

Statistical Applications in Genetics and Molecular Biology

Volume 4, Issue 1

2005

Article 5

Comparing Automatic and Manual Image Processing in FLARE Assay Analysis for Colon Carcinogenesis

| | | |
|-----------------------------------|-------------------------------|----------------------------------|
| Malgorzata Leyk [*] | Danh V. Nguyen [†] | Sanju N. Attoor [‡] |
| Edward R. Dougherty ^{**} | Nancy D. Turner ^{††} | Laura K. Bancroft ^{‡‡} |
| Robert S. Chapkin [§] | Joanne R. Lupton [¶] | Raymond J. Carroll |

^{*}Texas A&M University, gosia@tamu.edu

[†]University of California, Davis, ucdnguyen@ucdavis.edu

[‡]Texas A&M University, sanju@tamu.edu

^{**}Texas A&M University, e-dougherty@tamu.edu

^{††}Texas A&M University, n-turner@tamu.edu

^{‡‡}Albert Einstein Cancer Center, lkbancroft@yahoo.com

[§]Texas A&M University, r-chapkin@tamu.edu

[¶]Texas A&M University, jlupton@tamu.edu

^{||}Texas A&M University, rcarroll@tamu.edu

Copyright ©2005 by the authors. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher, bepress, which has been given certain exclusive rights by the author. *Statistical Applications in Genetics and Molecular Biology* is produced by The Berkeley Electronic Press (bepress). <http://www.bepress.com/sagmb>

Comparing Automatic and Manual Image Processing in FLARE Assay Analysis for Colon Carcinogenesis*

Malgorzata Leyk, Danh V. Nguyen, Sanju N. Attoor, Edward R. Dougherty, Nancy D. Turner, Laura K. Bancroft, Robert S. Chapkin, Joanne R. Lupton, and Raymond J. Carroll

Abstract

Measurement of the amount of oxidative damage to DNA is one tool that can be used to estimate the beneficial effect of diet on the prevention of colon carcinogenesis. The FLARE assay is a modification of the single-cell gel electrophoresis (Comet) assay, and provides a measure of the 8OHdG adduct in the cells. In this paper, we present two innovations to the existing methods of analysis. The first one is related to the FLARE assay itself. We describe automated image analysis techniques that can be expected to measure oxidative damage faster, reproducibly, with less noise, and hence achieve greater statistical power. The proposed technique is compared to an existing technique, which was more manual and thus slower. The second innovation is our statistical analysis: we exploit the shape of FLARE intensity histograms, and show statistically significant diet effects in the duodenum. Previous analyses of this data concentrated on simple summary statistics, and found only marginally statistically significant diet effects. With the new imaging method and measure of oxidative damage, we show cells in the duodenum exposed to fish oil as having more oxidative damage than cells exposed to corn oil.

KEYWORDS: colon carcinogenesis, comet assay, corn oil, fish oil, FLARE assay, hierarchical models, image analysis, single-cell gel electrophoresis.

*Our research was supported by grants from the National Cancer Institute (CA57030, CA59034, CA61750, CA82907), NASA (NPFR00202), and by the Texas A&M Center for Environmental and Rural Health through a grant from the National Institute of Environmental Health Sciences (P30-ES09106).

INTRODUCTION

Cancer usually develops slowly; hence, a reliable biomarker of oxidative damage would be beneficial. Altered guanine base 8-hydroxy-2'-deoxyguanosine (8OHdG) can be considered as one such biomarker (see Boiteux and Radicella, 1999; Halliwell, 2002; Jackson and Loeb, 2001; Loft and Poulsen, 1996). 8OHdG has been found to be highly mutagenic (see Boiteux and Radicella, 1999). Hence, an increase in the amount of 8OHdG damage to DNA should be correlated with increased rates of cancer (see Halliwell, 2002).

To be able to measure increased levels of 8OHdG, the FLARE assay can be utilized. The FLARE assay uses the single-cell gel electrophoresis assay, also known as Comet assay, to produce images of the amount of damage to DNA of individual cell nuclei (see Collins et al., 1995; Singh et al., 1988). Moreover, half of the nuclei are exposed to the repair enzyme formamidopyrimidine-DNA glycosylase enzyme (FPG). FPG is found in *Escherichia coli*, and it is part of the repair process by removing the oxidized guanine (refer to Boiteux and Radicella, 1999). DNA of cells that are exposed to FPG should contain both naturally occurring breaks in DNA, as well as breaks caused by FPG at places containing oxidized guanine base. DNA of cells not exposed to FPG can serve as a baseline for naturally occurring DNA breaks.

Existing methods of image analysis involve manually identifying each comet on an image, and utilizing a macro on each identified comet to compute a measure of DNA damage (see Bancroft et al., 2003). Doing this for thousands of comet images is slow, and prone to possible error. We propose an automated image analysis that will automatically identify comets in an image, as well as compute a desired measure of damage for each comet, which would be faster and more consistent than the current image analysis method.

In the paper, we compare the two techniques to determine how they differ by analyzing the amount of 8OHdG DNA damage in colonocytes from rats consuming diets containing either fish oil or corn oil (see Bartsch et al., 1999; Diggle, 2002; Hong et al., 2002; Sanders et al., 2002; Sugimura, 2000), and with or without dextran sodium sulfate (DSS) treatment (see Tardieu et al., 1998; Tardieu et al., 2000). Analysis of our data using existing methodologies yielded marginal results. The analysis using the proposed automated imaging algorithm produced more significant results.

We also hypothesize that an important component of the data collected is the image shape. This paper reports a new statistical variable that describes the shape of a comet, along with the automated imaging algorithm, and statistically significant diet effects.

MATERIALS AND METHODS

Two groups of Sprague-Dawley male rats were established. One group consumed a diet containing corn oil, while the other group of rats consumed a diet containing fish oil (see Bancroft et al., 2003). Three further subgroups were made in each of the diet groups. The first subgroup was treated with 3% DSS (ICN Biomedicals; Aurora, OH) in the drinking water for 48 hours. Another subgroup was also treated with DSS for 48 hours, but was then given 48 hours before euthanasia, to allow for recovery and DNA repair. The final subgroup was the control which received no DSS treatment. Each of the subgroups contained ten rats. Results from one rat fed corn oil with DSS treatment was excluded due to a sample preparation problem.

