

THIS IS A README FILE BELONGING TO THE SOFTWARE OF THE MANUSCRIPT:

"Likelihood ratio test for differentially expressed pathways"  
David Bakewell and Ernst Wit

There are two *main* Matlab 6 programs for evaluating P-values in our Bioinformatics submission. These are contained in the zipped folder “**pathprog**”. The first program is called “**moca.m**” for **MO**n**te CA**rlo simulation for generating ROCs Fig. 2 (a) – (d). Corresponding histograms and scatterplots of P-values are also generated, the histograms can be seen in Supplementary data.

The second Matlab program is called “**cadat.m**”, an acronym for **CA**n**c**er DA**t**a****. It uses breast cancer data data from West et al (2005) to calculate the likelihood ratio test p-value on the pathway identified by Akhurst and Derynck (2001). The results are plotted in Fig. 3. All programs were written in Matlab 6.

Both programs share a common set of program “subroutines” and functions. These are also included in the folder, and their nested relationship is on the last page of these notes. A number of in programs in the *pathprog* folder have prefix, “**nigml**”. This because the original name for the first main program was called *nigml*: **N** independent samples & **m** Independent **Genes** **Maximum Likelihood**. Other names of programs relate to mathematical modelling & are of no consequence.

The software has been written in a way that assumes the data input and output folders are located in folder *pathprog* on your desktop. **Therefore, please place folder *pathprog* on your desktop.** To tell *moca* and *cadat* the root directory suitable for your computer you will need to change the address, in program “**filoc.m**”. Currently the code appears as “c:/Documents and Settings/Dave/Desktop/pathprog/”.

#### **A. Simulation: moca**

To start type the main program name “moca” into the Matlab command window. The program is menu driven but for simplicity for first-users practically all menu options have been commented out. The menu options include t-test (regularised or normal), sample variance options (pairwise or non-pairwise NPW which is default), multiple t-test correction (Sidak, Hochberg or Bonferroni), Bartlett small sample correction for the LR test, etc.

For first-user convenience/speed the Monte Carlo runs have been set to 30 rather than 300 so the ROCs will be more coarse than shown in the journal paper.

#### **B. Experimental: cadat**

To start type the main program name “cadat” into the Matlab command window. Again, program is menu driven but for simplicity for first-users practically all menu options have been commented out.

For first-user convenience/speed, the sampling for the P-value combinations has been chosen to be 30 rather than 10 – used for the paper.

PTO

List of programs and functions (bulleted) called by *moca* and *cadat*. Programs are self documented.

Moca-nigm (main program)

- xyij\_setup
  - nigmTij\_Eval
- Skpa\_k\_jH0H1
- nigpreNRcp
  - cp\_nigllh
    - deltC
    - p\_xy\_delta
      - ❖ f\_delta
    - f\_delta
  - deltC (p\_xy\_deltaA, f\_delta)
  - pre\_delt\_NR (p\_xy\_deltaA, f\_delta)
- p\_xy\_deltaC
  - p\_xy\_delta0\_vec
- negLp\_xy
- Lp\_xy
- delt\_NR
  - g\_f\_delta\_Nkpa
  - Dg\_f\_Ddelta\_Npka
- deltaNRmpxy
- nigmlplot
- nigmlroc\_hbs

cadat nigm (main program)

- nigmTij\_Eval
  - Skpa\_k\_jH0H1
- etc