**ALMA MATER STUDIORUM UNIVERSITY OF BOLOGNA**

**FACULTY OF MATHEMATICAL, PHYSICIAL AND NATURAL SCIENCES**

**MASTER DEGREE IN MOLECULAR AND INDUSTRIAL BIOTECHNOLOGY**

**IN COLLABORATION WITH EVOLUTIONARY GENETICS**
**RIJKSUNIVERSITEIT GRONINGEN**

---

**MASTER DEGREE PROJECT**

# GENE EXPRESSION ANALYSIS OF CONDITIONAL LETHALITY IN DROSOPHILA MELANOGASTER INBRED LINES

**STUDENT**

***Kristina Gagalova***

**THESIS DIRECTOR**

***Prof. Alejandro Hochkoeppler***

**SUPERVISOR**

***Dr. Corneel Vermeule***

**3rd SESSION**
**ACADEMIC YEAR 2011/2012**

university of groningen

**MASTER DEGREE PROJECT**

# GENE EXPRESSION ANALYSIS OF CONDITIONAL LETHALITY IN DROSOPHILA MELANOGASTER INBRED LINES

**Student:** Kristina Gagalova

**Institute:** RijksuniversiteitGroningen

Evolutionary Genetics

**Thesis director:** Alejandro Hochkoeppler

**Supervisor:** Corneel Vermeulen

## ABSTRACT

Inbreeding depression is defined as the reduced fitness in closely related individuals. Deleterious effects of inbreeding depression are lower survival rate and lower individual fertility: these phenotypes are strongly enhanced under stressful environmental conditions. Characterization of inbreeding depression is an important issue in conservation biology, agriculture and ecology and the comprehension of undelaying mechanisms could lead to better understanding of a population's adaptive ability and extinction risk of endangered species.

Inbreeding depression is a complex phenotype, still poorly characterized at molecular mechanistic level. For the individuation of deregulated candidate genes in molecular and physiological patterns of inbreeding depression, *Drosophila melanogaster* model organism was used for the establishment of several inbred lines. From those, a line with lower survival rate after cold shock treatment was chosen as candidates for more detailed genetic and molecular investigation.

An explorative proteomic analysis, carried out through two dimensional gel electrophoresis and mass spectrometry, showed the differentially expressed proteins in the lethal inbred line, in comparison to the control lines, after exposure at harsh environmental conditions. In particular three proteins were expressed 48 hours after brief cold shock treatment (CG18067, β-Tubulin at 56 D, CG11796) and four proteins also at natural growth conditions (Angiotensin converting enzyme, Pro-phenol oxidase, Transaldolase, Phosphogluconate mutase). The main goal of this research project was the gene expression quantification through Real Time quantitative PCR (RT-qPCR) of the seven previously highlighted proteins and the correlation between protein and gene abundance. Missing molecular correspondence for most of the considered genes suggests a post-transcriptional and post-translational mechanisms of regulation involved in the inbreeding depression deregulated pathways.

RT-qPCR gene expression quantification was additionally used for validation of the six most significantly altered genes emerging from microarrays analysis, among the *D. melanogaster* lines exposed to stressful conditions. RT-qPCR results correlate positively with previous expression analysis.

# 1. INTRODUCTION

## 1.1 Genetics of inbreeding depression

### 1.1.1 Inbreeding in natural populations

### 1.1.2 Genetic mechanisms of inbreeding depression

### 1.1.3 Inbreeding depression and environment

## 1.2 Drosophila melanogaster models of inbreeding depression

### 1.2.1 Drosophila melanogaster as model organism

### 1.2.2 Drosophila melanogaster stocks

## 1.3 Molecular approaches for studying inbreeding depression

### 1.3.1 Proteome characterization

### 1.3.2 Transcriptome characterization

Inbreeding depression is generally defined as the reduced fitness in a given population, due to a high rate of breeding between related individuals: the result of inbreeding depression is documented and described in many species. The flowering plant species were the first to be identified as subjected to inbreeding depression for their self-fertilization reproduction. The advantage of wide spreading of plants in the environment is, in this case, associated with a drastic reduction of the offspring quality, if compared with cross-fertilization reproduction plant species (Charlesworth and Charlesworth, 1987). This effect of inbreeding depression in plants was first described by Charles Darwin, sustaining that cross-fertilization is favoured when inbreeding depression occurs. Inbreeding depression is broadly documented and described also in the animal kingdom, in particular for wild populations of endangered species. A representative example, regarding common species, are *Canis lupis*, *Cervus elaphus* and *Vipera berus* (Liberg et al, 2004; Slate et al, 2000; Madsen et al, 1996): field studies and direct observations show drastic weakening of adult specimens and lower litter size in inbred populations. Inbreeding was also frequent in human modern society around the 19$^{th}$ and the first half of the 20$^{th}$ century: also nowadays consanguineous marriages are common in archaic model based societies, because of religious and cultural traditions. Representative research was carried out on small populations, highly inbred because of historical and geographical reasons: some examples are the Hutterites, the Icelandic and Sardinian populations, involved in broad genetic studies regarding rare diseases (Hästbacka et al., 1992; deCODE company: Kim et al., 2009; Pardo et al., 2012). Highly inbred individuals in these areas carry lethal allele forms at high frequencies and this makes gene identification easier to seek.

Inbreeding depression characterization gains raising importance because of this evidence in natural populations (Keller and Waller, 2002). Moreover, characterization of common patterns in inbreeding depression could lead to improvement of prospective in several disciplines such agriculture, conservation biology and ecology, focusing, in particular, on the adaptive ability and extinction probability of populations.

# 1.1 GENETICS OF INBREEDING DEPRESSION

## 1.1.1 Inbreeding in natural populations

Alleles frequency in a wild population are subjected to mutation, genetic drift and purifying selection. These events gain capital importance of determining alleles fixation.

The phenomenon of inbreeding occurs for modifications in natural populations and it's guided by two mechanisms: the first is the subdivision of an original population and the drift to a smaller one, due to natural events (such migration, geographic barriers, etc). The lack of unrelated specimens induces mating with inbred partners. The second is a non-random mating, deriving from the fact that some individuals are more desirable to mate with for their phenotypic characteristics.

The main short term genetic characteristic of an inbred population is the increased rate of homozygosity, resulting from related individuals mating. Alleles, derived from the same ancestral copy, substitutes genetic variation of heterozygosity in few generations. Fixed homozygous genotype from common ancestor is added to naturally present homozygous genotype, increasing the probability of having homozygous individuals for all the alleles. The homozygous offspring carrying an allele from one of the ancestors, called identical by descent (IBD), is indicated as autozygous while individuals carrying outbred homozygous alleles are indicated as allozygous.

The quantification of inbreeding depression used for this project is the pedigree coefficient of inbreeding (*F*) and is defined as the probability that two genes in an individual are identical by descent. *F* is relative to a particular ancestral generation, assuming that at a given pedigree point individuals are unrelated (inbreeding equal to zero). Inbreeding coefficient is equal to coefficient of ancestry, ranging from 0 (unrelated individuals) to 1 (same individuals). This coefficient doesn't consider mutations and for this reason is partially overestimated.

Increasing coefficient of inbreeding is highly related to population size: if is the number of individuals is lower, less generation (*t*) are necessary for reaching a certain level of autozygosity. Figure 1.1 below shows that inbreeding coefficient grows at every *t* as a function of breeding adults (*N*), according to the following function:

$$Ft = 1 - (1 - 1 / 2N)^{t} \qquad (1)$$

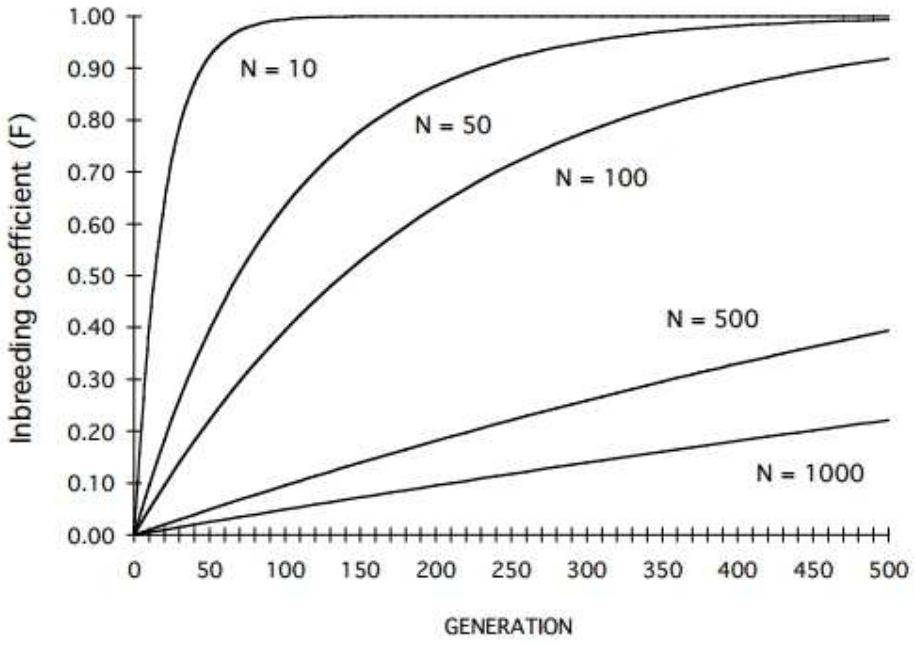



**Figure 1.1** *Autozygosity expressed in terms of inbreeding coefficient F in relationship with breeding individuals. For smaller populations (N= 10, 50) high inbreeding rate is reached after less generation than wider populations. "Genetics of Populations"- Hedrick PW (2005)*

Smaller is the population size and faster is the augment of inbreeding coefficient F. Heterozygosity is strongly related to this coefficient (see Figure 1.2) and the number of breeding adults in the population plays a fundamental role in both the mechanisms. Heterozygosity and *Ft* (inbreeding at every generation) are linked through the following function:

$$Ht = H0 \,.\, (1- Ft) \qquad (2)$$

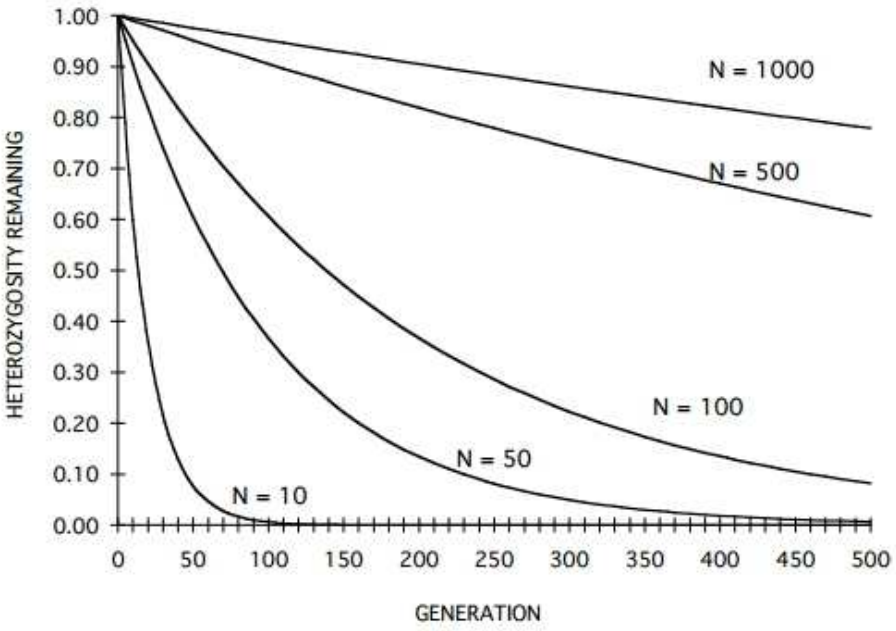



**Figure 1.2** *Heterozygosity in relationship with generations of inbred individuals mating. Smaller populations (N= 10, 50) decrease heterozygosity in few populations, on the contrary to wider populations (N=100, 500). "Genetics of Populations"- Hedrick PW (2005)*

## 1.1.2 Genetics mechanisms of inbreeding depression

The increased homozygosity in inbred populations is a mechanism that affects fitness in at least two genetically distinct ways. The first genetic mechanism involves the recessive deleterious mutations, present in outbred populations at low frequencies. Spontaneous mutations arise at both genomic and mitochondrial level (Keightley et al., 2009) and could be either eliminated through genetic drift or natural selection or drift to fixation by random chance. Deleterious mutations in small size populations, compounded by less than 100 individuals are more prone to lead to extinction due to the reduced efficiency of selection within individuals.

In this hypothesis, also called dominance hypothesis, only homozyotes, carrying the same two copies of a deleterious allele, are subjected to detrimental effects, while in heterozygosis, wild-type-allele compensate detrimental effects, maintaining the "healthy" phenotype because of the dominance over the mutated allele. Since

homozygosity increases during inbreeding, the expressions of deleterious alleles will too (inbreeding depression). In the partial dominance hypothesis, more supported than the pure dominance hypothesis, a combination of deleterious alleles are involved in the deleterious effect of inbreeding depression. Deleterious effects of inbreeding depression could be produced also at mild levels of dominance, on condition that deleterious alleles are at least partially recessive. The effects of deleterious recessive alleles have been broadly studied in *Drosophila melanogaster* (Crow, 1979) and in vertebrates such as *Danio rerio* (Zebrafish) (McCune et al., 2003). From these studies, involving several generations of siblings mating emerges the evidence of two distinct classes of deleterious alleles: one class of recessive lethal alleles is overt with high penetrance of lethal morphological abnormalities during early life stages. A second class of deleterious alleles is evident in late developmental stage, like partially dominant - later acting mild deleterious alleles, with unclear-cut penetrance; the phenotype of the second class lethal alleles is represented by morphologically cryptic phenotypes, afflicting fertility, reproduction and low adult survival rate. Both these classes of deleterious mutation carried by inbred individuals' lower population mean fitness.

Together with the lethal recessive alleles hypothesis, another genetic mechanism explains the lower fitness in inbred individuals, the overdominance hypothesis, due to the decrease of heterozygosity. This theory sustains that heterozygote genotype is superior to either homozygote dominant or homozygote recessive genotype, showing a higher relative fitness than parent individuals. This overdominance hypothesis was first supported by the plant geneticists George Shull and Edward East ("Hybrid Seed Corn", 1946;" "Studies on Size Inheritance in Nicotiana", 1916), upholding the concept that hybrid plants are more vigorous and prone to generate healthy offspring than the inbred ones, phenomenon described as heterosis or the increased function of any biological quality in hybrid specimens.

The following Figure 1.3 shows schematically the dominance hypothesis at single locus, the partial dominance hypothesis involving multiple deleterious alleles and the overdominance hypothesis, considered at single locus. In both the partial dominance

and the overdominance hypothesis heterozygote offspring shows a higher fitness than the parental homozygote genotype.

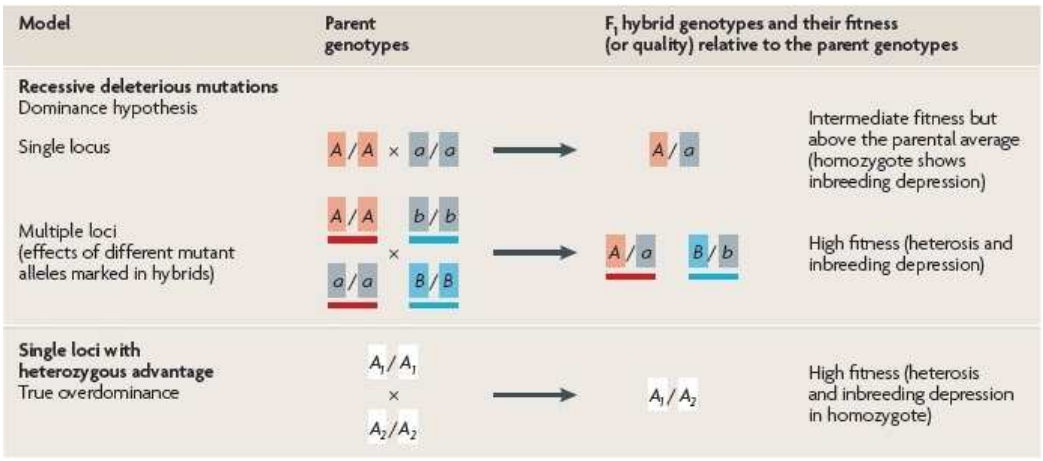



**Figure 1.3** *Main genetic hypotheses in inbreeding depression, involving single or multiple loci dominance hypothesis (partial dominance) and overdominanance hypothesis. Charlesworth and Willis, 2009*

Considering inbreeding at phenotypic level points out that both the genetic mechanisms above mentioned play important role for lower fitness. Even if deleterious mutations are more supported, overdominance plays an important role in some fitness reducing additive effect characteristics such early female fecundity, male mating success and viability observed in *D. melanogaster.* In fine, it's not possible to suggest leading mechanism acting at molecular level since some observed inbreeding depressed phenotypes are caused by alleles present at intermediate frequencies.

The approach for this kind of genetic studies is through quantitative trait locus mapping or QTL, for individuating genome regions that contain specific genes involved in the trait. However, QTL mapping, even though has become a sophisticated methodology of investigation, is still rough technique for his low resolution. Even if results lead to overdominance as evidence of reduced fitness, it's

not possible to leave out the hypothesis of linked recessive lethal mutations in repulsion phase. That's why identification of underlying mechanism is still unclear.

Our understanding of inbreeding depression is still poor: the innovative approach enclosing data of molecular population genetics and genetic variance in fitness is an effective contribute for shedding light on inbreeding depression complexity.

## 1.1.3 Inbreeding depression and environment

Environment plays a key role for inbred populations: inbreeding depression effects and gene expression could be highly different under various (in particular more severe) environmental conditions. Characterization of interaction between small size populations and anthropogenic environment changes, fundamental for conservation biology, aims to prevent extinction risk for both plants and animals in natural habitats.

Long term fragmentation of original populations with consequent arise of inbreeding coefficient limits the adaptive ability of subpopulations: the possible explanation of the missing adaptation are the variability at genetic level, the selection pressure and a high tolerance level at adverse environmental conditions of inbred individuals (Bell and Gonzalez, 2009).

