1 Title

2 Fatquant: An automated image analysis tool for identification of fat cells in heterogeneous
3 hematoxylin and eosin tissue sections.

4

5 Authors

6 Roshan Ratnakar Naik¹, Annie Rajan², Nehal Kalita³

7

8 Affiliations

9 ¹ Dr. Roshan Ratnakar Naik, Guest Faculty, Dnyanprassarak Mandal's College and Research
10 Centre, Assagao, Khorlim, Mapusa-Goa, 403507. ORCID ID: 0000-0003-4515-871X. Email
11 address: ros007swa@gmail.com

12

13 ² Mrs. Annie Rajan, Associate Professor, Department of Computer Science and Technology,
14 Dhempe College of Arts and Science, Miramar, Panaji-Goa, 403 001. ORCID ID: 0000-0001-
15 6201-2654. Email address: ann_raj_2000@yahoo.com

16

17 ³ Nehal Kalita, M.Sc. Computer Science. ORCID ID: 0000-0001-6997-5621. Email address:
18 nehalkalita94@gmail.com

19

20

21 Corresponding Author: Dr. Roshan R. Naik

22

23

24 **Author contributions**

25 R.R.N. conceptualized the idea and wrote the manuscript.

26 A.R. conceptualized the idea.

27 N.K. formulated the mathematical approach and wrote the Python programs.

28

29 **Abstract**

30 Hematoxylin and eosin (H and E) is one of the common histological staining techniques that
 31 provides information on the tissue cytoarchitecture. Adipose (fat) cells accumulation in pancreas
 32 has been shown to impact beta cell survival and its endocrine function. The current automated
 33 tools available for fat analysis are suited for white adipose tissue which is homogeneous and
 34 easier to segment unlike heterogeneous tissues such as pancreas where fat cells continue to play
 35 critical physiopathological functions. In the current study, we present an automated fat analysis
 36 tool, Fatquant, where mathematical formula to calculate diagonal of a square drawn inside circle
 37 is utilized for identification and analysis of fat cells in heterogeneous H and E tissue sections.
 38 Using histological images of pancreas from a publicly available database, we show an area
 39 accuracy overlap of 89-93% between manual versus automated algorithm based fat cell
 40 detection.

41

42

43 **Keywords**

44 Adipose or fat cells, Hematoxylin and Eosin, image processing, tissue segmentation.

45

46

47 **Author Summary**

48 We have developed an automated tool, Fatquant, for identification of fat cells in complex
49 hematoxylin and eosin tissue sections such as pancreas which can aid the pathologist for
50 detection and analysis. Fatquant is unique as current fat automated tools work well for
51 homogeneous white adipose tissue but not for other tissue samples.

52

53 **Introduction**

54 Hematoxylin and eosin (H & E) is a widely used histological tissue staining technique for
55 medical diagnosis and scientific research. Hematoxylin stains cell nuclei blue while eosin stains
56 the cytoplasm and connective tissue pink thus allowing microscopic differentiation of tissue
57 cytoarchitecture in sections. The analysis involves manual examination by pathologist to
58 ascertain presence/absence of disease markers and or grading of disease progression which is
59 semi-quantitative and subjective in nature [1]. Obesity is characterized by increased adipocyte or
60 fat cell numbers and enlargement (change in diameter) due to lipid accumulation leading to
61 increase in white adipose cellular mass [2, 3]. However, in recent years, the role of fat cells in
62 pathophysiology of liver, pancreatic and lung disease is increasingly recognized due to close
63 interaction with their corresponding endocrine cells [4-6]. To complement the current manual
64 assessment, several digital tools have been developed such as CellProfiler [7], Adiposoft [8], and
65 Adipocount [9]. However, they cannot efficiently identify fat cells in complex tissues such as
66 liver, pancreas, lung as white adipose tissue is generally homogeneous making tissue
67 segmentation relatively easier [10]. Usually, these algorithms involve splitting the image into
68 various colour channels with the red channel binarized using automatic thresholding method to
69 separate the bright pink fat areas from dark purple-bluish cell nuclei. Subsequently, a watershed

algorithm is applied to fillup missing fat cell membrane and improve cell count [8, 9]. The output of these processes include the labels and statistical analysis of individual cells. These tools have been successfully applied to white adipose tissue, however, other organs like liver, pancreas, lungs have been challenging due to heterogeneous cell types. Previously, we have used tile rendering approach for profile analysis from three dimensional spatial data [11] as it reduces time complexity by considering only a portion of an image at a time. In the current study, we have developed an automated tool, Fatquant, in which the fat cells were identified in processed images by calculating the diagonal of a square circumscribed by circle.

78

79 Results

The Table 1 provides fat cell area accuracy in percentage with yellow and blue denoting manual and algorithm based detection, respectively. While the green represents overlapping fat cell area accuracy.

83 **Table 1: Percent area accuracy of overlapping fat cells**

Figure	Threshold value	Min. diameter	Max. diameter	Area accuracy (%)
8 (c)	234	27	115	89.70
9 (c)	228	25	Null	93.04
10 (c)	228	29	Null	90.64

84

From Table 1, it can be noted that Figure 8 (c) and Figure 10 (c) show less accuracy as compared to Figure 9 (c) because fat cells in these two images have more overlaps with other artifacts or fat cells excluded when it touches the boundary.