FLARE Assay Preparation

For each rat, cells were obtained from both the colon (large intestine) and the duodenum (small intestine). There were usually more than 200 cells per rat in each location available for analysis. Half of the cells obtained from each location were treated with FPG, while the other half remained untreated. Since these locations are directly exposed to the oil effect, they should be more precise in predicting the direct effect of oxidative damage in the intestinal tract due to the oils, as opposed to a more indirect measurement of oxidative damage, for example, in blood or urine.

To measure the amount of DNA damage in each cell, the FLARE assay, also known as the FPG-comet assay, was utilized. The FLARE assays were prepared by first melting Comet LM Agarose (Trevigen; Gaithersburg; MD), and allowing it to cool in a 42°C water bath for 10 minutes. The cells were prepared at a concentration of 100,000 cells per 1mL in cold 1X PBS (calcium and magnesium free). In a microcentrifuge tube, 25µL of the cells and 250µL of the cooled agarose were mixed. Then 50µL of the mixture was quickly pipetted onto each sample area of the CometSlide (Trevigen). The mixture was then evenly spread onto the whole sample area of the slide and placed in a dark refrigerator at 4°C for 20 to 25 minutes to allow for the agarose to shrink. The lysis solution (1% sodium lauryl-sarcosinate, 2.5M NaCl, 100mM EDTA, 1% Triton X-100, 10mM Tris base; pH 10) was made same way as was described by Singh, et al. (1988). The slides were then immersed in the pre-chilled lysis solution, and sat on ice for 30 minutes. The slides were then transferred into 1X FLARE buffer (1mM HEPES-KOH at pH 7.4 100mM KCl) for 15 minutes. The buffer was then disposed, and this step was repeated three additional times. The area on the slides around the agarose was dried, and the FPG-treated cells were treated with a working enzyme solution (50% FPG activity in 800µL of the reaction buffer (0.04% 25X FLARE Buffer, 0.01% 100X BSA)). The untreated cells were only exposed to the reaction buffer. The treatment was added dropwise to the sample area. The slides were incubated in a humidity chamber at 37°C for 45 minutes. Then the slides were incubated in alkali solution (1mM EDTA; pH 12.5) at room

temperature in the dark for 15 minutes. The alkali solution was replaced and incubation was done for another 15 minutes. The horizontal gel electrophoresis apparatus was filled with enough alkali solution to cover the slides by 1cm. The slides were placed equidistance from the electrodes, and were allowed to electrophorise for 20 minutes at 25V in the dark at 4°C, as previously cited by Singh. The excess alkali solution was removed, and the slides were immersed in 70% ethanol for 5 minutes. The slides were allowed to air dry overnight in the dark. The slides were then stained with a diluted SYBR Green (0.1µL SYBR Green, 1mL TE(10mM Tris, 1mM EDTA; pH 7.4)). The slides were then viewed by epifluorescence microscopy.

A complete description of the procedure is available in Brancroft, et al. (2003).

As was described above, the FLARE assay is performed by running gel electrophoresis on DNA from a single isolated cells. Migration of the DNA is dependent on its size, with smaller pieces migrating farther in the gel than larger pieces, resulting in an image that gives the appearance of a comet. A comet shows the amount of damage to DNA of a single cell. Figure 1 contains examples of comets with varying levels of DNA damage. These comets were chosen to show the different levels of comet damage.

The comet is composed of two parts. The circular part to the left is called the head (or nucleus). When there is no damage, there will only be a head. However, when there is damage, the comet will also have a tail. Everything to the right of the head is called the tail. In Figure 1, the comet in the top left corner only has a head. Consequently, the DNA of this cell had only minimal level of damage. However, as you move down and from left to right, the amount of DNA damage increases. This is reflected in the fact that more of the intensity of the comet is contained in the tail. The comet showing most DNA damage is in the bottom right corner. Almost all of the intensity is in the tail, and the head is not very visible.

This method is able to detect 1 single-strand DNA break per 3×10^9 Daltons of DNA (refer to Singh, 2000).

After the preparation, images were obtained from the processed cells. An image algorithm was then used, which computed a measure of the amount of oxidative damage in each cell. The amount of 8OHdG adducts for each rat-location used in the analysis was the difference between the FPG measure and the no-FPG measure.

Existing Image Analysis and Calculations

The existing analysis used the Metamorph Imaging System (Version 4.6r3, Universal Imaging Corporation, Downingtown, PA) for capture and analysis of the image, as follows. Images of randomly selected comets were captured and the file was stored on the computer. Each comet image was then identified manually on the slide. For each comet, the head of the comet was identified by selecting a threshold which separates the head from the background.