The panorama of data included in complete studies, related to inbreeding depression and switch of environment for inbred individuals, is at the moment heterogeneous. A metanalysis, looking at an accurate coverage of the main studies relative to environment and inbreeding depression, was carried out (Armbruster and Reed, 2005): the considered parameter for inbreeding interaction with environment is the number of haploid lethal equivalents, B (Morton et al., 1956), calculated through the following formula:

$$B = -1/ F. \ln(wf/w0) \qquad (3)$$

F is the coefficient of inbreeding and wf – w0 are respectively the mean fitness for the inbred population and the fitness in outbred population. B is particularly suitable for correlation because of the relative comparison between studies, normalized for effects of inbreeding at different levels. The following Figure 1.4 shows the number of studies (count) under benign and stressful conditions, with a significant increase of B. Higher is this factor and more severe are the effects of inbreeding depression on fitness. Calculated median of 0.85 lethal equivalents, from selected studies, under benign environment and of 1.45 for stressful environment are significantly different: the median number, reflecting inbreeding depression at stressful environment, is calculated to be 69% greater than the effects under benign environment.

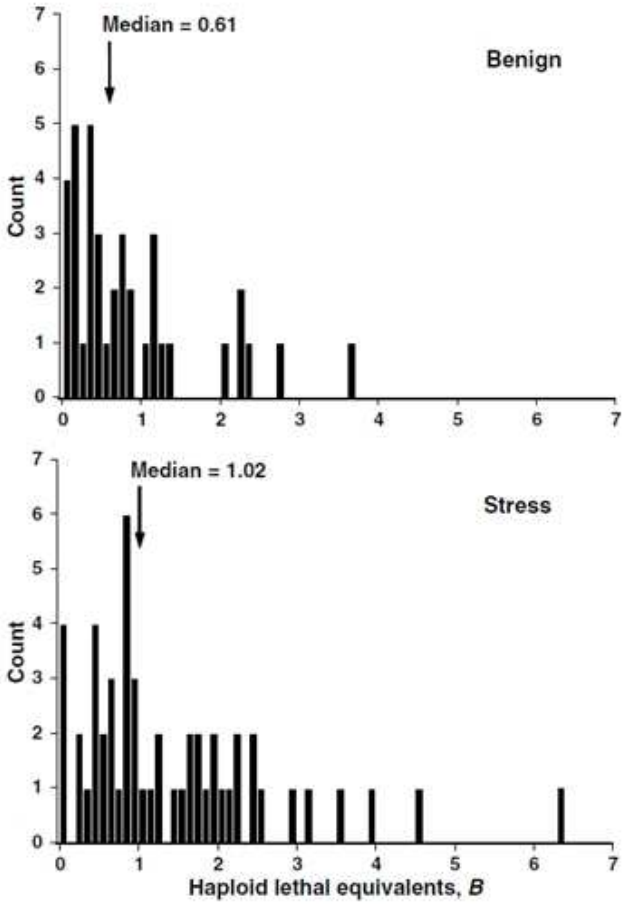



**Figure 1.4** *Metanalysis of 32 representative studies showing number of lethal equivalents of inbred populations under benign and stressful environmental conditions. Severe environment conditions have deeper negative impact on population fitness. Armbruter and Reed, 2005*

Further studies, focusing on inbreeding depression in *D. melanogaster*, show enhanced prevalence of lethal effects under stressful conditions (Bijlsma et al., 2000). Effects of inbreeding are indeed highly boosted by severe environment and fitness seems to be synergistically decreased, in interaction with these conditions. Expected pattern of this interaction is represented in Figure 1.5: fitness rate, at different quality of environment, shows more extreme pattern of interaction, given by the synergistic effects of I X E (inbreeding-by-environment).

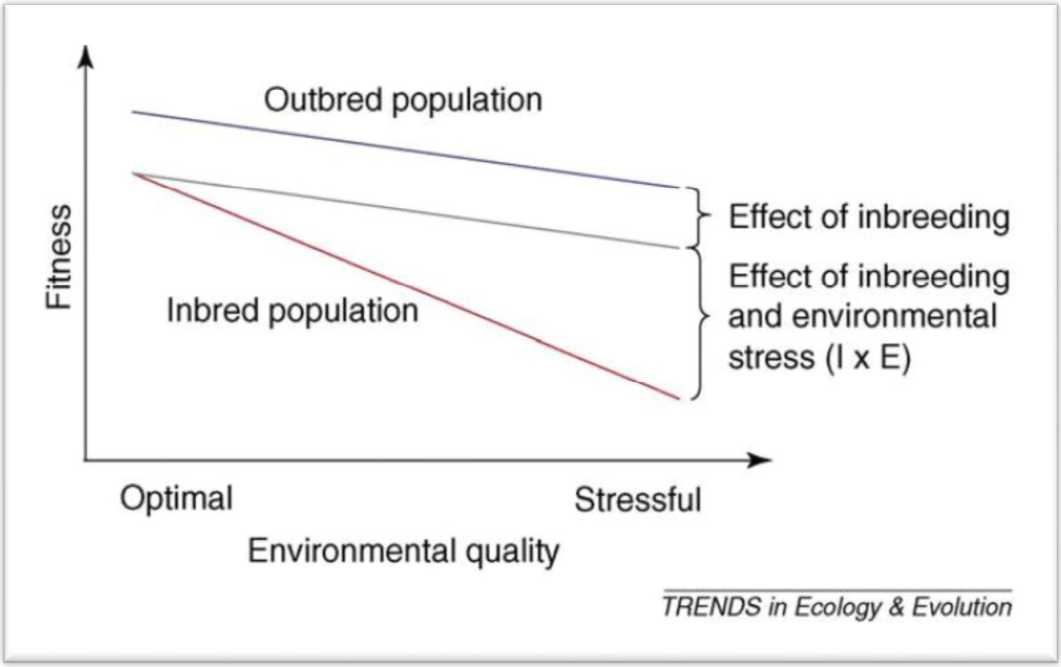



**Figure 1.5** *Hypothetical chart of fitness level, related to inbreeding and environment. Inbred populations are expected to have a synergistic interaction with stressful environment, leading to a steeper decrease of fitness in relation. Kristensen et al, 2009*

# 1.2 DROSOPHILA MELANOGASTER MODELS OF INBREEDING DEPRESSION

## 1.2.1 Drosophila melanogaster as model organism

*D. melanogaster* is widely recognized and studied model organism in conservation biology and environmental stress studies. Spontaneous highly recessive lethal mutations, considered fundamental mechanism in inbreeding depression, have been first characterized in this organism, considered particularly suitable for understanding complex interaction as environment and inbreeding (E X I) because of several reasons. First of all, obtaining small populations in laboratory from *D. melanogaster* wild type strains makes population analysis easier to handle (Frankham, 1995): in this way, survival is an quantifiable phenotype. Furthermore, another advantage of *D. melanogaster* handling is the possibility to generate multiple lines at different level of inbreeding, including also extreme ones. This experiment design allows calculation of genetic load, the rate at which fitness decreases in relationship of inbreeding coefficient, observed to follow a linear decline. At last, flies are easily submitted to harsh environmental conditions and to different experimental designs, varying intervals of treatment and level of harsh conditions. Fragmentation of original population into smaller ones coupled with changing in an environmental conditions in laboratory, regarding alteration in chemical (for example tolerance to ethanol) and physical conditions (such temperature variation), shows different survival rate in *D. melanogaster,* with the advantage of comparison between inbred populations submitted to environmental stress (Bakker et al. 2010). Even if, flies in nature are exposed to mild environmental changes, with fluctuating frequency, difficult to recreate with experimental conditions, laboratory investigation could lead to understanding of survival dynamics, testing hypothesis at different levels.

*D. melanogaster* populations make possible the evaluation of molecular response induced in inbred lines and also the comparison of differences and similarities for several environmental conditions. This approach makes the complexity of the problem more practical and possible to treat at quantitative level.

## 1.2.2 Drosophila melanogaster stocks

For analysing I X E interaction in mortality pattern, several *D. melanogaster* inbred lines were established in independent replicates, in parallel with their correspondent outbred controls (Vermeulen and Bijlsma, 2004). Diverse levels of inbreeding were reached through generations of brother-sister's mating from a *D. melanogaster* common base stock, G83. The coefficient of inbreeding of starting population was set at zero and F of inbred lines was estimated according to pedigree inbreeding. Few inbred lines, with adequate fecundity and viability, were chosen because of a highly marked effect on lifespan (described below) and maintained in culture.

Altering temperatures of growth in those lines and observing adult survival rate, conditionally lethal lines were identified, showing sensitivity to cold environment.

This line's origin from common base stock is a fundamental factor to reach an optimal experimental design. Genetic background or epistasis usually rules complex phenotypes, and in the case of inbreeding depression, survival. As the entire lines share a common origin and, in principle, have many alleles in common (excluding new mutations), lower complexity makes genotypic anomalies detection less complex. (Sarup et al., 2012)

Male adult mortality shows a temperature sensitive line's phenotype at low temperatures: in Figure 1.6 it's possible to observe a decreased survival rate in early stage (around 4 days after cold treatment) particularly for inbred L10 line, having 78% coefficient of inbreeding.

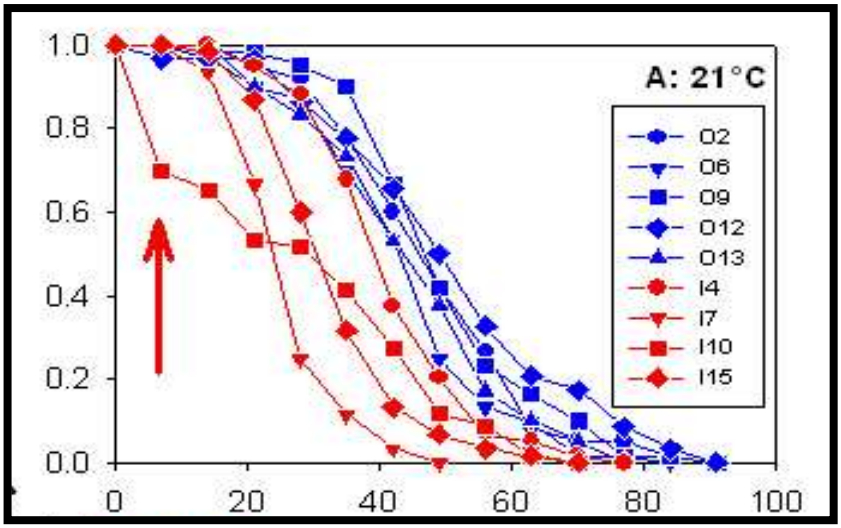



**Figure 1.6** *Cumulative survival probability (presented as percentage of live individuals), in relation to the age (expressed in days) of the individuals from different lines. Line L10 has a decreased survival rate in the early stage of life, in comparison with the others inbred lines (red lines). Blue lines represent outbred controls, independent populations of base stock. Vermeulen and Bijlsma, 2004*

In this L10 line specific inbreeding deleterious alleles were presumably fixed generation after generation. The inbred and the corresponding outbred control line, because of their characteristic lifespan pattern, are particularly suitable to be compared with the purpose to identify differences at both phenotypic and molecular level.

# 1.3 MOLECULAR APPROACHES FOR STUDYING INBREEDING DEPRESSION

Inbreeding depression is a well characterized mechanism at phenotypic level for decreased rate of fitness, involving survival rate and fertility. Anyway, molecular basis underlying this complex trait and molecular patterns are still unknown or poorly deepen.

Considering the difficulties related to genes interaction identification compounding the lethal *D. melanogaster* phenotype, a cross-analysis is necessary for investigating at different levels and pinpointing multiple genes interaction. Investigating around the three molecular levels, core part of the sequence hypothesis in molecular biology (information conveying from DNA, to RNA and finally to proteins), is a successful approach to identify possible key roles mechanisms involved in this phenotype. A full genomics characterization involves genomic DNA sequence, transcriptome and proteome, representing different aspects of this issue.

Several molecular mechanisms could affect fitness in inbred individuals if considering genetic mechanisms described before. In the dominance hypothesis at single or multiple loci, high resolution mapping highlights deleterious mutations at both coding and non-coding regions. Molecular population genetic analysis, including integrate data on markers about sequence divergence, discovers deleterious amino acids variants and mutated splicing consensus sequences on introns. This winning strategy is oriented for the recognition of structural modifications or eventual loss of protein functionality or stability. Proteins involved in altered signal transduction or biochemical processes could lead to reduced survival rate in *D. melanogaster* populations. The genome mapping for identification deregulated molecular pathways in case of dominance hypothesis, can't be applied to the overdominance hypothesis molecular mechanisms because of the difficult selection of molecular

markers tractable among crosses: this approach is indeed poorly accurate in this case.

## 1.3.1 Proteome characterization

Outlining a proteome expression is always been considered a priority in molecular biology. Protein profiling is defined as the characterization of the entire set of proteins present in specific conditions, identified with different techniques.

Protein profiling of *D. melanogaster* in our case compares different populations and is carried out with 2 Dimensional Gel- Electrophoresis (2-DGE), coupled with Mass-Spectroscopy. This technique is considered the only method currently available capable of simultaneously separating thousands of proteins, based on their mass and charge (Issaq et al, 2008).

2-DGE gel is used mainly for the detection of characteristic proteins expressed at population level from individuals carrying presumably the same homogeneous genetic make – up, under the same environmental conditions. Starting from this assumption it's possible to compare different protein expressions from several individuals representative of a population as a whole (Biron et al. 2006), at each environmental conditions and compare them to the corresponding controls. Together with the advantage of organism protein identification, this technique offers the benefit of a semi-quantitative analysis or the comparison of protein amount in relation with spot intensity.

In our research approach, proteome analysis is carried out to detect quantitative differences between lines and treatments. In the context of inbreeding depression, an explorative proteomic analysis was conducted in order to compare the proteome in Lethal Inbred Line (LI), and a Control Inbred Line (IC) with Outbred Line, submitting the tree lines to restrictive conditions in contrast to expression to permissive conditions (Vermeulen et al., unpublished). For setting up restrictive condition treatment, samples were exposed to a 30 minutes cold-shock treatment, consisting

of ice bath. As a control, a similar amount of fly samples were placed in a bath at growth temperature (25° C) for permissive condition. The table below shows experimental set-up of proteome profile: individuals from the tree lines undergo the same treatments, cold shock or temperature of growth.

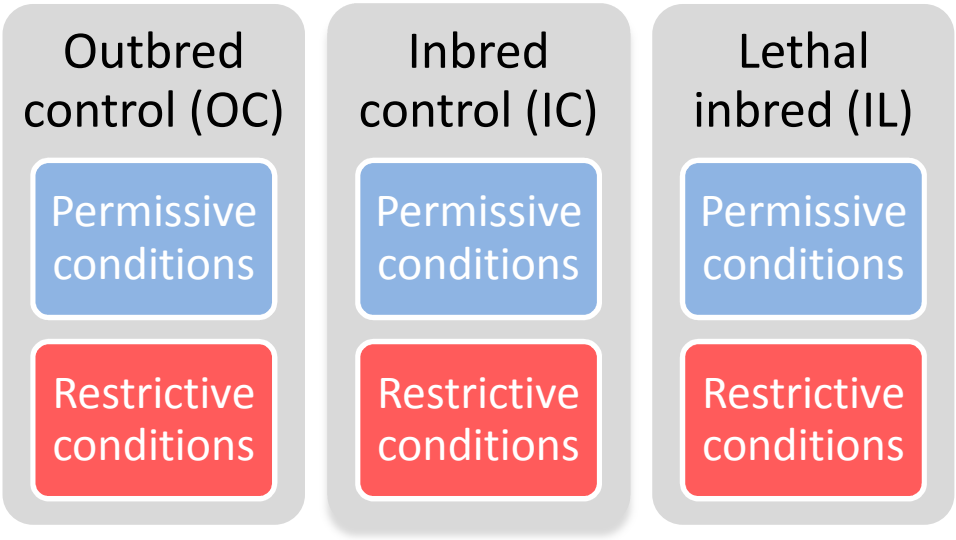



**Table 1.1** *Experimental set-ups for proteome 2DGE protein profile. Individuals from each line (coloured boxes) were submitted to different treatments, cold shock (restrictive condition) or temperature of growth (permissive condition)*

Proteins were extracted from samples, 48 hours after treatment, considered necessary time lapse to have the complete protein maturation, including protein assembly in complex domains and post transcriptional modifications. Intensity from gels, repeated tree times for each condition, compared to each other, shows characteristic protein expression for 48 proteins in the lethal line (IL), against inbred control (IC) and outbred line (OC).

In particular, exclusively focusing on lethal line protein expression, 7 proteins have a peculiar pattern. Here below the list of proteins with characteristic Lethal line expression, 3 expressed in IL at only restrictive condition and 4 expressed at both restrictive and permissive conditions:

| |
|---|
| CG18067 |
| •3’-5’ cyclic nucleotide phosphodiesterase |
| β- Tubulin 56 Dalton |
| Angiotensin converting enzyme |

**Table 1.2a** *Protein expression only at restrictive conditions (RC), exclusively in Lethal Line*

| |
|---|
| CG11796 |
| • 4 - Hydroxiphenil piruvate |
| Phosphogluconate muatase |
| Pro-phenol oxidase A1 |
| Transaldolase |

**Table 1.2b** *Protein expression at both permissive and restrictive conditions (RC - PC) exclusively in Lethal Line*

This characteristic expression pattern connects the proteins to a biological footprint, present in thermo-sensitive *D. melanogaster* lines.

The possible interpretation of these results splits in two the possible role of specifically expressed proteins: if considering different level of depression among inbred individuals and corresponding controls (Garcia et al., 2012), altered expression could be considered either dysfunctional or protective. Whether more inbred individuals show higher fitness than less inbred ones, altered expression could

play a protective role against effects of inbreeding: in this hypothesis, individuals failing to accomplish this change suffer from lower fitness. On the contrary whether expression of singular molecules increases with the augmentation in inbreeding depression, the explanation involves a non-protective hypothesis. Anyway this last pattern seems ambiguous because in contrast with conservation strategies.

Overlapping this hypothesis to our context, these seven proteins, not detectable in control lines and up-regulated characteristically in the lethal line, could trigger themselves deleterious effect if submitted to harsh environment, or could be direct response to stressful environment, in a way to contrast unsuccessfully the lethal effect.

## 1.3.2 Transcriptome characterization

Transcriptome analysis, also the quantitative detection of RNA messenger, is useful step in the characterization of dynamic processes as those involved in cold shock response. Even if complete transcriptome is not the final "player" involved in molecular pathway and it's only an intermediate measure of gene expression, it's a powerful tool to identify genes that are somehow involved in a specific process. Analysis of protein expression is an efficient technique to determine important changes at molecular level but, considering technical difficulties and awkward analysis, is still unable to carry on. Genome-wide expression studies could identify genes that are in some way involved in particular traits and are relatively easy and cheap. This field of analysis, in combination with the whole proteome expression analysis, enable us to gain a more complete and rounded view of the possible candidate genes involved in inbreeding depression.

