

# GENOME-WIDE ANALYSIS OF GENETIC PREDISPOSITION IN PATIENTS WITH THE HISTORY OF ACUTE MYOCARDIAL INFARCTION

*Presentation at ISCB30, Prague, Czech Republic*

Talk (short title) “*Genetic Predisposition in Patients with AIM History*” presented on AUGUST 26, 2009

Zdeněk Valenta, Michal Kolář, Hana Grünfeldová,  
Ivan Mazura, Petra Feglarová and Jana Zvárová

*Centre of Biomedical Informatics, Institute of Computer Science AS CR  
Institute of Molecular Genetics, Academy of Sciences of the Czech Republic  
Centre of Biomedical Informatics, Municipal Hospital in Čáslav, Czech Republic*

Genome-Wide  
Analysis of Genetic  
Predisposition in  
Patients with the  
History of Acute  
Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

Understanding the  
Data

Modelling Results

Discussion

References

# The Talk Outline

- 1 The Aim and Methods
- 2 Study Design
- 3 Sources of Variability
  - Background Noise
  - Target Populations
- 4 Handling Multiple Tests of Significance
- 5 Samples Description
- 6 Understanding the Data
- 7 Modelling Results
- 8 Discussion
- 9 References

Genome-Wide  
Analysis of Genetic  
Predisposition in  
Patients with the  
History of Acute  
Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise  
Target Populations

Handling Multiple Tests  
of Significance

Samples Description

Understanding the  
Data

Modelling Results

Discussion

References

- **The aim**

- 1 Identification of '*differentially expressed genes*' in patients undergoing the acute phase of '*primary myocardial infarction*' relative to their '*matched controls*'

- **Methods**

- 1 **Scope:** We have exploited the Illumina technology allowing for a '*genome-wide study of differential expression of genes*' within each study population
- 2 **Data source:** Data on both '*cases*' and '*controls*' were obtained in 2007–2009 from the Municipal Hospital in Čáslav, serving approximately an area of 30,000 inhabitants
- 3 **Entry criteria:** Cases were selected out of subjects not older than 80 years, who were admitted to the Municipal Hospital in Čáslav with the diagnosis of a primary '*acute myocardial infarction*' ('AIM')
- 4 **Target populations:** Focus on genome-wide assessment of differential gene expression (GE) in '*homogeneous subgroups of AIM cases*', relative to matched pair controls



- **Statistical design**

- 1 We used **matched case-control study design** allowing for modelling the matched pairs differences using '*Linear Models for Microarray Data Analysis*' implemented in the "limma" package in **R**, part of the Bioconductor project
- 2 This approach uses linear models generalisation of paired *t*-test scenario under the Bayesian framework, where the standard errors of **moderated *t*-statistic** are shrunk across the genes towards a common value.

- **Matching criteria:** '*controls*' were matched to '*cases*' based on:

- (a) **Gender**
- (b) **Age** – same age or older by max 5 years
- (c) **Smoking status**
- (d) **Status of Diabetes Mellitus**



- Data from microarray experiments are typically **noisy**, reflecting many possible **sources of variability**.
- They may include, for instance:
  - 1 **Extraction and storage** of genetic material from the peripheral blood of patients – adopted techniques for preserving blood samples over extended period of time
  - 2 **Repeated un-freezing** of stored genetic blood samples
  - 3 **Isolation and hybridisation** processes and their set-up
  - 4 **Experimental conditions** – Illumina chip series, chip expiration dates, chip flaws, number of beads representing each gene on a chip, etc.
  - 5 **Background noise** on Illumina chips
  - 6 **Chip-to-chip** signal distribution **differences**
  - 7 **Effectiveness of matching procedures** designed for paired comparisons
  - 8 **Patient's characteristics** associated with the gene expression left **unaccounted for**
  - 9 **Target populations** definitions – homogeneity of target groups to be compared
  - 10 **Random fluctuation** of gene expression intensities within each target group/population



- **Background noise on Illumina chips**

- 1 Throughout this project we have used **normal–exponential model** in correcting for background noise intensities on Illumina chips
- 2 The ‘*normexp*’ model assumes that we are dealing with a mixture of **normal background noise and exponentially distributed signal** of gene expression intensities
- 3 It offered a most realistic scenario for **separating the signal intensities** from those appearing in the background for our data
- 4 This follows recent developments published in **Silver JD, Ritchie ME and Smyth GK**: ‘*Microarray background correction: maximum likelihood estimation for the normal–exponential convolution*’ (Biostatistics. 2009 Apr;10(2):352-63).



- **Target populations**

- 1 Definition of appropriate *'target populations'* at the design stage of a project is an important way of reducing the overall variability among the cases
- 2 Goal: Reduce overall variability among the cases via *'identification of the most homogeneous subgroups'*. This may prove critical for the project
- 3 The cases who *'died from cardiovascular causes'* during the 6 months follow-up period following the primary event ('AIMD6') were expected to differ in terms of gene expression from the rest of the cases
- 4 Furthermore, those who *experienced a 'heart failure (HF) during the event'* were considered to possibly represent another transcriptionally distinct subpopulation in terms of GE. However, these cases from our sample were found to be part of the 'AIMD6' group
- 5 As a result, two distinct sub-populations of the cases were defined ('AIM' and 'AIMD6')