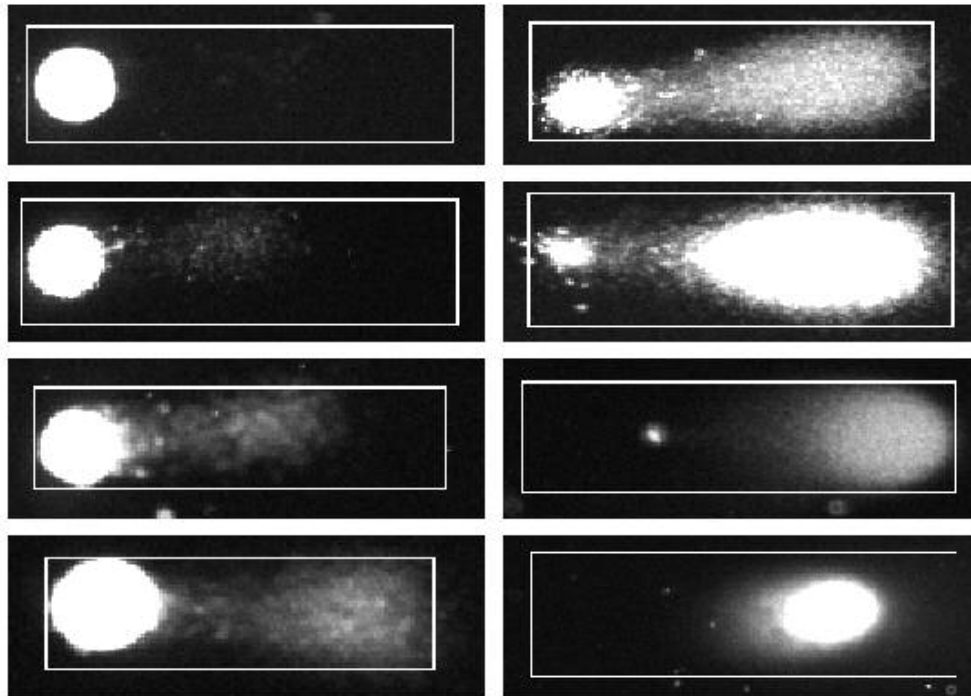


Figure 1: Different levels of damage in comets. Damage level increases for comets moving from top to bottom, and from left to right. The comet in the top left corner has the least amount of damage (it is almost all head). The comet in the bottom right corner has the most damage (it is almost all tail).

If pixels that were not part of the comet were selected, then a freeform tool was used to define the area of the image where the comet was located. If pixels in the tail were included in the head after thresholding, the user could create a two pixel gap between the head and the tail to help the program identify the end of the head. Any pixels still identified by thresholding but not part of the head could be removed from calculations by manually identifying them.

A box was then drawn that contains the comet, with the right side of the box extending 10 to 20 pixels beyond the end of the comet tail. A macro was executed that circled the thresholded head and found its center. Average intensity of the pixels to the right of the tail was used to estimate the background intensity, which was subtracted from the result. A measure of the amount of damage was then computed by the program for the comet.

The standard output from FLARE analysis is the relative tail moment (RTM) (see Hellman et al., 1995; Morris et al., 1999; Riso et al., 1999), defined as

$$\text{RTM} = 100 * (\text{tail moment}) / (\text{tail moment} + \text{head moment})$$

where head or tail moment is defined as the sum of the distance away from the center of the head times the amount of DNA in the head or tail at that distance, respectively. The advantage of this variable is that it does not depend on the scale used to measure the distance (as long as the same scale is used for both the head and the tail moment), or whether the intensity of the comet is standardized.

Proposed New Image Analysis

The existing image analysis method did not have the benefit of a standardized routine of data generation, making it less objective than desired. Therefore, the possibility exists for inconsistent comet identification and processing. The new method of image analysis uses a set algorithm for identifying a comet and its components, thus reducing the variability introduced by human error in the existing image analysis system. It provides a methodical approach that is reproducible, and concentrates on eliminating as much of the background noise as possible before processing.

The processing of the FLARE images was done in the following manner. First, a grayscale FLARE image was converted to a zero-minimum-intensity image by subtracting the minimum intensity from all the pixels. The image was then enhanced by a nonlinear scaling of the pixel intensities: each pixel intensity was raised to the power 1.05. The enhanced image was converted to a binary image by thresholding at a grayscale level of 50. This binary image had legitimate comets together with abnormal non-comet areas of staining on a black background with speckled noise. From observation it was found that comets had an area of less than 11000 pixels. Anything larger than that was considered non-specific, and was removed by using a sequence of morphological operations (refer to Dougherty, 1992; Dougherty and Astola, 1999; Giardina and Dougherty, 1988; Serra, 1982).

Incomplete comet images (those partially obstructed by the image boundary) were eliminated using morphological operators for removing boundary regions (see Dougherty, 1992). The speckled noise in the resulting binary image was removed by applying morphological opening on the image using a disk of radius 1. A grayscale image corresponding to the original enhanced image was created using the morphological operations of thinning and reconstruction and using the above binary image as the mask. Morphological methods for reconstruction and thinning may be found in Dougherty (1992) and the SDC morphology toolbox for matlab.

The resulting image, J , was then used to construct markers for the potential comet regions, which could be head, tail, head with tail, or abnormal shaped regions. First, markers for heads were constructed through a sequence of

morphological operations. The image J was negated and basins with contrast greater than 170 were found using the SDC morphology toolbox for matlab. These regions were labeled as the heads since it was found that the head region had a very high contrast (greater than 170) compared to tails. To remove noise, shapes with a disk radius of less than 3 were removed by morphological opening. This image was converted into a binary image via thresholding as discussed earlier. A thinning operation on the binary image provided the markers for the heads. The head was reconstructed from the original enhanced image via reconstruction using morphological operators and the markers for the head. The process of thinning and reconstruction removed some of the noise not removed by the morphological opening used earlier.

A similar operation for reconstructing the tails was used. Regions corresponding to the head were subtracted from the original image and markers for the tail were generated as in the case of the head after cleaning the image. The tails were then reconstructed. It should be noted that the regions detected may not necessarily be tails; they could be comets complete with a head and tail or heads which are weak in signal intensity. The two binary images of the head and the tail regions were combined and each region was labeled. Thus the outline of all clearly visible heads and tails of comets in an image were obtained. Tails with weak signal intensities and low density may not be detected and were specially processed later on.

Before further processing of the head and tail regions, regions with abnormal shapes should be removed. To do this, an intensity vector was constructed for each region by summing along the vertical length of the region for each pixel along the horizontal length of the image. The shape of the intensity vector was used to determine if the region is a legitimate head/tail region. The intensity vector was smoothed using an FIR filter of order 25 and cut-off frequency 0.1. A measure, defined as the square root of the weighted sum of the squares of the intensity values and the respective index values (of the intensity vector), was then computed. This measure was indicative of the amount of matter around a peak: the region was declared as abnormal if either this measure was too low or if the measure was too large and the number of peaks was very high. This method was not 100% effective; however, it removed most of the abnormal regions. Categorization specified below was used to identify additional abnormal regions (anomalies).