Microarray technique is today's wide spread technique used for the simultaneous detection of transcripts; in particular, DNA microarray makes possible to identify thousands of transcripts and detect transcriptome variation in particular conditions. This throughput technique, consisting of immobilized DNA-oligo probes on silicon surface, detects through fluorescence the quantity of specific mRNA, characterized *a*

*priori*. This complex gene expression footprint, induced by the interaction of an organism with its environment, finds out specific genes involved in a trait and also provides valued information regarding the groups of genes altered after a particular environmental condition. For this reasons genome wide expression studies are particularly suitable applied to inbreeding-by-environment (I X E) as the biological response of *D. melanogaster* conditionally lethal lines.

Expression study (Vermeulen et al, unpublished) compares the previously described IC line with IL line, submitted to the same conditions described before in the proteome profile. Either a significant line-by-treatment interaction or a difference between lines in transcript abundance was taken criteria for selecting interesting candidate genes. The snapshot obtained from the microarray analysis presents the expression level variance of candidate genes involved in inbreeding depression, in response to harsh environment.

A DNA microarray is a technique that combines information from 40.000 probes and noisy signals and artefacts often occur in this complexity. This reasons lead to a validation of microarray's results for establishing the "reliability" of genome-expression wide study. The "gold standard" for nucleic acid quantification is today considered to be the Real Time quantitative PCR (RT-qPCR) because of the high sensitivity and reproducibility (VanGuilder et al, 2008). A typical genome-wide expression experiment usually follows a common outline, associated with validation. Following Figure 1.7 shows all the important steps conducted before microarray validation.

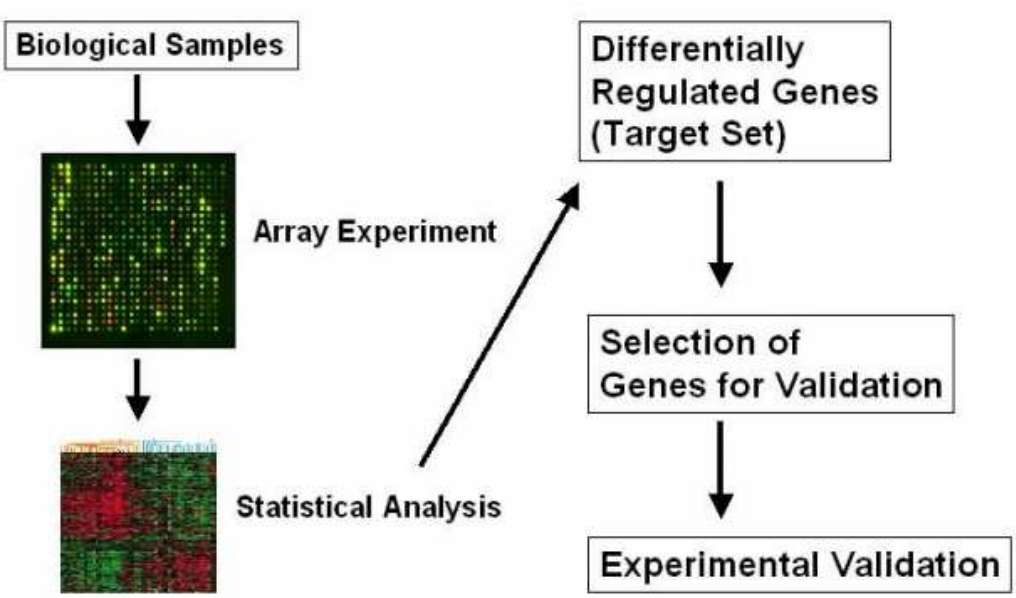



**Figure 1.7** *Schematic workflow of a microarray experiment with subsequent result's validation. RT-qPCR is considered the "gold standard" for expression validation. "Experimental validation of microarray data"- Patrick Tan*

Validation is usually thought to be executed on the most representative results in terms of difference in expression: these significant results emerge after line-by-treatment statistical analysis; anyway the same candidate genes were confirmed also within the two treatments. Here below the list of the affected genes, selected for RT-qPCR validation.

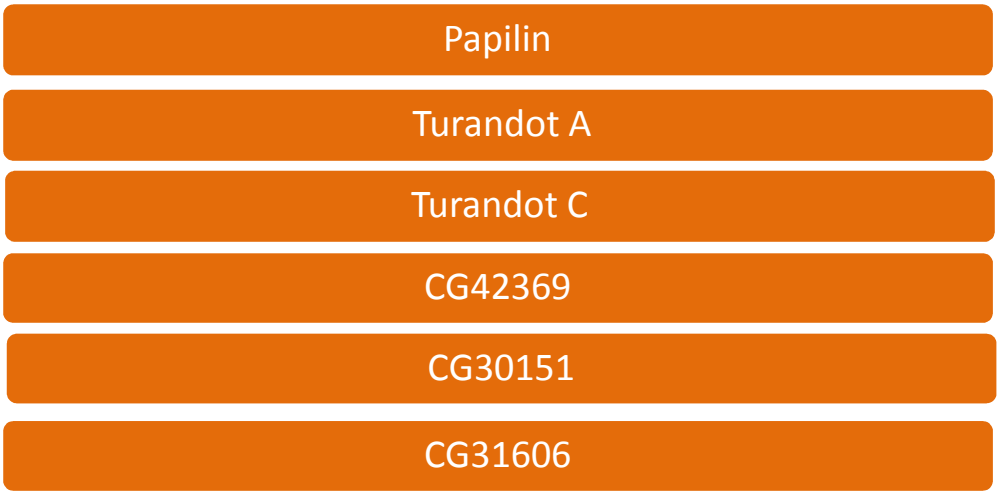



**Table 1.3** *Selected genes from microarray analysis for RT-qPCR validation. The genes are the most representative results, emerging from line-by-treatment analysis. (Vermeulen et al, unpublished)*

In conclusion, global profiling technologies, as microarrays, lead to important goal for the simultaneous quantification of hundreds of messenger RNA and a global response to treatment at system level. However, these approaches need a continuous optimization and technical hurdles could partially falsify obtained results. The use of RT-qPCR for subsequent main data validation gives a higher reliability, at least for the most significant results.

# 2. PROJECT GOALS

Quantitative real-time polymerase chain reaction (RT-qPCR) is an efficient technique to quantify gene expression: it is widely used because of its sensitivity and accuracy. RT-qPCR was here used for relative quantification of gene expression differences in inbred *D. melanogaster* lines, that were submitted to benign and stressful environmental conditions, line growth temperature around 25° C (permissive conditions) and brief cold shock treatment at 0°C (restrictive condition) respectively. In particular, I focused on expression for several selected candidate genes which were obtained in previous experiments, described before.

The first part of the project focused on the gene expression quantification of proteins that were differentially expressed in a study using proteome profiling, carried out with two dimensional gel electrophoresis (2DGE) coupled with mass spectroscopy technique. Singular protein pattern was observed in conditional inbred line (IL), showing lower survival rate after a brief cold shock (see more details below). Seven proteins were identified as only expressed in the lethal line, 3 only present at restrictive condition and 4 expressed at both conditions. RT-qPCR is here used to determine whether protein expression is correlated with corresponding gene expression level after cold shock or if protein abundance is independent from transcript abundance.

The second part of the project was performed to validate transcriptome results focusing on several of the most differentially expressed genes, shown by microarray data analysis. Differences in gene expression were estimated between permissive and restrictive conditions for IL, in relation with inbred control (IC) with the same coefficient of inbreeding but lacking lethal effect after cold-shock. Six genes, showing highly increased or decreased gene expression in IL only after cold-shock, were chosen for RT-qPCR validation.

Candidate gene expression was monitored at two time points, to see how expression changed during time, at 24 and 48 hours after treatment. The 24 hours' time-point integrates quantitative analysis of 48 hours from both proteome and microarrays results. Earlier time-point provides additional data, enriching further characterization of gene's expression after cold shock. The quantitative comparison between time-

points in this case is an advantageous strategy for distinguishing whether gene is involved in an inducible gene expression or whether is part of a more complex secondary activation. Inducible gene expression is directly activated from changes in the environment: a typical example is the heat-shock protein class, widespread in many species including *D. melanogaster*. When an organism is submitted to high temperature, heat-shock proteins modify their structure, due to new intra-molecular bonds, and expose domains of the protein that favour oligomerization. The finally formed molecule is an active transcription factor assembled after short period of time. Faster gene expression monitored at 24 hours after cold shock is probably a process triggered by similar transcription factors, active immediately after stimulus: furthermore also chromatin modelling via epigenetics mechanisms is considered to be fast acting mechanism. The later 48 hours gene expression contrarily considers slower acting mechanisms, involving multiple components, for instance transcription factors made by protein complexes. Including this goal in the project may allow considering in more detail gene expression processes underlying deregulated genes.

# 3. MATERIALS AND METHODS

## 3.1 Drosophila melanogaster stocks

## 3.2 RNA extraction and cDNA conversion

### 3.2.1 RNA extraction

### 3.2.2 Genomic DNA digestion

### 3.2.3 RNA to cDNA conversion

## 3.3 RNA quality check

### 3.3.1 RNA quantification and purity control

### 3.3.2 RNA integrity verification

### 3.3.3 cDNA conversion and gDNA presence check

## 3.4 RT-qPCR primers design and optimization

### 3.4.1 Primer design

### 3.4.2 Reference genes

### 3.4.3 Candidate genes

## 3.5 RT-qPCR optimization and sample processing

### 3.5.1 RT-qPCR method of detection: SYBR Green - ROX

### 3.5.2 RT-qPCR quantification strategy

### 3.5.3 RT-qPCR set-up and technical details

### 3.5.4 RT-qPCR optimization and Standard Curves

### 3.5.6 Primer sets efficiency

### 3.5.7 Reference genes stability

### 3.5.8 Sample reactions and experimental design

### 3.5.9 Melting curves analysis

## 3.6 Data processing

### 3.6.1 Fluorescence data processing

### 3.6.2 Data Normalization

## 3.7 Statistical analysis

### 3.7.1 Oneway ANOVA

### 3.7.2 Twoway ANOVA

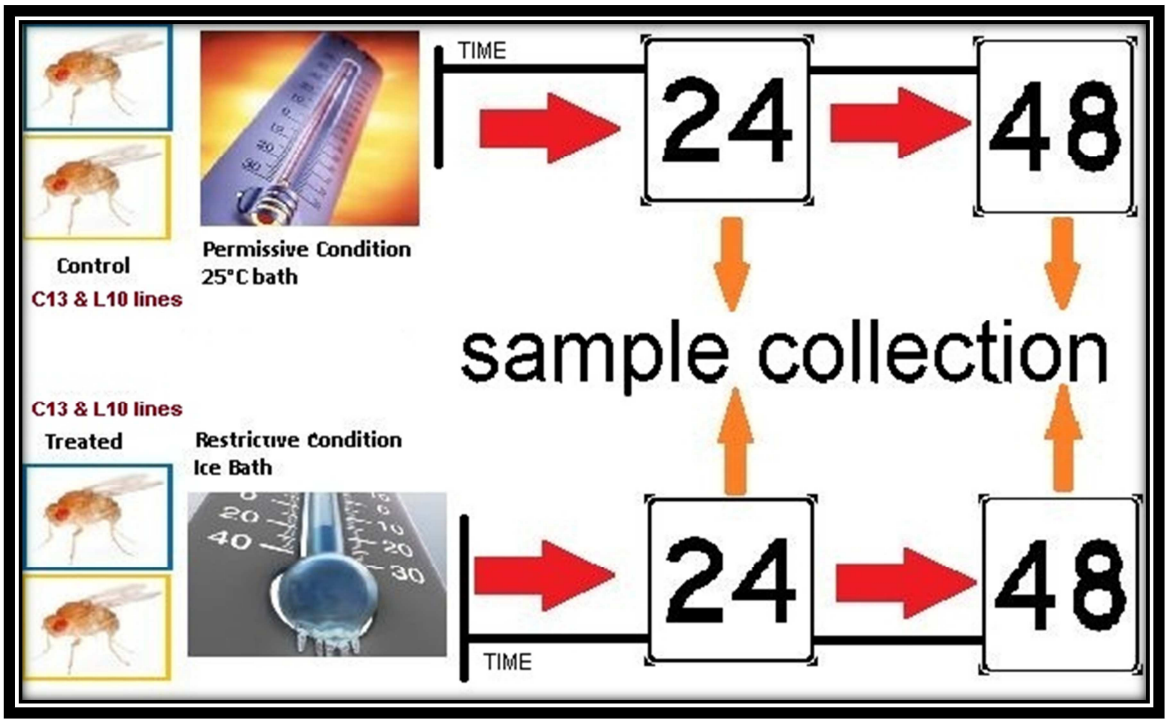
### 3.7.3 Fold Change: expression variation between treatments

# 3.1 DROSOPHILA MELANOGASTER STOCKS

RNA expression of target genes was compared between two *D. melanogaster* lines: Control Inbred Line (IC) and Lethal Inbred Line (IL). Both lines have high coefficient of inbreeding (F ~ 95%) and line L10, which is our line of interest, is a suitable candidate for lethal effect characterization in inbred lines because of the high level of adult mortality at low temperatures.

IL and IC derive from the same base stock, Groningen 83 wild type (G83), and were obtained after 14 generations of brother-sister mating: because of this experimental organization, a reliable comparison between lines is possible due to the same genetic background.

The protocol used for treating samples, the same used for the proteome profiling and the DNA microarray, consists of collecting eggs from the two lines. Adult flies were kept for three days in large number in quarter-pint bottles (more than 300 individuals) with 30 mL standard medium (26 g dead yeast, 54 g sugar, 17 g agar and 13 mL nipagine solution per litre) in addition with live yeast. Temperature of growth was 25 °C, at 40-60% relative humidity. Eggs were transferred in low density vials (100 flies/ vial) and at 25° C developmental temperature. Fly collection was executed at pupal eclosion and for this experiment lethal effect was tested only for male individuals because of the higher lethal effect penetrance. Individuals were moved to fresh medium vials with 100 mg/L ampicillin, five individuals each; lines were submitted to specific treatment of 30 minutes consisting of 25° C treatment, permissive condition (PC), or 0 °C cold shock, restrictive condition (RC) (see Figure 3.1). After treatment, samples were collected in 24 and 48 hours and five flies were pooled together in 1.5 mL Eppendorf tubes and treated as single sample. No dead flies were sampled. After snap freezing in liquid nitrogen, samples were stored at -80° until RNA extraction. From each particular condition were collected six samples, treated as an independent biological replicate. The table 3.1 below shows a schematic overview.




| | IC line (control inbred) samples | | IL line (lethal inbred) samples | |
|---|---|---|---|---|
| Treatment (Time 0) | PC – cold shock | RC + cold shock | PC – cold shock | RC + cold shock |
| After 24h | 6 | 6 | 6 | 6 |
| After 48h | 6 | 6 | 6 | 6 |

**Figure 3.1** *Schematic overview of samples collection, 24 and 48 hours after treatment at permissive or restrictive condition*

**Table 3.1** *Samples collected for the RT-qPCR expression quantification, divided by Line (IC or IL), Treatment (P or R) and hours after treatment (24h and 48h)*

In parallel, flies-batch collected for the RT-qPCR experiment, were used to estimate lethal effect and score survival rate, confirming the mortality trend. Figure 3.2 shows cumulative mortality at the fifth day after treatment for lines in proteome profile at restrictive and permissive conditions; mortality for IL is about 80% while the same line shows only 1% mortality at permissive conditions. For both the treatments, inbred control shows low mortality. The lethal effect, here presented in the

proteome profile, was successfully confirmed for samples used for RT-qPCR expression quantification: estimated mortality in the IL at restrictive condition was 98%, 2% for the same line at permissive condition and 0% for the IC at either condition.

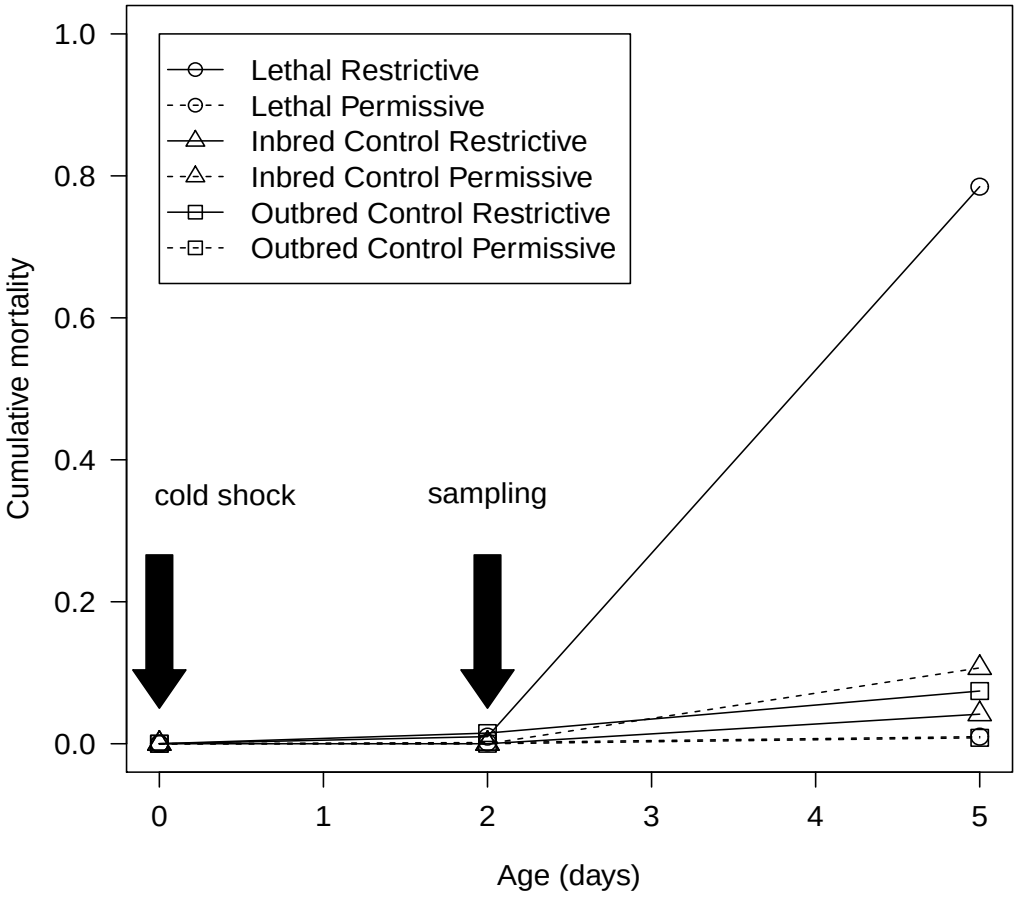



**Figure 3.2** *Cumulative mortality in OC, IC and IL line, at permissive and restrictive conditions, for samples in proteome profile. Collected samples for RT- qPCR expression experiment, show the same lethal effect for IL line at restrictive conditions*

# 3.2 RNA EXTRACTION AND cDNA CONVERSION

## 3.2.1 RNA extraction

Total RNA extraction was performed on samples while still frozen, to avoid damage of RNA. The technique used for extraction is Guanidinium thiocyanate-phenol-chloroform or TRIzol® extraction (Chomczyński and Sacchi, 1987).

