Universitätsklinikum Hamburg-Eppendorf

Department of Andrology, UKE Hamburg

H. Cappallo-Obermann, W. Schulze and A.-N. Spiess

By grant Sp721/1-4 of the Deutsche Forschungsgemeinschaft

A cross-platform/cross-laboratory microarray study as a powerful tool to reveal gene expression signatures of male infertility

Aims & Approach

The molecular basis of idiopathic male infertility is largely unknown. Gene expression profiling of normal and pathological human ejaculates/spermatozoa has been shown to be a vital tool to identify causes on a molecular level. **We present a cross-laboratory/cross-platform microarray study with gene expression profiles of 127 human ejaculates/spermatozoa.** Involved are donors/patients belonging to different groups in respect to fertility status and spermiogram parameters (according to WHO guidelines, 2010).

25 ejaculates with different outcomes of IVF treatment (fertilization rates, pregnancy rates) were collected at the Fertility Center Hamburg. RNA was isolated and whole genome microarrays (Codelink, 55k) were hybridized. For cross-platform analysis, seven sets of raw data from 5 publications were additionally downloaded from the GEO database (NCBI): Platts *et al*., 2007; Linschooten *et al*., 2008; Lalancette *et al*., 2009; Pacheco *et al*., 2011; Jodar *et al*., 2012. **Overall, this resulted in a final dataset of 127 samples from 8 investigations, 6 laboratories and 5 microarray platforms.**

All data were background corrected, log-transformed and quantile normalized (*Affy* package, Bioconductor). Datasets were merged by a set of 13751 EntrezID's present in all platforms. In case of multiple probes targeting one EntrezID, the one with highest MAD (Median absolute deviation) was chosen. Batch effects were eliminated using the *ComBat* package for the R statistical programming language.

Results

The 127 samples obtained from 8 different microarray investigations of human spermatozoa (including our own) gave an overall hybridization pattern as shown in Figure 1. As typical for the different microarray platforms (Affymetrix, Codelink, Agilent, Illumina), a significant difference in the magnitude (y-value) and dynamic range (length of boxes) is noticable.

The complete microarray dataset was transformed by quantile normalization (Figure 2), which normalizes all fluorescence values to a common range. This procedure is a prerequisite in all common microarray studies.

However, when analyzing this complete dataset using standard clustering methods such as Principle Component Analysis (PCA) or Hierarchical Clustering (HCL), one observes that the „batch effect“, i.e. the dominant effect of microarray platform/laboratory has not been adequately removed: In the PCA (Figure 4) as well as in the HCL (Figure 6), the samples are separated clearly by the platform/study from which they were derived.

Contrasting this, a removal of the „batch-effect“ (Figure 3) results in a complete mixture of samples in which the effect of microarray platform/laboratory has been successfully eliminated and is not evident in clustering by PCA (Figure 5) or HCL (Figure 7). This modified dataset was used to investigate gene expression signatures in respect to potential targets of male infertility.

n a first step, we filtered the top 200 variant genes across all 27 samples, an approach usually conducted to enrich for genes with potential correlation to some outcome without imposing a pre-defined grouping structure. Interestingly, the most significantly enriched functional category (GO-Terms) was „Translation“ (Table 1), consisting mainly of transcripts for ribosomal proteins of the large/small ribosomal subunits and elongation factors/co-factors.

n a next step, we filtered differential genes in those samples for which data for fertility outcome was available (94 of 127, top color bar in Figure 7). By this approach we obtained 383 transcripts which were highly significant even with the most conservative Bonferroni correction ($p_{bonf} < 0.05$). Clustering these genes by PCA resulted in a good separation of the fertile (coded in green) and the infertile (coded in red) samples (Figure). Again, a following analysis for functional enrichment of these differential genes indicated a prevalent role of translation-associated transcripts (Table 2). Consequently, we further interrogated a subset of 19 transcripts for ribosomal proteins in respect to their correlation with fertility outcome. Although these genes exhibited a highly differential pattern, the data was heterogeneous (Figure 9): while in some investigations (UKE, Platts Affy) ribosomal transcripts were downregulated, the converse was true for others (Jodar and Platts Illumina data).

Conclusions and Perspectives

When merging datasets from different microarray platforms/laboratories the main challenge is to overcome non-biological technical bias while keeping an optimum of biological information. By eliminating this "batch-effect", we were able to extract vital information in respect to fertility outcome from a cohort of 94 samples that were derived from different investigations. The results suggest that the ribosomal compartment may play an essential role in disturbing the fertility outcome, which tallies with our observations on the RNA level (Cappallo-Obermann et al., 2011).

ferences:

Platts AE, Dix DJ, Chemes HE, Thompson KE, Goodrich R, Rockett JC, Rawe VY, Quintana S, Diamond MP, Strader LF, Krawetz SA.
ccess and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum Mol Genet. 2007 Apr 1;16(7):763-73.

