SD<sup>3</sup> consists of four main steps (Fig. 2A): user input, image segmentation, molecule detection and dot detection. Using the Graphic User Interface (GUI) of SD<sup>3</sup>, the user selects input images and



**Fig. 2.** A) The SD<sup>3</sup> pipeline contains 4 steps, described in detail in Section 2.2. B) Histogram of edge scores for a sample experiment together with an automated threshold (see Section 2.2.3) C) Image showing a zoom-in region containing two detected DNA molecules that pass the automated threshold. D) DNA molecule and dot intensity plots of the detected DNA molecule 2 in C (marked). E) Dots on the DNA molecule marked with 'x' after LoG filtration. Scale bars = 1  $\mu$ m.

parameters for the analysis (Figure S1), which are then used to generate the results. The results from the software include visual output with score histograms (Fig. 2B), detected molecules (Fig. 2C), and their intensities in both channels (Fig. 2D). Fig. 2E visualises the dot positions along the DNA molecule. SD<sup>3</sup> also generates a tabular output of molecule statistics, such as molecule lengths (in micrometers,  $\mu\text{m}$ ), eccentricity and convexity ratio, which define the straightness of the stretched DNA molecules, as well as the number of dots and their intensities along each single DNA molecule. The details of the four steps in SD<sup>3</sup> are described in subsections 2.2.2 – 2.2.5.

### 2.2.2. User input and parameter definitions

In the first step (Fig. 2A) the user input is processed. The input parameters can be divided into three separate groups: image processing, molecule score filtering, and dot score filtering. Table 1 shows a summary of the default parameters used in the SD<sup>3</sup> pipeline. For the image processing, the user is asked for channel masks, camera pixel size nm/px, and width of Laplacian of Gaussian (LoG) filter logSigmaNm. The channel masks separate the molecule and dot channels, and the other two parameters help convert filter size from dimension (micrometers,  $\mu\text{m}$ ) to image dimension (pixels, px). Molecule score filtering parameters widthLims, lengthLims, elim and ratlim are used to filter out DNA molecules of poor quality. Eccentricity (elim) describes the elongation of a molecule, with values near 0 indicating a circular shape and values close to 1 representing a straight-line shape. The convexity ratio (ratlim) reflects how convex the region of a molecule is. Molecules that overlap or appear branched tend to have a lower convexity ratio and can be excluded from the analysis. Dot score filtering parameters minE and lineMethod are related to the dot quality and numSigmasAutoThresh is a single control parameter for the dot score thresholding. Detected dots with positions less than minE from the edge of the molecule are not considered. Dots at DNA ends could be due to mechanical fragmentation resulting from DNA extraction and hence need to be excluded. To exclude such dots from the analysis, SD<sup>3</sup> uses minE to specify how many pixels from the end to exclude. Section 3.1 demonstrates how the user interface can be used to get the desired output.

All the parameters in Table 1 can be varied by modifying values in the GUI of SD<sup>3</sup>. Additional parameters such as minimum distance of a molecule from the image edge and minimum branch length are hard-coded in the script, as they are in general and not changed for the analysis.

### 2.2.3. Image segmentation

In the second step, we segment the YOYO-1 image into DNA molecules and background (Fig. 2A). We first apply a Laplacian of Gaussian (LoG) filter using MATLAB's function imfilter (with filter width  $\sigma$  in

**Table 1**  
List of the default parameters in the SD<sup>3</sup> pipeline.

| Name                | Value           | Explanation   |
|---------------------|-----------------|---|
| nmpx                | 130             | Pixel size of camera in nanometers (nm/px)            |
| logSigmaNm          | 300             | Width of Laplacian of Gaussian filter in nanometers   |
| CH0                 | C= 0            | YOYO-1 channel flag                                   |
| CH1                 | C= 1            | dUTP-ATTO-647N channel flag                           |
| widthLims           | [1 Inf]         | Minimum / maximum molecule width in pixels (px)       |
| lengthLims          | [1 Inf]         | Minimum / maximum molecule length in pixels (px)      |
| minE                | 2               | Margin at the edge of molecule for dot detection (px) |
| elim                | 0.8             | Minimum molecule eccentricity                         |
| ratlim              | 0.4             | Minimum molecule to convexity ratio                   |
| lineMethod          | spline/<br>line | Molecule detection method                             |
| numSigmasAutoThresh | 5               | Number of sigmas for dot autothresh method            |

pixels, defined as logSigmaNm/nmpx) to get a LoG filtered image, logI. After filtering with a LoG filter, the edges (boundaries) of the molecules appear as zero-crossings (i.e. points where the LoG filtered image changes sign). We get a binarized image which we refer to as 'thedges', using MATLAB's imbinarize function with the threshold parameter as 0. Finally, we trace the exterior boundaries of objects and boundaries of holes inside the binarized image 1-thedges using MATLAB's bwboundaries (using the 'noholes' option). The first object detected using bwboundaries represents background (as given by definition of this function). Background pixels are assigned the original intensities of these pixels in the YOYO-1 image. For the rest of the objects, we calculate edge scores from the intensities of the LoG image logI by summing the values of  $n$  points (half the distance between minimum and maximum of the 1D LoG filter) from the zero crossing in the gradient direction, i.e.:

$$score_{edge}(k) = \frac{1}{N_k} \sum_{i=1}^{N_k} \left( \sum_{d=1}^n \log I(x_i - d \cdot g_{x_i}, y_i - d \cdot g_{y_i}) - \sum_{d=1}^n \log I(x_i + d \cdot g_{x_i}, y_i + d \cdot g_{y_i}) \right)$$

$N_k$  is the number of pixels in the object, and  $g_{x_i}, g_{y_i}$  are the slope coefficients in x and y directions along the molecule. To get a threshold on edge scores, we create a background score histogram by calculating edge scores for randomly sampled subsets of 20 pixels from the background and repeating this 1000 times. We set an auto-threshold  $score_{thresh}$  on the edge scores  $score_{edge}$  by calculating a mean and 3 times the standard deviation for the background score histogram. In Fig. 2B, we show an example of the edge score histogram together with an auto-threshold value " $score_{thresh}$ ". Finally unsuitable DNA molecules are filtered out based on  $score_{thresh}$ , elim, ratlim, lengthLims, widthLims. Binary mask matrices with the molecule positions are then saved for further steps.

### 2.2.4. Molecule detection

To extract the molecule length from the molecule masks of the YOYO-1 image, there are two options: either to fit a spline, or a simple line. To extract the molecule length from the molecule mask matrix using the spline method, we first create a skeleton (1-pixel wide line along the molecule) using MATLAB's bwskel with minimum branch length = 20 (so that short branches would be removed). We determine if the molecule is more vertical or horizontal (by taking the direction which covers more unique pixels). We smoothen pixels of the skeleton using a quadratic fit method with MATLAB's smooth function using a smoothing parameter of 0.1 and "loess" smoothing method. We fit a smoothing spline using MATLAB's fit function. We sample equidistant points along the spline and extract their intensity values using MATLAB's 2D interpolation function interp2 (using linear interpolation). Sampling these points over the YOYO-1 image (distance between two sampling points is 1 pixel) gives an emission intensity vector along the DNA molecule. The length of this vector gives the length of the molecule. We utilize the YOYO-1 intensity profile as a template for length detection. It is important to note that these intensity profiles provide a one-dimensional intensity trace of the molecule, which can be used i.e. to detect intensity variations along the molecule [33] or to discriminate linear and circular DNA [32].

### 2.2.5. Dot detection

Next, the dots are detected in the ATTO-647N channel (dots image) utilizing the molecule position information obtained from the YOYO-1 image. We calculate the dots intensities in a similar fashion as described in the molecule detection section. Then, to detect dot positions along the dot-channel intensity profile  $B(i)$ , we first subtract the median of the background pixels (background pixel positions are the same for both YOYO-1 and dots image) from the dot-channel intensity profile. We then filter with a 1D LoG filter using MATLAB's imfilter using 1D LoG

filter with parameter  $\sigma_2 = sF \times \sigma$  ( $sF$  approximately 1.31, since the dots channel color has a longer wavelength). For the case that the dots are well-separated the intensity profile is locally well-approximated by a Gaussian distribution. For such a case, the minima of the LoG filtered image (local minima of the LoG output is the local maxima) gives the locations  $p_L(i)$  of the dots  $i = 1, 2 \dots N$  along the molecule, where  $N$  is the number of dots on each molecule. We estimate the peak depth along the molecule by finding the minima between  $(p_L(i) \pm \sigma_2)$  to the boundaries of the molecule. As the dot score, we take the sum over the peak and a few surrounding pixels, i.e.

$$dot_{score}(i) = \frac{1}{2\sigma_2 + 1} \sum_{k=-\sigma_2}^{\sigma_2} (B(p_L(i) + k) - I_{bg})$$

where  $\sigma_2$  is rounded to the nearest integer, and  $I_{bg}$  is background intensity per pixel.

We now want to determine an auto-threshold parameter for the dot scores in order to discard false peaks. Since we have estimated the intensities of background pixels, we calculate dot scores for a randomly permuted vector of background pixels. The threshold is then taken to be `numSigmasAutoThresh` times the standard deviation of the background pixel dot scores plus mean of the background peak intensities. We take `numSigmasAutoThresh = 5` sigma in order to be strict when detecting dots. How strict we set this parameter depends on how important it is to avoid false positives. We note that some dots could be missed since 1) background intensity could be over-estimated if some signal pixels are included in background calculation, and 2) even though if we are using spline detection, the detected curve might not pass through the "center" of the dot, thus resulting in a lower score. The detected dots are then visualized, and the results saved in an output.txt file.