Classification of the labeled regions as head, heavy tail or head with tail was necessary to identify the comets in the given image. For this purpose, four measures were computed for each labeled region. Two of these measures were the area and the width-to-height ratio of the bounding box containing the labeled region. Two other measures, r_o and r_i , were computed. The measure r_o was computed by flipping the left half of the region onto the right and finding the excess of the original region on the right half with respect to the flipped one. The measure r_i , on the other hand, was computed by flipping the left half of the region

onto the right and finding the excess of the flipped half with respect to the original right half. The four measures form a parameter space, and the space was split into the following six classes: head, small head, significant tail, heavy tail, head and tail, or anomalous.

Once the labeled regions were classified, the comets were reconstructed. First solitary heads and tails were considered and matched to see if they form a head-tail pair. The matching conditions were straightforward: the tail should be to the right of the head (in the direction of the current flow) for an unpaired head and the head to the left of the tail for the unpaired tail. The region to the right of the unpaired head was masked: 130 pixels to the right of the head were considered. This region was enhanced by thresholding the particular region in the original image with a low threshold level of 30. Post processing on the region was done to extract the signal from the noisy background.

For the unpaired tail, the region to the left was masked. Here again, re-thresholding at a lower threshold value was done and the density of signal (the sum of the signal intensities, in the bounding box placed left of the tail with a length of at most 65 pixels and a height equal to that of the tail, divided by the area of the bounding box) to the left of the tail was computed. If this density was greater than 1 (determined via trial and error runs), then further checking was done to ensure that the region does have a head. This involves identifying the peak of the head from the intensity vector. If the region about the peak was symmetric, then the region was classified as a head, and a mask, i.e. a new labeled region, which contains both the head and the tail was drawn. The labeled regions were thus classified as heads with tails, head only or tail only. The anomalous regions were neglected and masked out.

Following identification of the comets, the intensities for each comet were computed. Before the actual raw intensities were computed, the background intensities were computed locally for each region in the following manner. The mean intensities on the four sides of the labeled region viz., the top, bottom, right and left were computed. The median of these four background estimates was then taken as the background estimate for the labeled region.

The intensity estimation was done for the head and the tail separately. For the new labeled regions, which have been classified as head with tail or just head, the head intensity was calculated by creating a mask symmetric about the vertical axis passing through the center of the head. This mask was created by replacing the right half of the head mask by the left half. The original image under the mask was reconstructed and the pixel intensities along the vertical axis were summed up for each pixel along the length of the head. The rest of the region in the bounding box was considered as tail, and the intensities were integrated as for the head.

For the regions classified as not having a head or a very small head, the processing was different. The head mask was identified by defining a box around it and the intensities were integrated. The rest of the region was considered as tail,

and the intensities for the tail were integrated. Finally, the background intensities computed locally for each region were subtracted from the intensities computed above. Various measures could be computed from these raw intensities.

The above description of the algorithm includes several set parameters, like raising each pixel intensity to the power 1.05. Small subset of slides was used to help in the selection of these parameters. Through iteration, it was ensured that this small subset of slides was being processed correctly. Once this subset of slides was being processed correctly, the same parameters were used for the entire set of slides. The entire set of slides was processed only once. Since the parameters were the same for all the slides, this ensured that all of the slides were being processed in the same way, and by a known algorithm. By manually comparing the slides and the results obtained by this proposed algorithm, we know that 90% of the time the slides are being processed correctly. The reason for incorrect specification of the 10% of the slides is described in the discussion. Consequently, even though the proposed algorithm is not entirely objective, it is a set algorithm which ensures that all of the comet slides are processed using exactly the same criteria.

Figure 1 shows FLARE images isolated by the program. After the head and tail intensities were summed over in the vertical direction, an intensity histogram for the FLARE was created. It should be noted that summing intensities in the vertical direction is feasible since the electric current only has a left-right effect, so the vertical position of a stained particle does not provide more information about the damage in a comet. An example of such a histogram is shown in Figure 2. This histogram corresponds to the first comet in the second column of Figure 1. The horizontal axis is the number of pixels in the comet image. The vertical axis in Figure 2 corresponds to the intensity of the comet for a given distance in the horizontal direction. Note that for the analysis, the vertical axis will be standardized so that the area under the function will be one. The solid line is the tail, while the dashed line corresponds to the head of the comet.

The Matlab code used for processing the slides, along with a few example FLARE slides, is available at the following website: stat.tamu.edu/~mleyk in the "Comet" section.

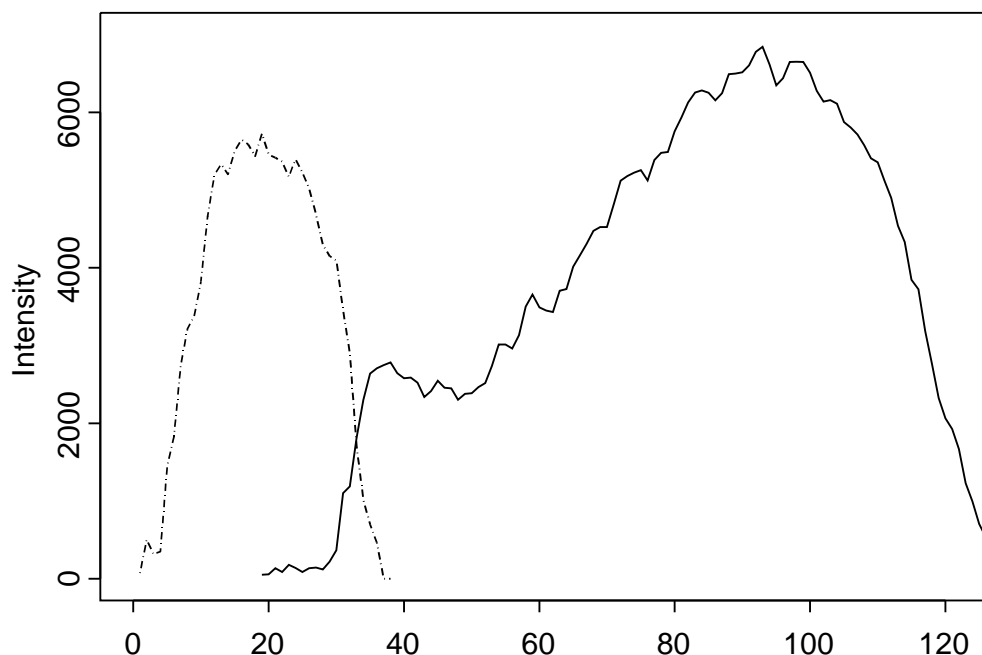


Figure 2: Example of intensity histogram for a comet, with head (dashed) and tail (solid) intensities.