Linschooten JO, Van Schooten FJ, Baumgartner A, Cemeli E, Van Delft J, Anderson D, Godschalk RW.
of spermatozoal mRNA profiles to study gene-environment interactions in human germ cells. Mutat Res. 2009 Jul 10;667(1-2):70-6.

Lalancette C, Platts AE, Johnson GD, Emery BR, Carrell DT, Krawetz SA.
ntification of human sperm transcripts as candidate markers of male fertility. J Mol Med (Berl). 2009 Jul;87(7):735-48.

Pacheco SE, Houseman EA, Christensen BC, Marsit CJ, Kelsey KT, Sigman M, Boekelheide K.
egrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. PLoS One. 2011;6(6):e20280.

Jodar M, Kalko S, Castillo J, Ballescà JL, Oliva R.
ferential RNAs in the sperm cells of asthenozoospermic patients. Hum Reprod. 2012 May;27(5):1431-8.

ppallo-Obermann H, Schulze W, Jastrow H, Baukloh V, Spiess AN.
ghly purified spermatozoal RNA obtained by a novel method indicates an unusual 28S/18S rRNA ratio and suggests impaired ribosome assembly. Mol Hum Reprod. 2011 v;17(11):669-78.

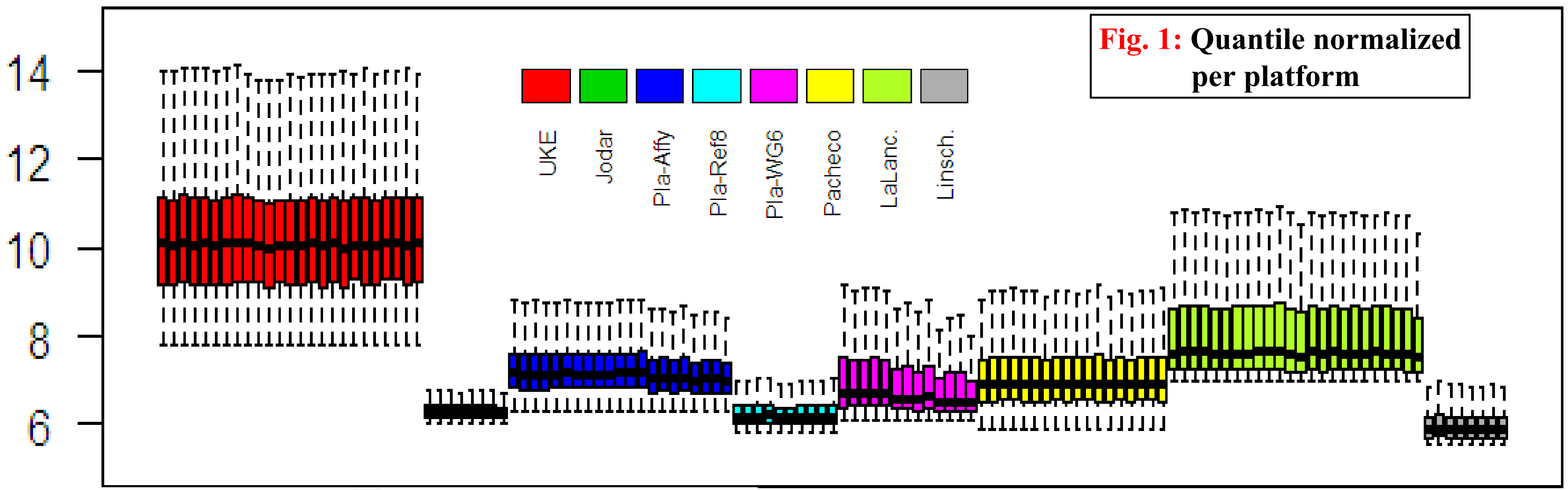


Fig. 1: Quantile normalized per platform

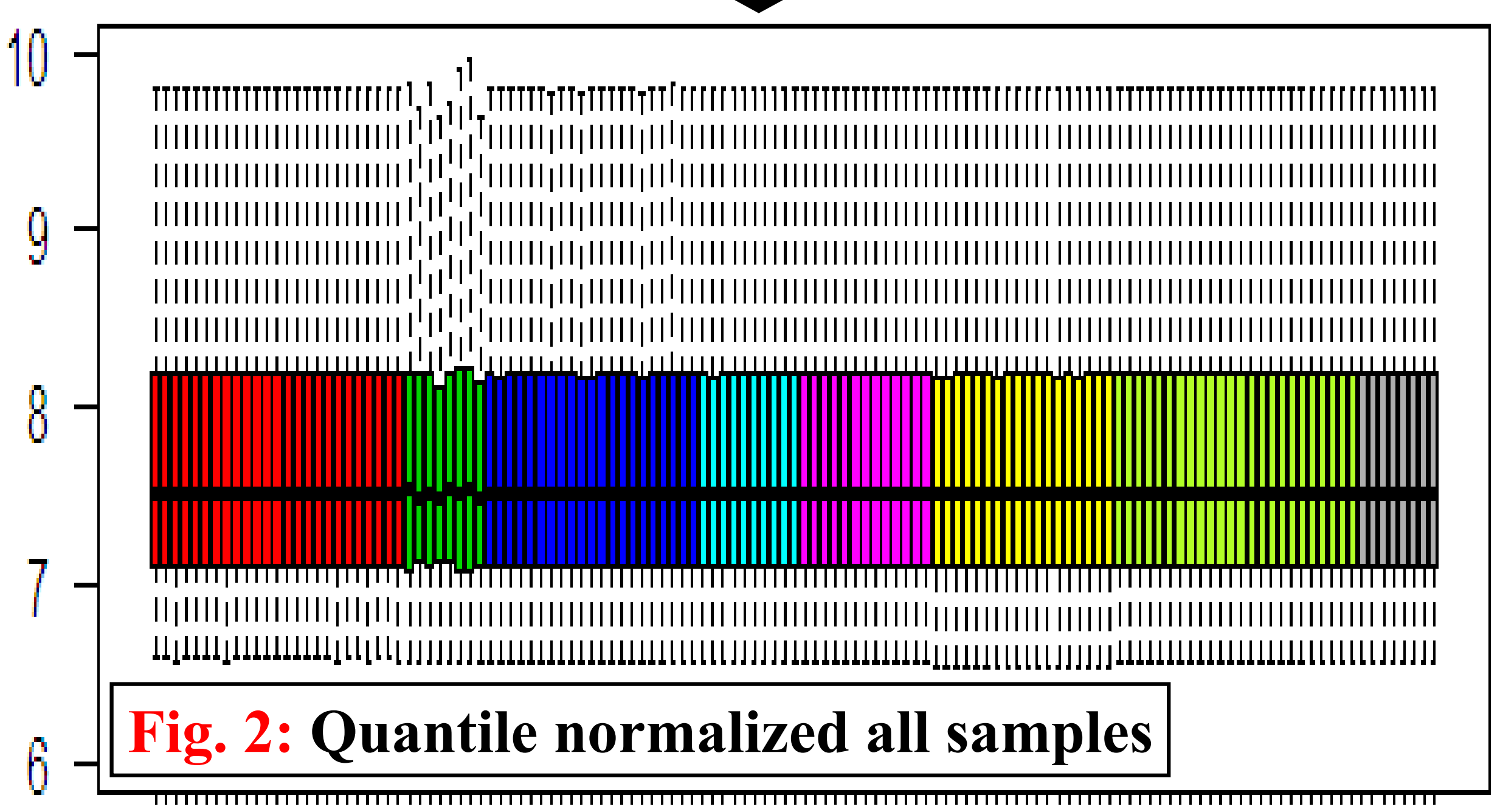


Fig. 2: Quantile normalized all samples

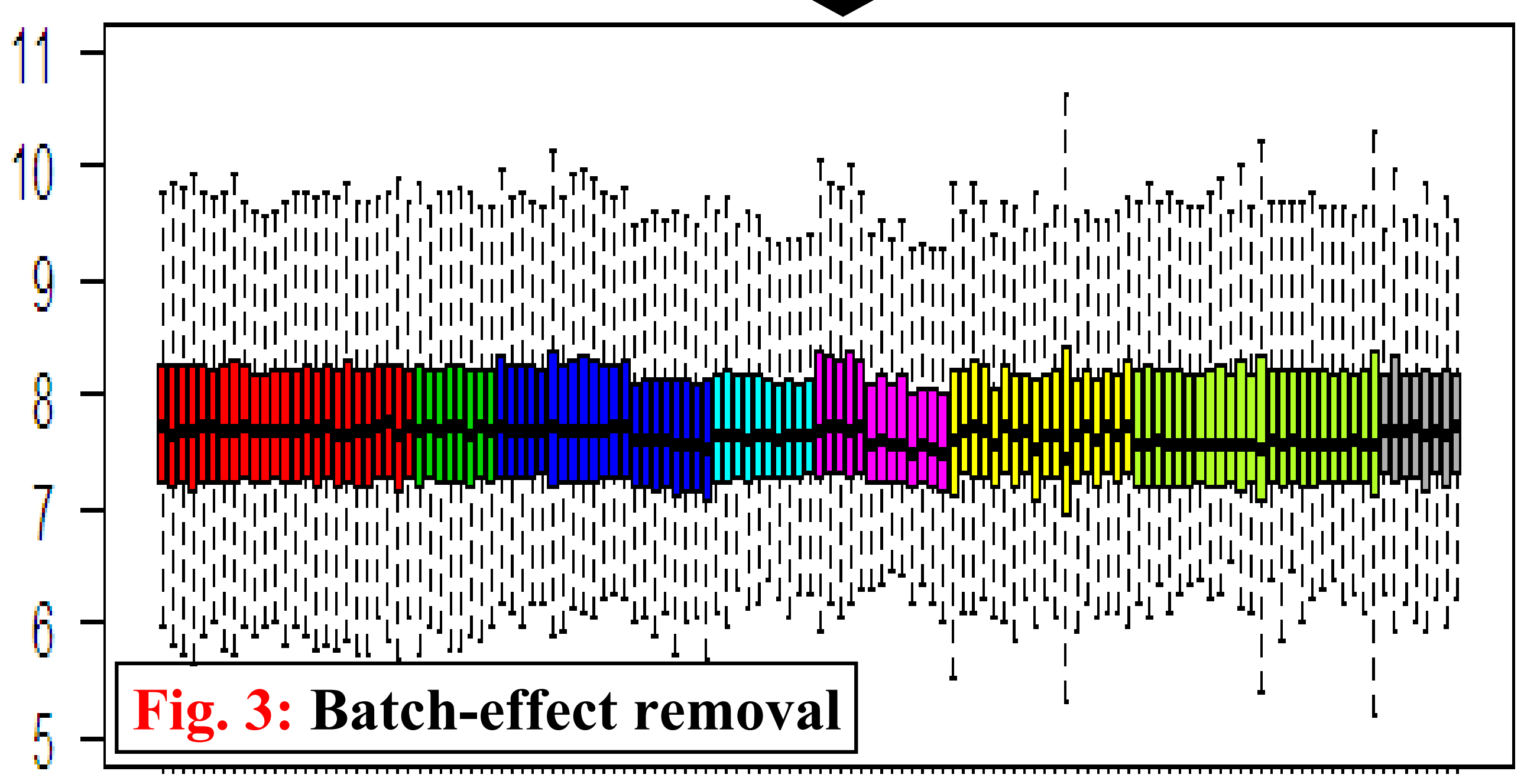


Fig. 3: Batch-effect removal