The intensity of each dot is calculated as

$$dot_{intensity}(i) = \frac{dot_{score}(i) \cdot (2\sigma_2 + 1)}{N}$$

where

$$N = normcdf(1) - normcdf(-1) \approx 0.6827$$

and *normcdf* is the cumulative distribution function of the standard normal distribution. The denominator in the expression for the dot intensity above corrects for the intensity that we miss when we sum over a region with left and right boundaries at  $\pm \sigma_2$  in the definition of the dot score. Note that multiple dots that could not be resolved due to the diffraction limit are counted as one dot.

### 2.2.6. Manual length and dot analysis using ImageJ

Manual data analysis was performed and compared with  $SD^3$  to evaluate the performance of  $SD^3$ . For manual DNA length measurements, a line was drawn along each DNA molecule using ImageJ straight line selection tool. Fluorescent dots were also counted manually by eye to estimate how many fluorescent dots that are present on each stretched DNA molecule.

## 3. Results

To evaluate the performance of  $SD^3$ , we analyzed images from experiments described in Section 2.1 where PBMCs were treated with  $H_2O_2$  before DNA extraction. We assessed the accuracy of  $SD^3$  in detecting DNA molecules and estimating the number of dots along the DNA length and compared it to manual scoring which was done using ImageJ (see Materials and Methods). The software is fast, yielding the data (molecule lengths and intensities, number of dots, dot intensities) at a rate of on average 10 images ( $1028 \times 1028$  pixels) analyzed on a standard laptop computer per minute (1.4 GHz Quad-Core Intel Core i5). A step-by-step guide with example data and results output is available as [Supporting Information](#).

### 3.1. Automated selection of DNA molecules

$SD^3$  offers several options that give the user flexibility on how images are analyzed. When stretching DNA on functionalized glass coverslips, capillary forces pull the DNA solution between the two glass surfaces, resulting in DNA molecules being stretched and aligned on the coverslip. Not all DNA molecules are perfectly straight, rather different molecule morphologies of the stretched DNA, from straight to highly bent DNA molecules, can be observed (Figure S2) and the bent/coiled DNA molecules are often not suitable for data analysis. In Fig. 3A, we show an output of  $SD^3$  after detecting DNA molecules.  $SD^3$  analyses the convexity ratio and eccentricity of each DNA molecule detected. Molecules 26 and 28 with convexity ratios of 0.4 and 0.8 are examples of bent and straight molecules, respectively (Fig. 3B–C). An increased convexity ratio corresponds to straight molecules, therefore by increasing the cut-off for the convexity ratio, bent molecules can be excluded from the analysis. Since the user defines these parameters, they can be applied during batch analysis.

### 3.2. Automatic DNA molecule length measurement

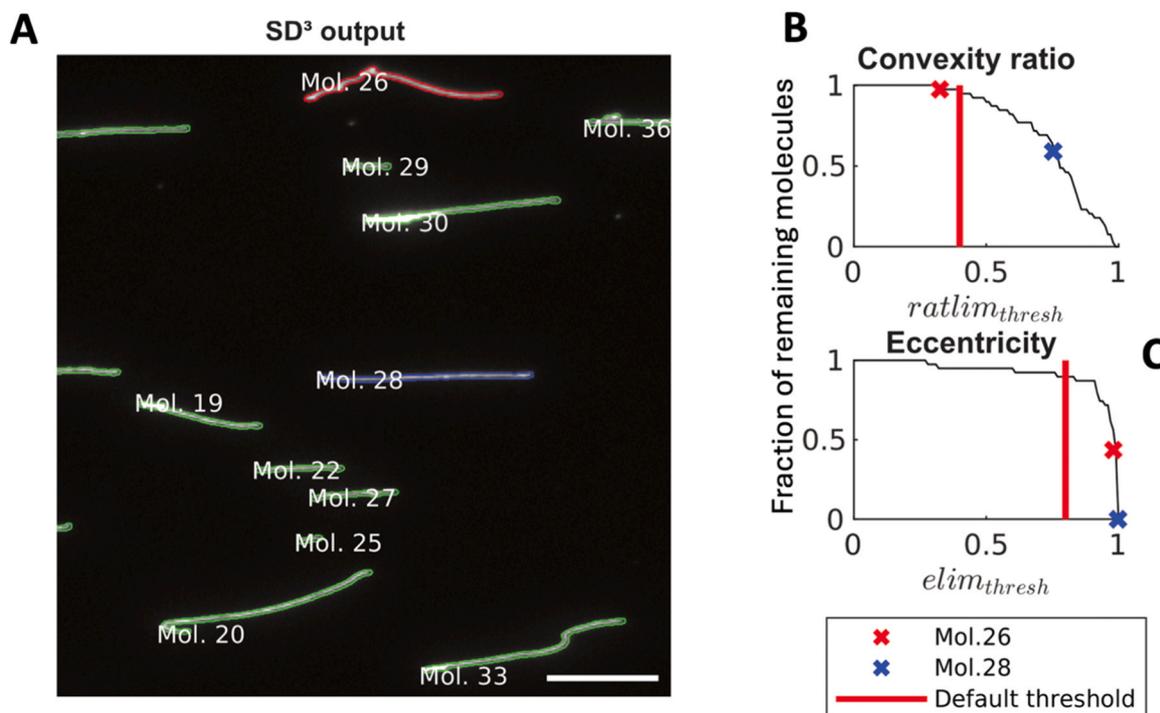
The next step of the analysis is to estimate the length of each DNA molecule. DNA was extracted from untreated PBMCs and stained with YOYO-1. The stained DNA was stretched on two separate functionalized coverslips and imaged. We compared the DNA length estimation using  $SD^3$  to manual length measurement using ImageJ where a line was manually drawn along the DNA. We observed a strong correlation between the DNA length measurement by  $SD^3$  and manual measurement for each of the two sets of images (Fig. 4A–B). The correlation was statistically significant indicating that the molecule length measurement by  $SD^3$  very closely matched that done manually using ImageJ. Using  $SD^3$ , we also compared DNA lengths in an untreated sample to an  $H_2O_2$ -treated sample. Treatment with  $H_2O_2$  resulted in significantly shorter DNA molecules (Fig. 4C and Figure S3) which is consistent with previous reports where  $H_2O_2$  has been demonstrated to fragment DNA [34].

