# Fighting breast cancer with mathematics

# Instructor's notes for "Automatic alignment"

This exercise is a big simplification from the original approach, but can be scaled up in complexity if desired. The point is to add a bit of quantitative reasoning in what to many students might be a different kind of context... we aren't solving an equation to get an answer, but making our own equation that allows us to describe something of interest, in the service of automation and saving our own time.

**Basic example:**

The example asks students to compare three possible alignments with a base image. A simple way of quantifying difference in pixel values between subsequent slices is defined, and students are asked to practice this. There is an excel file [**here**](http://www.radford.edu/~jmwojdak/AIMS_Cancer_InstructorResources/) with the values for each pixel in each image, and the sums of squared deviations. The images score a 126, 12, and 108, respectively, left to right. Notice that none matches perfectly, as the new slice is not exactly the same as the previous, just like in real life; the tissue is not perfectly uniform in the vertical direction. (Also available at that link are files for each of the gridded slice examples, if you wanted to make handouts or transparencies.)

An equation for this process might look like:

where *v* is the value of intensity in the base image and *u* is the value of intensity in the second slice, *x* and *y* are coordinates, *i* and *j* are subscripts for rows and columns...

The very general bit of math used here is a sum of squared deviations - it is probably helpful to point out to students that have had some statistics that this is a very familiar calculation. Moreover, while it may seem like a different context, we are using it for exactly the same conceptual reason as in ordinary statistics - quantifying differences or matching.

The last question asks students to think through the consequences of registering with full color images or after they were made binary black and white images. There is no "right" answer per se here, but rather just an interesting thought question to get students thinking through what is happening inside the software.

If not used on a handout, this could be a good discussion question to spur students to consider whether this protocol is optimized... many labs will have step-by-step recipes that seem like they HAVE to be in that order. Here that is not the case. There are lots of other possible protocols, and considering why one is good or another is better is worthwhile. Another consideration is that one step affects the other... if you register first, with full color images, it will introduce black patches to image edges where there is no image information (think of a corner where the image is rotated out of, and the software has no information to fill in.. it just puts black). Now, if you go from this step to auto brightness/contrasting, the black will register, and the images won't be effectively equalized.

**Extensions**

One complication we skipped over is that Kerri and colleagues also normalized across differences in overall brightness/intensity between images (think of one image being more exposed, in the photographic sense, than another). That could be a "challenge" problem for a homework assignment, after students finished the core exercise equation and algorithm defined. It *could* be done with the simple gridded examples here, or presented as a "write a pseudo-code" algorithm to do it for real images (e.g., take the average intensity across the first slice, again for the second slice, find the difference, and adjust the intensity of all pixels in the second slice to achieve equal average intensities.)

Here, we also skipped over rotational alignment. This requires calculating the centroid, and that is interesting math in itself, and is related to the center of mass calculations used in physics. Refer to [Norton et al. (2012)](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0044011) for details. Again, showing how the same math can be used in many different contexts is often powerful to students - they often feel like they are more likely to use something, and therefore that it is more valuable, if the uses are manifold.