New Statistical Methods

The previous statistical analysis was based on data where the relative tail moment (RTM) was computed for each cell, and then a summary measure of the RTMs for all cells was computed at the rat-level, see Figure 3. The summary measure was either the mean, median, 1st quartile, or 3rd quartile of RTM values. Since the quantity of 8OHdG is the variable of interest, the difference between the results with and without FPG needs to be calculated at the rat-location level. The statistical analysis is based upon two steps: (a) the measure of oxidative DNA damage computed for each cell, and (b) the aggregation of cell-level summaries to rat-level summary statistic.

First consider the measure of DNA damage. While RTM does contain information about the amount of oxidative damage, additional information about the level of damage is contained in the shape of the comet. When presented with an image of the comet, severity of damage can be easily assessed just by looking at the shape of the comet. By describing the shape, and not just means, the differences between comets should be more apparent than what is provided by

some of the previous estimators. In our work, in addition to RTM, we used skewness to test for differences in comet shape, which we denote by \tilde{u}_3 . Formally, this is defined as

$$\text{skewness} = \tilde{u}_3 = u_3 / (u_2)^{3/2}$$

where

$$u_i = \sum_j (x_j)^i \times (\text{proportion of intensity at point } x_j).$$

In the above equation, x_j is the value on the horizontal axis of the intensity histogram, with the center of the head of the comet being 0 on the horizontal axis. Note that to compute \tilde{u}_3 , the comet does not need to be separated into a head and a tail part.

The second part of the statistical analysis is based on aggregating the cell-level measurement (RTM or skewness) to form a summary measure for each rat. In addition to the mean and the median, we also used the first and third quartiles (25th and 75th percentiles, respectively). The mean and median are two ways of describing the distributions of the RTM or skewness. The median describes only what has happened in the middle of the distribution. However, the first and third quartiles describe what has happened to the left and the right of the center. These three are not necessarily dependent on each other. In particular, for the same median, there can be many possible values for the two quartiles. Therefore, additional information can be obtained about the distributions by examining the first and third quartiles.

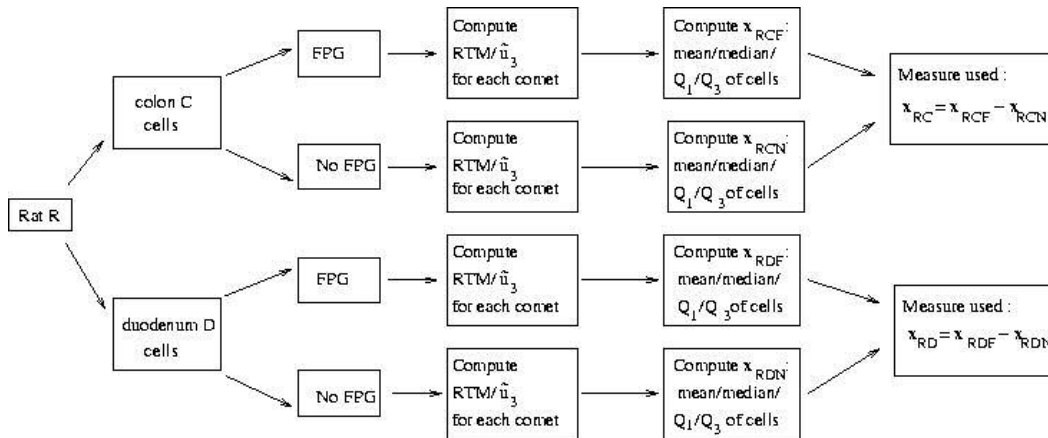


Figure 3: Representation of how the final measure of oxidative damage for each rat is computed in the new analysis.

It should be noted that the mean will depend on the median, first and third quartile values. If the distribution of RTM or skewness is skewed to the right, the location of the mean will be to the right of the median, and to the left of third quartile. Therefore, any significant results seen in the mean may be explained by what is happening with the quartiles.

RESULTS

As described above, the previous technique used a manual image analysis, computed the RTM for each comet, where each comet represents DNA damage in a single cell, and aggregated it to the rat-level by computing the mean, median, first and third quartile RTM over all cells in a rat. The level of 8OHdG was then assessed as the difference between the aggregated value for FPG and no-FPG cells. A graphical representation of this process is similar to the process given in Figure 3, except only RTM is computed for each comet.

With previous imaging and summary statistics, the only significant diet effect was found in the colon when aggregation was performed using the first quartile RTM (p-value of 0.044, see Table 1). Here, the corn oil diet estimate was higher than the fish oil diet estimate, indicating more oxidative damage to cells exposed to corn oil (Table 2). The only other significant effect was in the colon when aggregation was done using the third quartile (p-value of 0.030, Table 1). It indicated that administration of DSS resulted in a statistically significant increase in oxidative damage over control or DSS with a recovery period. The estimates for the statistically significant treatment and diet effects are given in Table 2.

In contrast with the previous imaging and statistical analysis, the new imaging and new statistical analysis showed stronger diet effects, especially in the duodenum.

In the new image analysis, we first considered RTM. Here we see a statistically significant result for diet in the duodenum when the summary measure aggregated for the rat is the third quartile of the RTM (p-value of 0.007, Table 1). The fish oil caused more oxidative damage than the corn oil (see Table 2). This finding suggests that diet influences the distribution of DNA damage, particularly in the amount of severe DNA damage. This is the only statistically significant effect found when RTM was examined.