### 3.3. Automatic dot detection by $SD^3$

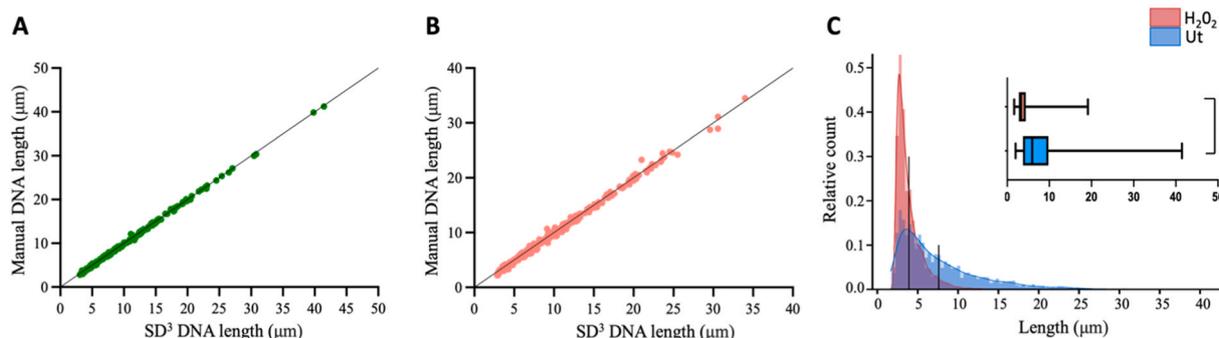
The next step in the data analysis is to detect the fluorescent dots. PBMCs were again treated with  $H_2O_2$ , and the extracted DNA samples were labeled using the procedure described in Section 2.1.3. Data quantification using  $SD^3$  reports damage as dots/ $\mu m$ , which later can be converted to dots/Mb. As shown in Fig. 5A, treatment with  $H_2O_2$  resulted in increased DNA damage, in agreement with previous reports that show  $H_2O_2$  to cause oxidative DNA damage [30].

Again, we tested the accuracy of  $SD^3$  in counting the number of dots along the DNA and compared it to manual scoring. A positive correlation was obtained when comparing dot estimation by the two methods and this was statistically significant (Fig. 5B).  $SD^3$  generates a result output that includes the intensity of each individual dot as well as the total intensity of all detected dots. The dot intensity measurements can be used where the actual amount of intensity is relevant for an assay and could help identify sites that are so close to each other that they cannot be distinguished as two (or more) dots.

$SD^3$  has the option to specify how many pixels from the molecule end to exclude dots. This is particularly important for DNA damage where the repair enzymes used can label damaged DNA ends that are formed when handling the extracted DNA. Based on this consideration, we performed analysis by excluding dots that were a certain number of pixels away from the ends of the molecules. In untreated and  $H_2O_2$ -treated samples, we demonstrated that if not using this criterion at all, the number of dots per DNA length are higher than if a certain number of pixels are excluded from the ends (Fig. 5C–D), but it is enough to remove two pixels from each end to avoid this effect.