- Another important issue affecting the results from microarray data analysis is the way the **multiple tests of significance** across individual genes are being handled
  - 1 Since *Bonferroni* and other techniques controlling the *family-wise error rate* (FWER) are rather conservative, methods designed by **Benjamini & Hochberg** to control “**False Discovery Rate**” (FDR) were used instead. FDR approach provides greater power, the measure represents the expected false positive rate
  - 2 Another measure, ***q*-value** (Storey 2001), is defined to be the FDR analogue of the *p*-value. The *q*-value of an individual hypothesis test is the minimum FDR at which the test may be called significant
  - 3 In our analyses, the method by **Benjamini & Hochberg** and Storey’s ***q*-value** approach are being used simultaneously
- **Statistical vs. clinical significance** – only the genes passing the  $\alpha = 0.05$  level of statistical significance **and** a fold-change of at least 2 ( $|\log_2 FC| \geq 1$ ) were considered differentially expressed

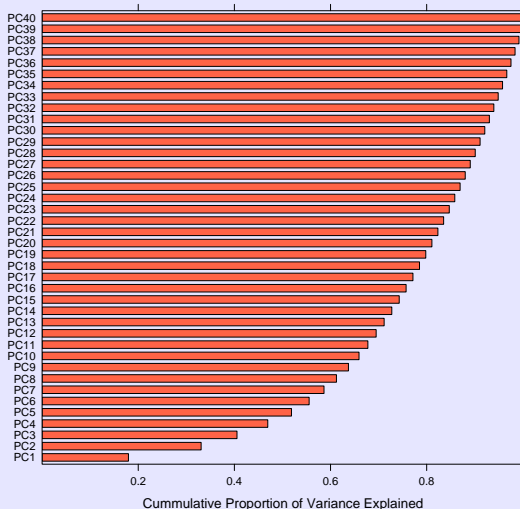


- **Six samples** were placed on each HumanWG-6 v2 Expression Illumina BeadChip
  - ① They involved peripheral blood of **AIM cases** obtained shortly (usually within 6 hours) after the event, and that of their **matched controls**
  - ② Depending on their availability, 6 months follow-up samples of the AIM cases were also placed on a chip
  - ③ All (correlated) **follow-up samples** of the AIM cases **were excluded** from the subsequent “*limma*” analyses discussed in this presentation
- The final dataset involved **nine Illumina chips** with valid data on both AIM cases and their matched controls
- Altogether **20 pairs of AIM cases and their matched controls** entered the final “*limma*” model
- Gene expression intensities were subject to non-specific **IQR-based filtration** and **quantile normalisation** before entering the model on the **log<sub>2</sub>**-scale



# FIGURE 1: DATA PREVIEW USING PRINCIPAL COMPONENTS: SCREE PLOT OF THE PRINCIPAL COMPONENTS ANALYSIS RESULTS

## Principal Components Analysis Summary



Genome-Wide  
Analysis of Genetic  
Predisposition in  
Patients with the  
History of Acute  
Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRŮNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

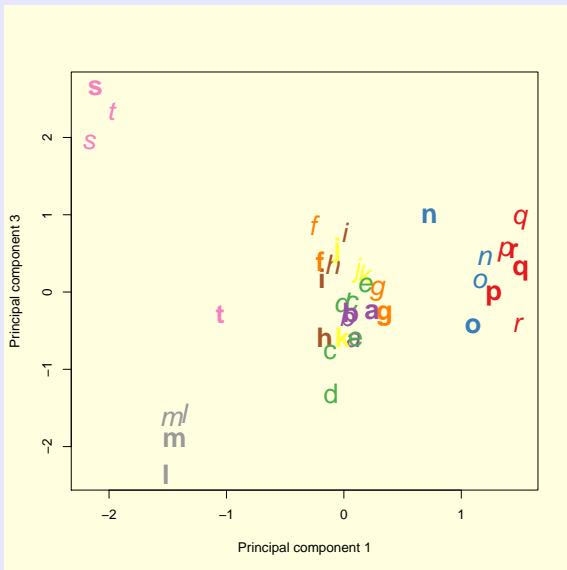
Understanding the  
Data

Modelling Results

Discussion

References

FIGURE 2: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY ILLUMINA CHIP ID (CONT'D)



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

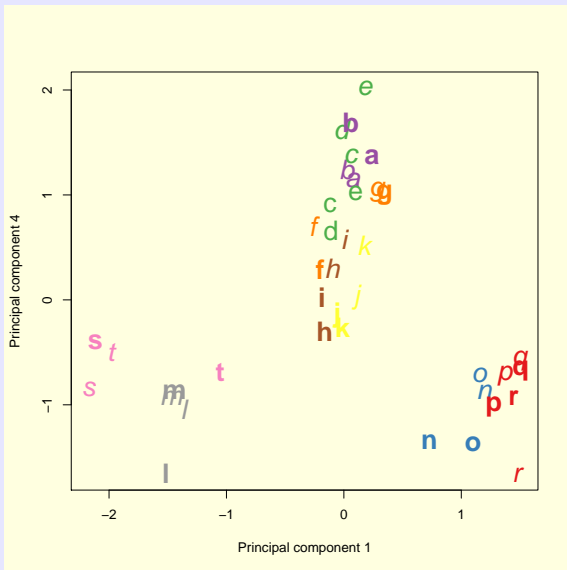
Understanding the  
Data

Modelling Results

Discussion

References

FIGURE 3: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY ILLUMINA CHIP ID (CONT'D)



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

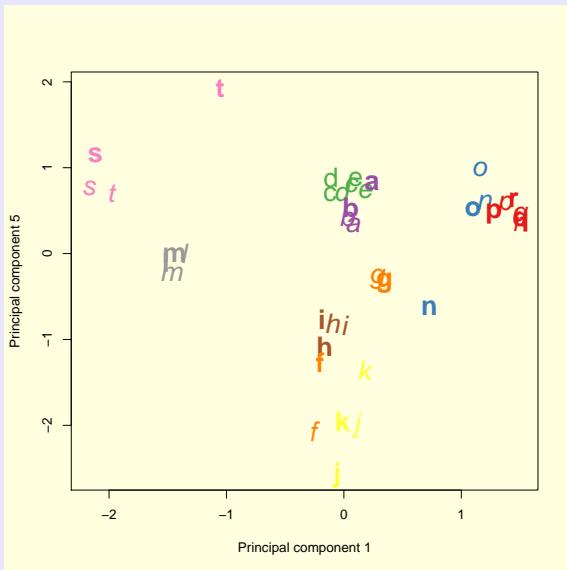
Understanding the  
Data

Modelling Results

Discussion

References

FIGURE 4: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY ILLUMINA CHIP ID (CONT'D)