Figure 12 (a)

Figure 12(b)

Figure 12 (a) and Figure 12 (b) are Adiposoft and Adipocount outputs, respectively obtained with pancreas H and E stained image. Note, the corresponding output obtained for same image with Fatquant tool in Figure 9 (b). AdipoCount does not require any input parameter. Parameters for Adiposoft used were minimum diameter (25) and maximum diameter (3000).

Discussion

We have developed a tool, Fatquant, to identify fat cells in heterogeneous histological tissue sections that can complement work of pathologist in identification and analysis of fat cells due to their growing importance in pathophysiological functions and with the availability of whole slide digitized images it can save significant amount of time. We observed an overlapping area accuracy ranging between 89- 93% in manual versus automated Fatquant based detection indicating good concordance with variation due to possible artifacts during tissue processing, or peripheral cells excluded due to incomplete cells. The 10% loss in accuracy is mostly because of two reasons. One reason is valid fat cells getting overlapped with other white coloured tissue material or the background which adds extra area to detected fat cells. Another reason is some fat cells having nucleus or other dark coloured tissue inside them which restricts fat cells getting detected as per minimum diameter.

116 **Materials and Methods**

117 The H and E images for the analysis were downloaded from a publicly accessible Genotype-
 118 Tissue Expression (GTEx) Portal (Broad Institute, Cambridge, MA, USA) using the Histology
 119 Viewer tool [12]. The GTEx tissue image library contains high-resolution histology images for
 120 various tissue types from several postmortem donors. Spherical or oval white spaces were
 121 categorized as fat (adipose) cells while large and irregular white spaces were grouped as
 122 artifacts. A sample size of three high resolution pancreas histology images were analyzed with
 123 sample IDs GTEx-11DXZ-0826, GTEx-1122O-0726, GTEx-1117F-1726. The images have
 124 magnification of 20x (0.4942 mpp—microns per pixel) [13]. The source code was written in
 125 Python version 3.8.0 (Python Software Foundation, Beaverton, Oregon, USA) with Core i7 3rd
 126 Generation CPU, 8 GB RAM and Intel HD graphics 4000 GPU.

127

128 **Image processing**

129 The following steps were utilized for image analysis:

- 130 1. Binary thresholding of input image.
- 131 2. Segmentation of white pixels from thresholded image.
- 132 3. Selection of segments for fats as per range of diameter.
- 133 4. Removal of fats from boundary.
- 134 5. Analysis.

135

136

137

138 These steps are described as follows:

139 1. Binary thresholding of input image

141 **Figure 1:** a) Input image, b) binary thresholded image.

142 Binary thresholding is performed on input image (**Figure 1a**) by referring the grayscale values of
 143 each pixel. The system relies on a parameter value which ranges from 0 to 255 to identify the
 144 limit of each color. Pixels with grayscale values below the parameter are considered as black and
 145 above the parameter as white with the binary thresholded image (**Figure 1b**) having parameter
 146 value set as 230.

147 2. Segmentation of white pixels from thresholded image

148 White pixels are initially segmented by combining tile rendering with scanline rendering and
 149 then identifying possible merge of segments in a tile with their immediate neighbours. The input
 150 image is divided into multiple square tiles each with length 100 pixels and then scanline
 151 rendering is performed within each tile to identify white pixels and form possible segments with
 152 their neighbours on left or top (**Figure 2a, 2b**). Once scanline rendering is performed till the last
 153 row of a tile then the identified segments are merged on the basis of their horizontal and vertical
 154 neighbors (**Figure 2c**).

Figure 2 (a)

Figure 2 (b)

Figure 2 (c)

Figure 2: Segmentation of white pixels in thresholded image with possible segments with neighbours on (a) left, (b) top, (c) identified segments merged.

Segments which have only diagonal neighbours with another segment and not horizontal or vertical, are not merged. **Figure 3** and **4** shows segments within tiles and fully merged segments of the thresholded image respectively. Varying colours in the images represent different segments.

Figure 3: Segments within tiles.

Figure 4: fully merged segments of thresholded

images with variegated colours as different segments.

3. Selection of segments for fats as per range of diameter

White segments are considered as probable fats if a square kernel of diagonal length equal to length of assigned diameter value fits anywhere within their boundary. This square can be

assumed as the largest square that can get inscribed in a circle of input diameter values. Side length of a square, can be calculated as:

$s = \frac{d}{\sqrt{2}}$, where s = square side length; d = diameter. The system makes use of minimum and maximum diameter values to select fats. The system first refers minimum value to select segments whose area need not be narrow everywhere (**Figure 5a**) and then refers maximum value to discard the segments of the selected list having a very wide area (**Figure 5b**).

Figure 5: Selection of fat using a) minimum and b) maximum diameter values.

Before traversing by a square kernel to find probable valid segments, the system determines the smallest dimension that encompasses all of the image's segments (**Figure 6a, 6b**). Because the kernel's position is only updated by one pixel (horizontally or vertically) every iteration, determining the smallest dimension for traversal reduces the time complexity.