**Fig. 3.** A) An output of SD<sup>3</sup> after analysis. Two molecules (Mol. 26 and Mol. 28) are color coded (red and blue) in order to show specific values of the estimated parameters. B-C) The proportion of detected molecules 26 and 28 in (A) kept by SD<sup>3</sup> as a function of the parameters: convexity ratio (B) and eccentricity (C). Note that Mol. 28 has a larger convexity ratio and eccentricity than Mol. 26, and Mol. 26 would be discarded by SD<sup>3</sup> if the convexity ratio was set to the default value (0.4). Scale bar = 10  $\mu\text{m}$ .



**Fig. 4.** Comparison of DNA molecule length between manual measurement and SD<sup>3</sup>. A-B) YOYO-stained DNA extracted from untreated PBMCS stretched on two separate functionalized glass coverslips. Spearman's rank correlation shows a strong positive correlation between DNA molecule length given by SD<sup>3</sup> and manual measurement ( $r_s=0.99$ ,  $n = 289$ ,  $p < 0.0001$ ) and ( $r_s = 0.99$ ,  $n = 200$ ,  $p < 0.0001$ ) for A) and B) respectively). C) Frequency distribution of DNA molecule lengths of untreated (Ut) and H<sub>2</sub>O<sub>2</sub> - treated PBMCS. Statistical analysis in C) was performed using Mann-Whitney-Wilcoxon test. \*\*\*\* represents  $p < 0.0001$ .