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

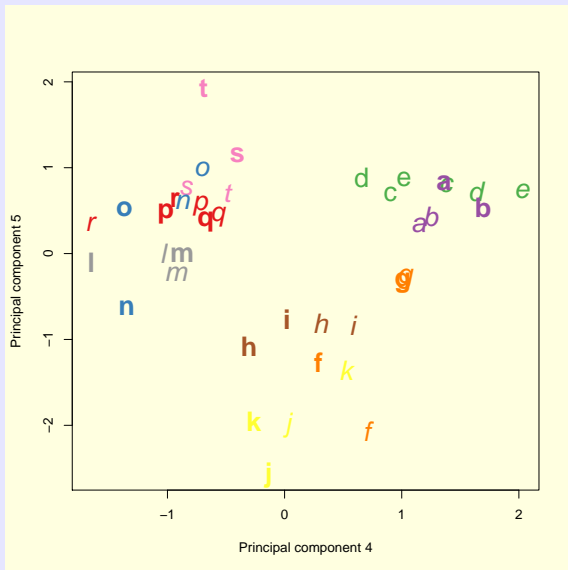
Understanding the  
Data

Modelling Results

Discussion

References

FIGURE 5: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY ILLUMINA CHIP ID (CONT'D)



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

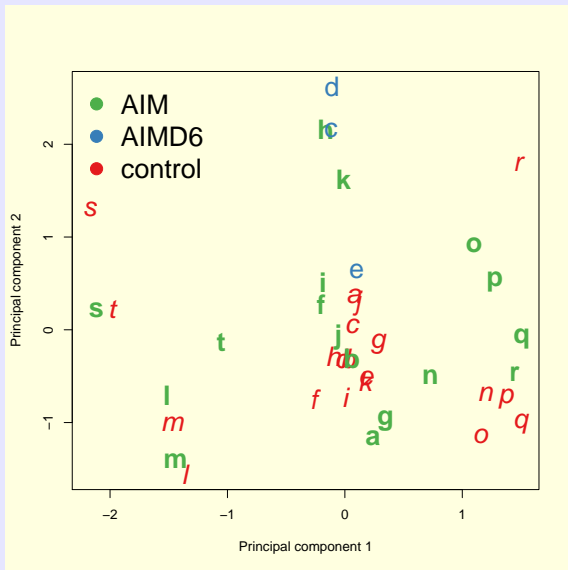
Understanding the  
Data

Modelling Results

Discussion

References

## FIGURE 6: DATA PREVIEW USING PRINCIPAL COMPONENTS: BIPLOT OF PCs BY GROUP



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

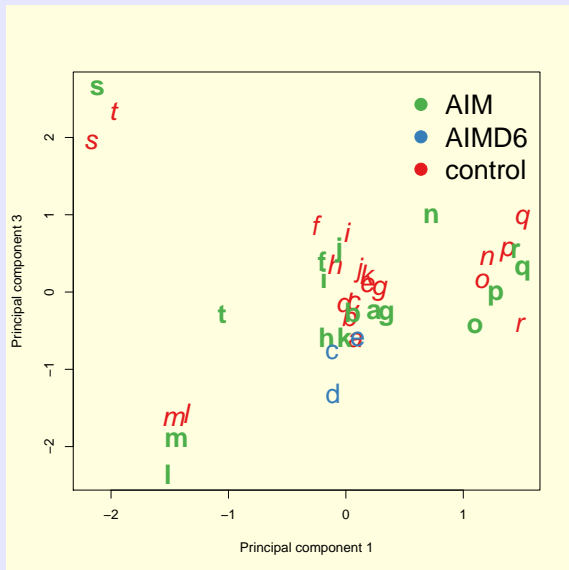
Understanding the  
Data

Modelling Results

Discussion

References

FIGURE 7: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY GROUP (CONT'D)



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

Understanding the  
Data

Modelling Results

Discussion

References

FIGURE 8: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY GROUP (CONT'D)

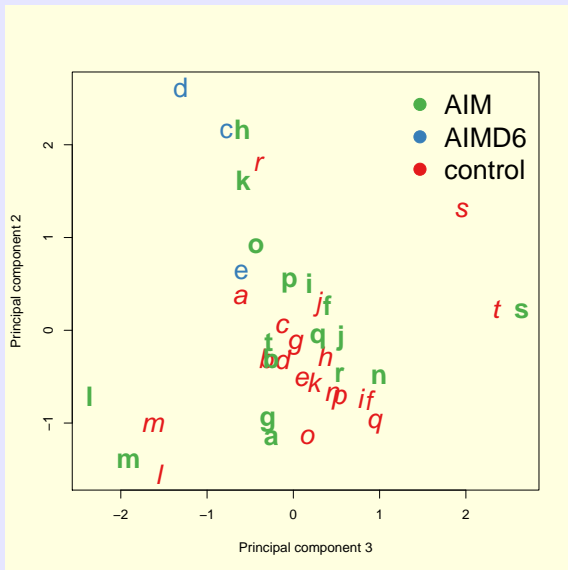
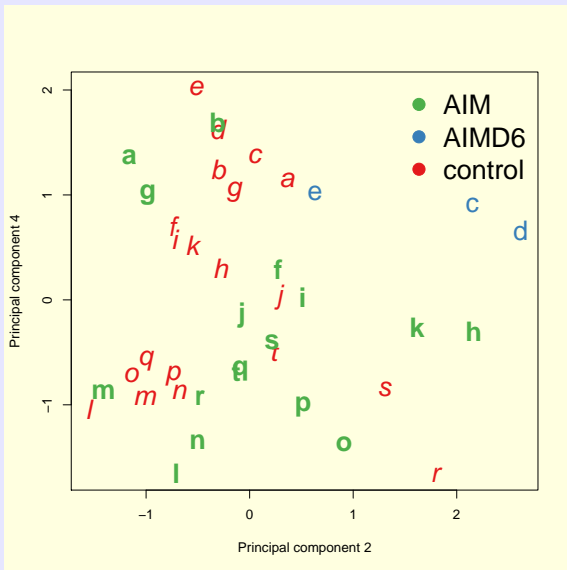