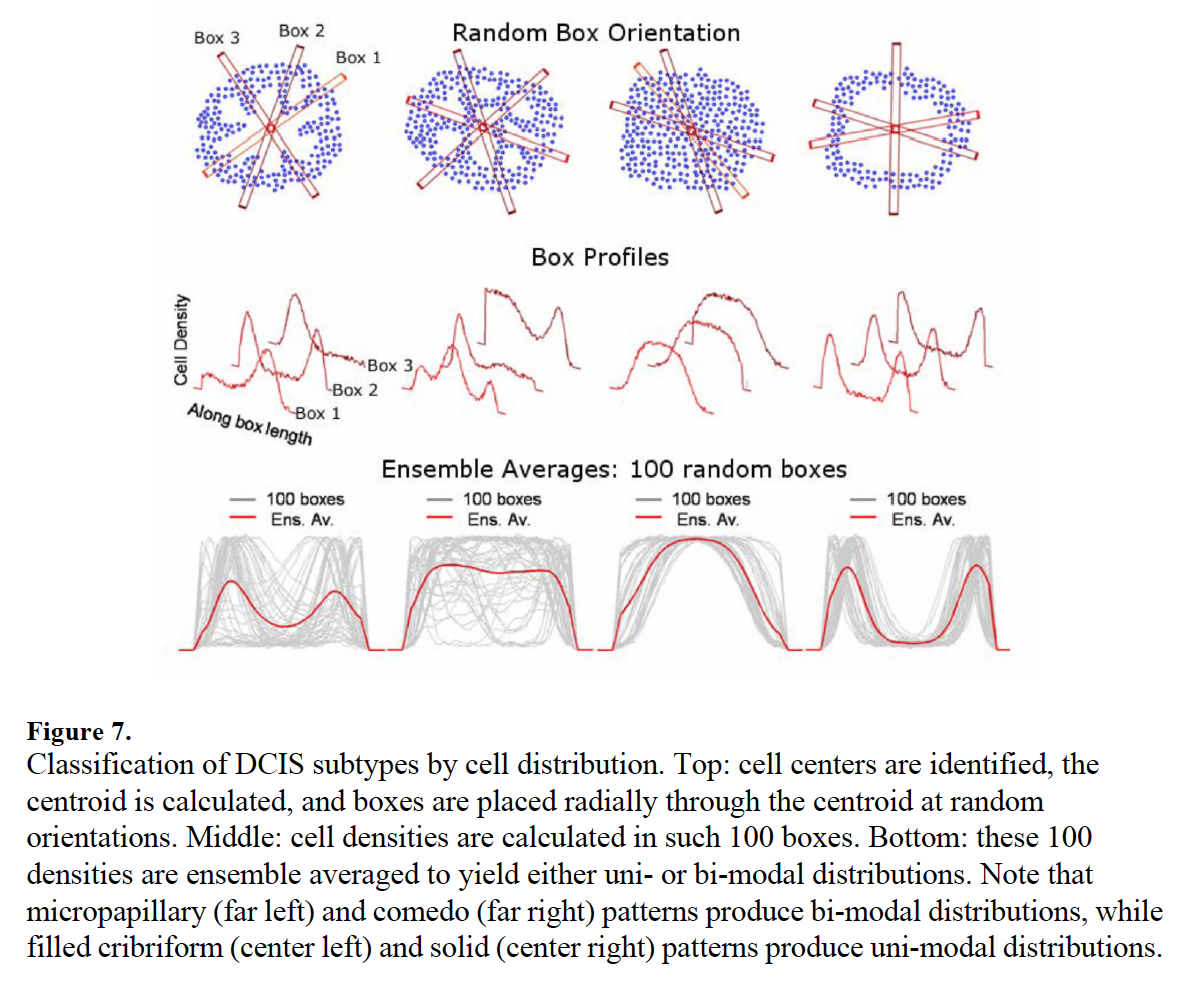
For students familiar with a bit of programming, there are many possible extensions. One could have students write Matlab/R/ImageJ macros or scripts to mimic the automation process used here, in whole or in part. For example, one could have students import the matrix of first and second slice pixel intensity values, then write a loop that translates (or even rotates) the values around a plausible range, calculating the sum of squared deviations at each position. (e.g., start with a given position, calculate the difference compared to the base image, shift position to right, re-calculate, etc. looking for the minimum). Starting with binary slices might make things easier.

# Instructor's guidance for "Quantifying complex spatial arrangements, simply "

Below is Figure 7 from [Norton et al. 2012](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0044011). It shows examples of the four main DCIS tissue types and the process described in the student exercise applied to each.

Things to watch out for: students may say a healthy duct, a comedo duct, and a micropapillary duct are all "bimodally" distributed, which is good... but there should be noticeable quantitative differences... the cell density in the interior bins should be lowest in a healthy duct, then perhaps higher for the comedo, then perhaps some intermediate, positive, low number of cells in the middle for the micropapillary. The shapes of the distributions should be different, even if of the same overall form.