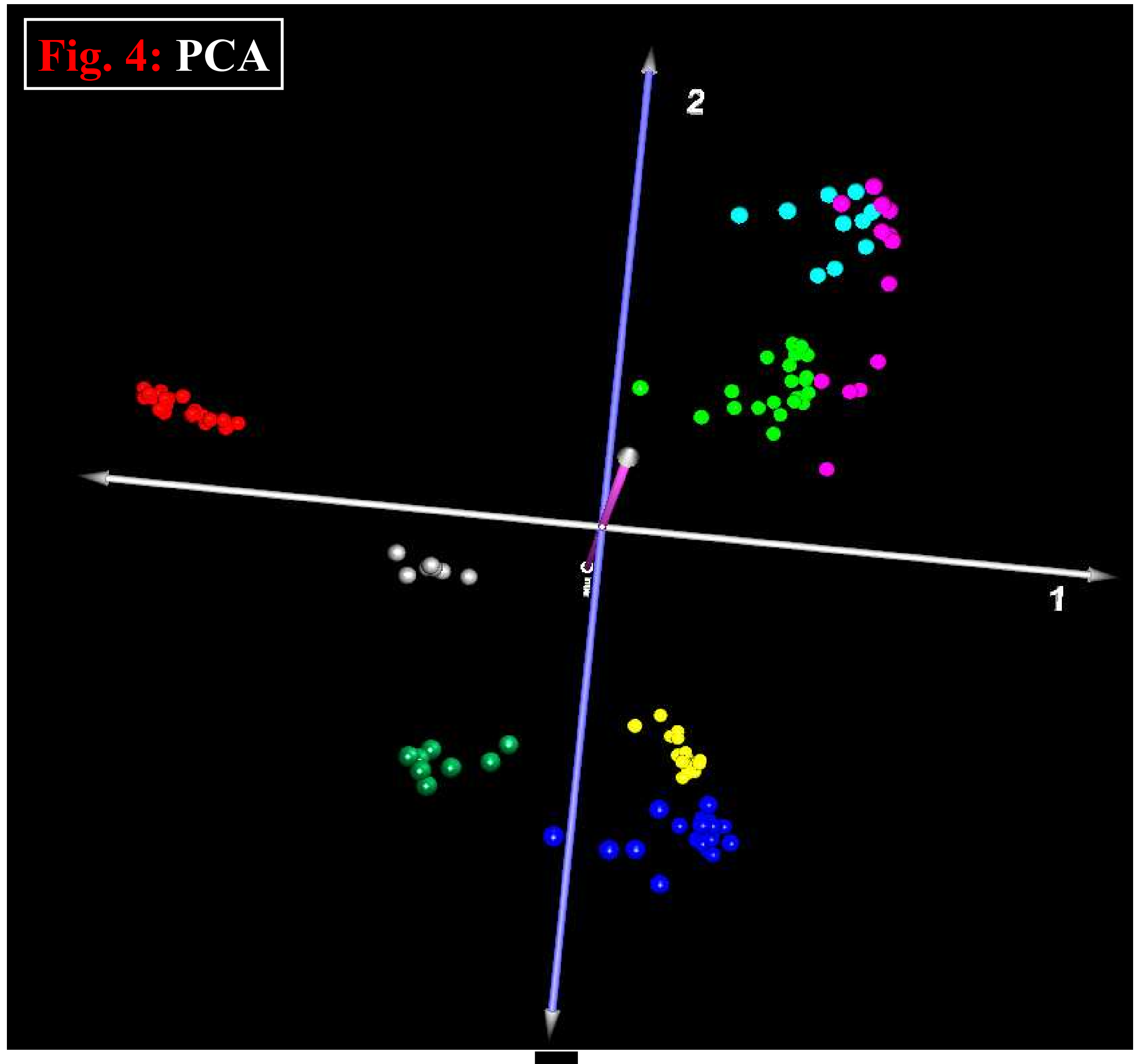


Fig. 4: PCA

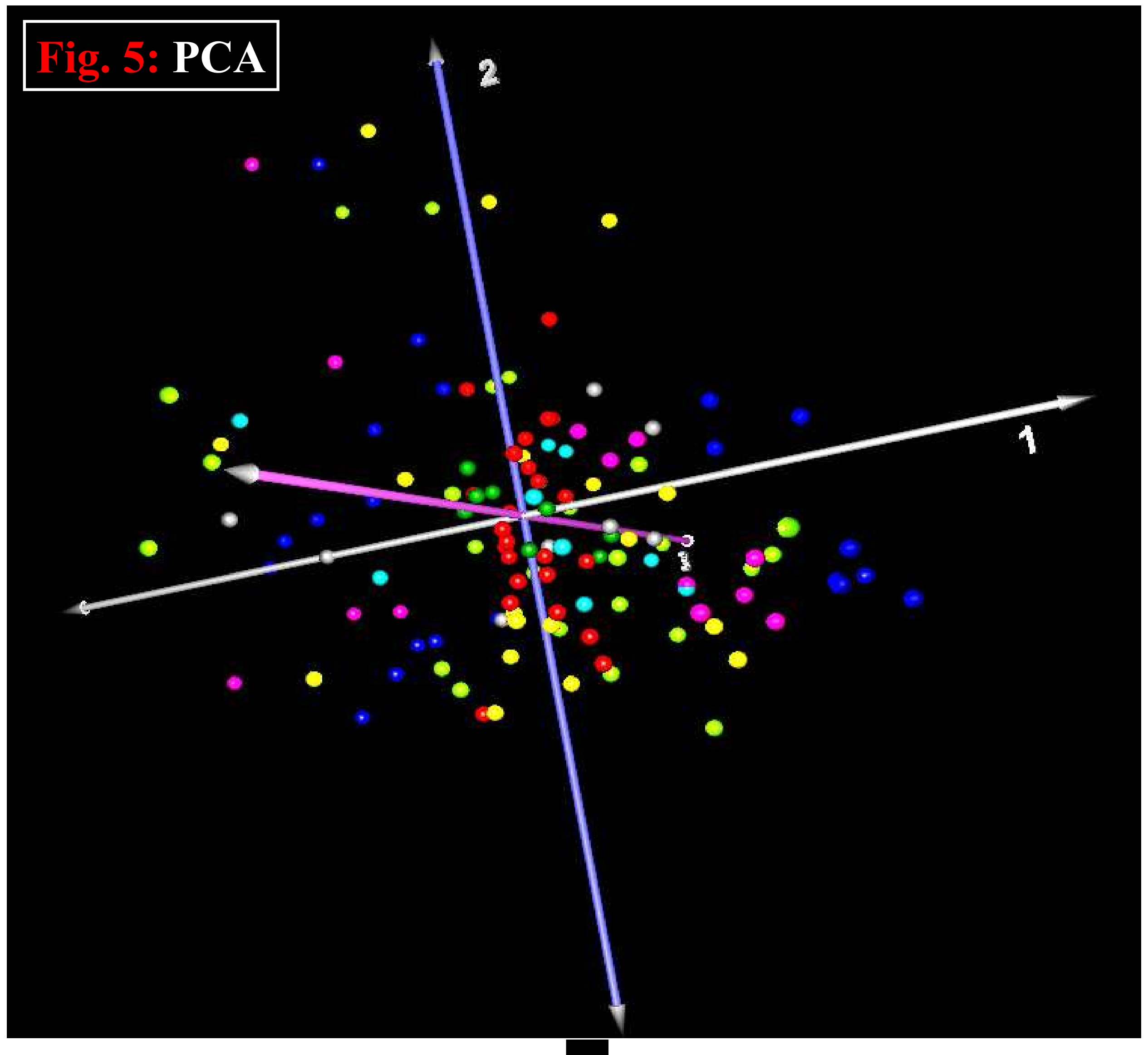


Fig. 5: PCA

Fig. 6: HCL

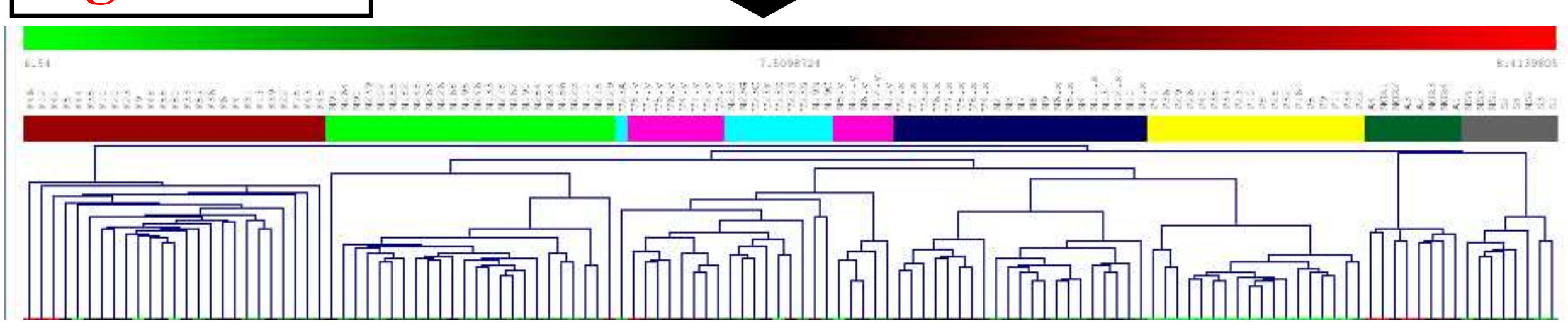


Fig. 7: HCL

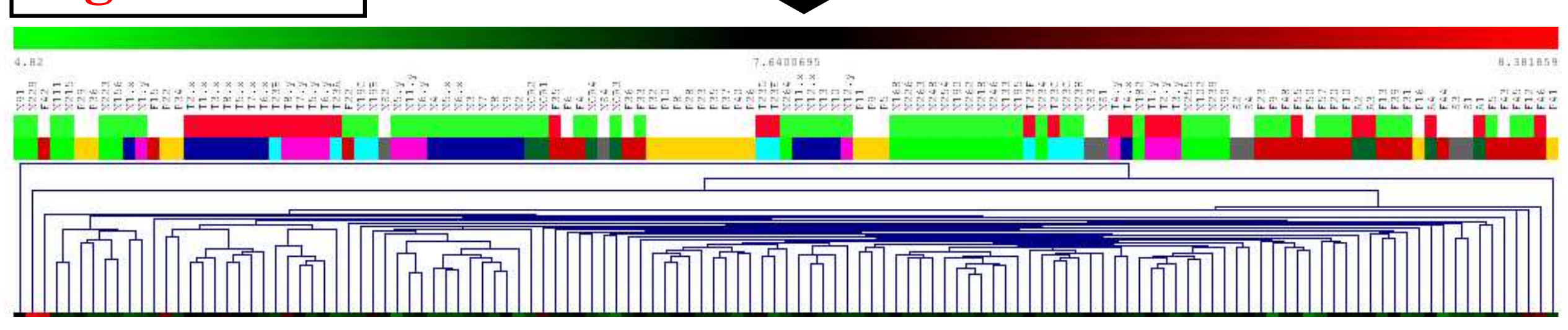


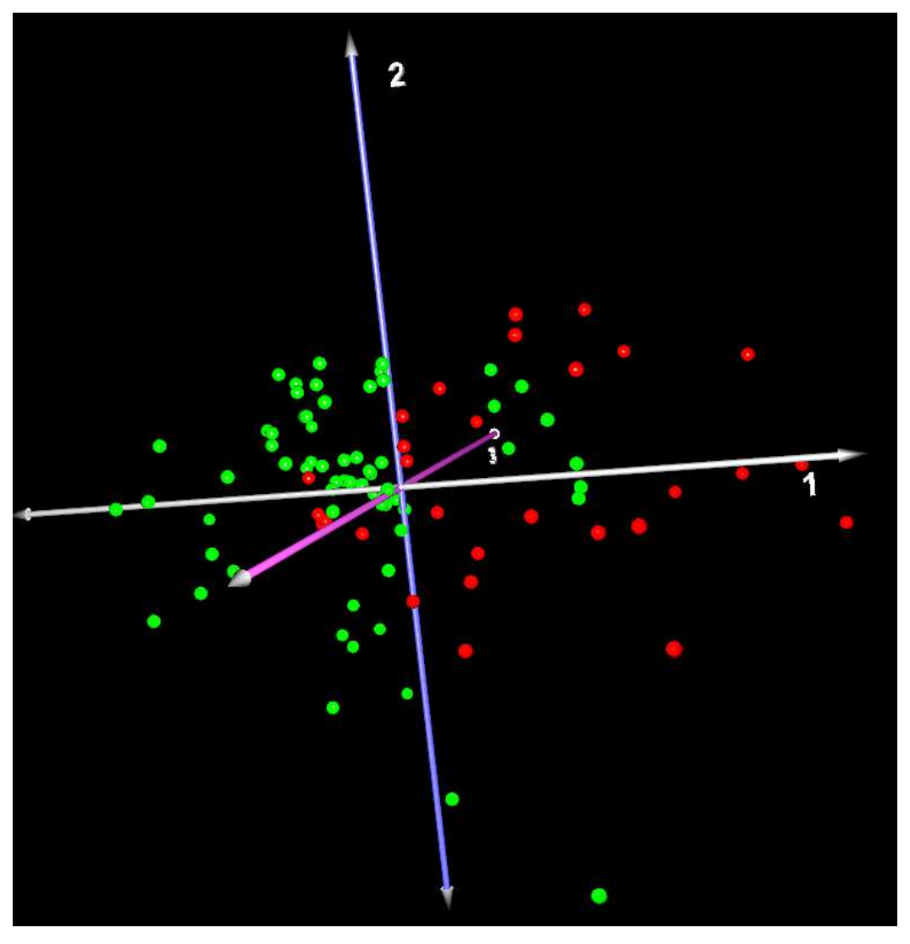
Table 1: Top-variant categories

Go_Term (BP_FAT)	p.value	p.Bonferroni	Genes
Translation	6.1E-09	9.4E-06	EEF1A1, MRPS15, COPS5, RPL35, RPL27, RPL24, RPS6, RPS3, GSPT1, RPS16, RPS3A, RPL13A, RPL6, EIF4A2, EIF3E, RPS14, GSPT2, RPL3, RPL4, RPL7A
Response to Inorganic Substance	5.4E-04	5.6E-01	AQP9, DUSP1, TFRC, HMOX1, GPX4, ANXA11, NDRG1, MT1H, CALM2, SOD2