FIGURE 9: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY GROUP (CONT'D)



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

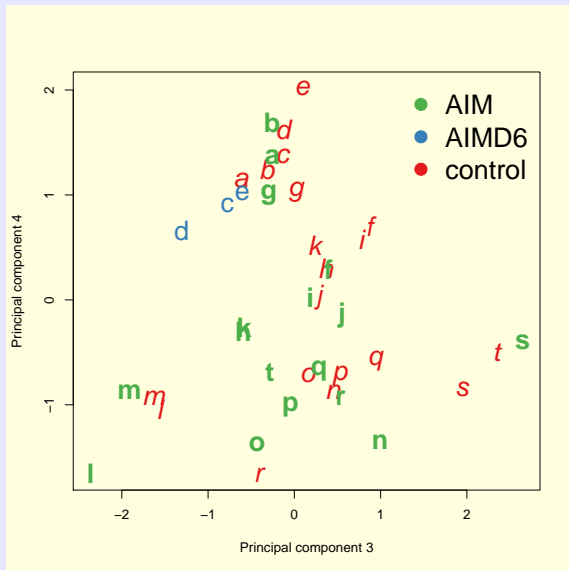
Understanding the  
Data

Modelling Results

Discussion

References

## FIGURE 10: DATA PREVIEW USING PRINCIPAL COMPONENTS: BIPLOT OF PCs BY GROUP (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

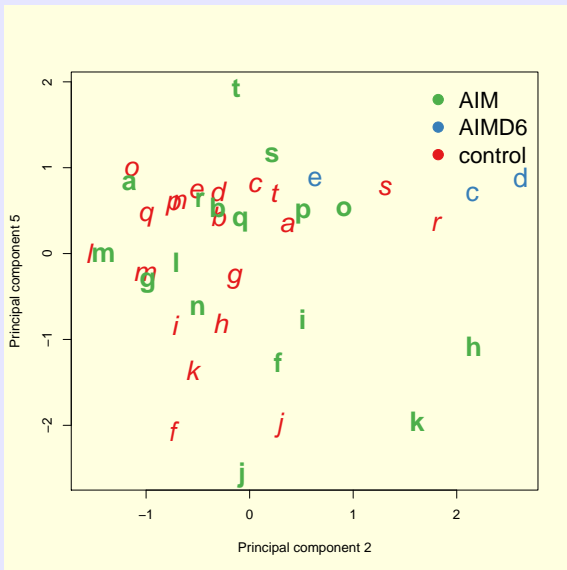
Understanding the  
Data

Modelling Results

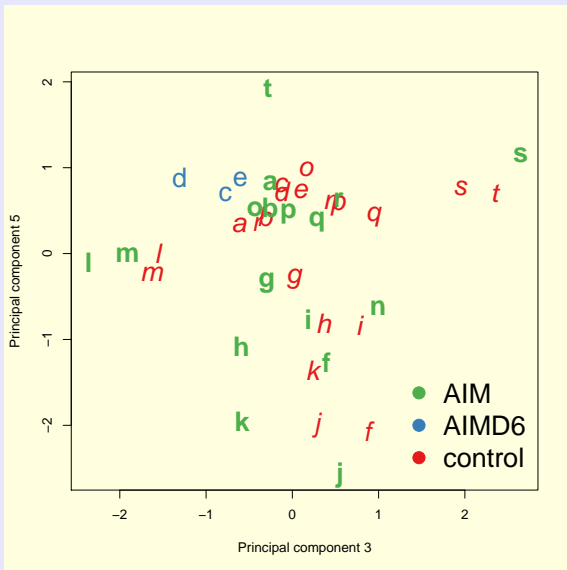
Discussion

References

## FIGURE 11: DATA PREVIEW USING PRINCIPAL COMPONENTS: BIPLOT OF PCs BY GROUP (CONT'D)



**FIGURE 12: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
BIPLOT OF PCs BY GROUP (CONT'D)**



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

[Background Noise](#)

[Target Populations](#)

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

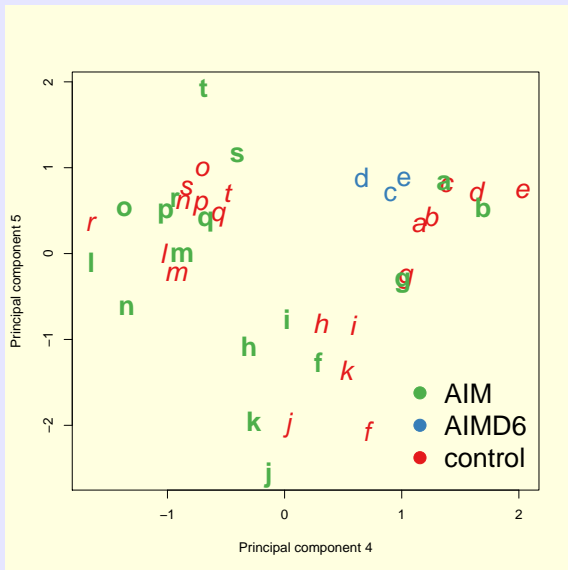
**[Understanding the  
Data](#)**

[Modelling Results](#)

