Methods to integrate multiple tables in biomedical studies to detect biomarkers and stratify individuals

Instituto de Salud Carlos III. Centro Nacional de Epidemiología
September, 2017

Juan R Gonzalez

BRGE - Bioinformatics Research Group in Epidemiology
Barcelona Institute for Global Health (ISGlobal)
e-mail:juanr.gonzalez@isglobal.org
http://www.creal.cat/brge
and Departament of Mathematics, UAB

Outline

5-days course aiming to introduce statistical methods and tools to discover new biomarkers and stratify individuals with similar profile from one ore more tables of data

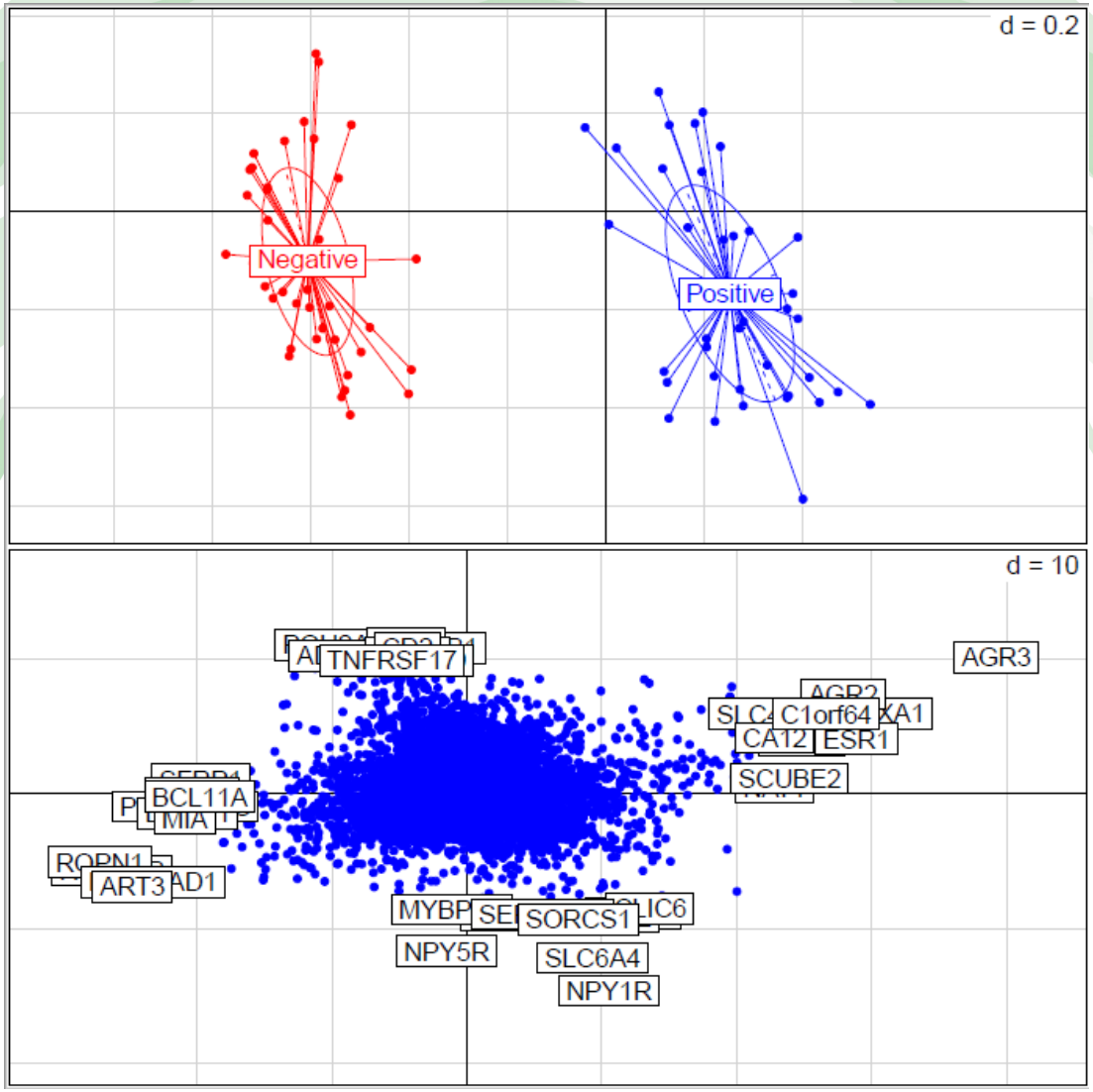
- **Day 1:** Introduction to R
- **Day 2:** Multivariate methods for one table (non-supervised / supervised)
- **Day 3:** Multivariate methods to integrate multiple tables (I)
- **Day 4:** Multivariate methods to integrate multiple tables (II)
- **Day 5:** DAGs, Mendelian Randomization and mediation analysis to integrate multiple tables

