

TD 3 Bioinformatique – M1 SPGF

Traitement bioinformatique de puces ADN

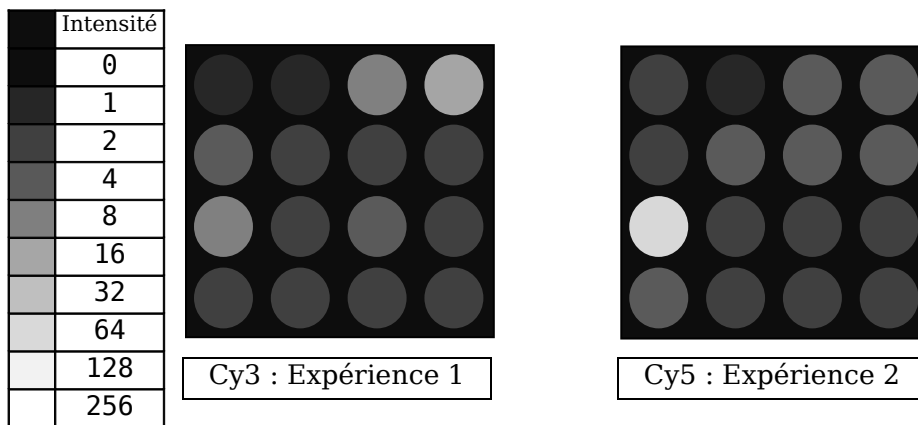
Octobre 2007

1) Introduction

Une équipe de chercheurs tente de comprendre le mécanisme d'action intracellulaire du virus de la fièvre de la vallée du Rift (VFVR). Il s'agit d'un arbovirus de la famille des *Bunyaviridae* (genre *Phlebovirus*) provoquant une fièvre hémorragique mortelle. A partir de biopsies de tissu sain et infecté par le virus, l'équipe décide de réaliser des expériences de puces à ADN.

2) Représentation des données et normalisation intra-puce:

Deux blocs présentant 2 expériences de puce à ADN sont représentés. Un marquage Cy3 (vert) est réalisé sur les ARNm de tissus sains, tandis que les ARNm issus de tissus infectés sont marqués par Cy5 (rouge). Après les expériences, puis acquisition des images les données brutes sont reportés dans un tableau.



	Exp 1A	Exp 2A
Gène 1,1	1,6	2,1
Gène 1,2	1,5	1,4
Gène 1,3	6,9	3,0
Gène 1,4	17,8	4,8
Gène 2,1	5,5	2,9
Gène 2,2	2,3	4,8
Gène 2,3	2,3	3,1
Gène 2,4	2,7	2,0
Gène 3,1	10,1	121,1
Gène 3,2	1,7	2,6
Gène 3,3	4,3	2,3
Gène 3,4	1,7	1,9
Gène 4,1	2,3	3,5
Gène 4,2	2,1	1,8
Gène 4,3	1,8	2,7
Gène 4,4	1,5	1,7

- Réalisez un graphique cy5 vs cy3.
- Calculez le \log_2 des valeurs et représentez le graphique $\log_2(\text{cy5})$ vs $\log_2(\text{cy3})$. Que constatez vous ?

c) Réalisez le graphique M vs A (avec $M = \log_2(\text{cy5}) - \log_2(\text{cy3})$ et $A = (\log_2(\text{cy5}) + \log_2(\text{cy3}))/2$)

d) Normalisez les données par ajustement linéaire, puis retracez le graphique.

3) Normalisation inter-puces :

Pour plus de fiabilité dans les analyses chaque expérience est répétée 3 fois. Les 3 puces correspondantes sont notées A, B et C.

Il y a deux types de population de cellules :

- Population non infecté par le virus (Série d'expériences 1 (Cy3) : 1A, 1B, 1C)
- Population infectée par le virus (Série d'expériences 2 (Cy5) : 2A, 2B, 2C)

Les résultats de ces expériences ($\log_2(\text{intensités})$) sont reportés sur le tableau suivant (nous supposons les données normalisées) :

	Exp 1A	Exp 1B	Exp 1C	Exp 2A	Exp 2B	Exp 2C
Gène 1,1	0,7	2,1	2,1	1,1	0,7	1,1
Gène 1,2	0,5	0,4	0,7	0,5	0,4	0,6
Gène 1,3	2,8	2,1	2,9	1,6	1,3	1,4
Gène 1,4	4,2	1,7	1,0	2,3	1,8	2,1
Gène 2,1	2,5	2,0	2,1	1,5	1,4	1,7
Gène 2,2	1,2	1,7	1,2	2,3	2,0	2,2
Gène 2,3	1,2	1,1	0,7	1,6	0,3	1,2
Gène 2,4	1,5	1,1	1,0	1,0	0,9	0,7
Gène 3,1	3,3	2,2	1,5	6,9	1,5	1,8
Gène 3,2	0,7	1,5	1,5	1,4	1,0	1,6
Gène 3,3	2,1	3,2	2,0	1,2	1,0	0,8
Gène 3,4	0,8	0,6	0,4	0,9	1,1	1,0
Gène 4,1	1,2	1,1	0,8	1,8	1,7	1,1
Gène 4,2	1,1	1,1	0,9	0,9	1,0	1,2
Gène 4,3	0,9	0,6	0,6	1,4	1,1	1,3
Gène 4,4	0,6	0,6	0,7	0,8	0,6	1,1

a) Représentez chaque expérience par un « box-plot ».

b) Réalisez une normalisation inter-puces par ajustement de la moyenne et tracez les « box-plots ».

c) Réalisez une normalisation inter-puces par moyenne centré réduite et tracez les « box-plots ».

4) Analyse des données

a) Pour déterminer les gènes sous ou sur exprimés selon la condition, réalisez un test t pour chaque gène.

b) Pour regrouper les gènes par profil d'expression selon les deux conditions, réalisez un classification hiérarchique ascendante.

1. Annexe : Table de la loi de Student

	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001	0.0001
1	1.000	3.078	6.314	12.706	31.281	63.657	127.32	318.31	636.62	6366.2
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	34.599	99.992
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924	28.000
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	15.544
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	11.178
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	9.082
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408	7.885
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	7.120
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781	6.594

10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587	6.211
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437	5.921
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318	5.694
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221	5.513
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140	5.363
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073	5.239
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015	5.134
17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965	5.044
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922	4.966
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883	4.897
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850	4.837
21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819	4.784
22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792	4.736
23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768	4.693
24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745	4.654
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725	4.619
30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646	4.482
35	0.682	1.306	1.690	2.030	2.438	2.724	2.996	3.340	3.591	4.389
40	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551	4.321
45	0.680	1.301	1.679	2.014	2.412	2.690	2.952	3.281	3.520	4.269
50	0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496	4.228
60	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460	4.169
70	0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435	4.127
80	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416	4.096
90	0.677	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402	4.072
100	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390	4.053
150	0.676	1.287	1.655	1.976	2.351	2.609	2.849	3.145	3.357	3.998
200	0.676	1.286	1.653	1.972	2.345	2.601	2.839	3.131	3.340	3.970
300	0.675	1.284	1.650	1.968	2.339	2.592	2.828	3.118	3.323	3.944
500	0.675	1.283	1.648	1.965	2.334	2.586	2.820	3.107	3.310	3.922
1000	0.675	1.282	1.646	1.962	2.330	2.581	2.813	3.098	3.300	3.906
infini	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291	3.891