You could include a random generator for the angle of the diameter placement, or pose this as a question to students. Clearly with some of the tissue types, you will get quite different profiles depending on arrangement of the box. Another "thought" question for discussion might be to ask what effect the degree of roundness/ovalness will have on the frequency histogram's shape.



# Instructor Guidance for "Evaluating our automated segmentation algorithm"

This is a great quantitative lesson for students - the concepts of sensitivity and specificity (or precision and recall, as defined here) are broadly applicable, a bit counter-intuitive, and very important. Many instructors are probably familiar with the very strange results you get when considering medical testing - a test with error rates of 1% false positives and 1% false negatives, for a rare disease (say, 1 in 10,000 people have the disease), given to a population of one million people, will produce:

10,000 false positive test results! (lots of worried people, for no reason!)

Only 1 false negative test result (despite the error rates "looking" the same in both directions!)

...and given a positive test result, the probability that you have the disease is 1%!

Besides medical testing, the same mathematics is used in many fields to assess procedures. This is an excellent chance to work on our students' understanding of probability.

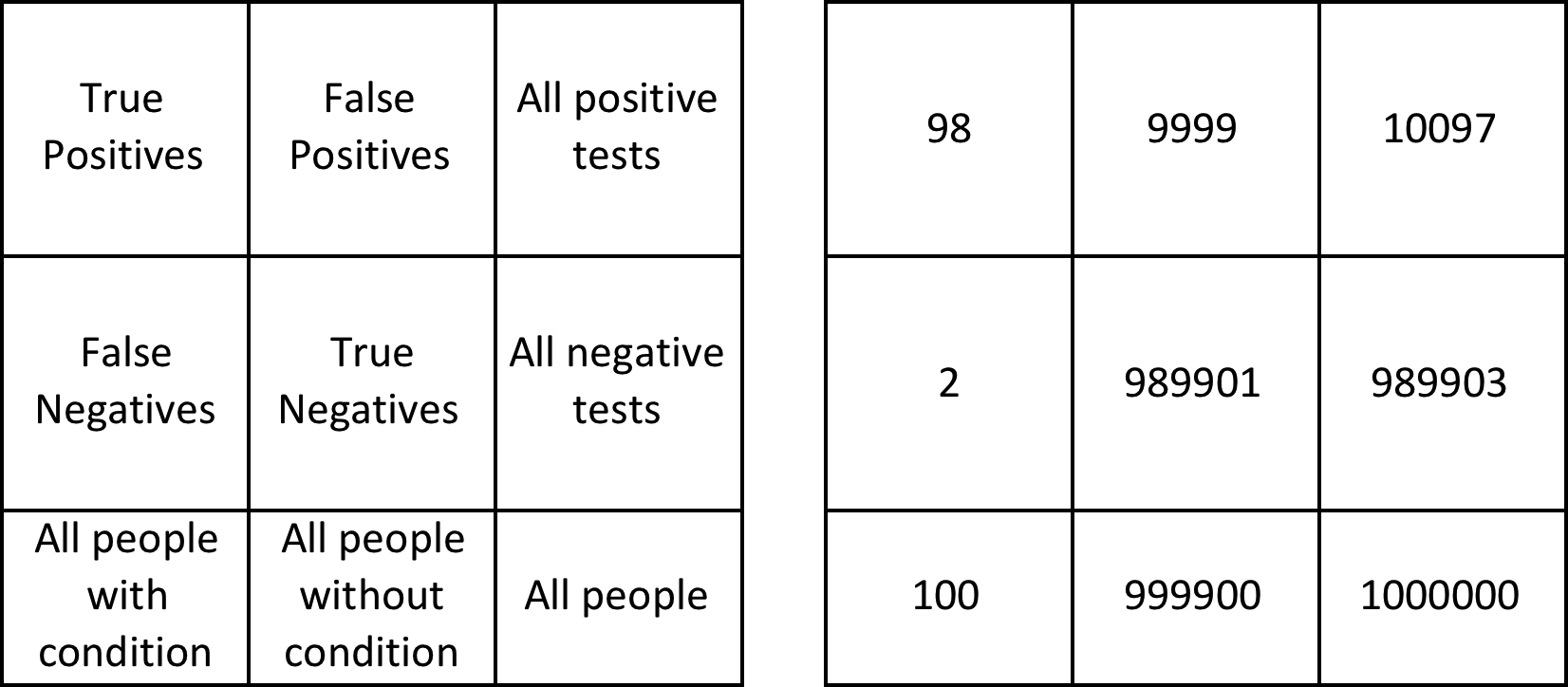
The steps to get the number of false positive, false negative, true positive, etc. pixels is a bit convoluted, and I would suspect many students will lose the plot along the way. Reference back to the visual circle example might help. Going through the whole process with very simplified images (try these here) can help too... when it is easy to see the "subtraction" steps of comparing two images for differences, the process may be less mysterious.