Negative Regulation of Transcription Factor Activity	1.6E-03	9.2E-01	FOXJ1, HMOX1, NFKBIA, RPS3, TRIB1
Cell Cycle	2.9E-03	9.9E-01	MAEA, RABGAP1, IL8, ANXA1, RPL24, MLF1, SESN3, KIF2B, CCNB2, PSMA6, DUSP1, GSPT1, NSL1, PSMC2, GSPT2, G0S2, PPP1R15A, CALM2, CCAR1
Response to Metal Ion	3.4E-03	9.9E-01	AQP9, DUSP1, TFRC, ANXA11, NDRG1, MT1H, CALM2
Inflammatory Response	3.9E-03	1.0E+00	CEBPB, S100A8, IL8, TFRC, CD44, CCL20, CXCR4, HMOX1, ANXA1, NFKB1, ITCH

Table 2: Top categories Fertile vs. Infertile

GO_Term (BP_FAT)	Count	%	p.value	p.Bonf.	Genes
Translation	21	16.9	7.0E-14	5.9E-11	EEF1A1, MRPS15, COPS5, NARS, RPL27, RPL24, RPS6, KARS, RPS3, EIF4G3, RPL32, RPS3A, RPLP0, EIF3E, RPL3, EIF3L, RPS13, RPL5, RPL4, RPL7A, UBA52
Ubiquitin-Dependent Protein Catabolic Process	10	8.1	4.2E-05	3.4E-02	PSMB7, PSMB1, UBE3A, PSMC2, SKP1, TCEB1, CUL4B, UBA52, BUB3, CUL1
Ribosome Biogenesis	6	4.8	1.5E-03	7.2E-01	RPLP0, RPL5, RPL24, RPL7A, RPS6, NSA2
Fertilization	5	4.0	2.1E-03	8.3E-01	PLCZ1, ZPBP, SMCP, KLHL10, SPA17
Glucose Metabolic Process	6	4.8	4.1E-03	9.7E-01	LDHC, LDHA, PDK4, PDHA2, PRKAA1, PPP1CC
Binding of Sperm to Zona Pellucida	3	2.4	5.9E-03	9.9E-01	ZPBP, SMCP, SPA17
Cell Cycle	13	10.5	6.7E-03	1.0E+00	CCNH, RPL24, SKP1, PPP1CC, SESN3, PSMB7, PSMB1, PSMC2, CUL4B, BUB3, CUL1, UBA52, CALM2
Ribonucleoprotein Complex Biogenesis	6	4.8	8.0E-03	1.0E+00	RPLP0, RPL5, RPL24, RPL7A, RPS6, NSA2
Single Fertilization	4	3.2	8.1E-03	1.0E+00	PLCZ1, ZPBP, SMCP, SPA17
Cell-Cell Recognition	3	2.4	8.2E-03	1.0E+00	ZPBP, SMCP, SPA17
Hexose Metabolic Process	6	4.8	1.0E-02	1.0E+00	LDHC, LDHA, PDK4, PDHA2, PRKAA1, PPP1CC
Negative Regulation of Neuron Differentiation	3	2.4	2.1E-02	1.0E+00	CNTN4, CD24, TTC3
Sexual Reproduction	8	6.5	3.8E-02	1.0E+00	PLCZ1, ZPBP, SMCP, KDM3A, SPATA4, KLHL10, SPA17, TBPL1

