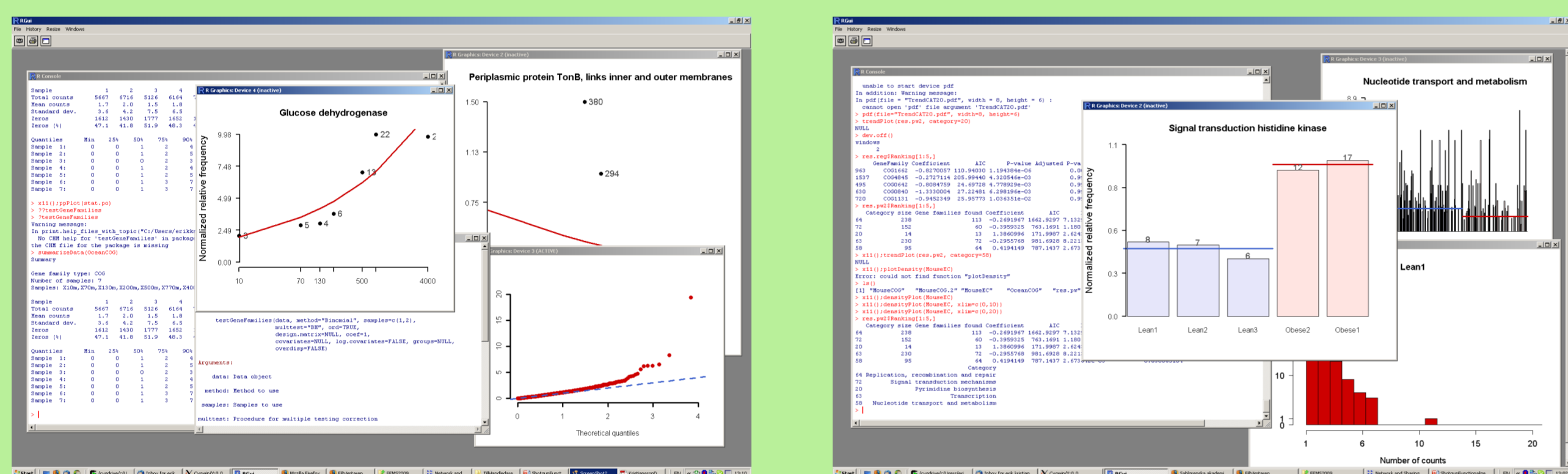


# ShotgunFunctionalizeR

- an R-package for functional comparison of metagenomes

## Introduction

- ▶ **ShotgunFunctionalizeR** is an R-package for functional comparison of metagenomes.
- ▶ **ShotgunFunctionalizeR** contains tools for annotation, visualization and statistical analysis of data generated from high-throughput shotgun sequencing.
- ▶ **ShotgunFunctionalizeR** can be used to identify both genes and pathways that are differentially abundant between metagenomes.
  - ▶ *Gene-centric analysis* uses gene-families (e.g. COGs and EC-numbers).
  - ▶ *Pathway-centric analysis* uses sets of gene-families (e.g. KEGG and COG pathways).



Screenshots showing **ShotgunFunctionalizeR** in action!

**ShotgunFunctionalizeR** is available at  
<http://shotgun.zool.gu.se>

## Implementation

- ▶ **ShotgunFunctionalizeR** contains several standard and previously published statistical testing procedures.
  - ▶ Procedures for gene-centric analysis include binomial and hypergeometric tests (derived from Poisson and multinomial sampling assumptions respectively) and a test using Gaussian approximations.
  - ▶ Procedures for pathway-centric analysis include overrepresentation test (using Fisher's exact test), a  $\chi^2$  test of independence and the Gaussian sum test (also known as the d-statistic on the IMG/M web site).
- ▶ We have also implemented a novel method based on a Poisson model. In contrast to most other procedures, the Poisson model is highly flexible and thus applicable to a wide range of experimental designs, e.g. regression analysis.