Aims of the course

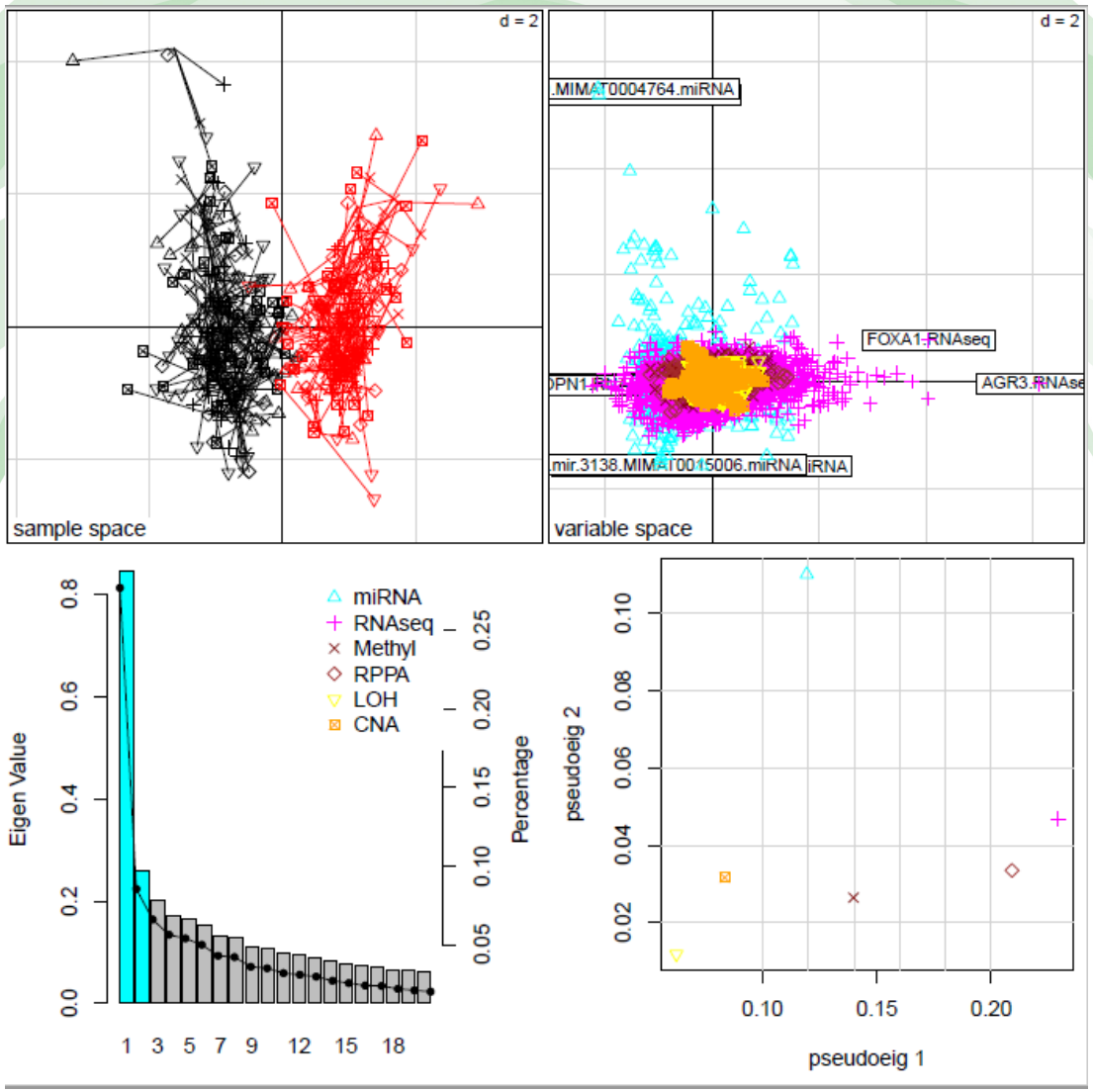
	Cases	Controls	Unadjusted OR		
			OR	95% CI	*p*
Reproductive factors					
Age at menarche (years)*	15.8	15.7	1.03	0.92–1.14	0.64
Parity*					
0	10 (5.3%)	4 (2.1%)	1		
1	16 (8.2%)	7 (3.6%)	0.83	0.17–4.05	0.82
2	110 (56.1%)	96 (49.0%)	0.38	0.10–1.47	0.16
3	39 (19.9%)	64 (32.7%)	0.19	0.05–0.74	0.02
4+	14 (7.1%)	24 (12.2%)	0.25	0.06–1.09	0.06
Age at 1st parity	24.2	23.7	1.03	0.97–1.09	0.33
Total months breastfeeding	36.1	37.4	1	0.99–1.02	0.81
Anthropometric factors					
BMI at age 18*					
<18 kg/m^2	42 (21.5%)	60 (31.6%)	1		
18–20 kg/m^2	64 (32.8%)	80 (42.1%)	1.16	0.69–1.94	0.44
>20 kg/m^2	89 (45.6%)	50 (26.3%)	2.73	1.56–4.79	0.0006
Hormonal factors					
OCP Use*					
No	173 (89.2%)	184 (94.9%)	1		
Yes	21 (10.8%)	10 (5.2%)	2.2	1.01–4.89	0.05

∗ denotes the variables used for multivariate analysis to calculate the adjusted OR.