After fly homogenization, standard RNA extraction TRIzol® protocol from Evolutionary Genetics group was followed (see appendix protocol 1). RNA extraction was executed with 1 mL TRIzol® per sample, for total amount of 5 mg tissue. At the end of RNA extraction, total RNA volume was diluted in 27 µL nuclease free water.

## 3.2.2 Genomic DNA digestion

Genomic DNA (gDNA) was digested using DNA-free ™ Kit according to Applied *Biosystems* protocol for routine DNase digestion (see appendix protocol 2 for more details). This procedure involves the elimination of DNase/reagents through precipitation and makes possible preventing chemical/protein contaminants, ensuring the adequate purity of samples.

## 3.2.3 RNA to cDNA conversion

After DNA digestion, 10 µL purified RNA was immediately converted to complementary DNA (cDNA): this approach is a more convenient way to work because RNA could be easily degraded by RNases and the immediate cDNA conversion avoids risk of degradation. RevertAid H minus 1st strand Kit from *Fermentas* was used, following proposed company protocol: oligo-dt primers in combination with random hexamer primers were used in proportion 1:6 for

conversion reaction. Total volume of obtained cDNA is 20 µL and samples were stored at -20°C.

# 3.3 RNA QUALITY CHECK

One of the most important steps in RT-qPCR workflow is the RNA purity and integrity check-up (Fleige and Pfaffl, 2006). Contaminants in samples could lead to reaction inhibition, both in RT-qPCR and in Reverse Transcriptase-PCR. Ascertaining RNA integrity and purity is fundamental to ensure the correct RT-PCR quantification (Taylor et al, 2010).

## 3.3.1 RNA quantification and purity control

RNA concentration was quantified using the NanoDrop 2000, *Thermo Scientific*: concentrations were estimated from the 230 nm wavelength absorbance.

Samples purity was determined using the ratio of two absorbance readings at different wavelength. The first presented by $OD_{260/280}$ wavelength ratio to check for protein presence: 1.8-2.0 $OD_{260/280}$ wavelength ratio shows absence of proteins in RNA extracted samples. The second measure of nucleic acid purity, which indicates TRIzol® and Ethanol presence, both used during RNA extraction, is $OD_{230/260}$ wavelength ratio. Optimal $OD_{230/260}$ for RNA extraction is around 2.0-2.2, for extremely pure RNA.

1 µL RNA volume from each extracted sample was quantified with NanoDrop before and after DNA digestion. Generally $OD_{260/280}$ wavelength ratio for all samples was around optimal values, while $OD_{230/260}$ wavelength ratio showed decrease after DNA digestion, maybe because of DNA-free™ Kit "Inactivation Reagent" presence; usually values were around 1.5 - 1.8 ratio. Samples with lower $OD_{230/260}$ wavelength ratio were discarded from analysis and only five samples were considered for RT-qPCR.

## 3.3.2 RNA integrity verification

After RNA extraction and DNA digestion, native RNA was run on Agarose Gel TBE 0.5X. After dilution with DEPC water 1:1, 4 µL RNA was loaded on Agarose TBE 0.5X Gel 1%. To avoid degradation of RNA, because of the presence of RNases, all the electrophoresis equipment was treated with 0.5 M NaOH.

Before loading RNA on gel, samples were submitted to high temperature treatment (70°) for few minutes, breaking molecular boundaries: after that, samples were immediately placed on ice, for maintaining RNA in linear condition.

Integrity was roughly based on absence of evident smears on Agarose Gel 1% TBE 0.5X. Presence of rRNA 18S (length about 2 Kbp) and rRNA 28S (length about 4.1 Kbp) was considered less important because fragment 28 S in Drosophila is usually processed in two minor fragments that migrate in similar manner to the 18S rRNA: not always was it possible to observe different rRNA bands. Picture 3.3 shows distinct bands, but only in less concentrated samples (for example sample 1): RNA integrity was tested only through the presence of clear bands on gel.

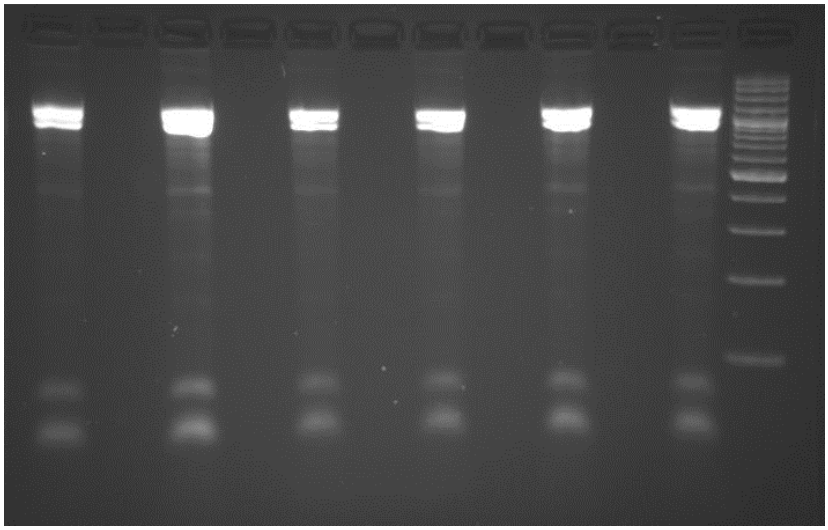


**Figure 3.3** *RNA integrity check: native RNA agarose gel electrophoresis 1%, TBE 0.5X for six different samples.100 bp ladder at the right. At around 2Kbp rRNA 16S and 28S are present in different quantity (usually 2:1 ratio), indicating RNA integrity (not evident in all the samples).*

## 3.3.3 cDNA conversion and gDNA presence check

To ensure succesful cDNA conversion, standard PCR was performed using the fd68Aq2 primer set which amplifies part of the forkhead domain 68A transcript (Ponton et al., 2011). Picture 3.4 shows amplified genomic region containing an intron of 144 bp. Thus, the lenght of amplified region differs between gDNA and mRNA.

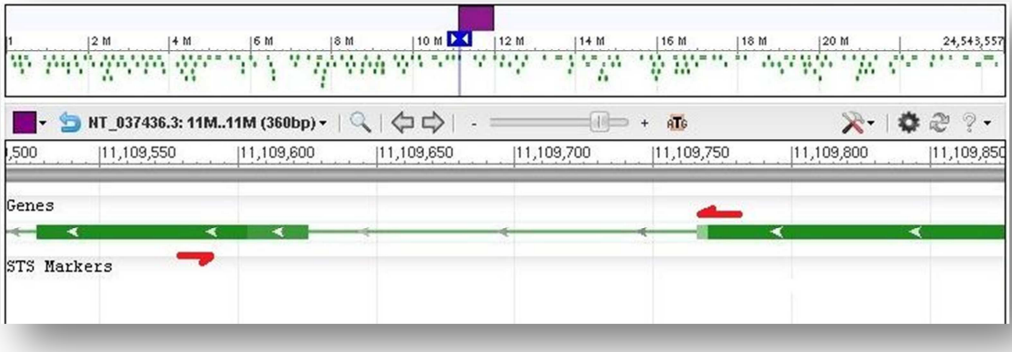



**Figure 3.4** *Primer set amplifying forkhead domain 68A region containing an intron of 144 bp in Drosophila melanogaster. For cDNA the region amplified is 66/86 bp while for genomic DNA amplified fragment is 230 bp. The primer set is used to check gDNA presence and successful cDNA conversion. NCBI graphics.*

For a sucessfully converted cDNA samples, which are free of gDNA, the expected bands on agarose gel are 66/86 bp (two alternative splicing forms are present). On the contrary, if a band of 230 bp is present on gel, the sample contains a fragment of gDNA (genomic DNA), which has an intron present. Standard 10 µL PCR was executed with 1 µL cDNA, diluted 1:1 with MQ water. All the samples were tested for sucessfull cDNA conversion (Figure 3.5).

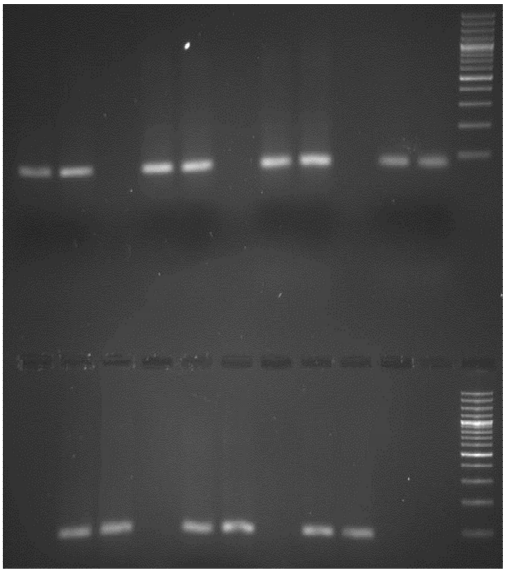


**Figure 3.5** *Standard PCR product, fd68Aq2 primer set. Six samples tested for successful cDNA conversion, including positive (last two bands) and negative control (last loaded reaction).*

# 3.4 RT-qPCR PRIMER DESIGN AND OPTIMIZATION

## 3.4.1 Primer design

Candidate gene and reference gene primer sets were designed with NCBI Primer-Blast (Ye et al., 2012; http://www.ncbi.nlm.nih.gov/tools/primer-blast) and PerlPrimer v 1.1.21. Good design is critical for reducing chance of primers dimerization and non specific amplification. Optimal primer length is around 18-24 bp, to reduce time of hybridization and removal; amplicon lenght is between 80-150 bp to reach higher amplification efficiency and high TAQ-enzyme processivity, annealing temperature was set on 60° C. Primer design follows specific standards to assure optimal RT-qPCR amplification process (Thorton and Basu, 2010). Primers were designed to be specific for mRNA if possible, by including exon junctions regions.

Different softwares programs were used to *in silico* determine properties of designed primers and expected amplicon. To check absence of self complementarity between primers, PerlPrimer softwere and Beacon designer free online edition (http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1) were employed. mFold (http://eu.idtdna.com/UNAFold) software was used to check for amplicon secondary structures. Best primer sets were chosen, based on thermodynamics characteristics, amplicon secondary structures and specificity.

## 3.4.2 Reference genes

Reference genes are fundamental to normalize RNA quantity variation; after RNA extraction, all samples, because of variability among biological replicates and variability in extractraction procedure, contain different amount of RNA. RT-qPCR is, by definition, a quantitative technique and requires an internal standardization. Five reference genes were designed for RT-qPCR to ensure suitable comparison between

samples. All reference genes were choosen because of constant expression patterns across treatments, also if exposed to stressful conditions (Ponton et al., 2011).

Primer sets for all reference genes are quoted below.

| Gene | Reference ID | Primer set |
|---|---|---|
| **α-Tubulin 84B (αTub84)** | FlyBase: FBgn0003884 Uniprot: P06603 | F: GTTTGTCAAGCCTCATAGCC<br>R: TGGATAGAGATACATTCACGCA |
| **Forkhead domain 68A (fd68A)** | FlyBase: FBgn0036134 Uniprot: Q0E8F8 | F: GCTAGTCCACGTCAGGGTTTC<br>R: GTCTGGAACAGATCCTGTGTATTG |
| **Elongation factor 1α2 100E (EF)** | FlyBase: FBgn0000557 Uniprot: P05303 | F: CGTCTACAAGATCGGAGGCA<br>R: CATCTCCACAGACTTTACTTCGG |
| **Eukaryotic initiation factor 1A (eIF)** | FlyBase: FBgn0026250 Uniprot: Q9VEA1 | F: TCACATTCGGGGGAAACTTCG<br>R: GTACGTCTTCAGGTTCCTGGC |
| **RNA polymerase II 215kD subunit (RpII)** | FlyBase: FBgn0003277 Uniprot: B4R347 | F: GTGACAAGACTGGTGGTTCGG<br>R: CCACACAAGCAATAACCTGGGA |

**Table 3.2** *Reference genes used for RT-qPCR normalization with corresponding FlyBase, Uniprot ID and primers sequences*

## 3.4.3 Candidate genes

Candidate genes primer sets from proteome profile and microarrays are show below.

| Candidate Gene | Reference ID | Primer set |
|---|---|---|
| **CG18067** | FlyBase: FBgn0034512 Uniprot: A1ZBU8 | F: ATAAATTAGAGGCGGAGGCA<br>R: GATCCCTTTGAATATCAGCCA |
| **β-Tubulin 56D (βTub56D)** | FlyBase: FBgn0003887 Uniprot: Q24560 | F: GTTTGTGTTGTGTTCGACTGC<br>R: GATGATCTCCCAGAACTTGGC |
| **Angiotensin converting enzyme (Ance)** | FlyBase: FBgn0012037 Uniprot: Q10714 | F: TTACCAAGCTGCCACAGGAC<br>R: GGTGAGATAAAAGTCCCAGGCG |
| **CG11796** | FlyBase: FBgn0036992 Uniprot: Q9VPF3 | F: GTGATAATGACCAGCTATACTGAC<br>R: GCCAACGTAGAAGGTTAAGTG |
| **Phosphogluconate mutase (Pgm)** | FlyBase: FBgn0003076 Uniprot: Q9VUY9 | F: GCTTTGGGCGGCATTGTTTTG<br>R: GTCCTCCGTTCTCGCAGTTG |
| **Pro-phenol oxidase A1 (PPOA)** | FlyBase: FBgn0261362 Uniprot: Q27598 | F: CGTTTACACCAAGGATCGTG<br>R: AGGCGGAAGGTTCTACAAAG |
| **Transaldolase (Tal)** | FlyBase: FBgn0023477 Uniprot: Q9W1G0 | F: TTGGGCAATGGGAAGTGATCG<br>R: TGTTGATGGCTTCAAAGTCGC |

**Table 3.3** *Target genes from proteome used in RT-qPCR with corresponding FlyBase, Uniprot ID and primers sequence*

| Candidate Gene | Reference ID | Primer set |
|---|---|---|
| **Turandot A (TotA)** | FlyBase: FBgn0028396 Uniprot: Q8IN44 | F: ACTGCTCTTATGTGCTTTGC R: GATTTTGGAGTCATCGTCCT |
| **CG42369** | FlyBase: FBgn0259715 Uniprot: Q9VME1 | F: GGACCTGTTCGTCATCTTTC R: CTCGTCGTATAGCAAACTGG |
| **Turandot C (TotC)** | FlyBase: FBgn0044812 Uniprot: Q8IN43 | F: TCCATTTCTCTACTATGCCTT R: TGTCAGATTCCCTTTCCTC |
| **CG31606** | FlyBase: FBgn0051606 Uniprot: Q8IPG7 | F: AGTCGTTTGGAAGATAAGATG R: GATTGGAAATGGTTTAGAGGT |
| **Papilin (Ppn)** | FlyBase: FBgn000313 Uniprot: AQ868Z9 | F: TCACATGGATCAAGGACGAC R: GGTACAACAGATTGAATGGACAC |
| **CG30151** | FlyBase: FBgn0050151 Uniprot: A1ZBU6 | F: GCAAACAGAAATCAACTTGGAG R: CTTCGATTAGACCACCTTCAC |

**Table 3.4** *Target genes from microarray used in RT-qPCR with corresponding FlyBase, Uniprot ID and primers sequence.*

Before RT-qPCR experiment, all primer sets were tested for specificity and optimal melting temperature, in standard gradient PCR. Amplification efficiency was checked in a range of temperatures and 56°C was choosen as optimal annealing temperature for RT-qPCR for most of the candidate genes; only Pro-phenol oxidase A1 and CG42369 needed increased annealing temperature, 58° C, because of aspecific binding.

# 3.5 RT-qPCR OPTIMIZATION AND SAMPLE PROCESSING

## 3.5.1 RT-qPCR method of detection: SYBR Green - ROX

The RT-qPCR chemistry chosen for the experiment is SYBR Green, an intercalating aspecific dye for double stranded DNA (ds-DNA). Quantification of amplification in RT-qPCR reactions is based on dyes with diferent wavelength detection: SYBR-Green I, an assymmetrical cyanine dye, also defined an active component and exibiting low florescence if free in solution while generates higher fluorescence, up to 1000 times, only if binding ds-DNA; ROX$^{TM}$, called also passive reference dye, compounded by the 5-carboxy-X-rhodamine fluorescent molecule that doesn't change fluorescence properties during PCR reaction (PCR reaction independent). The use of ROX reference dye allows compensation of variations in fluorescence between wells: well-to-well difference can be due to the instrument design or pipetting errors. Because of the constant ROX fluorescence, reference dye provides a stable baseline which allows sample normalization. Fluorescence considered at the end of every cycle is determined by the ratio of SYBR Green and ROX.

Figure 3.6 shows schematically properties of of active and passive dyes and Figure 3.7 shows SYBR Green and Rox components amplification dynamics for single reaction. While ROX, red line in the picture, is always constant in fluorescence, SYBR Green (blue line) increases during exponential phase and decreases after denaturation step (see below). SYBR Green fluorescence is directly proportional to ds-DNA, obtained through PCR reaction.