With this background, we now consider the replacement of the RTM by the skewness (\tilde{u}_3). Because we conjecture that shape is critical in elucidating diet effects, we would expect to see diet effects to become more obvious. This happens in the duodenum since both the mean and the third quartile of the shapes (skewness) show highly statistically significant diet effects (Tables 2 and 3).

Table 1: Comparing p-values for RTM when the previous or new imaging methods were used

| Method | Location | Effect | Mean p- value | 1 st Quartile p-value | Median p-value | 3 rd Quartile p-value |
|----------|----------|----------------|---------------------|--|-------------------|--|
| Previous | Colon | Treatment | 0.266 | 0.991 | 0.613 | 0.030 |
| Previous | Colon | Diet | 0.119 | 0.044 | 0.151 | 0.690 |
| Previous | Colon | Diet*Treatment | 0.650 | 0.664 | 0.417 | 0.766 |
| Previous | Duodenum | Treatment | 0.337 | 0.383 | 0.470 | 0.555 |
| Previous | Duodenum | Diet | 0.805 | 0.869 | 0.059 | 0.806 |
| Previous | Duodenum | Diet*Treatment | 0.858 | 0.894 | 0.696 | 0.566 |
| New | Colon | Treatment | 0.608 | 0.855 | 0.626 | 0.090 |
| New | Colon | Diet | 0.961 | 0.583 | 0.587 | 0.778 |
| New | Colon | Diet*Treatment | 0.435 | 0.526 | 0.175 | 0.858 |
| New | Duodenum | Treatment | 0.474 | 0.159 | 0.919 | 0.724 |
| New | Duodenum | Diet | 0.275 | 0.396 | 0.721 | 0.007 |
| New | Duodenum | Diet*Treatment | 0.417 | 0.463 | 0.557 | 0.180 |

Note: there were 10 rat-level observations for each location (colon, duodenum), diet (fish oil, corn oil), and treatment (control, DSS, DSS with recovery), except for the corn oil with DSS treatment there were only 9 rat-level observations for both the colon and the duodenum. Each rat-level observation is usually based on at least 100 cell-level observations.

Table 2: Estimates of significant diet and treatment effects

| Imaging Method | RTM/skewness | Location | Diet/Treatment Effect | To Rat Level Aggregation Type | Estimate \pm Standard Error |
|----------------|--------------|----------|-----------------------|-------------------------------|--------------------------------|
| Previous | RTM | Colon | Control | Q ₃ | 2.44 \pm 3.14 ^{a1} |
| Previous | RTM | Colon | DSS | Q ₃ | 12.64 \pm 3.23 ^{a2} |
| Previous | RTM | Colon | Recovery | Q ₃ | 1.42 \pm 3.14 ^{a1} |
| Previous | RTM | Colon | Corn | Q ₁ | 8.79 \pm 2.48 ^{b1} |
| Previous | RTM | Colon | Fish | Q ₁ | 1.62 \pm 2.43 ^{b2} |
| New | RTM | Duodenum | Corn | Q ₃ | -1.03 \pm 1.06 ^{c1} |
| New | RTM | Duodenum | Fish | Q ₃ | 3.13 \pm 1.05 ^{c2} |
| New | skewness | Duodenum | Corn | Mean | -0.05 \pm 0.05 ^{d1} |
| New | skewness | Duodenum | Fish | Mean | -0.21 \pm 0.05 ^{d2} |
| New | skewness | Duodenum | Corn | Q ₃ | -0.13 \pm 0.09 ^{e1} |
| New | skewness | Duodenum | Fish | Q ₃ | -0.41 \pm 0.09 ^{e2} |

Effects with the same letter superscript (a, b, c, d, or e) are compared for statistical significance. Different superscript number (1, 2) following the same letter denotes statistically significant difference.

Table 3: P-values for skewness, when the new imaging methods are used

| Location | Effect | Mean p-value | 1 st Quartile p-value | Median p-value | 3 rd Quartile p-value |
|----------|----------------|--------------|----------------------------------|----------------|----------------------------------|
| Colon | Treatment | 0.441 | 0.928 | 0.852 | 0.254 |
| Colon | Diet | 0.385 | 0.205 | 0.202 | 0.915 |
| Colon | Diet*Treatment | 0.223 | 0.412 | 0.226 | 0.347 |
| Duodenum | Treatment | 0.412 | 0.244 | 0.671 | 0.519 |
| Duodenum | Diet | 0.024 | 0.482 | 0.159 | 0.027 |
| Duodenum | Diet*Treatment | 0.847 | 0.700 | 0.792 | 0.811 |

Note: there were 10 rat-level observations for each location (colon, duodenum), diet (fish oil, corn oil), and treatment (control, DSS, DSS with recovery), except for the corn oil with DSS treatment there were only 9 rat-level observations for both the colon and the duodenum. Each rat-level observation is usually based on at least 100 cell-level observations.

To explain the negative skewness results, Figure 4 will be used. It uses the intensity histograms for all comets. The center of the head was set as zero on the horizontal axis so that they would be aligned. The intensities for all the comets were standardized so that the area under the intensity histogram was one. Let the

intensity histograms of these comets be denoted by $f_{DRLF}(x_i)$, where D denotes the diet type, R the rat given diet D , L the location in the rat, F the FPG type, and C a comet obtained from cell in location L from rat R . Then for all the comets for rat R , location L , and FPG type F , the mean intensity for each horizontal axis location was found. In the above notation, this means that:

$$f_{DRLF}(x_i) = \sum_{C \in RL} f_{DRLFC}(x_i) / n_c$$

for each value of x_i . Above, n_c is the number of comets for rat R and location L . The average is then taken for all the rats with the same location and FPG type for each horizontal axis location:

$$f_{DLF}(x_i) = \sum_{R \in DL} f_{DRLF}(x_i) / n_r$$

for each value of x_i . Here, n_r is the number of rats with diet D and location L . From the resulting function $f_{DLF}(x_i)$, Figure 4 was created.