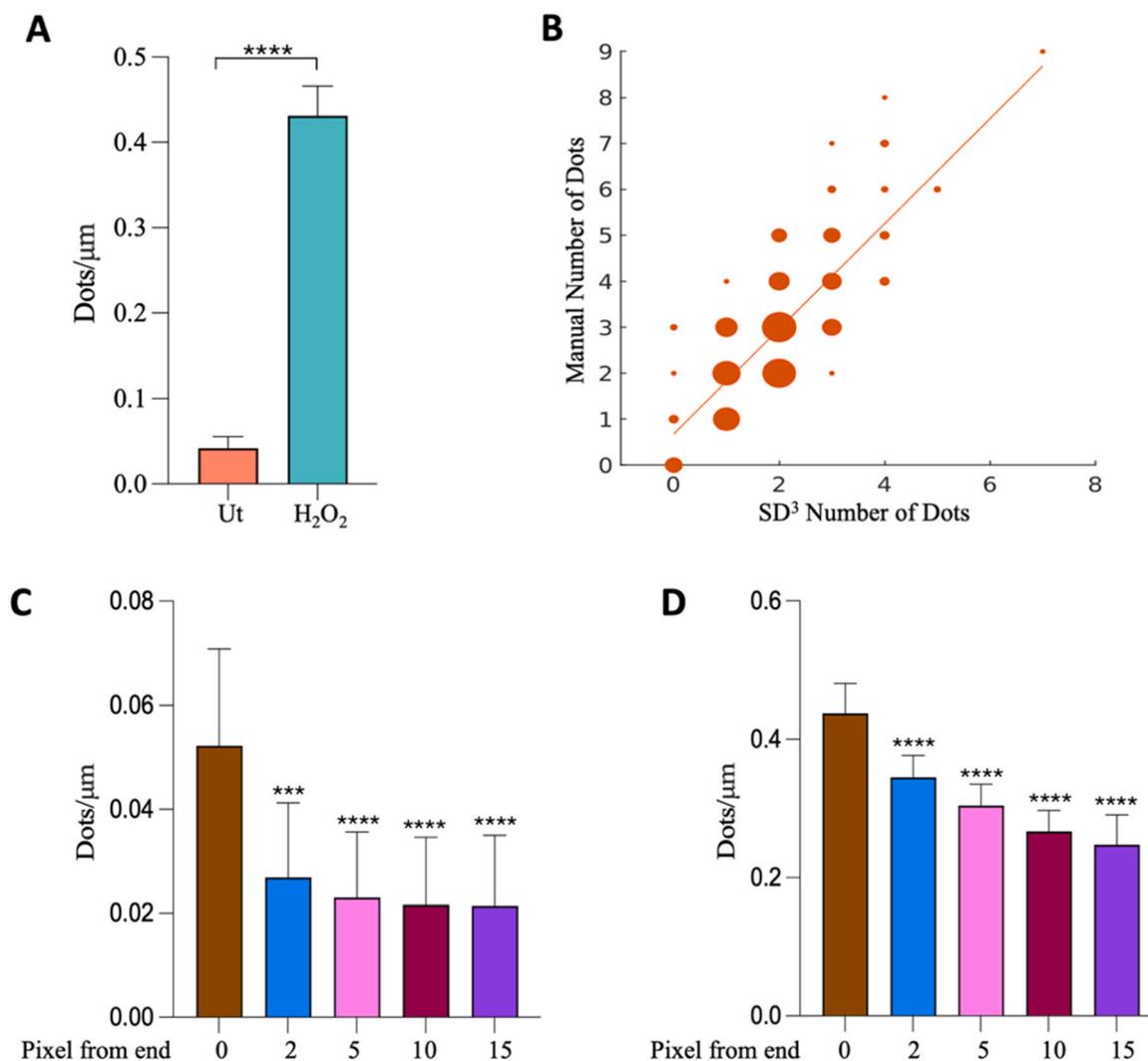
#### 4. Discussion and conclusion

Detection and quantification of fluorescent dots along DNA at the single DNA molecule level is important in many different aspects, such as for DNA damage analysis [5,6] and in epigenetics [3,4]. We have here presented an automated software, SD<sup>3</sup>, that can effectively analyze such images with high precision and reproducibility. This significantly improves the workflow compared to previous manual analysis, which is laborious, time consuming and limits the number of images that can be processed [4,7]. Since the software also detects where along a DNA molecule a dot is located, it can also be used for optical DNA mapping analysis where enzymes are used to position a fluorophore at a specific sequence motif [35].

SD<sup>3</sup> performs an accurate length measurement of detected DNA molecules and counts the number of dots along the molecule length. SD<sup>3</sup> has a user interface that allows the user to define the parameters to

include in the analysis. We recommend users to apply the automatic threshold of dotscore and molecule edgescore to get the most out of molecule and dot detection. In comparison to conventional analysis of drawing lines across the molecule and visually counting dots on the molecules, SD<sup>3</sup> significantly reduces the analysis time (10 1028x1028 pixel images per minutes) while achieving highly accurate results. In comparison to the existing custom-made image processing and analysis software Tiff16\_Analyzer [36,37], it is important to note that while SD<sup>3</sup> gives data on each individual molecule detected along with a summary text file of the analysis, Tiff16\_Analyzer generates a result file which gives a summary of all DNA molecules detected per field of view.

Single-molecule methods demand precise and careful sample preparation. Mechanical shearing during preparation such as pipetting and vortexing can lead to DNA fragmentation or unintended damage unrelated to the experiment. To avoid counting DNA strand breaks caused by shear forces and non-specific labeling by DNA polymerase I, fluorescent



**Fig. 5.** A) DNA damage detection in PBMCs from pooled blood left untreated (Ut) or exposed to  $H_2O_2$  using the enzyme cocktail. Statistical analysis was performed using Mann-Whitney-Wilcoxon test. \*\*\*\* represents  $p < 0.0001$ . B) Comparison of number of dots detected between manual estimation and  $SD^3$  ( $r_s = 0.73$ ,  $n = 325$ ,  $p < 0.0001$ ). The circle size scales linearly with the number of overlapping datapoints. C and D) DNA damage detection after excluding datapoints located at specified pixel distances from the DNA molecule ends in untreated (C) and  $H_2O_2$ -treated (D) samples. Dots as well as DNA molecule lengths where dots were located were excluded from the analysis. One-way ANOVA statistical significance was determined using Tukey's model for multiple comparisons with a family-wise alpha threshold and confidence level of 95 % (confidence interval). \*\*\* represents  $p < 0.001$  and \*\*\*\* represents  $p < 0.0001$ .

dots at the ends of the DNA can be excluded from the analysis. We typically exclude dots two pixels from both ends of the molecule. Furthermore,  $SD^3$  has an option to specify the minimum length of each DNA molecule to include in the analysis for which we generally recommend molecule length to be at least greater than 50 pixels. On the downside, since conventional fluorescence microscopes are used for image acquisition, which typically have diffraction limit of a few hundred nanometers, two or more damage sites positioned within the diffraction limit cannot be spatially resolved. Instead, they overlap and appear as a single dot. Note, however, that  $SD^3$  records the emission intensity of each dot. Therefore, if several damage sites are within one dot, emission intensity will be proportionally higher compared to a single damage site. This could potentially be used for extracting information about the number of sites per dot.

To conclude,  $SD^3$  is an easily accessible open-source software that can be used to investigate damage on stretched individual DNA molecules. In a bigger picture, the various features in  $SD^3$  will enable a vast range of research on stretched DNA molecules to be performed. This will in turn improve biophysical and biochemical image analysis.

#### CRediT authorship contribution statement

**Ambjörnsson Tobias:** Writing – review & editing, Supervision, Software, Methodology, Funding acquisition, Conceptualization. **Johansson Pegah:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Westerlund Fredrik:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Nyblom My:** Writing – review & editing, Conceptualization. **Dvirnas Albertas:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Carlson Johanna:** Writing – review & editing, Formal analysis. **Krog Jens:** Writing – review & editing, Software. **Aning Obed A.:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

#### Declaration of Competing Interest

The authors declare no conflicts of interest

## Acknowledgements

We acknowledge funding from the Swedish Cancer Foundation (2017/654 and 20 1145 PjF to F.W.) the Swedish Child Cancer Foundation (PR2019–0037 and PR2022–001 to F.W. and PR2020–0105 to P. J.), Familjen Erling Perssons Stiftelse to F.W. and T.A. and the Wenner-Gren Foundation (UPD2021–0036 to O.A.).