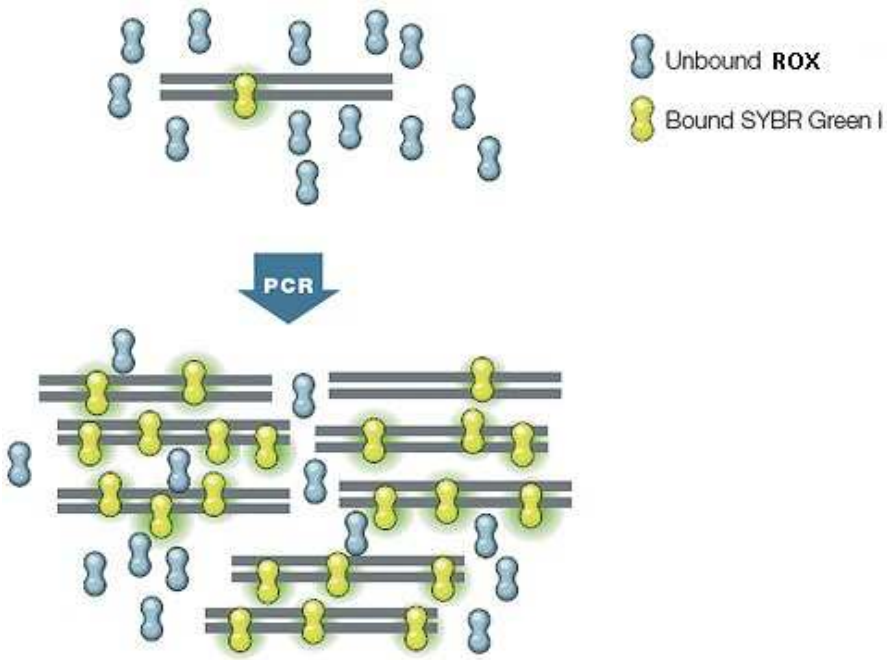



**Figure 3.6** *Model for SYBR Green DNA binding-dye and ROX. SYBR Green binds double stranded DNA while ROX is always free in solution, for baseline normalization.*

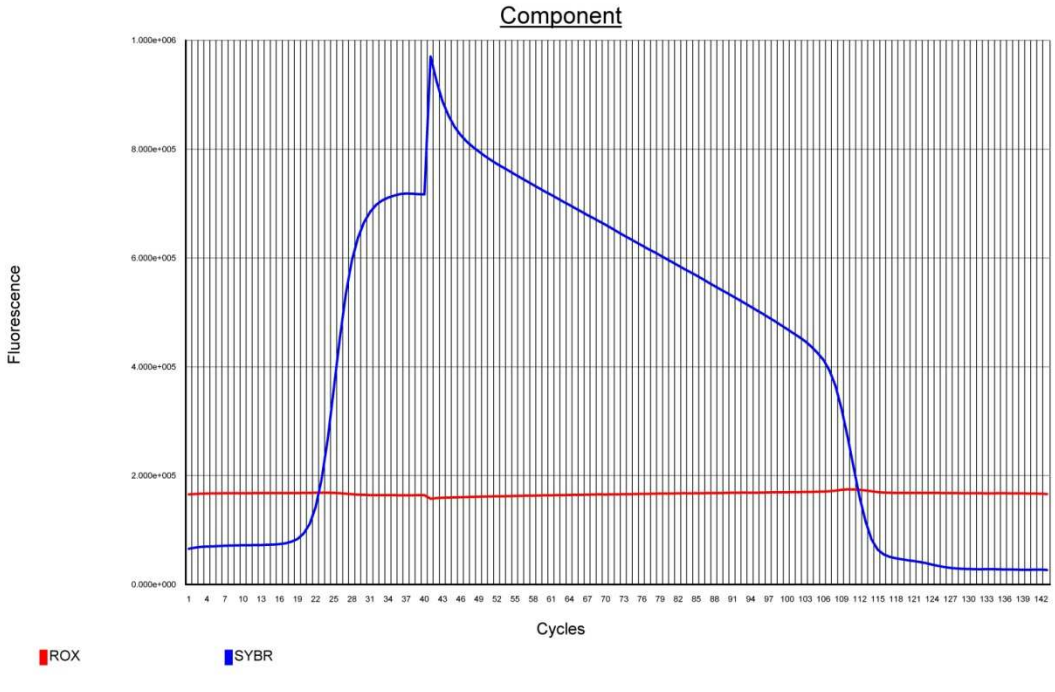



**Figure 3.7** *Single RT-qPCR reaction by component: SYBR Green (blue line) fluorescence increases during amplification a decrease after denaturation step while reference ROX dye (red line) remains always constant during cycles. SDS version 1.4, component curve*

## 3.5.2 RT-qPCR quantification strategy

Project design focus attention on comparison between two different conditions (cold shock -, cold shock +), through the lines: for this reason, RT-qPCR relative quantification or comparative quantification among samples was choosen to determine changes in mRNA expression. The strategy is suitable to investigate also small physiological changes in gene expression.

## 3.5.3 RT-qPCR set-up and technical details

RT-qPCR reactions were run on ABI RT-qPCR cycler 7300, in standard PCR 96-wells plates. Standard RT-qPCR protocol from Evolutionary Genetics group was followed for technical part and for global experimental setting up; for plate organization, was followed the online protocol: "Quantification of mRNA using Real-Time RT-PCR"; Nolan, Hands and Bustin, 2006 (appendix tables).

## 3.5.4 RT-qPCR optimization and Standard Curves

To test all primer set efficiencies and see whether primer sets have high rate of amplification, standard dilution curves were set up. Five dilution series were made from unique standard "test" sample (starting dilution 1:3 with nucleasy free water to reduce viscousity of cDNA samples) with the following dilutions: 1, 1/5, 1/25, 1/125, 1/625. Primer sets concentration was 7.5 µM, which showed immediately excellent efficiency of amplification for all primer sets except for PPOA1 primer set. This set was amplified with a lower primer set concentration (1.25 µM), to avoid primer dimers in reaction and consequent aspecific fluorescence.

## 3.5.5 Primer sets efficiency

For estimating primer set global efficiency, five dilution series from the same cDNA template were performed, each dilution with three technical replicates. This approach aims to assess accuracy and reproducibility through three hallmarks (Bio-Rad laboratories, 2006): consistency across replicates reaction, global linearity of the standard curve (both evaluated by $R^2$ higher than 0.980) and amplification efficiency (optimal values around 90-105 %). Primer sets with lower amplification efficiency were used for quantification on condition that their linearity and specificity were suitable. Standard curves were constructed from amplification curves, plotting Ct values against log10 of template dilution; efficiency (E) is drawn from the standard curve slope, using the following formula:

$$\text{Efficiency} = 10^{-1/\text{slope}} \qquad (4)$$

Figure 3.8 shows a representative standard curve obtained from the TotC primer set and five dilution series template.

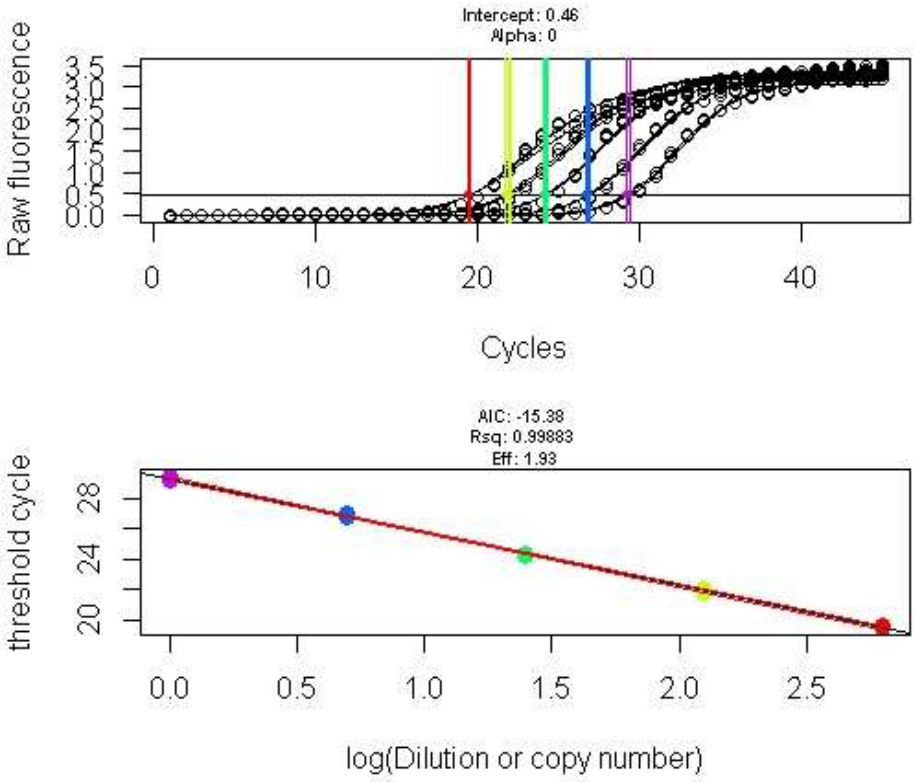



**Figure 3.8** *Upper picture: amplification curves of serial template dilutions; Lower picture: corresponding standard curve, compounded by calculated Ct values plotted against the log10 of dilutions. "R" and "qpcR" package, calib function*

Here below a table containing $R^2$ and efficiencies (%) for all the primer sets used for the gene expression RT-qPCR quantification.

| *Proteome profile* | *Efficiency* | *$R^2$* |
|---|---|---|
| **CG18067** | 98% | 0.99 |
| **βTub56D** | 98% | 0.99 |
| **CG11796** | 97% | 0.99 |
| **Ance** | 94% | 0.98 |
| **PPOA** | 99% | 0.99 |
| **Pgm** | 97% | 0.99 |
| **Tal** | 93% | 0.99 |

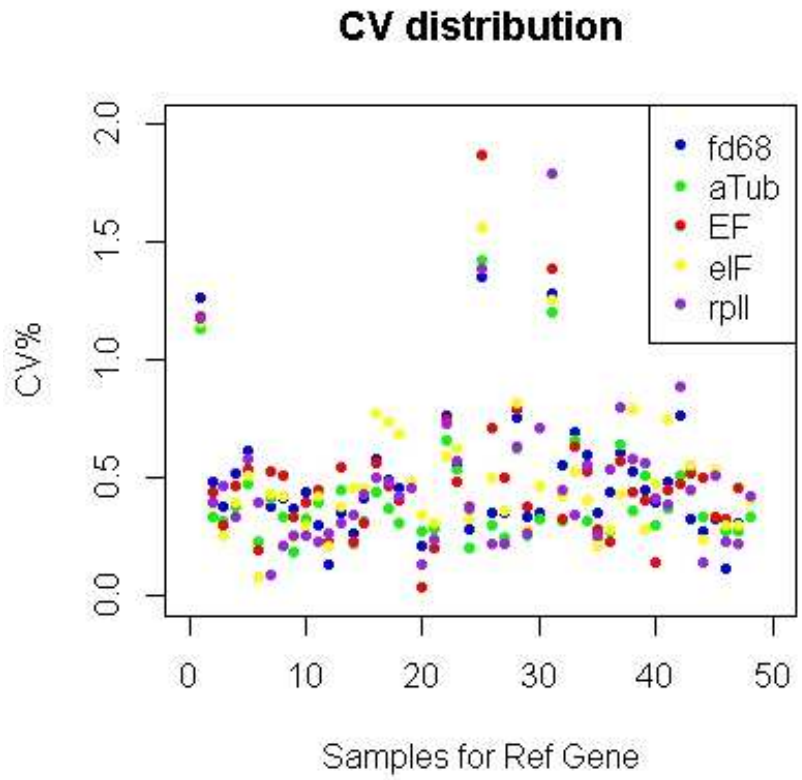
| Microarray Validation | Efficiency | $R^2$ |
|---|---|---|
| ***TotA*** | *90%* | *0.99* |
| ***CG42369*** | *92%* | *0.99* |
| ***TotC*** | *93%* | *0.99* |
| ***CG31606*** | *87%* | *0.99* |
| ***Ppn*** | *88%* | *0.99* |
| ***CG30151*** | *95%* | *0.99* |

| Reference Genes | Efficiency | $R^2$ |
|---|---|---|
| ***Fd68A*** | *98%* | *0.99* |
| ***bTub84D*** | *97%* | *0.99* |
| ***EF*** | *88%* | *0.99* |
| ***eIF*** | *97%* | *0.99* |
| ***RpII*** | *94%* | *0.99* |

**Table 3.5** *$R^2$ and efficiencies (%) for the primer sets used qRT-PCR expression quantification*

## 3.5.6 Reference genes stability

For assessing the stability of reference genes, the Coefficient of Variation (CV) was calculated singularly (Hellemans et al., 2007) for each reference gene and conditions. Figure 3.9 shows CV% values: variation expressed in percentage is mainly concentrated around 0.5%. Range of CV % values for constantly expressed genes is considered to be around 3-5% or below (de Jonge et al., 2007).




**Figures 3.9** *Calculated Coefficient of Variation (CV) for the five reference genes: forkhead domain 68A, α-Tubulin 84D, elongation factor 1 α, eukaryotic initiation factor eIF-1 and RNA polymerase II. Values are mainly concentrated around 0.5 – 1%, indicating stable expression*

## 3.5.8 Sample reactions and experimental design

Optimal sample dilution for RT-qPCR amplification was based on standard curves. Perfect sample dilution should cross fluorescence threshold between 20$^{th}$ and 30$^{th}$ cycle because of correct data setting against background. For most of the amplified genes, selected dilution was 1/25, with the exception of two genes which had 1/125 for CG42369 and 1 for PPOA1. Samples dilution is highly dependent on absolute gene expression: higher samples concentration used for RT-qPCR reflects low abundance of transcription product. RT-qPCR cycles were set to 45 to permit amplification also of lowly expressed genes. Temperature profile was 95° C for 15 minutes, followed by

45 cycles at 95° C for 15 seconds, 56/58°C for 30 seconds and 72°C for 30 seconds. Mix for single RT-qPCR reaction contains 1 µL cDNA and 24 µL 1X SYBR Green.

For all the biological replicates, a single RT-qPCR reaction was set up so no technnical replicates were included. Contrary to standard RT-qPCR protocols, no technical replicates were included because this would have been at the cost of biological replication. By including the maximal number of biological replicates, the statistical power to detect differences between experimental groups was maximised.

Samples were run in RT-qPCR plates with balanced design, that is to say an equal number of samples, belonging to the same treatment group, were run together in the same plate. Balanced design is fundamental to avoid block effects and to carry out a proper relative quantification. Plate design was organized in a way to have all the GOI reactions besides all the reference genes reactions, for each biological replicate

## 3.5.9 Melting curves analysis

A standard ABI 7300 dissociation step was added after RT-qPCR reactions: during the dissociation step, the temperature is raised and double strand DNA begins to dissociate and releases SYBR Green dye. Fluorescence, measured at each temperature point, evaluates dissociation characteristics of amplicon: by observing dissociation curve it's possible to check whether the desired product is exclusively amplified. Through denaturation curve it's also possible to detect primer dimers and contaminating DNA. Figure 3.10 shows denaturation curves from four RT-qPCR products, amplified by the same primer set: presence of a single peak indicates amplification of single RT-qPCR product.

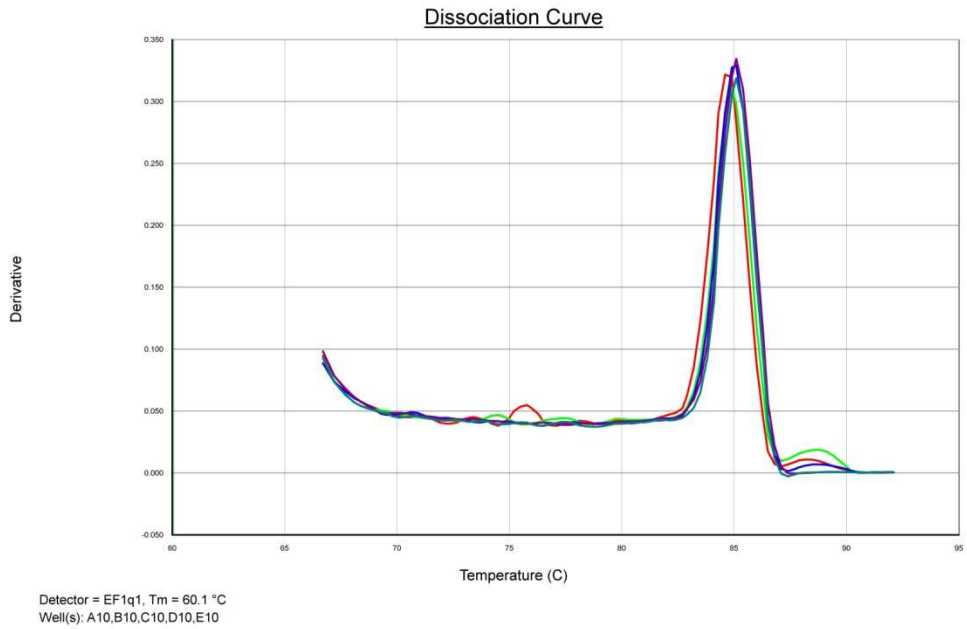



**Figure 3.10** *Derivative plot of a dissociation curve: Four samples dissociation curves, from the same primer reactions. A single peak indicates single amplicon product presence and ensures specificity of primer set. SDS version 1.4, dissociation curve.*

# 3.6 DATA PROCESSING

## 3.6.1 Fluorescence data processing

Fluorescence from RT-qPCR reactions was quantified with "Sequence Detection Software" SDS, version 1.4 and was exported as measure of fluorescence quantity per cycle. Values were exported as an Excel Spreadsheet, containing SYBR Green and ROX fluorescence values.

At first, SYBR Green's values were normalized with ROX fluorescence, at each cycle point ("baseline correction"). After that, data were processed with LinReg PCR, version 11.0.

LinReg algorithm is an assumption-free approach aiming to determine starting mRNA concentration and individual efficiency for each reaction. LinReg determines values at the maximum of the second derivate (SDM) of observed fluorescence values in a log-linear phase; the longest straight line of data points is choosen down from the terminal plateau-phase with constant fluorescence. This computation calculates first the background, a level of fluorescence measured before any specific amplification can be detected (Ruijter et al., 2009); than Window-of-Linearity" (W-o-L) is determined, a subset of four points in the exponential phase, indicative for single reaction efficiency. Optimal efficiency is equal to 2 because PCR optimal dynamics provides doubling of molecules at every cycle; LinReg algorithm determins efficiency of reaction considering calculated W-o-L. Because of the assumption that initial ammount of starting target DNA is proportional only to the exponential phase in PCR reaction, a threshold cycle (Ct) is calculated for each reaction and represented by the fluorescence on y-axis in logaritmic phase, set arbitrary in LinReg software between the first and the second point of the W-o-L. Linking together the mean efficiency (Eff_mean), including only values whitin the range of 5 % around the median efficiency, and Ct values, initial fluorescence (N0) values are computed with the following formula:

$$N0 = \text{threshold} / (\text{Eff\_mean}^{\text{Ct}}) \qquad (5)$$

N0 is considered as the starting concentration of target DNA, expressed in arbitrary fluorescence units, used to compare expression in this experiment. (Karlen et al, 2007). Figure 3.11 below shows LinReg properties: at the left multiple RT-qPCR amplification reactions with corresponding Ct values in W-o-L; at the right a single RT-qPCR reaction with underlined W-o-L (blue lines) and Ct (green line), computed throuh LinReg algoritm on curve's properties (plateau, exponential phase and background).