Figure 6a

Figure 6 b

Figure 6: a) and b) Determination of smallest dimension that covers all image segments.

4. Removal of fats from boundary

Segments which are identified as fats but touch pixels in boundary are discarded because their entire size is not known within the dimension of input image (**Figure 7**). Some of these segments may even get discarded while selecting segments as per diameter.

Figure 7: Fat cell removed from peripheral boundary.

5. Analysis

The machine annotated fat segments were compared with human annotated fat segments to check the validity of our image analysis algorithm. If machine and human annotated segments have pixels in common then, those pixels are considered as valid. The accuracy of the output in terms of area covered is calculated as:

$$AC = \frac{Av}{(Av + (Ah - Av) + (Am - Av))}$$

, where AC = valid area accuracy; Av = total valid area; Ah = total human annotated area; Am = total machine annotated area.

Acknowledgments

There is no potential conflicts of interest and no funding to declare. The datasets generated are not publicly available as they are part of ongoing work to identify more cell types in complex H and E histology sections. The source code can be made available based upon reasonable request to the authors.

References

- [1] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: a review. *IEEE Rev Biomed Eng*, 2, 147-171. doi: 10.1109/RBME.2009.2034865.
- [2] Jo, J., Gavrilova, O., Pack, S., Jou, W., Mullen, S., Sumner, A. E., . . . Periwal, V. (2009). Hypertrophy and/or Hyperplasia: Dynamics of Adipose Tissue Growth. *PLoS Comput Biol*, 5(3), e1000324. doi: 10.1371/journal.pcbi.1000324.
- [3] Longo, M., Zatterale, F., Naderi, J., Parrillo, L., Formisano, P., Raciti, G. A., . . . Miele, C. (2019). Adipose Tissue Dysfunction as Determinant of Obesity-Associated Metabolic Complications. *Int J Mol Sci*, 20(9). doi: 10.3390/ijms20092358.
- [4] Ntikoudi, E., Kiagia, M., Boura, P., & Syrigos, K. N. (2014). Hormones of adipose tissue and their biologic role in lung cancer. *Cancer Treat Rev*, 40(1), 22-30. doi: 10.1016/j.ctrv.2013.06.005
- [5] Pacifico, L., Di Martino, M., Anania, C., Andreoli, G. M., Bezzi, M., Catalano, C., & Chiesa, C. (2015). Pancreatic fat and beta-cell function in overweight/obese children with nonalcoholic fatty liver disease. *World J Gastroenterol*, 21(15), 4688-4695. doi: 10.3748/wjg.v21.i15.4688

219 [6] Gerst, F., Wagner, R., Oquendo, M. B., Siegel-Axel, D., Fritsche, A., Heni, M., . . . Ullrich,
220 S. (2019). What role do fat cells play in pancreatic tissue? *Mol Metab*, 25, 1-10. doi:
221 10.1016/j.molmet.2019.05.001.

222 [7] Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., . . .
223 Sabatini, D. M. (2006). CellProfiler: image analysis software for identifying and
224 quantifying cell phenotypes. *Genome Biol*, 7(10), R100. doi: 10.1186/gb-2006-7-10-r100.
225

226 [8] Galarraga, M., Campion, J., Munoz-Barrutia, A., Boque, N., Moreno, H., Martinez, J. A., . . .
227 Ortiz-de-Solorzano, C. (2012). Adiposoft: automated software for the analysis of white
228 adipose tissue cellularity in histological sections. *J Lipid Res*, 53(12), 2791-2796. doi:
229 10.1194/jlr.D023788.
230

231 [9] Zhi, X., Wang, J., Lu, P., Jia, J., Shen, H. B., & Ning, G. (2018). AdipoCount: A New
232 Software for Automatic Adipocyte Counting. *Front Physiol*, 9, 85. doi:
233 10.3389/fphys.2018.00085.

234 [10] Glastonbury, C. A., Pulit, S. L., Honecker, J., Censin, J. C., Laber, S., Yaghootkar, H., . . .
235 Lindgren, C. M. (2020). Machine Learning based histology phenotyping to investigate
236 the epidemiologic and genetic basis of adipocyte morphology and cardiometabolic traits.
237 *PLoS Comput Biol*, 16(8), e1008044. doi: 10.1371/journal.pcbi.1008044.

238 [11] Maurya, R.K., Kulkarni, S.T., & Kalita, N. (2018). Projection-Based Spatial Morphology
239 for Extracting Ridge and Valley Profiles of Mountains from 3D Amorphous Data. *2018*
240 *4th International Conference on Computing Communication and Automation (ICCCA)*,
241 1-6.

- 242 [12] Consortium, G. TEx. (2015). Human genomics. The Genotype-Tissue Expression (GTEx)
 243 pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235), 648-660. doi:
 244 10.1126/science.1262110.
- 245 [13] Badea, L., & Stanescu, E. (2020). Identifying transcriptomic correlates of histology using
 246 deep learning. *PLoS One*, 15(11), e0242858. doi: 10.1371/journal.pone.0242858.