[Discussion](#)

[References](#)

## FIGURE 13: DATA PREVIEW USING PRINCIPAL COMPONENTS: BIPLOT OF PCs BY GROUP (CONT'D)



### Genome-Wide Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

[Background Noise](#)

[Target Populations](#)

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

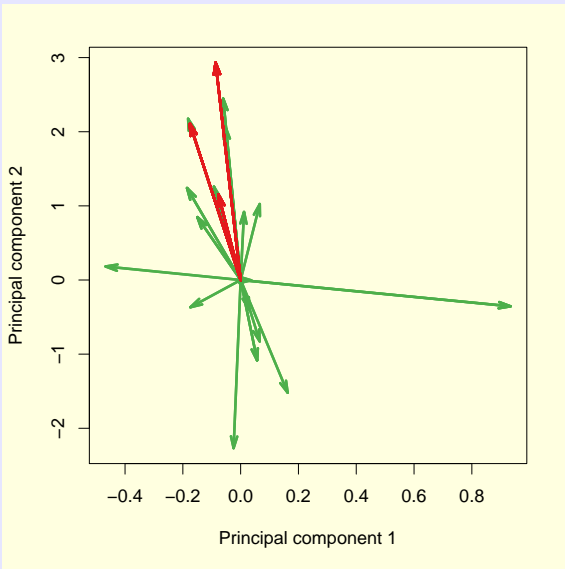
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

## FIGURE 14: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

[Background Noise](#)

[Target Populations](#)

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

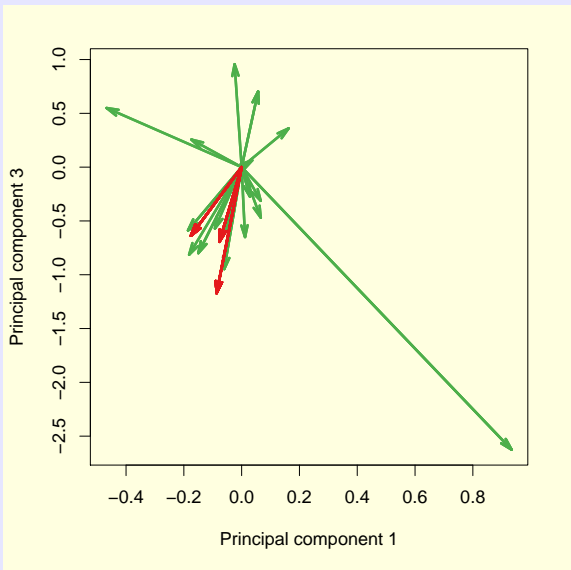
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

## FIGURE 15: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

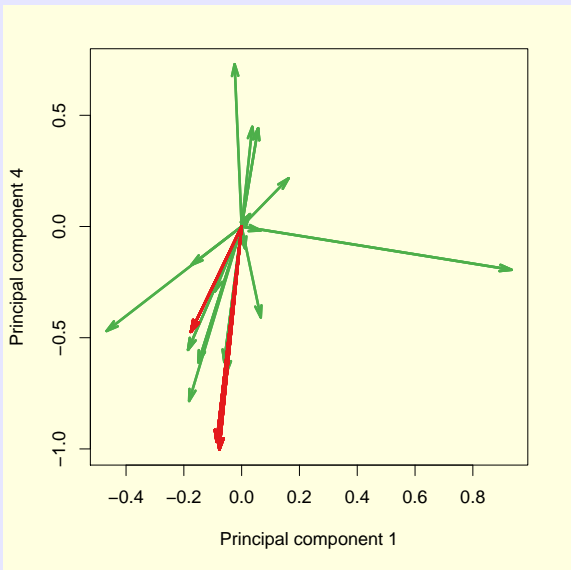
Understanding the  
Data

Modelling Results

Discussion

References

## FIGURE 16: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

[Background Noise](#)

[Target Populations](#)

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

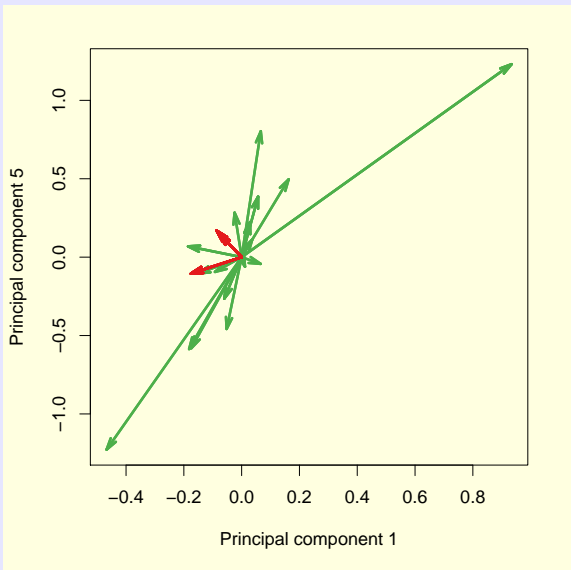
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