When students get their values for precision and recall, they are asked to compare against Norton et al. 2012's values. Kerri automated this procedure so she could look at the distribution (and central tendency) of precision/recall across lots and lots of image slices (rather than just comparing a single slice as we do here). Students may get confused by the mention of mean and median in Kerri's language.

Given the values for FP, TP, and FN, the value for TN can be obtained, and the more commonly used specificity metric could be calculated, rather than precision. Good ole' Wikipedia has a [good primer](https://en.wikipedia.org/wiki/Sensitivity_and_specificity) on sensitivity and specificity, and the numerous other metrics one is likely to see. These concepts are taught to many medical professionals, so they should be both relevant and interesting to many biology students.

**Problem set answers:**

1. **This will depend on the image selected, the quality of the work, and the algorithm itself. We might expect to not do quite as well as Kerri did... but close?**
2. If a test always gave a positive result, it would be perfect at diagnosing people with the disease as positive (100% sensitive). Likewise, if a test always gave a negative result, it would be perfect at diagnosing people without the disease as negative ( 100% specific).
3. The answer is "c", not very specific. You would like it if some people without the condition accidentally got a false positive. You hope your positive result was a false positive.
4. The answer is "a", very specific. A highly specific test correctly identifies people without the condition as negatives. you hope your negative is a true negative.
5. I would start with a grid of all the possible scenarios... TP, TN, FP, FN, and the row and column sums. If there are 1,000,000 people and 0.01% of them have the disease, we know that there are 100 people with the disease, and 999900 without. Of those 100, 98 should be correctly identified as such (true positives), given the 98% sensitivity. Hence, 2 should be false negatives. Again, if 100 people have the disease, 999900 don't. Of those, 99% should be correctly diagnosed as not having the disease, given the specificity of 99%. The remainder should be false positives.



Note here that of those with a positive test result, the VAST majority will not have the disease. Thus, you should most likely shrug off a positive test result!

**Extensions**: As in many other parts of the module, a group of students who are adept at programming could automate this part of the process too. In addition to looking at precision/recall across many or all the image slices, one could look for the minimum and maximum values, and see what conditions, image qualities, etc. affect accuracy.

Besides sensitivity/recall, specificity, and precision, there are lots of other metrics, concepts, and vocabulary to cover here... (e.g, odds ratios, likelihood ratios, false discovery rates, type-I and type-II errors). Connecting these concepts here in an information retrieval context to medical diagnoses and hypothesis testing frameworks could prove useful (or, if it doesn't go well, confusing :-) ). Even having students think about why information folks use precision and recall, while medical folks use sensitivity and specificity (among others), could be an interesting "thought question"... my guess is it depends on the relative importance of population prevalence, and the number of true negatives (information retrieval fields have lots of true negatives... here, pixels that aren't cribra).