Motivation

Genetical genomics provided world of life science with powerfull tools for revealing secrets underlying DNA to phenotype interactions. There is however a trap in using GG, because it’s main source of power is also its weakest point – obtaining GG data means conducting two separated experiments for every object – genotype and phenotype study. Not only it means doubling efforts and costs, but is a major source of mistakes. Latest reviews suggests, that in mice, almost 40% of probes may be wrongly matched. This number is shocking and may introduce doubts in practically every research using GG in mice. Obviously, this doesn’t prove GG is useless. Rather shows how extremely powerful technique could be easily affected by simple human mistakes.

Although our possibilities in managing huge amounts of data are insufficient and microarray researches are expensive, there is constant tendency to conduct researches screening whole genome even when inspecting only few genes. This leads to dropping most of the data and putting them into public databases as useless for research during which it was obtained.

Abovementioned problems lead to proposing novel methodology in dealing with microarray data. GG could be done using only a phenotypic data. The only requirement is possessing data for father mother and their children, which is usually met in GG experiments. Then, using parental we could recreate genotype map for children. Afterwards analysis is the same as in normal GG case. Additionally to standalone analysis our approach could be used as test for obtained data, big differences showing somebody messed something up.

PLAN

|  |  |  |
| --- | --- | --- |
| **March** | | |
| **Naïve beginnings with RIL.** | | |
| Week 1 | Present project plan to Ritsert. If we’re not immediately kicked out of his office, acquire data, format it probably, get to know what the hell is inside. | Project plan approved by Ritsert. |
| Week 2 | First analysis on public data. Starting with RIL, only few genes, that can be easily divided in expressing/non-expressing and are evenly distributed throughout genome. Trying different thresholds – median, mean etc. Recreating simple map using our algorithm and one form r/qtl. | Fancy charts (sexiness of project grows by 1000%!). Proves that this approach is working. |
| Week 3 | Still working only on RILs and assuming that there is only cis regulation, but inputting also genes, that are evenly expressed in both parents. Recreating gene maps using our algorithm, comparing real map. | First gene map. |
| Week 4 | Constructing first QTL map. Working on introducing trans regulation. | First QTL map w/out trans effects taken into calculation. |
| Week 5 | Final introducing trans regulation, working algorithm for RILs. | First QTL map. |
| **summary** | First test (it’s working, jiiiiiiiiii!/why the fuck this shit is not working?). In RILs there are strong arguments for this | First complete QTL profile. |

|  |  |  |
| --- | --- | --- |
| **April** | | |
| **Crossing swords with Back Cross.** | | |
| Week 1 | Starting from very limited data (as in week 1 and 2 with RILs). |  |
| Week 2 |  |  |
| Week 3 |  |  |
| Week 4 | EASTER |  |
| summary |  |  |
| **May** | | |
| **Crossing swords with Back Cross/F2** | | |
| Week 1 |  |  |
| Week 2 |  |  |
| Week 3 |  |  |
| Week 4 |  |  |
| summary |  |  |
| **June** | | |
| **Making it generally OK.** | | |
| Week 1 |  |  |
| Week 2 |  |  |
| Week 3 |  |  |
| Week 4 |  |  |
| Week 5 |  |  |
| summary |  |  |
| **July** | | |
| **Summary.** | | |
| Week 1 |  |  |
| Week 2 |  |  |
| Week 3 |  |  |
| Week 4 |  |  |
| summary |  |  |