## FIGURE 17: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

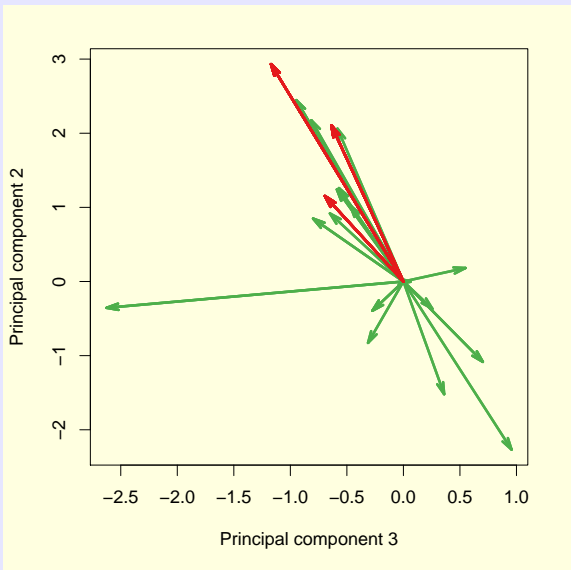
Understanding the  
Data

Modelling Results

Discussion

References

## FIGURE 18: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

Background Noise

Target Populations

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

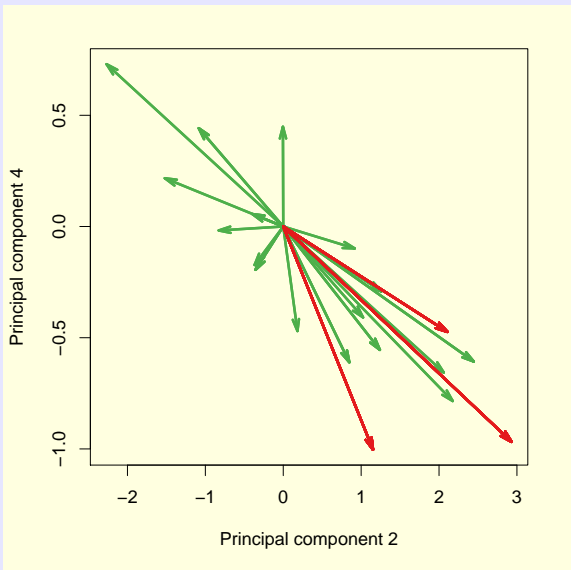
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

## FIGURE 19: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

[Background Noise](#)

[Target Populations](#)

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

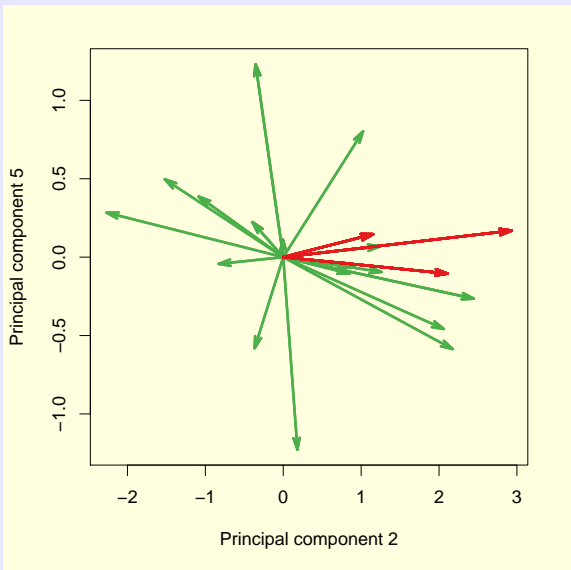
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

**FIGURE 20: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
PCA RELATIVE TO CONTROLS (CONT'D)**



**Genome-Wide  
Analysis of Genetic  
Predisposition in  
Patients with the  
History of Acute  
Myocardial Infarction**

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

Background Noise  
Target Populations

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

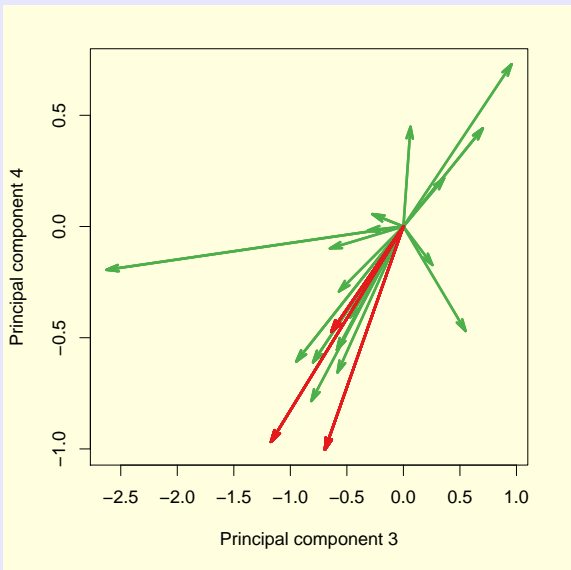
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

## FIGURE 21: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

Background Noise

Target Populations

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

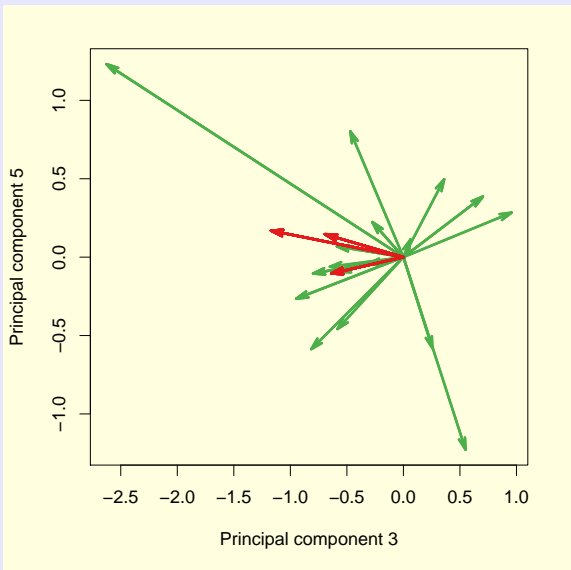
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