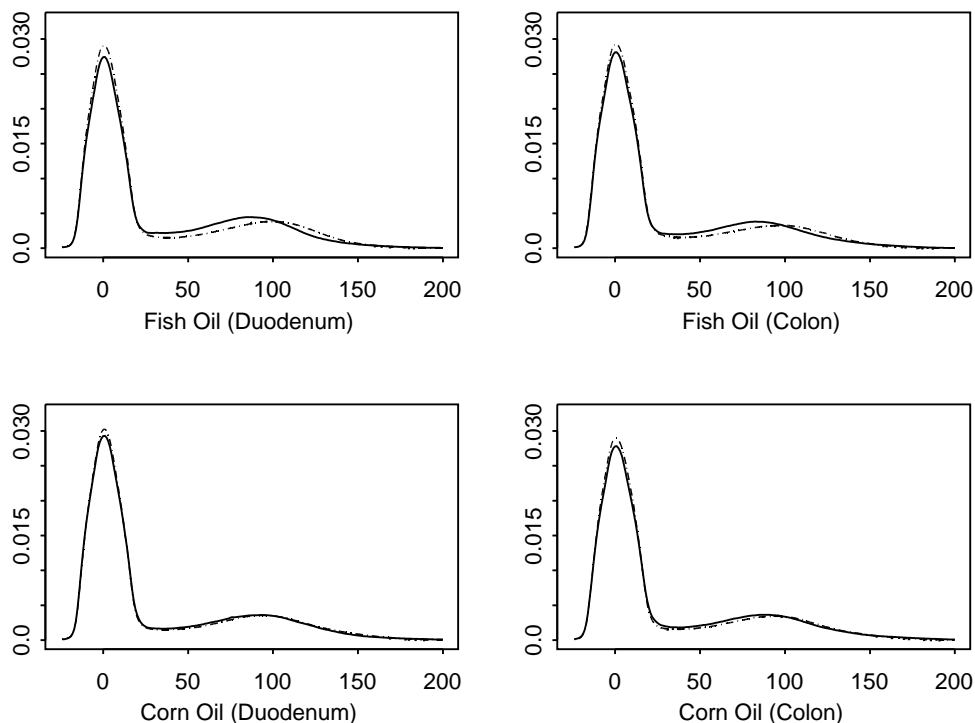


Figure 4: Representative curves for FPG (solid) and no FPG (dashed) in the duodenum and colon for rats fed a fish and corn oil diets.

DISCUSSION

The main advantage of automated image analysis is that it is faster than processing 35,000 comets manually. It is also more consistent in selecting and identifying comets. While the current image analysis is not perfect, it identifies parts of the head and tail the same way for every comet. There is no variability or bias due to a technician error. It is also reproducible, meaning that if the algorithm was performed again on the same FLARE images, it would produce the same results. This is not true with an analysis that involves extensive human intervention.

While the new image analysis approach is beneficial in processing large amounts of data, it also has a few drawbacks. The program processing the images finds it hard to distinguish a comet with a small head from clustering of the fluorescent material that is not a comet. This type of clustering seems to occur naturally, appearing as little specs on the image. One solution to this problem is not to consider “comets” that are smaller in height than a certain threshold. In this analysis, the threshold of 12 pixels was established by comparing images and their corresponding intensity histograms. Specs that were too small to be nuclei of cells were identified. The height of these specs was considered, based on which threshold was established. The comets below that threshold (2.4%) were not considered in the analysis.

Another problem that occurred, similar to the above problem, was that specs were identified as small heads, and a comet with a head right behind the spec was identified as all tail. This occurred in 4.4% of the comets, with too much of the comet being attributed as tail. While this type of problem could be fixed by adjusting the threshold for the smallest amount of intensity in an area that could be considered as a head, doing this would also eliminate from the analysis some comets with a substantial amount of damage. The last major problem with the analysis was when two comets were too close together. In 3.5% of the cases, the comets were overlapping enough to be identified as one comet.

Some of the problems mentioned above might be corrected by adjusting the thresholds (parameters) on the comet specification. However, this will also cause a certain percentage of good comets not to be processed because they did not meet the specifications. This will most likely hold true for comets that had a substantial amount of damage. Eliminating these types of comets would be counterproductive, so the thresholds on the comet specification were not adjusted further. The current imaging algorithm correctly processed 90% of the images, while at the same time yielding gains in processing speed, precision, and reproducibility.

Using the new image analysis, the diet effects turned out to be stronger than in the previous image analysis (see Bancroft et al., 2003). The two imaging techniques can be compared based on the results obtained for RTM values. The RTM p-value was more significant when the new imaging method was used rather than when the previous imaging method was used. However, the location of

the diet differences was not the same. In the previous imaging method, the marginally statistically significant diet effect was only seen in the colon. Using the new imaging method, the statistically significant diet effect was seen only in the duodenum, but not in the colon. Different results in the two methods must be due to difference in image processing since the same set of images was used for both procedures. We prefer the new imaging method since it is reproducible, and puts more emphasis on eliminating noise.

From this new data analysis it can be concluded that fish oil causes more oxidative damage than corn oil in the duodenum. No significant diet effects were found in the colon. While there is also a strong diet effect for the skewness (\tilde{u}_3) in the duodenum, the result is not as easily interpretable. The difference between corn and fish oil estimates is positive, but each of the individual diet estimates is negative (see Table 2). The corn oil diet estimate showed no statistical difference between data with and without FPG incubation, so these values could not be concluded to be different from zero (p-values 0.3576 and 0.1421 for mean and third quartile, respectively). However, for fish oil the difference in FPG types was negative (p-values less than 0.0001 for both mean and third quartile).