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.dnarep.2025.103836](https://doi.org/10.1016/j.dnarep.2025.103836).

## Data availability

Data will be made available on request.

## References

- [1] K. Long, L. Cai, L. He, DNA sequencing data analysis, *Methods Mol. Biol.* 1754 (2018) 1–13.
- [2] M.L. Metzker, Sequencing technologies - the next generation, *Nat. Rev. Genet.* 11 (2010) 31–46.
- [3] Y. Michaeli, T. Shahal, D. Torchinsky, A. Grunwald, R. Hoch, Y. Ebenstein, Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases, *Chem. Commun. (Camb.)* 49 (2013) 8599–8601.
- [4] G. Nifker, M. Levy-Sakin, Y. Berkov-Zrihen, T. Shahal, T. Gabrieli, M. Fridman, Y. Ebenstein, One-Pot Chemoenzymatic Cascade for Labeling of the Epigenetic Marker 5-Hydroxymethylcytosine, *ChemBiochem* 16 (2015) 1857–1860.
- [5] V. Singh, P. Johansson, E. Ekedahl, Y.L. Lin, O. Hammarsten, F. Westerlund, Quantification of single-strand DNA lesions caused by the topoisomerase II poison etoposide using single DNA molecule imaging, *Biochem Biophys. Res Commun.* 594 (2022) 57–62.
- [6] V. Singh, P. Johansson, Y.L. Lin, O. Hammarsten, F. Westerlund, Shining light on single-strand lesions caused by the chemotherapy drug bleomycin, *DNA Repair (Amst.)* 105 (2021) 103153.
- [7] S. Zirkov, S. Fishman, H. Sharim, Y. Michaeli, J. Don, Y. Ebenstein, Lighting up individual DNA damage sites by in vitro repair synthesis, *J. Am. Chem. Soc.* 136 (2014) 7771–7776.
- [8] J.H. Hoeijmakers, Genome maintenance mechanisms for preventing cancer, *Nature* 411 (2001) 366–374.
- [9] R. Hakem, DNA-damage repair; the good, the bad, and the ugly, *EMBO J.* 27 (2008) 589–605.
- [10] S.P. Jackson, J. Bartek, The DNA-damage response in human biology and disease, *Nature* 461 (2009) 1071–1078.
- [11] R. Abbots, D.M., 3rd Wilson, Coordination of DNA single strand break repair, *Free Radic. Biol. Med* 107 (2017) 228–244.
- [12] E.C. Friedberg, DNA damage and repair, *Nature* 421 (2003) 436–440.
- [13] "??(0) = "!count(descendant::ce:intra-ref)"??(0) = "!count(descendant::ce:doi)"??(0)[?pdfitagstart "Link"] > A. Tubbs, A. Nussenzweig, Endogenous DNA damage as a source of genomic instability in cancer, *Cell* 168 (2017) 644–656.
- [14] J. da Silva, DNA damage induced by occupational and environmental exposure to miscellaneous chemicals, *Mutat. Res Rev. Mutat. Res* 770 (2016) 170–182.
- [15] G. Figueroa-Gonzalez, C. Perez-Plasencia, Strategies for the evaluation of DNA damage and repair mechanisms in cancer, *Oncol. Lett.* 13 (2017) 3982–3988.
- [16] A.R. Collins, The comet assay for DNA damage and repair: principles, applications, and limitations, *Mol. Biotechnol.* 26 (2004) 249–261.
- [17] I.J. Chopra, D.H. Solomon, G.N. Beall, Radioimmunoassay for measurement of triiodothyronine in human serum, *J. Clin. Invest* 50 (1971) 2033–2041.
- [18] T.R. Berton, D.L. Mitchell, Quantification of DNA photoproducts in mammalian cell DNA using radioimmunoassay, *Methods Mol. Biol.* 920 (2012) 177–187.
- [19] M.S. Cooke, R. Olinski, S. Loft, European Standards Committee on Urinary Lesion, A. Measurement and meaning of oxidatively modified DNA lesions in urine, *Cancer Epidemiol. Biomark. Prev.* 17 (2008) 3–14.
- [20] T. Detinis Zur, J. Deek, Y. Ebenstein, Single-molecule approaches for DNA damage detection and repair: a focus on Repair Assisted Damage Detection (RADD), *DNA Repair (Amst.)* 129 (2023) 103533.
- [21] B. Cao, X. Wu, J. Zhou, H. Wu, L. Liu, Q. Zhang, M.S. DeMott, C. Gu, L. Wang, D. You, et al., Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage, *Nucleic Acids Res* 48 (2020) 6715–6725.