**ShotgunFunctionalizeR - an R-package for functional comparison of metagenomes**

Erik Kristiansson<sup>1,2</sup>, Philip Hugenoltz<sup>3</sup>, Daniel Dalevi<sup>4</sup>

<sup>1</sup>Department of Zoology - University of Gothenburg, <sup>2</sup>Department of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, <sup>3</sup>Microbial Ecology Program, DOE Joint Genome Institute, <sup>4</sup>Department of Computer Science and Engineering - Chalmers University of Technology and University of Gothenburg  
Contact: [erik.kristiansson@zool.gu.se](mailto:erik.kristiansson@zool.gu.se) or [daniel.dalevi@chalmers.se](mailto:daniel.dalevi@chalmers.se)

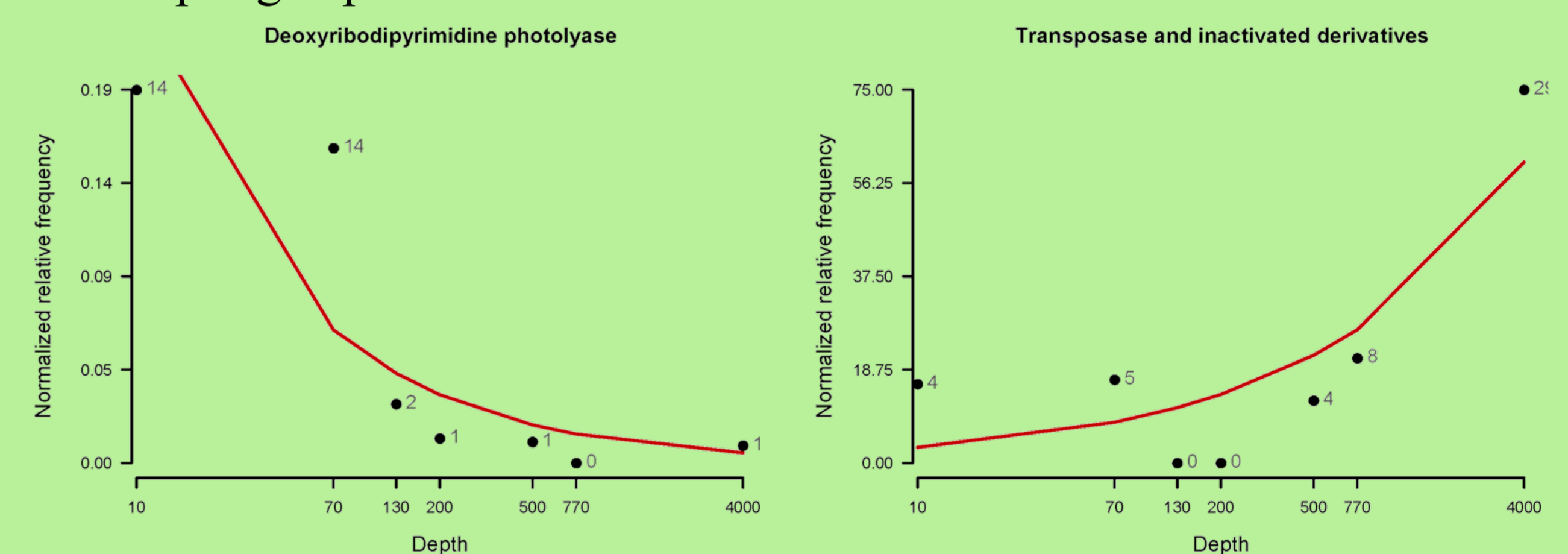
## Example I: Ocean's Interior

- ▶ Metagenomes sampled seven depths in the Pacific Ocean (data from DeLong *et al.* Science 2006).
- ▶ Poisson regression analysis revealed several gene-families have a significant correlation to sample depth.