In order to explain the negative values obtained, Figure 4 was created. It shows the FPG (solid) and no FPG (dashed) for fish and corn oil diet in the duodenum and colon. The curves can be thought of as representative curves for the comets in the given location, diet, and FPG type groups. If \tilde{u}_3 is computed for both the fish FPG curves for duodenum in Figure 4, and then the difference is taken, this difference is negative. From this figure, it can be concluded that there is more damage when FPG is present since the level of intensity (vertical axis) in the head is lower for FPG, and thus higher intensity must be in the tail than when FPG is not present. Similar behavior can be seen for the two fish FPG curves in the colon, but to a lesser extent. For the corn oil curves, there is almost no difference between the FPG curves. To summarize, differences in FPG curves for fish oil are more pronounced than for corn oil. Also, the negative difference between skewness of the FPG curve and the no-FPG curve for fish oil implies higher level of damage for the FPG curve than the no-FPG curve, as is expected.

Consequently, the data suggests that fish oil causes more oxidative damage than corn oil. However, further studies on the colon from the same rats found that fish oil enhances apoptosis (see Bancroft et al., 2003). From the FLARE assay, it could not be determined which cells were in the process of apoptosis, and would consequently have more DNA fragmentation. However, if the cells exposed to fish oil and an oxidizing agent were more likely to undergo apoptosis, then this would imply that the larger amount of damage seen when fish oil is in the diet is not necessarily harmful. The cells that were in the process of apoptosis would show more damage, but they would not pose a risk of increased digestive tract cancer since they would be eliminated.

REFERENCES

- Bancroft, L.K., J.R. Lupton, L.A. Davidson, S.S. Taddeo, M.E. Murphy, R.J. Carroll, and R.S. Chapkin.** 2003. Dietary fish oil reduces oxidative DNA damage in rat colonocytes. *Free Radical Biology and Medicine* 35:149-159.
- Bartsch, H., J. Nair, and R.W. Owen.** 1999. Dietary polyunsaturated fatty acids and cancers of the breast and colorectum: emerging evidence for their role as risk modifiers. *Carcinogenesis* 20:2209-2218.
- Boiteux, S. and J.P. Radicella.** 1999. Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress. *Biochimie* 81:59-67.
- Collins, A.R., A.G. Ma, and S.J. Duthie.** 1995. The kinetics of repair of oxidative DNA damage (strand breaks and oxidised pyrimidines) in human cells. *Mutation Research* 336:69-77.
- Diggle, C.P.** 2002. In vitro studies on the relationship between polyunsaturated fatty acids and cancer: tumor or tissue specific effects? *Progress in Lipid Research* 41:240-253.
- Dougherty, E.R.** 1992. *An Introduction to Morphological Image Processing.* SPIE Press, Bellingham, WA.
- Dougherty, E.R. and J. Astola.** 1999. *Nonlinear Image Filtering.* SPIE and IEEE Presses, Bellingham, WA.
- Giardina, C.R. and E.R. Dougherty.** 1988. *Morphological Methods in Image and Signal Processing.* Prentice-Hall, Englewood Cliff, NJ.
- Halliwell, B.** 2002. Effect of diet on cancer development: is oxidative DNA damage a biomarker? *Free Radical Biology & Medicine* 32:968-974.
- Hellman, B., H. Vaghef, and B. Bostrom.** 1995. The concepts of tail moment and tail inertia in the single cell gel electrophoresis assay. *Mutation Research* 336:123-31.
- Hong, M.Y., R.S. Chapkin, R. Barhoumi, R.C. Burghardt, N.D. Turner, C.E. Henderson, L.M. Sanders, Y.Y. Fan, L.A. Davidson, M.E. Murphy, C.M. Spinka, R.J. Carroll, and J.R. Lupton.** 2002. Fish oil increases mitochondrial phospholipid unsaturation, upregulating reactive oxygen species and apoptosis in rat colonocytes. *Carcinogenesis* 23:1919-25.
- Jackson, A.L. and L.A. Loeb.** 2001. The contribution of endogenous sources of DNA damage to multiple mutations in cancer. *Mutation Research* 477:7-21.
- Loft, S. and H.E. Poulsen.** 1996. Cancer risk and oxidative DNA damage in man. *Journal of Molecular Medicine* 74:297-312.
- Morris, E.J., J.C. Dreixler, K.Y. Cheng, P.M. Wilson, R.M. Gin, and H.M. Geller.** 1999. Optimization of single-cell gel electrophoresis (SCGE) for quantitative analysis of neuronal DNA damage. *Biotechniques* 26:282-9.
- Riso, P., A. Santangelo, and M. Porrini.** 1999. The comet assay for the evaluation of the cell resistance to oxidative stress. *Nutrition Research* 19:325-333.
- Sanders, L.M., C.E. Henderson, M.Y. Hong, R. Barhoumi, R.C. Burghardt, C.M. Spinka, N. Wang, R.J. Carroll, N.D. Turner, R.S. Chapkin, and J.R.**

- Lupton.** 2002. Dietary fish oil and pectin protect against oxidative DNA damage in rat intestinal epithelial cells due to heightened apoptosis induced by reactive oxygen species. *FASEB Journal* 16:A371.
- Serra, J.** 1982. *Image Analysis and Mathematical Morphology*. Academic Press, New York, NY.
- Singh, N.P., M.T. McCoy, R.R. Tice, and E.L. Schneider.** 1988. A simple technique for quantitation of low levels of DNA damage in individual cells. *Experimental Cell Research* 175:184-191.
- Singh, N.P.** 2000. Microgels for estimation of DNA strand breaks, DNA protein crosslinks and apoptosis. *Mutation Research* 455:111-127.
- Sugimura, T.** 2000. Nutrition and dietary carcinogens. *Carcinogenesis* 21:387-395.
- Tardieu, D., J.P. Jaeg, J. Cadet, E. Embvani, D.E. Corpet, and C. Petit.** 1998. Dextran sulfate enhances the level of an oxidative DNA damage biomarker, 8-oxo-7,8-dihydro-2'-deoxyguanosine, in rat colonic mucosa. *Cancer Letters* 134:1-5.
- Tardieu, D., J.P. Jaeg, A. Deloly, D.E. Corpet, J. Cadet, and C.R. Petit.** 2000. The COX-2 inhibitor nimesulide suppresses superoxide and 8-hydroxy-deoxyguanosine formation, and stimulates apoptosis in mucosa during early colonic inflammation in rats. *Carcinogenesis* 21:973-976.