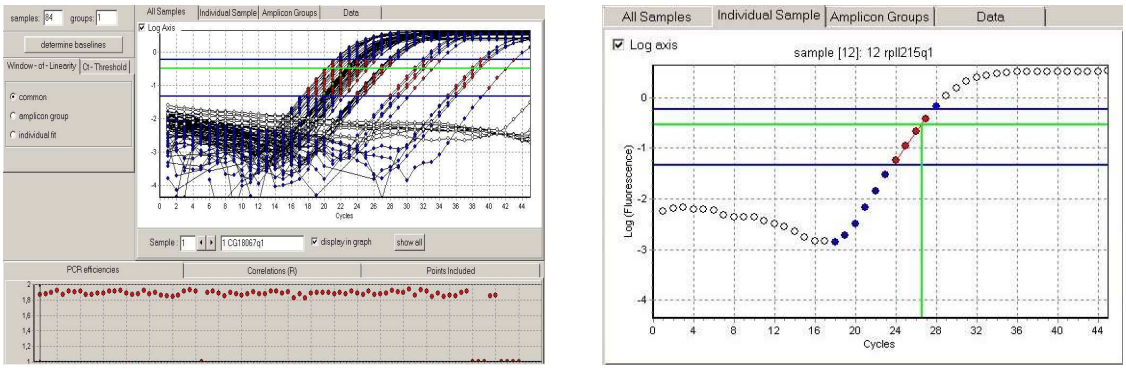



**Figure 3.11** *LinReg graphic interface, every spot is a detected fluorescence value detected in RT-qPCR reactions. At the left: multiple RT-qPCR reactions, presented with log-linear phase. Y-axis presents log of Fluorescence and X –axis RT-qPCR cycles. At the right: single RT-qPCR reaction, underlying W-o-L (blue lines) and determined Ct (Green Line).*

## 3.6.2 Data Normalization

For final results, all N0 values were normalized with the geometrical mean of reference genes N0 (Vandesompele et al., 2002); obtained values, called relative expression (R.e.), are used for relative quantification.

# 3.7 STATISTICAL ANALYSIS

## 3.7.1 Oneway ANOVA

Calculated "Relative expression" values (R.e.) were transformed logarithmically (log2) for correcting heterogeneity of variance within the treatments (Gomez and Gomez, 1984). Mean R.e. values were compared among treatments and statistical significance between differences was assigned through oneway ANOVA. Each time point was considered separately and ICP (inbred control – permissive treatment), ICR

(inbred control – restrictive treatment), LP (inbred lethal line - permissive treatment), LR (inbred lethal line - restrictive treatment) were considered as levels of fixed factor "experimental group". In case P value was significantly low (< 0.05), Tukey HSD (Honestly Significant Difference) at 95 % family- wise confidence level was used to carry out a simultaneous comparison. Both the data from the proteome profile and the microarray validation were analysed with the same method.

## 3.7.2 Twoway ANOVA

Twoway ANOVA is used on R.e. dependent values for detecting interaction between categories: values are divided in line (IC, L) and treatment (P, R) categories, and treatments, lines and categories interaction were considered as a fixed factors.

Interaction refers to the combined effect of factors on measures, that is to say that effect of one factor is subjected to the level of the other.

## 3.7.3 Fold change: expression variation between treatments

Quantification between expression differences is presented with Fold Change. These values are obtained by dividing mean of the five R.e. for restrictive condition by the mean of R.e. at permissive condition of corresponding time point and line. For Fold Change < 1, the following correction was applied: - (1/Fold Change). Standard error showed in figures is the mean of standard error of treatments (Ivo Rieu and Stephen Powers, letter to the editor 2009).

Difference in Fold Change between 24 and 48 hours are statistically validated through twoway ANOVA: results are grouped by time point (24, 48) and treatment (P, R).

### 3.7.4 Correlation between RT-qPCR and microarrays results

Correlation between the two data sets was evaluated through the Pearson product-moment validation, considering four groups, based on treatment (permissive or restrictive, P - R) and line ( inbred control and lethal, IC - L ). Correlation was estimated first between residuals of $\log_2$ N0 values from RT-qPCR and residuals of mean fluorescences (n = 44), after overall mean correction. Further correlation was also calculated with the $\log_2$ of R.e. normalized values (n = 24).

The statistical tests were executed using “R”, a free software for statistical computing and graphics, version 2.14.2.

# 4. RESULTS

## 4.1 Candidate genes from proteome profile

### 4.1.1 Genes and proteins expression comparison

### 4.1.2 Fold change and time points difference expression

## 4.2 Candidate genes from microarray

### 4.2.1 RT-qPCR and microarrays results comparison

### 4.2.2 Fold change and time points difference expression

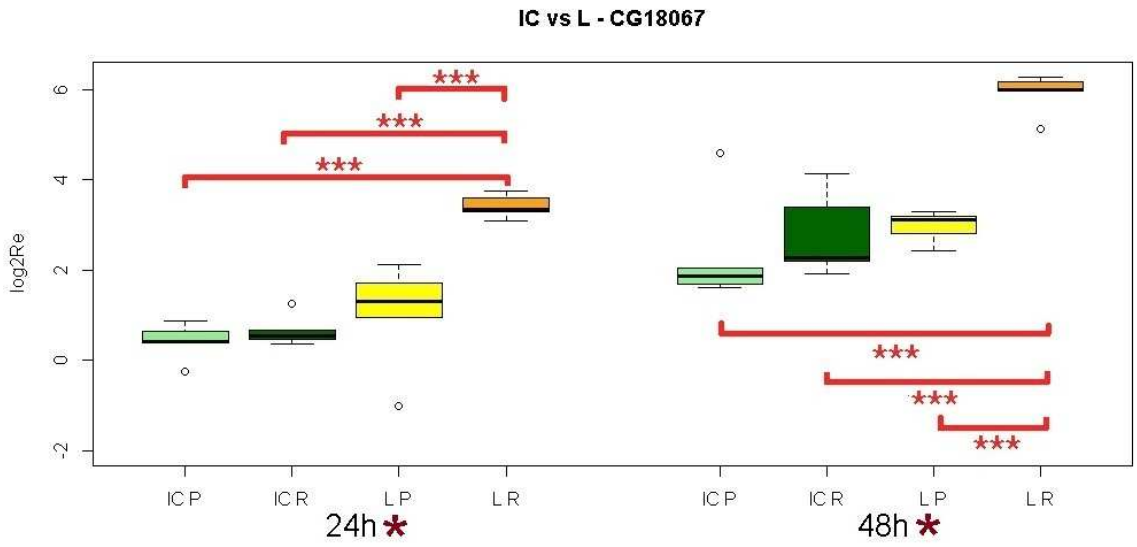
# 4.1 CANDIDATE GENES FROM PROTEOME PROFILE

## 4.1.1 Genes and proteins expression comparison

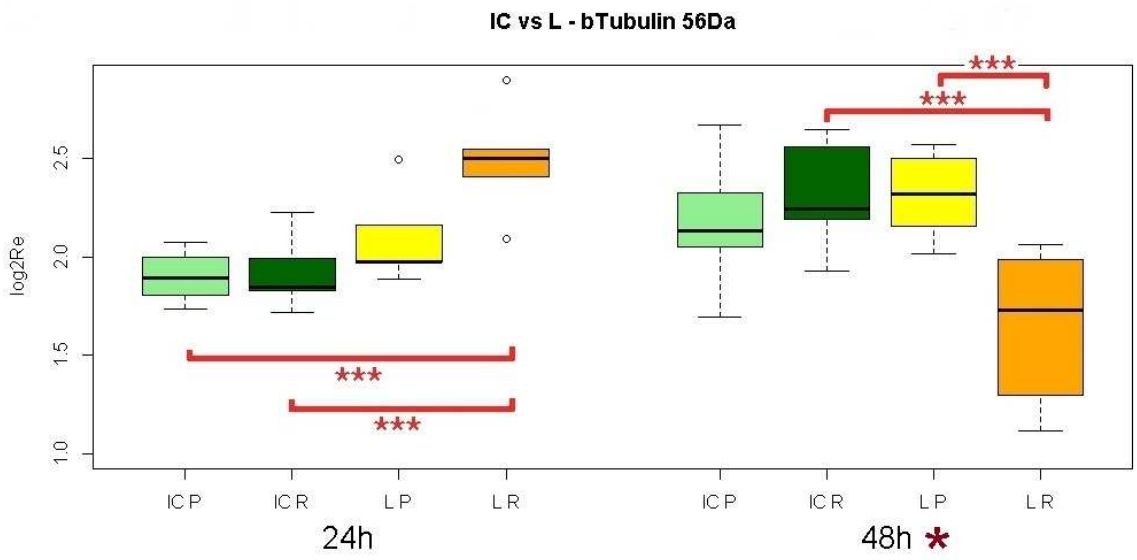
2 dimensional gel electrophoresis proteome profile was executed on inbred lethal (L), inbred control (IC) and outbred control (OC) lines of *D. melanogaster*, after exposure of a brief cold shock treatment (restrictive condition, R) or maintained at natural growth temperature (permissive condition, P) in order to provide a semi-quantitative comparison between protein expression levels (see details before). Proteins were considered 48 hours after treatment. Seven differentially expressed proteins were identified exclusively in L line, of which three exclusively at restrictive conditions (R) while other four present at both the conditions.

This project aimed to check whether patterns of gene expression at 48 hours, quantified with RT-qPCR, matched with the protein expression. Even if we found differential expression for all the genes at 48 hours, except for Ance and CG11796 (c and d in Figure 4.1 below), missing correspondence between protein and gene expression is immediately evident. Restrictive conditions enhance protein expression while genes expression is subjected mainly to downregulation if compared to corresponding controls: most of the genes, indeed, exhibit a lower gene expression in the L-R (genes βTubulin 56D, Prophenol-oxidase A1, Phosphogluconate mutase and Transaldolase, in Figure 4.1: b, e, f, g) and only CG18067 has higher gene expression (Figure 4.1a). 24 hours gene expression reflects 48 hours deregulation only in CG18067 gene, while all the others genes show deregulation only at the 48 hours' time point. Following figure presents data as boxplots, red bars highlight significant pairwise difference between lines and treatments, stars near time points indicate overall differential expression. Appendix table 1 and 2 contain ANOVA test values for details.

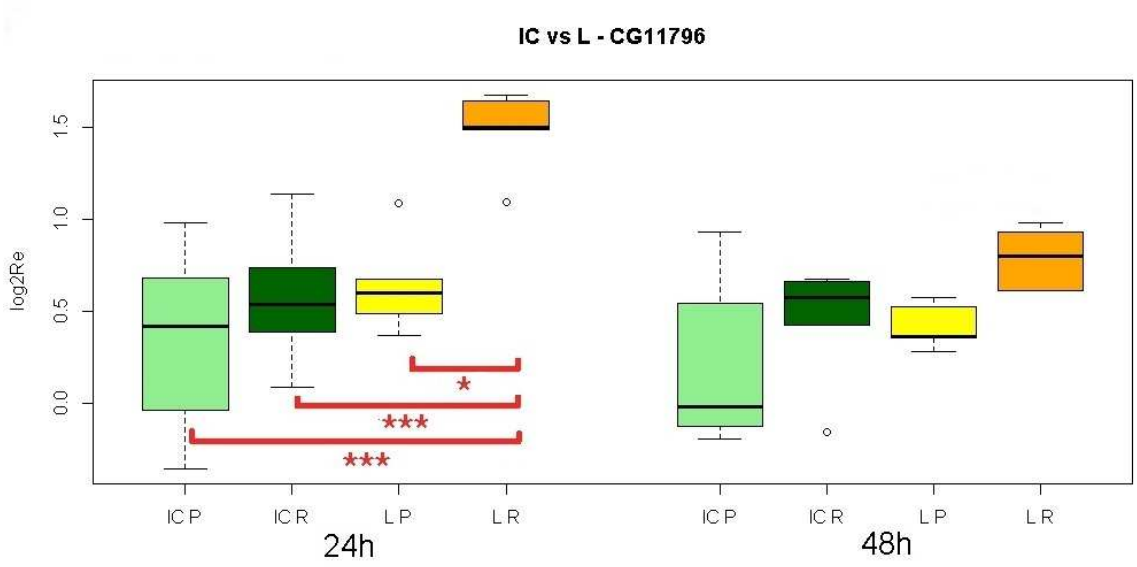
a)


IC vs L - CG18067
log2Re
6
4
2
0
-2
***
***
***
***
***
***
ICP
ICR
LP
LR
24h *
ICP
ICR
LP
LR
48h *

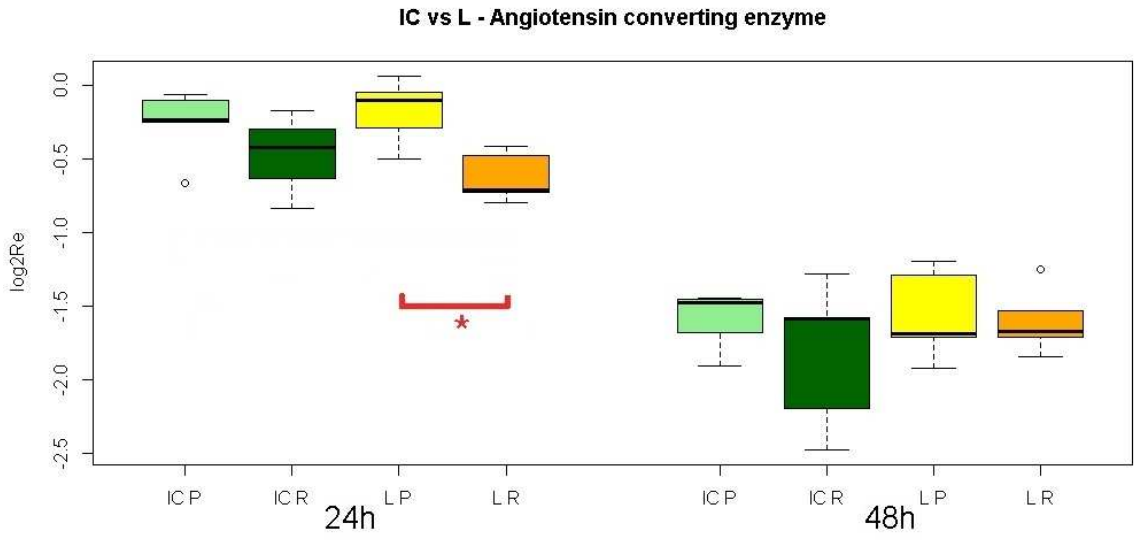
b)


IC vs L - bTubulin 56Da
log2Re
2.5
2.0
1.5
1.0
***
***
***
***
ICP
ICR
LP
LR
24h
ICP
ICR
LP
LR
48h *

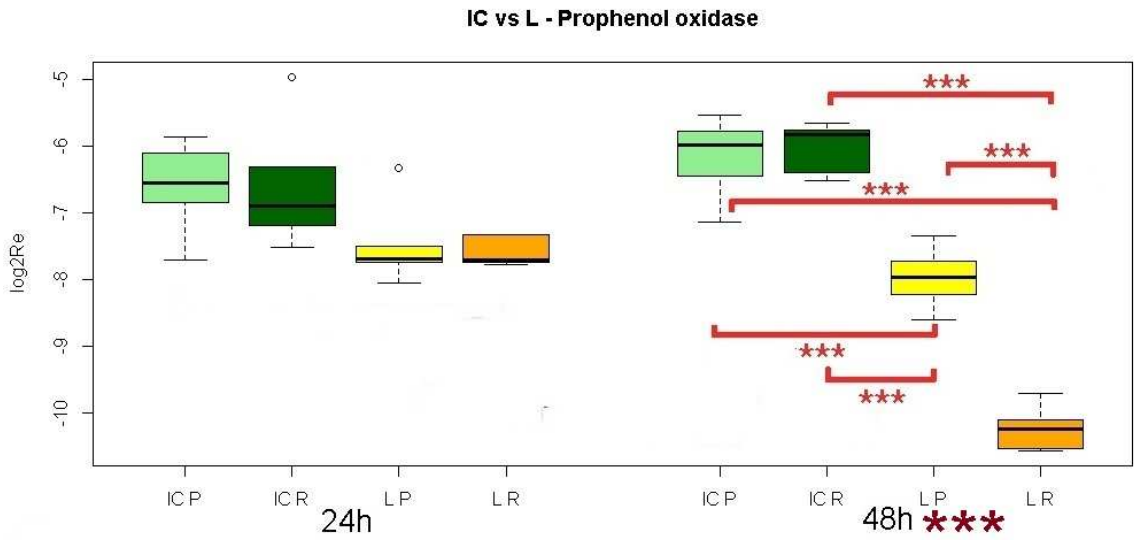
c)


IC vs L - CG11796
log2Re
1.5
1.0
0.5
0.0
*
***
***
IC P
IC R
L P
L R
24h
IC P
IC R
L P
L R
48h

