

## Laboratory Worksheet, Monday, Sept. 19.

### A. Generalities, Monday, Sept. 19.

**I. Rooms and times.** We have switched the lab time to Mondays, 2-3:30 PM, in the Bioinformatics Student Lab (2036 Palmer Common). We also have available to us the labs of the Mathematics Department in the basement of East Hall. These are Mac/Unix labs, the Palmer Commons lab is Linux/Unix for now. These two facilities are on two different local networks, and because of limits of site licences, they are NOT equivalent. The math labs have licences, not surprisingly, for all the basic computational packages (Mathematica, Matlab, Maple, etc.). Both have various basic BI tools and modules, though I have to check up on the math labs to make sure the requests sometimes haven't expired. This usually applies to Perl modules and molecular visualization programs, such as Rasmol.

Since these are two local networks, you will have to set up two different accounts. This is relatively easy for the math labs, since there are a lot of people around to assist you, even after hours, and because we will set those accounts up in class. (You probably have already done this, if you are reading this document.) We will discuss in class about how to get the accounts in PC set up, when that becomes clear. For now you will all be sharing one account on the white board in 2036 PC.

**II. 548 Lab Worksheets.** This is the webpage which led you to this document: <http://www.math.lsa.umich.edu/dburns/547/548labworksheets.html>.

Ideally, meaning assuming I get the necessary time, there will be a worksheet for each lab session detailing what is expected for that day. This usually means "for that week" since, like most labs, we get things started but often not finished in the ninety minutes. This may change this term, depending on your computing skills. That is, we may make the lab more independent in its workings, setting you off to work in teams on longer term projects, in which case the labs would become team meetings primarily to compare notes on the progress of your projects. That is for later, however.

There is also a 548 resource directory which should be accessible from the main page sidebar. The URL is <http://www.math.lsa.umich.edu/dburns/548/>.

**III. Perl.** Perl is a high level scripting language which is widely used for data handling in this field. It is relatively easy to get started in it. If you are unfamiliar with it, there is a tutorial available at <http://www.math.lsa.umich.edu/dburns/548/perl tut2003/>. This is a tutorial in four lessons reproduced from summer 2003's BI shortcourses offered here at UM. There is also a zipped file of these lectures for convenient download to your personal computing directory: `perlfreight.tar.gz`. There are also annotated, illustrative examples of Perl routines from the book *Beginning Perl for Bioinformatics*, by James Tisdall (O'Reilly). This is a well-written book, and recommended if you have little computing experience. If you have experience, you may find *Learning Perl* by Schwartz and Phoenix (O'Reilly) more to your taste and speed. The illustration scripts are in Begin-

PerlBioinfo.pm, a ready to use module, or file of scripts.

While gearing up for our Perl exercises, we will need to streamline programs by using sub-routines. There is an open library of Perl scripts, and you should look at the archive of Perl modules available at the **Comprehensive Perl Archive Network** (<http://www.cpan.org/>). Look especially at the BioPerl package of modules. The web resources page

<http://www.math.lsa.umich.edu/dburns/547/547webres.html>

now has a link to the search page for CPAN. You should take a couple of minutes to explore this feature.

The 548 resources directory also contains a directory bioperl-1.2/ containing these files. I have zipped these up for convenience. The file name is “bperlfreight.tar.gz”.

We will have to discuss whether separate meeting times will be required for a startup in Perl, or any of the other languages or programs we may come across during the term. At the moment, these will be Perl, Matlab and R, a statistical package. If you refer using a more advanced language (Perl is largely for historical reasons, and ease of start-up for the unfamiliar), speak to us about using it.

**IV. Background to an Exercise.** There are simple but illustrative calculations behind a paper of Karlin and Mrázek on predicting highly expressed genes. The paper is available from the 548 resource directory, where it is listed under KMjbac.pdf. You should begin having a look at this paper, especially the first two or three pages.

**V. Related Seminars.** There is a high level seminar tomorrow in the biological chemistry department on chromatin structure. That will feature at the outer edge of the course late this term, where we will get to open problems. Here is the description: “De novo Formation of Microdomains of Heterochromatin and Regulation of Gene Expression by KRAB Zinc Finger Proteins”, by Dr. David C. Schultz, Case Western Reserve University. It is at noon, tomorrow, Tuesday, September 20, 2005 in Room 6311, Medical Science I Building.

## **B. Specifics, Monday, Sept. 19.**

**Guest Lecturer: Richard McEachin, MCBI.**

**VI. CFTR: Review of This Morning’s Lecture.** For this part of the lab I would like you to go to NCBI and do a brief write-up about CFR. You should answer the following questions: what kind of protein is CFTR? What ion’s transport is it related to? What disease is it related to? What is the disease state of the protein (what does that mean)? Parse the structure of the gene: that is: how many exons does it have? How large are the intronic regions? Where are the nucleotides which code for the trans-membrane alpha helices? (We will have to go to SwissProt for this.)

Next, let us look up the family of proteins GPCR. What does this stand for? Where do these proteins live in the cell? What function do they serve? What is a transduction pathway? What is a kinase protein? Is there a single function for GPCR’s? Can you follow the signal pathway starting from one GPCR? This may not all get done during the lab period today. Save your work and finish this over the course of the next week. Due September 26.