- [22] J. Lee, H.S. Park, S. Lim, K. Jo, Visualization of UV-induced damage on single DNA molecules, *Chem. Commun. (Camb.)* 49 (2013) 4740–4742.
- [23] S. Poole, O.A. Aning, V. McKee, T. Catley, A.Y. Nielsen, H. Thisgaard, P. Johansson, G. Menounou, J. Hennessy, C. Sator, et al., Design and in vitro anticancer assessment of a click chemistry-derived dinuclear copper artificial metallo-nuclease, *Nucleic Acids Res.* 53 (2025).
- [24] I. Horvath, O.A. Aning, S. Kk, N. Rehnberg, S. Chawla, M. Molin, F. Westerlund, P. Wittung-Stafshede, Biological amyloids chemically damage DNA, *ACS Chem. Neurosci.* 16 (2025) 355–364.
- [25] T. Detinis Zur, S. Margalit, J. Jeffet, A. Grunwald, S. Fishman, Z. Tulpova, Y. Michaeli, J. Deek, Y. Ebenstein, Single-molecule toxicogenomics: Optical genome mapping of DNA-damage in nanochannel arrays, *DNA Repair (Amst.)* 146 (2025) 103808.
- [26] D. Torchinsky, Y. Michaeli, N.R. Gassman, Y. Ebenstein, Simultaneous detection of multiple DNA damage types by multi-colour fluorescent labelling, *Chem. Commun. (Camb.)* 55 (2019) 11414–11417.
- [27] N. Gilat, D. Torchinsky, S. Margalit, Y. Michaeli, S. Avraham, H. Sharim, N. Elkoshi, C. Levy, S. Zirkov, Y. Ebenstein, Rapid Quantification of Oxidation and UV Induced DNA Damage by Repair Assisted Damage Detection-(Rapid RADD), *Anal. Chem.* 92 (2020) 9887–9894.
- [28] N. Gilat, D. Fridman, H. Sharim, S. Margalit, N.R. Gassman, Y. Michaeli, Y. Ebenstein, From single-molecule to genome-wide mapping of DNA lesions: repair-assisted damage detection sequencing, *Biophys. Rep. (N. Y.)* 1 (2021) (None).
- [29] V. Singh, P. Johansson, D. Torchinsky, Y.L. Lin, R. Oz, Y. Ebenstein, O. Hammarsten, F. Westerlund, Quantifying DNA damage induced by ionizing radiation and hyperthermia using single DNA molecule imaging, *Transl. Oncol.* 13 (2020) 100822.
- [30] J. Lee, Y. Kim, S. Lim, K. Jo, Single-molecule visualization of ROS-induced DNA damage in large DNA molecules, *Analyst* 141 (2016) 847–852.
- [31] Q. Wei, W. Luo, S. Chiang, T. Kappel, C. Mejia, D. Tseng, R.Y. Chan, E. Yan, H. Qi, F. Shabbir, et al., Imaging and sizing of single DNA molecules on a mobile phone, *ACS Nano* 8 (2014) 12725–12733.
- [32] G. Goyal, E. Ekedahl, M. Nyblom, J. Krog, E. Frobrant, M. Brander, T. Sewunet, T. Tangkoskul, C.G. Giske, L. Sandegren, et al., A simple cut and stretch assay to detect antimicrobial resistance genes on bacterial plasmids by single-molecule fluorescence microscopy, *Sci. Rep.* 12 (2022) 9301.
- [33] A.N. Nilsson, G. Emilsson, L.K. Nyberg, C. Noble, L.S. Stadler, J. Fritzsche, E. R. Moore, J.O. Tegenfeldt, T. Ambjornsson, F. Westerlund, Competitive binding-based optical DNA mapping for fast identification of bacteria–multi-ligand transfer matrix theory and experimental applications on *Escherichia coli*, *Nucleic Acids Res* 42 (2014) e118.
- [34] Y. Fujii, K. Tomita, H. Sano, A. Yamasaki, Y. Hitsuda, I.M. Adcock, E. Shimizu, Dissociation of DNA damage and mitochondrial injury caused by hydrogen peroxide in SV-40 transformed lung epithelial cells, *Cancer Cell Int* 2 (2002) 16.
- [35] E.T. Lam, A. Hastie, C. Lin, D. Ehrlich, S.K. Das, M.D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, et al., Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly, *Nat. Biotechnol.* 30 (2012) 771–776.
- [36] N. Jain, T. Shahal, T. Gabrieli, N. Gilat, D. Torchinsky, Y. Michaeli, V. Vogel, Y. Ebenstein, Global modulation in DNA epigenetics during pro-inflammatory macrophage activation, *Epigenetics* 14 (2019) 1183–1193.
- [37] D. Torchinsky, Y. Ebenstein, Sizing femtogram amounts of dsDNA by single-molecule counting, *Nucleic Acids Res* 44 (2016) e17.