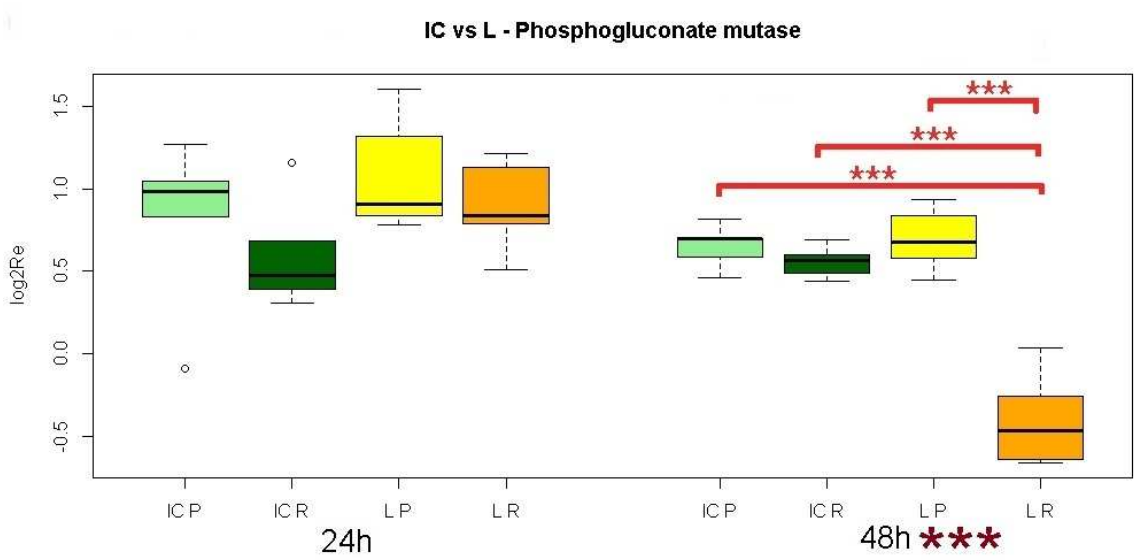
d)


IC vs L - Angiotensin converting enzyme
log2Re
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
*
IC P
IC R
L P
L R
24h
IC P
IC R
L P
L R
48h

e)


IC vs L - Prophenol oxidase
log2Re
-5
-6
-7
-8
-9
-10
***
***
***
***
***
IC P
IC R
L P
L R
24h
IC P
IC R
L P
L R
48h ***

f)


IC vs L - Phosphogluconate mutase
log2Re
1.5
1.0
0.5
0.0
-0.5
***
***
***
IC P
IC R
L P
L R
24h
IC P
IC R
L P
L R
48h ***

g)

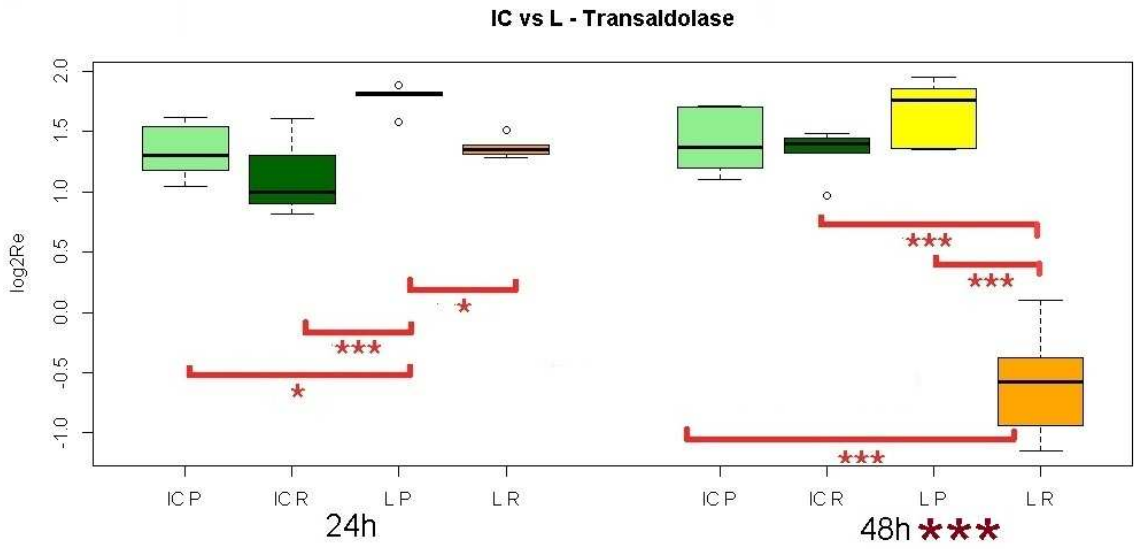



**Figure 4.1 a), b), c), d), e), f), g)** *R.e. values by levels for CG18067, βTub56D, CG11796, Ance, PPOA, Pgm, Tal genes, showed as boxplot graphs. Red bars highlight significant difference in one way ANOVA and Tukey HSD paired comparison. Dark red stars near time point show line by treatment interaction, twoway ANOVA. P-value < 0.05 (*), P-value < 0.01 (***)*

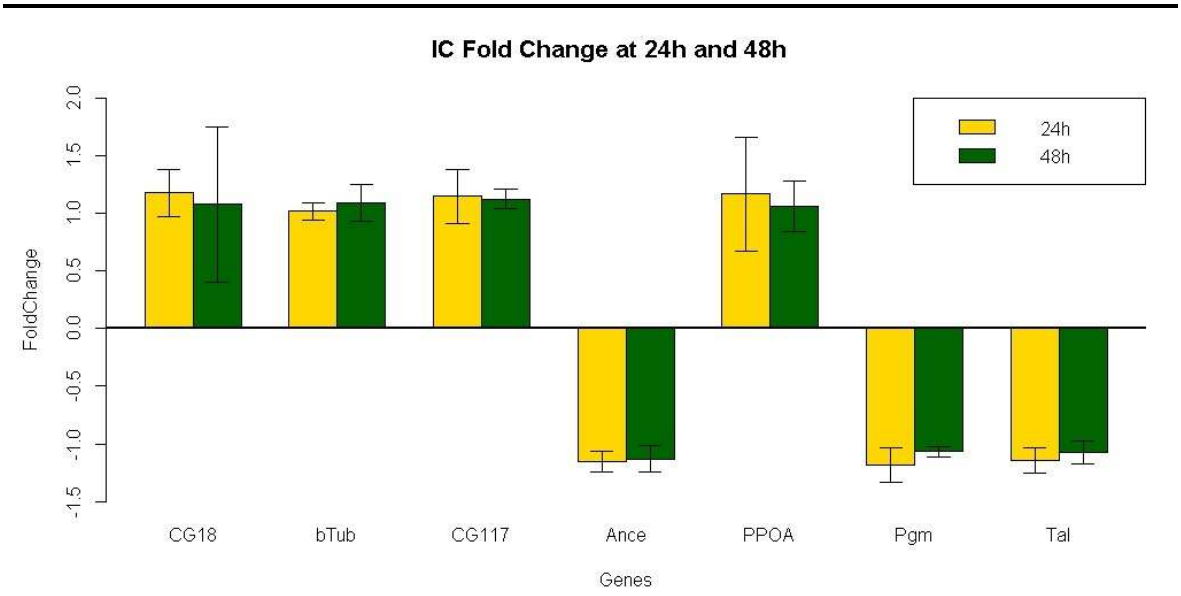
## 4.1.2 Fold change and time points difference expression

Quantification of the level of differential expression is here presented through Fold Change absolute values; these values consider restrictive condition gene expression deregulation within the two lines.

Fold change variances are only present in the lethal inbred line (L) (Figure 4.2 a), while gene expression deregulation is absent for all the tested genes in the inbred control (IC) line (Figure 4.2 b). Excluding Ance and CG11796 where fold change values are higher at the earlier time point, L shows significant gene expression differences at 48 hours: indeed most of the tested genes reach the maximum deregulation effect 48 hour after the treatment (see fold change values in Appendix table 4).

Fold Change values were compared between time points for checking whether transcript abundance differs between the 24 and the 48 hours: three genes show decreased genes expression at 48 hours (βTub56D, PPOA and Tal in Figure 4.2 b) respect to earlier time point (more details in Appendix Table 3).

a)




b)

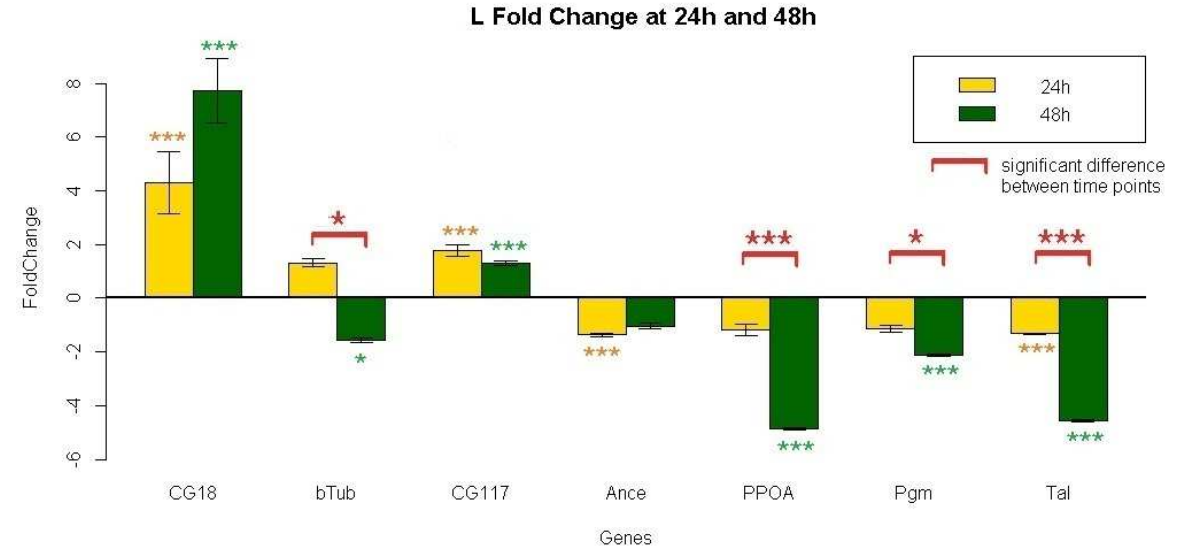
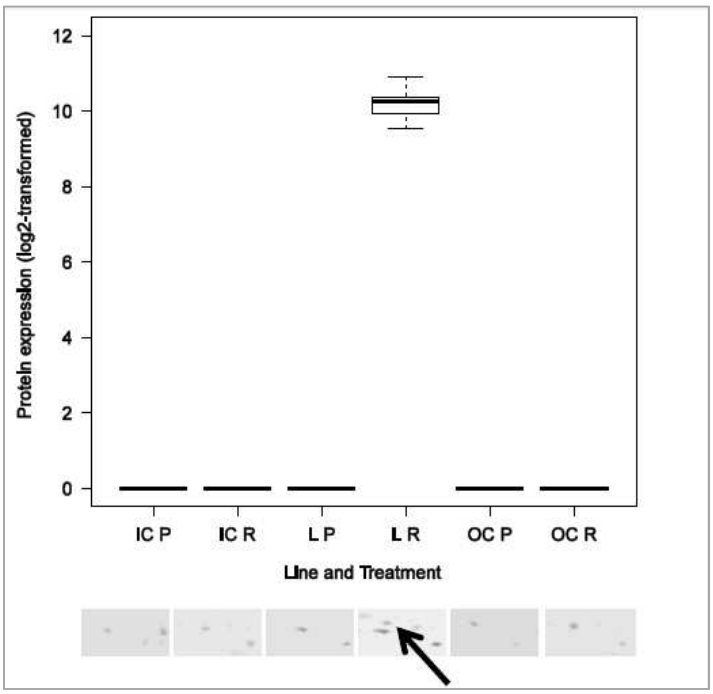



**Figure 4.2 a), b)** *Fold change values at 24 and 48 hours with mean standard error for IC and L line. Stars on the top of the bars show significant fold change. Red lines underline difference in expression between time points. P-value < 0.05 (*), P-value < 0.01(***)*

Even if accurate comparison between proteins and gene's expression is not merely possible, due to differences in the two used techniques, only rough comparison were executed for the most important data. In the proteome profile the most dramatic protein change expression was observed for CG18067 protein. Protein abundance at restrictive conditions was evaluated to be 10 times higher than expression in all the others conditions where CG18067 protein abundance was equal to zero. Figure 4.3 illustrates protein upregulated expression for CG18067: this value was roughly confirmed at transcript level through high fold change value observed only in the L-R. Data from RT-qPCR show 7.7 fold change variation at 48 hours.




**Figures 4.3** *Boxplot graphs for CG18067 protein abundance at 48 hours. The protein is present only in lethal line and at restrictive conditions. Beneath, 2-DGE stained protein spots. CG18067 protein expression is missing at all the other conditions. Protein quantification was established through normalization against 10 randomly chosen spots on gel with constitutive expression between lines.*

# 4.2 CANDIDATE GENES FROM MICROARRAY

## 4.2.1 RT-qPCR and microarrays results comparison

Gene expression microarray was executed in order to obtain an overview of differentially expressed genes in the three mentioned before *D. melanogaster* lines, after cold shock treatment. The most representative genes, showing higher expression variance in L line at restrictive conditions, were validated in the current project through RT-qPCR.

Table 4.1 shows most representative genes and their normalized microarrays fluorescence values, evaluated for each line and treatment; direction indicates deregulation of L-R values respect to the other conditions L-P, IC-R and IC-P.

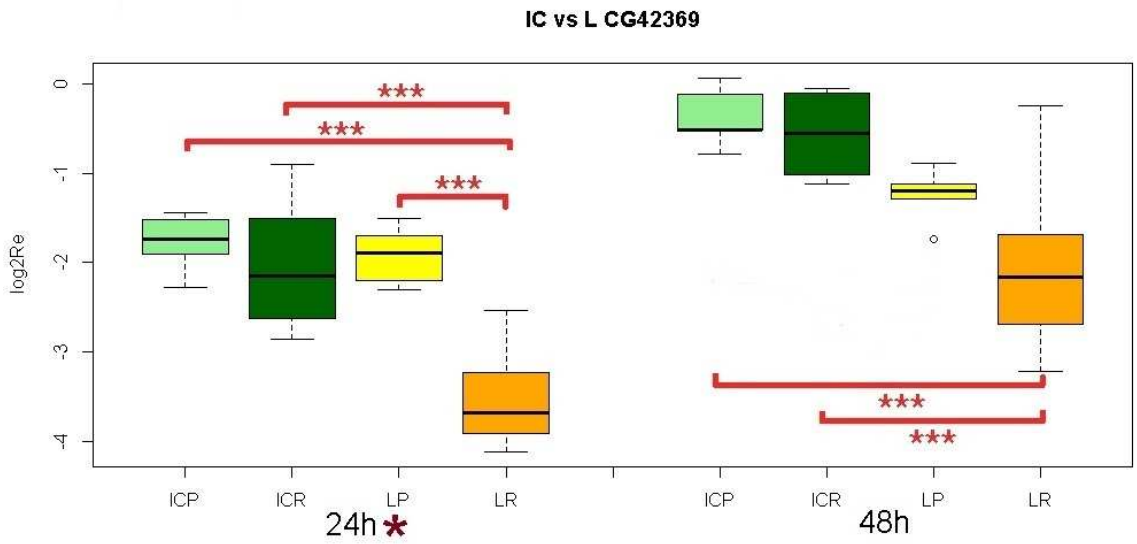
| Gene | L-R | L-P | IC-R | IC-P | Direction |
|---|---|---|---|---|---|
| CG42369 | 6,55 | 8,32 | 7,65 | 8,16 | - |
| Ppn | 7,49 | 10,11 | 8,94 | 9,8 | - |
| CG31606 | 4,28 | 7 | 6 | 6,8 | - |
| CG30151 | 7,85 | 7,21 | 6,66 | 7,24 | + |
| TotA | 14,68 | 13,97 | 12,45 | 13,14 | + |
| TotC | 13,55 | 11,21 | 8,66 | 9,67 | + |

**Table 4.1** *Most representative deregulated genes from microarrays analysis. Deregulation is present in the lethal line and under restrictive conditions (L-R) when compared to the other conditions (L-P, IC-R, IC-P). Vermeulen et al. unpublished*

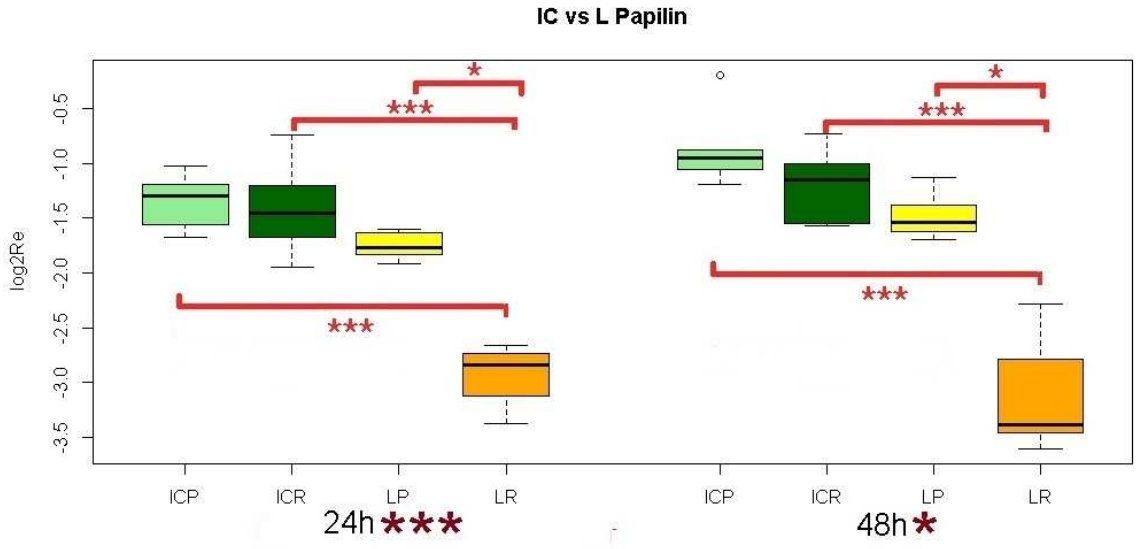
RT-qPCR values at 48 hours show significant deregulation in the L-R for all the genes, with the exception of CG42369 (a in figure 4.4, for more details see Appendix Tables 5 and 6). When compared to microarrays results, estimated correlation between data sets is significantly positive: r= 0.84 ($t_{42}$= 10.22, p-value = << 0.01) if correlating residuals of N0 values for gene of interest and reference genes and r= 0.91 ($t_{22}$= 10.66, p-value =

<< 0.01) if considering the relative expression (R.e) of the six gene of interest. 24 hours expression values confirm L-R deregulation for all the tested genes, with exclusion of TotA and TotC (e and f in figure 4.4).