**FIGURE 22: DATA PREVIEW USING PRINCIPAL COMPONENTS:  
PCA RELATIVE TO CONTROLS (CONT'D)**



**Genome-Wide  
Analysis of Genetic  
Predisposition in  
Patients with the  
History of Acute  
Myocardial Infarction**

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

[Background Noise](#)

[Target Populations](#)

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

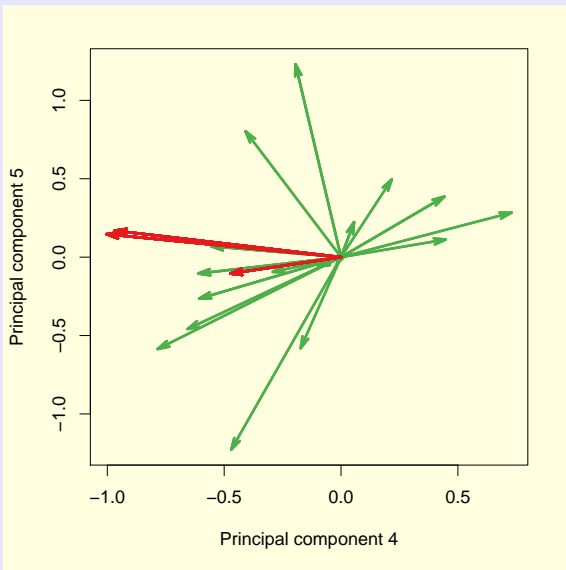
[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

## FIGURE 23: DATA PREVIEW USING PRINCIPAL COMPONENTS: PCA RELATIVE TO CONTROLS (CONT'D)



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



[The Aim and Methods](#)

[Study Design](#)

[Sources of Variability](#)

[Background Noise](#)

[Target Populations](#)

[Handling Multiple Tests  
of Significance](#)

[Samples Description](#)

[Understanding the  
Data](#)

[Modelling Results](#)

[Discussion](#)

[References](#)

### • PC-based Data Preview

- 1 Principal Components (PC) plots were used to identify distinct clusters among available cases
- 2 On the PC-scale cases were characterised in terms of vector distance from their matched control counter-parts
- 3 Clearly identifiable separate clusters of cases could indicate existence of heterogeneous groups in the data, either due to differences among cases or perhaps due to selection of “inappropriate” controls (e.g. only 4 parameters matched)
- 4 Some controls may actually be similar to cases in terms of GE, even though they did not experience an event by a set time limit. Vice versa, some cases may in fact be more similar to controls in terms of GE (e.g. event occurrence unrelated to genetic profile)
- 5 Existence of clusters lacking clinical interpretation may indicate potential problems at the analysis stage
- 6 Important: PC-based clusters should not be used for a definition of the subgroups. Subgroups need to be defined independently of GE data based on clinical characteristics



## • Statistical Model

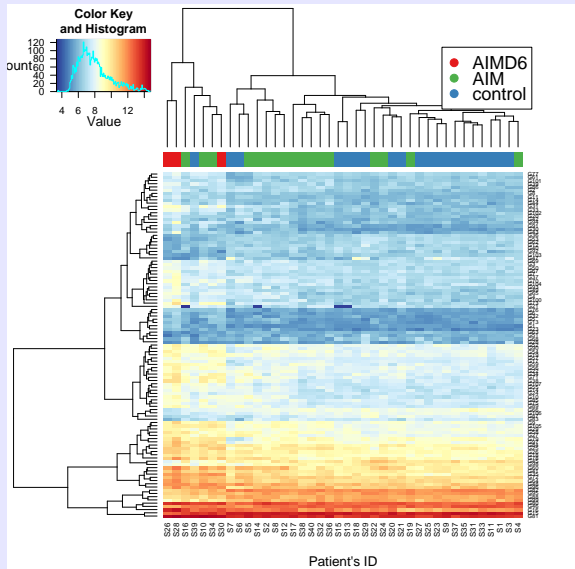
- 1 “limma” model:

$$\log_2(\textit{intensity}) \sim \textit{group} + \textit{pair} + \epsilon, \epsilon \sim N(0, \sigma^2)$$

- 2 No differentially expressed genes were identified between the cases and controls ‘*when AIM and AIMD6 groups were combined*’. This was due to **large overall variability among the cases** in our sample, even relative to their matched controls (vector PCA)
- 3 **Identifying homogeneous underlying groups** was found to be **critical** for the assessment of our data in terms of GE
- 4 Definition of underlying groups was **based on clinical information** which was available before the GE data were analysed (6-months survival, heart failure during the event)
- 5 **AIMD6 group**: Sufficiently strong differential GE signal despite the small number of cases enrolled
- 6 **AIM group**: Large variability among the cases still remains. Many factors are involved: validity of controls (some may be similar to cases), chip effects (known artefact of Illumina technology, not estimable due to financial constraints)