Gene-family	Coefficient	P-value	FDR	Annotation
COG0810	-0.37	$4.5 \times 10^{-20}$	$1.5 \times 10^{-16}$	Periplasmic protein TonB, links inner and outer membranes
COG0415	-1.47	$4.7 \times 10^{-9}$	$8.0 \times 10^{-6}$	Deoxyribodipyrimidine photolyase
COG2141	0.46	$2.8 \times 10^{-7}$	$3.2 \times 10^{-4}$	F420-dependent tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases
COG2801	1.14	$5.2 \times 10^{-7}$	$3.6 \times 10^{-4}$	Transposase and inactivated derivatives
COG5178	-0.36	$7.1 \times 10^{-7}$	$3.6 \times 10^{-4}$	U5 snRNP spliceosome subunit

The five most significant gene-families in the Poisson regression analysis. The sign of the coefficient indicates whether their abundance increase or decrease with increasing sampling depths.

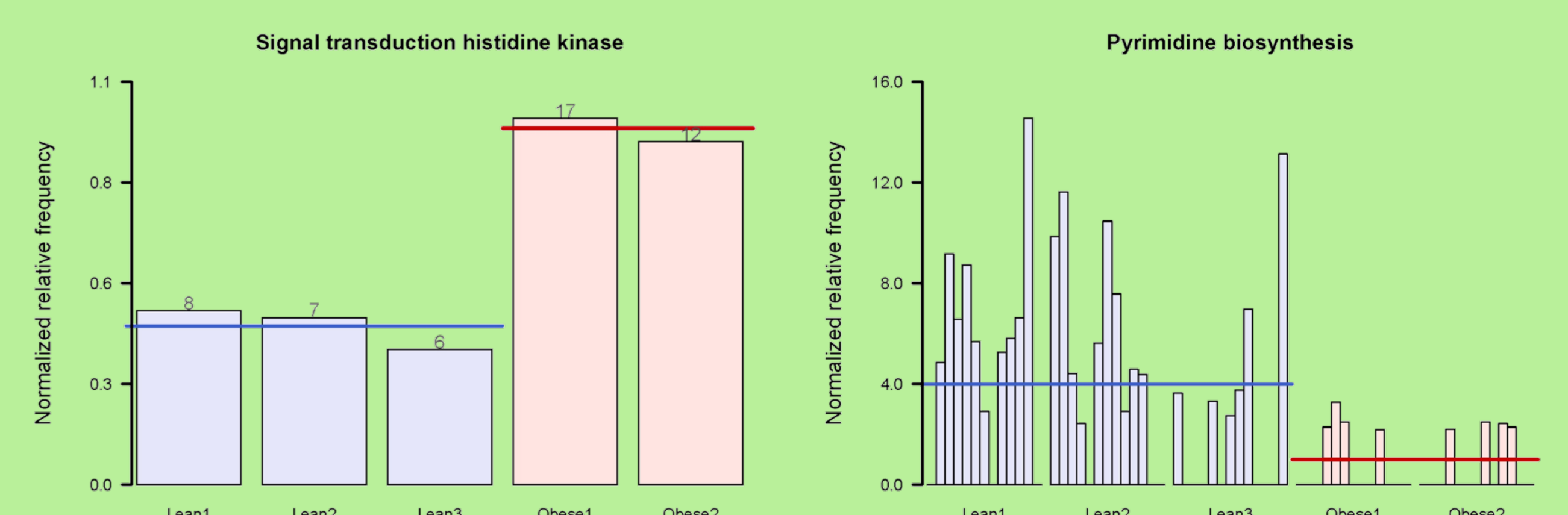
- ▶ The highly significant gene-family deoxyribodipyrimidine photolyase (COG0415) is involved in repair of UV radiation-induced DNA damage. The abundance of this gene-family decreased with increasing sampling depth.



The estimated trends for two significant gene-families, deoxyribodipyrimidine photolyase (COG0415) and transposase and inactivated derivatives (COG2801).

## Example II: Mouse Gut

- ▶ Gut metagenomes samples from lean and obese mice (Turnbaugh *et al.* Nature 2006).
- ▶ Gene and pathway-analysis in **ShotgunFunctionalizeR** using direct comparison of two groups (Poisson model).
- ▶ Several genes and pathways with significant difference in abundance between the lean and obese mice could be identified.



To the left a gene family (signal transduction histidine kinase, COG0642) and to the right a pathway (pyrimidine biosynthesis) with a significant difference in abundance between the two groups.