Fig. 8: PCA Fertile vs. Infertile



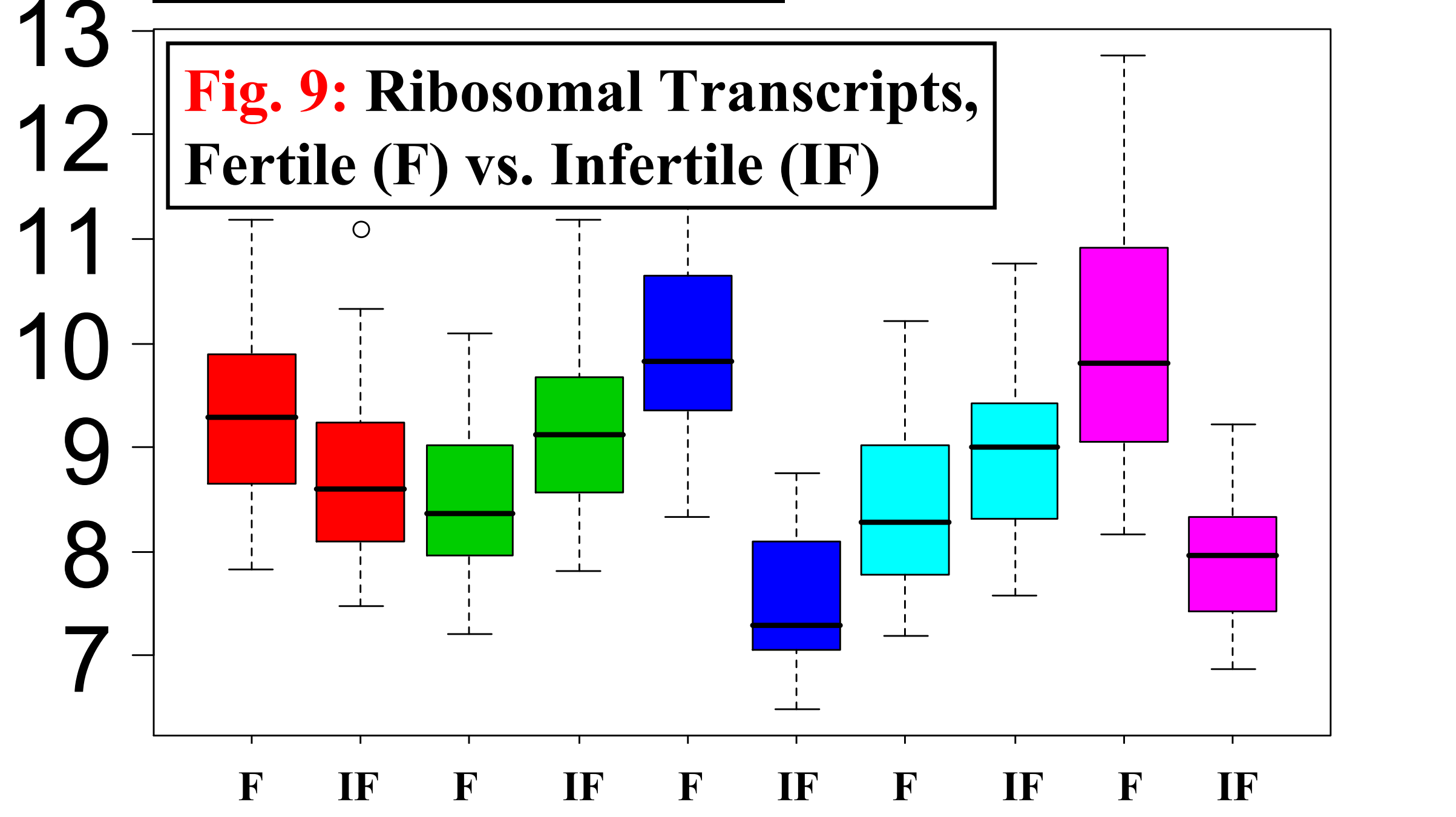


Fig. 9: Ribosomal Transcripts, Fertile (F) vs. Infertile (IF)