## FIGURE 24: DATA PREVIEW USING HEAT MAP OF DIFERENTIALLY EXPRESSED GENES



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

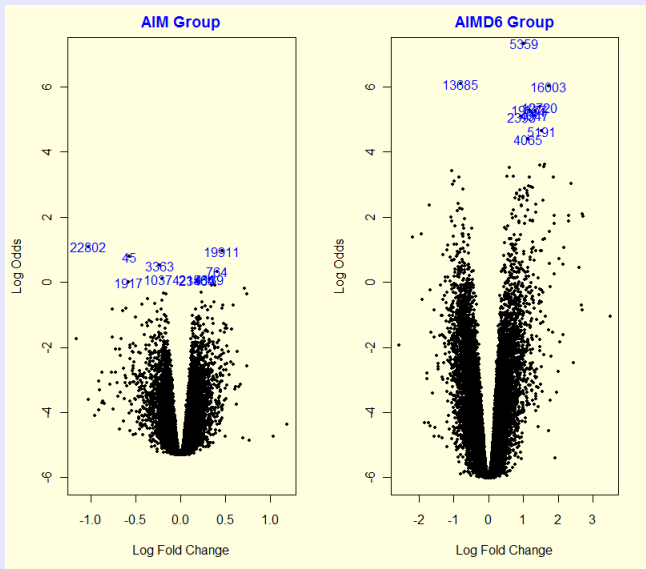
Understanding the  
Data

Modelling Results

Discussion

References

**FIGURE 25: VOLCANO PLOTS FOR THE AIM AND AIMD6 GROUP**



**Genome-Wide  
Analysis of Genetic  
Predisposition in  
Patients with the  
History of Acute  
Myocardial Infarction**

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

Understanding the  
Data

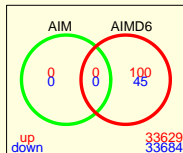
Modelling Results

Discussion

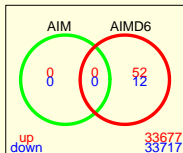
References

## FIGURE 26: MODELLING RESULTS SUMMARY USING VENN DIAGRAMS

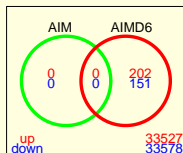
**Differentially Expressed Genes**  
 $|\logFC| \geq 0$  & FDR < 0.05



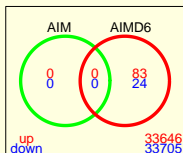
**Differentially Expressed Genes**  
 $|\logFC| \geq 1$  & FDR < 0.05



**Differentially Expressed Genes**  
 $|\logFC| \geq 0$  & q-value < 0.05



**Differentially Expressed Genes**  
 $|\logFC| \geq 1$  & q-value < 0.05



### Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
 MICHAL KOLÁŘ, HANA  
 GRÜNFELDOVÁ,  
 IVAN MAZURA, PETRA  
 FEGLAROVÁ AND JANA  
 ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
 of Significance

Samples Description

Understanding the  
 Data

Modelling Results

Discussion

References

- **Statistical Design**

- 1 Issues related to statistical design are of high importance.
- 2 Six months follow-up embedded in the design paid heavily, but also incurred high costs in terms of an advanced protocol adherence (e.g. developing technology for preparation and storage of samples at low temperatures for extended period of time before analysis, etc)
- 3 Definition of target groups proved crucial in terms of our ability to tackle heterogeneity among the samples
- 4 Target groups need to be medically/ biologically interpretable











### • Shortcomings

- 1 Controlling for variability among Illumina chips appeared outside financial scope of our project.
- 2 Since the project focused on the differences between the cases and their matched controls, much of this bias is believed to be removed via “limma” modelling
- 3 Large variability among the cases and controls points to a possible existence of subgroups within the two populations, particularly ‘*controls more similar to cases*’ and ‘*cases more like controls*’
- 4 Due to insufficient (missing) clinical data a more elaborate classification of the cases and controls could not be applied



# References

-  R Development Core Team.  
*R: A Language and Environment for Statistical Computing*.  
R Foundation for Statistical Computing, Vienna, Austria, 2008.  
ISBN 3-900051-07-0.
-  Gordon K. Smyth.  
*Limma: Linear Models for Microarray Data*.  
In R. Gentleman, V. Carey, S. Dudoit, and W. Huber R. Irizarry, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420.  
Springer, New York, 2005.
-  Robert C Gentleman, Vincent J. Carey, Douglas M. Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, Laurent Gautier, Yongchao Ge, Jeff Gentry, Kurt Hornik, Torsten Hothorn, Wolfgang Huber, Stefano Iacus, Rafael Irizarry, Friedrich Leisch, Cheng Li, Martin Maechler, Anthony J. Rossini, Gunther Sawitzki, Colin Smith, Gordon Smyth, Luke Tierney, Jean Y. H. Yang, and Jianhua Zhang.  
*Bioconductor: Open Software Development for Computational Biology and Bioinformatics*.  
*Genome Biology*, 5:R80, 2004.
-  Gordon K. Smyth.  
*Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments*.  
*Statistical Applications in Genetics and Molecular Biology*, 3(1):3, 2004.
-  Matthew E. Ritchie, Jeremy Silver, Alicia Oshlack, Melissa Holmes, Dileepa Diyagama, Andrew Holloway, and Gordon K. Smyth.  
*A Comparison of Background Correction Methods for Two-colour Microarrays*.  
*Bioinformatics*, 23(20):2700–2707, 2007.
-  Mark J. Dunning, Mike L. Smith, Matthew E. Ritchie, and Simon Tavaré.  
*beadarray: R Classes and Methods for Illumina Bead-based Data*.  
*Bioinformatics*, 23(16):2183–4, 2007.
-  John D. Storey.  
*The Positive False Discovery Rate: a Bayesian Interpretation and the q-value*.  
*The Annals of Statistics*, 31(6):2013–2035, 2003.
-  Y. Benjamini and Y. Hochberg.  
*Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing*.  
*J. R. Statist. Soc. B*, 57:289–300, 1995.

## Genome-Wide Analysis of Genetic Predisposition in Patients with the History of Acute Myocardial Infarction

ZDENĚK VALENTA,  
MICHAL KOLÁŘ, HANA  
GRÜNFELDOVÁ,  
IVAN MAZURA, PETRA  
FEGLAROVÁ AND JANA  
ZVÁROVÁ



The Aim and Methods

Study Design

Sources of Variability

Background Noise

Target Populations

Handling Multiple Tests  
of Significance

Samples Description

Understanding the  
Data

Modelling Results

Discussion

References