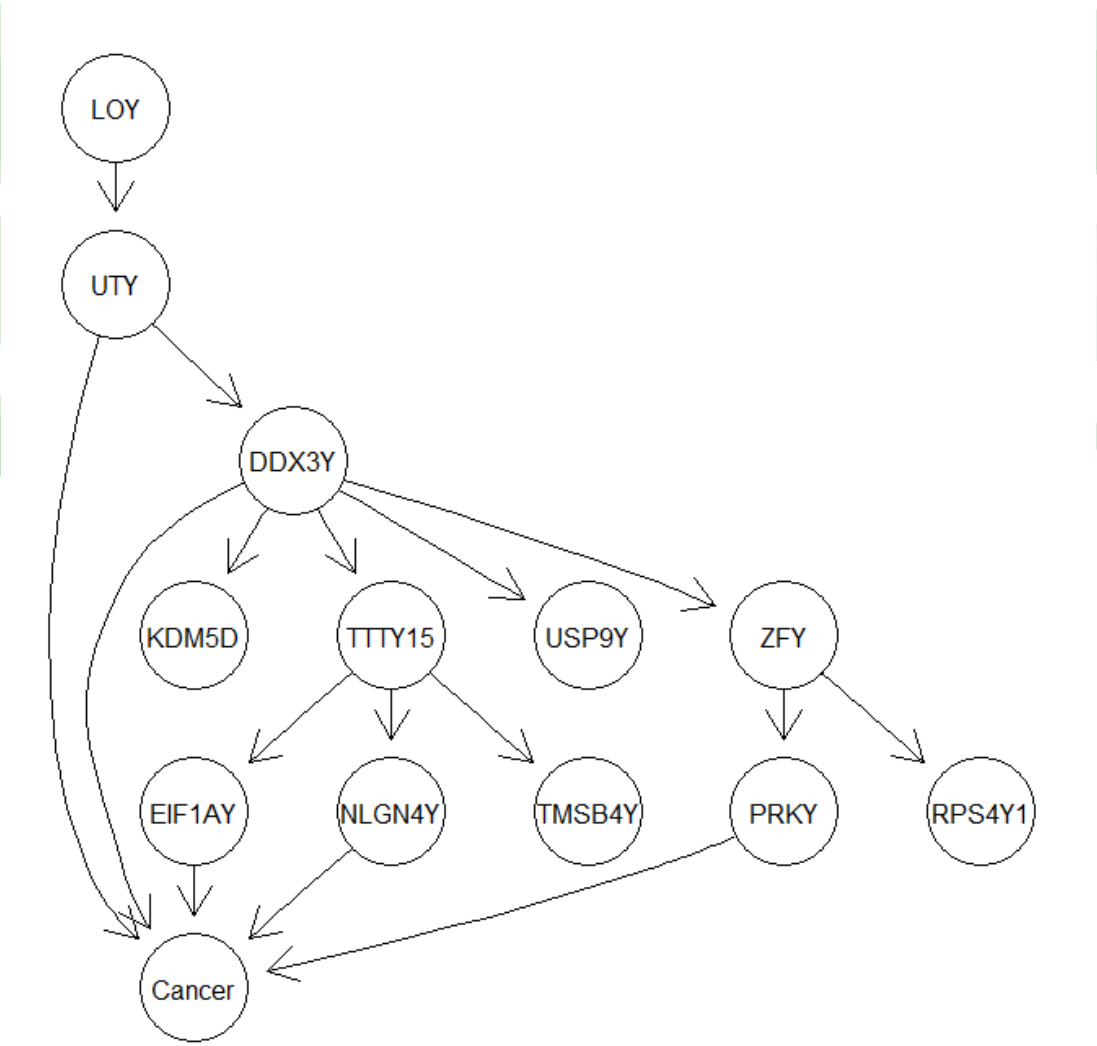
Aims of the course



Aims of the course



Aims of the course



- Lectures will introduce statistical methods and how to analyze real data (from different settigs including nutrition, air pollution, genetic, genomics and other biomedical datasets using R packages
- Excersises will analyze data from a case/control study in cancer setting were controls and 4 types of cancer (colorectal, stomach, breast and prostate) were studied
- Datasets will include a set of variables that are considered as confounding variables, another set about nutrients and a third one ecoding food compsumption variables
- Material (including answers to exercises) is available at `https://github.com/isglobal-brge/biomarkers_multiple_tables`