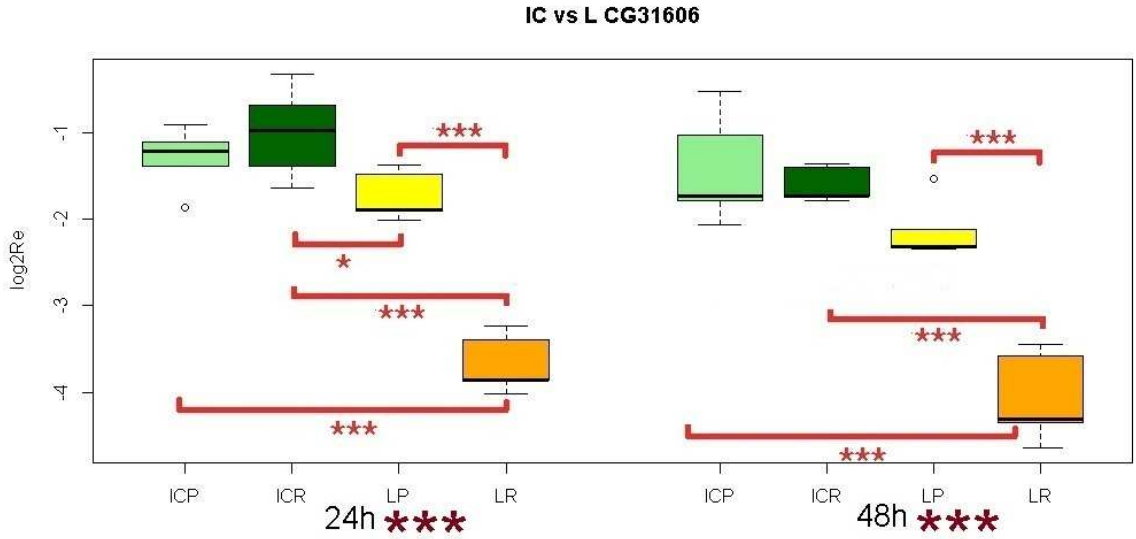
a)



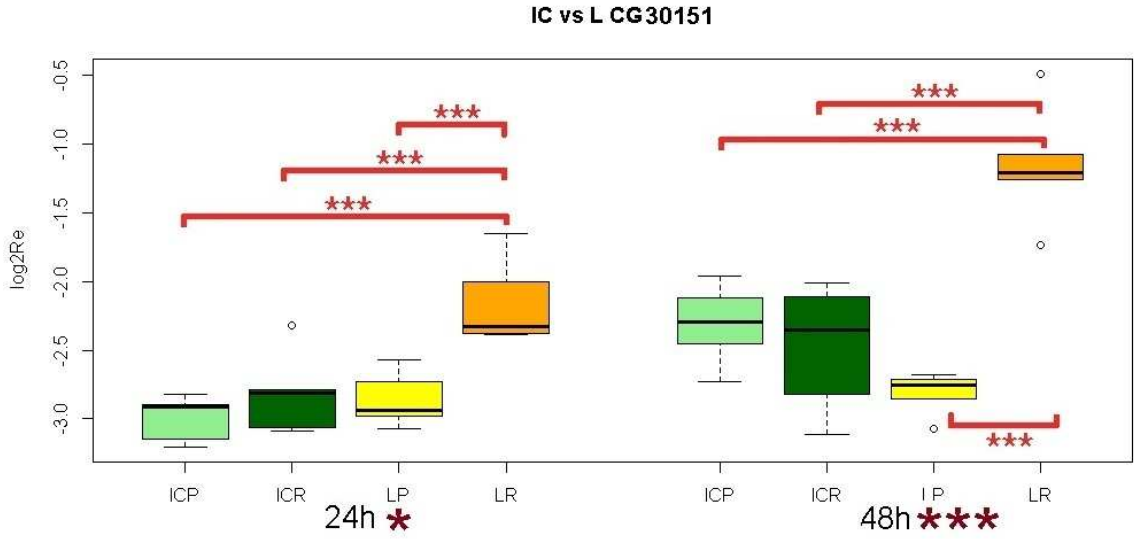

b)



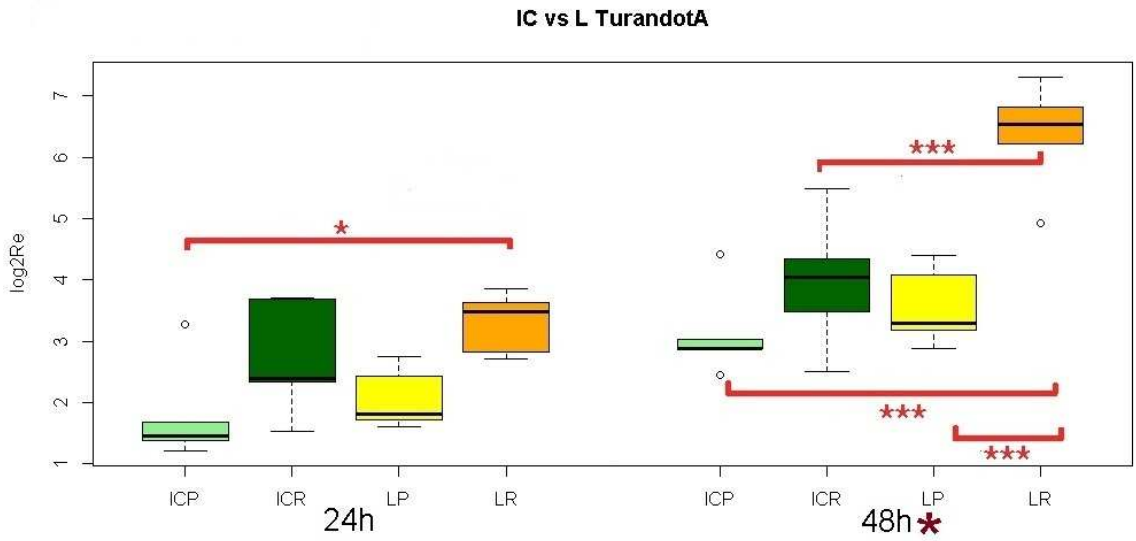
c)


IC vs L CG31606
log2Re
-1
-2
-3
-4
***
*
***
***
***
***
***
ICP
ICR
LP
LR
24h ***
48h ***

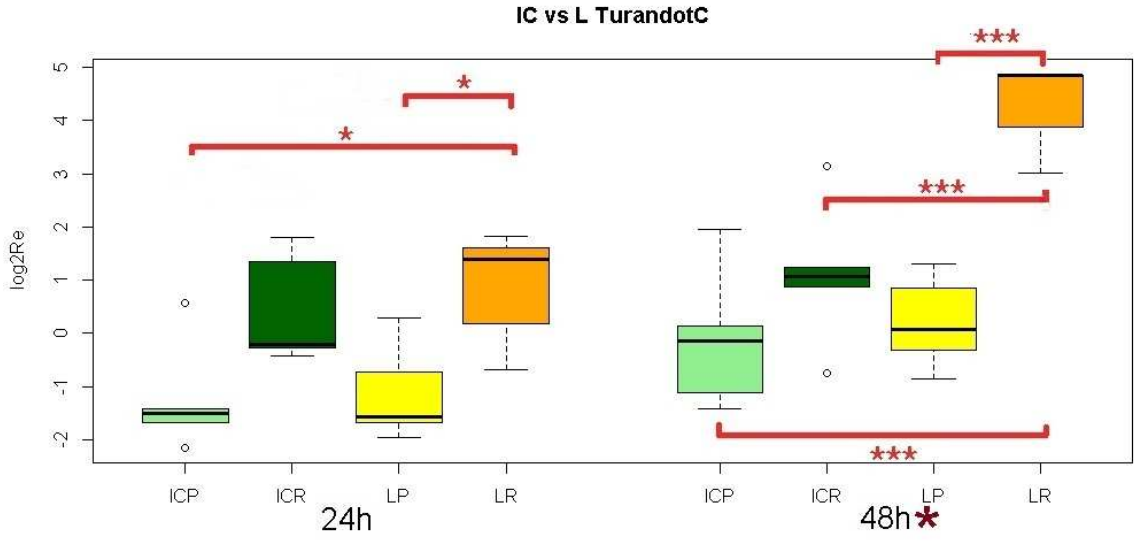
d)


IC vs L CG30151
log2Re
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
***
***
***
***
***
***
ICP
ICR
LP
LR
24h *
48h ***

e)




f)



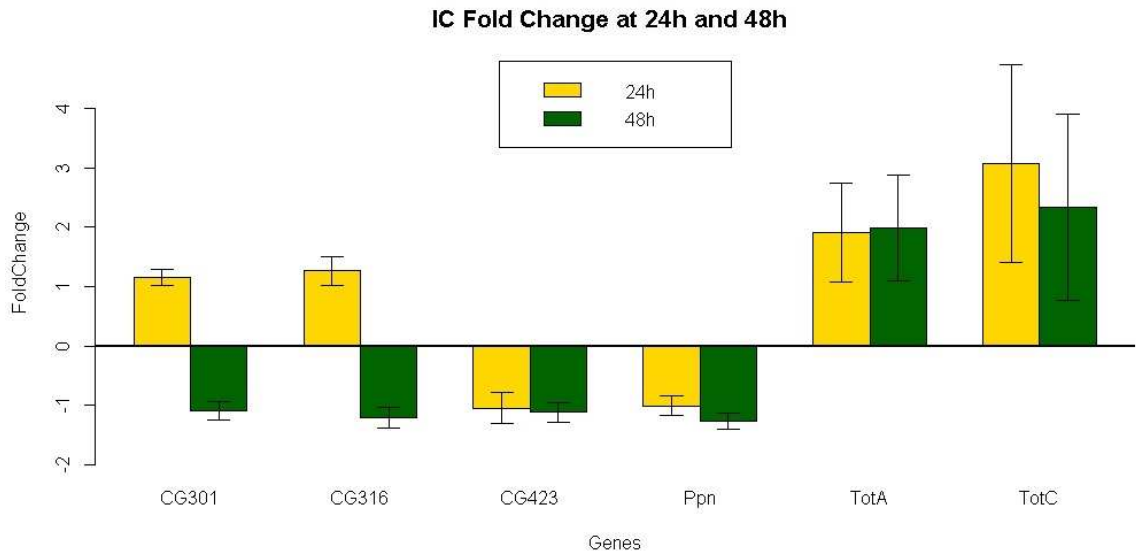

**Figures 4.4 a), b), c), d), e), f)** *R.e. values by levels for CG42369, Papilin, CG31606, CG30151 Papilin, TurandotA, TurandotC genes, showed as boxplot graphs. Red bars highlight significant difference in Tukey HSD paired comparison. Dark red stars near time point show line by treatment interaction, twoway ANOVA P-value < 0.05 (*), P-value < 0.01 (***)*

## 4.2.2 Fold change and time points difference expression

The same fold change method used in paragraph 4.1.2 for quantification of differential genes expression was also used here for the candidate genes from microarrays.

According to Figure 4.5, similarly to the proteome profile, only L line shows deregulation in genes expression when exposed to cold shock. Differential expression is observed for all the candidate genes in L: excluding CG42369, fold change is significant for all the tested genes at 48 hours. From the following results it's also possible to recognize two directions of deregulation: when upregulated, genes expression in L-R is boosted at 48 hours if compared to the 24 hours' time point. Downregulated genes, on the other hand, show similar expression for the two time points. Different fold change between 24 and 48 hours was observed in the upregulated genes TotA, TotC and CG30151 (see Appendix Table 7 and 8 for more details).

a)




b)

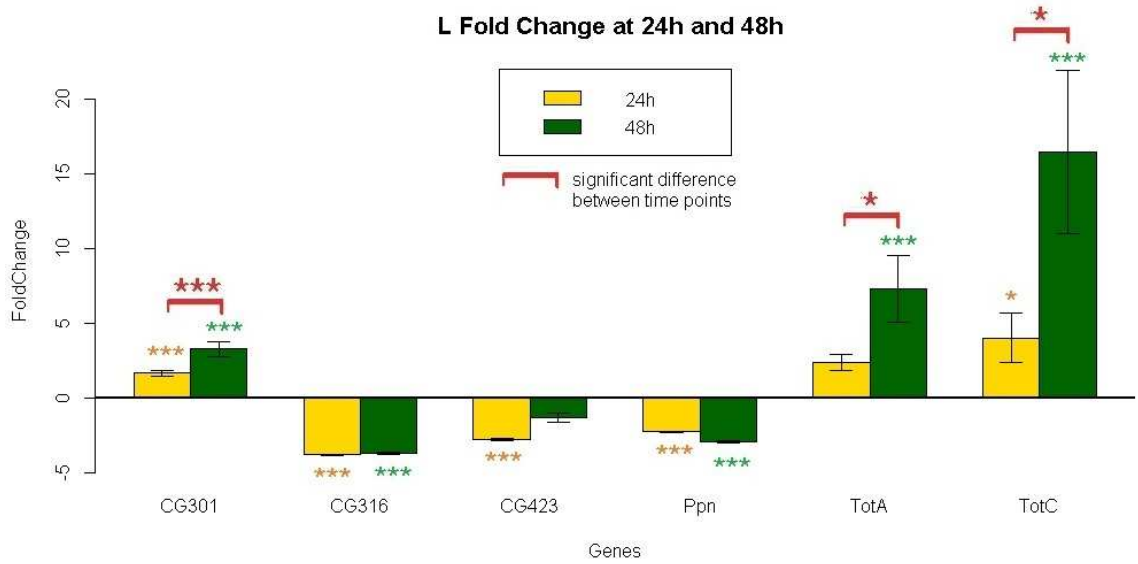



**Figure 4.5 a), b)** *Fold change values at 24 and 48 hours with mean standard error for IC and L line. Stars on the top of the bars show significant fold change. Red lines underline difference in expression between time points. P-value < 0.05 (*), P-value < 0.01(***)*

# 5. DISCUSSION AND CONCLUSIONS

## 5.1 Candidate genes from proteome profile

### 5.1.1 Gene and protein abundance comparison

### 5.1.2 Candidate genes functions

### 5.1.3 Gene expression characteristics

### 5.1.4 Stress response and heat shock proteins expression

## 5.2 Candidate genes from microarrays

# 5.1 CANDIDATE GENES FROM PROTEOME PROFILE

## 5.1.1 Gene and protein abundance comparison

This research project aims to assess transcript levels of proteins exclusively expressed in *D. melanogaster* conditional lethal inbred line. Compared to a control inbred line, seven differentially expressed candidate proteins, identified through previous explorative proteomic analysis, were detected 48 hour after cold shock: three of these proteins were exclusively expressed in the lethal inbred line after cold shock treatment and four were present also at natural temperature of growth. Here presented gene expression analysis of the candidate proteins, quantified through real time quantitative PCR (RT-qPCR), was observed at two time point after cold shock to see whether transcript correlates with protein expression and how expression differs for following two days after cold treatment. This approach, that combines multiple level investigations, involving parallel research on protein and gene expression, offers a higher level comprehension of the molecular processes affecting *D. melanogaster* inbred lines, exposed to harsh environmental conditions (Inbreeding X Environment interaction). The I X E synergy contributes to a multifactorial and complex phenotype that results still poorly characterized at physiological level. Inbreeding depression is a complex trait and causes of this phenomenon are various across populations: for this reason drawing a general conclusion it's a difficult task to carry on. Anyway the approach here used of a direct comparison between molecular expressions in related populations is a pilot study for the possible identification of key role deregulated genes.

## 5.1.2 Candidate genes functions

The seven candidate genes highlighted through proteome profile are involved in different cellular pathways, comparing their functions in "Gene Ontology" (http://www.geneontology.org). Some of these proteins, such transaldolase and phosphogluconate mutase, are involved in glucose metabolism processes, β-Tubulin covers a cellular modelling and structural role in several cellular functions. These results outline an activation of biochemical responses after a stress exposure: in the inbred line the demand of energy is hypothetically increased for coping with inner stress effects activated by the high level of inbreeding and the outer harsh

environment. β-Tubulin 56 Da is probably required for cellular dynamic processes involving microtubules kinesis and cell structures fashion. CG11796, a 4-hydroxiphenil pyruvate enzyme, and the angiotensin converting enzyme are proteins involved in more broad physiological processes such muscle twitch and blood stream regulation. The pro-phenol oxidase A1 is involved in the immune system response. The most dramatic overexpression is observed for CG18067, a 3'-5'-cyclic-nucleotide phosphodiesterase: the protein is still uncharacterized at structural level but homology with other similar protein sequences outlines several molecular functions covered by this protein. Large scale two yeast-hybrid experiments prove that CG18067 is physically in contact with other proteins involved in pathways such hemocyte proliferation, ubiquitin activation and immune response. The protein is also more generally involved in signal transduction. The molecular functions of CG18067 discovered by *in silico* methods and by high throughput experiments suggest a pivot role for this protein, observed only in the inbred line and exclusively after cold shock. The current results support the important function of this candidate gene during stress response in inbred lethal line and for this reason should be considered an excellent starting point for further future investigation.

### 5.1.3 Gene expression characteristics

The results from the proteome profile outline a differential expression of seven proteins, detected only in the lethal inbred line while poorly expressed in all the other lines. After a RT-qPCR gene quantification of the candidate genes, what mainly emerges, differently from the protein expression, is the fact that expression has a base level for all the considered genes and transcript is naturally present at the control conditions within the two analyzed lines. The reported base gene expression highlights their general role in molecular or physiological pathways and doesn't point to an exceptional function undertaken at specific conditions. Indeed inbreeding depression is a result of a global molecular deregulation in affected individuals. The current expression analysis of the affected genes shows altered product quantity for the stressful environment conditions in sensible lines and not extraordinary response pathways activation.

Excluding CG18067, transcript abundance doesn't mirror corresponding protein expression and for most of the examined genes, expression is lower after cold shock. Gene expression after stress exposition is well characterized to be lower due to a

global reduced RNA-polymerase activity (Bougdour et al., 2004): the overexpression of detected proteins combined with the poor transcript presence is indicative of post-transcriptional, post translational or slow protein turnover regulation mechanisms. The protein expression of the candidate genes uptakes a “ready-to-use” strategy in where proteins abundance is maintained high not through expression but by the upkeeping of partially mature protein forms. Molecular cold stress cell response of these genes in the lethal inbred line appears regulated at protein level: fast protein expression is triggered through the final steps of protein maturation such transport, assembly and peptides degradation. Current microarrays expression studies in *D. melanogaster* inbred lines (Garcia et al. 2012) identify a striking expression pattern for proteins associated with the endoplasmic reticulum and Golgi’s apparatus, involved in the protein synthesis, maturation and transport. These genes family results upregulated in *D. melanogaster* lines showing lower grade of inbreeding depression while in highly inbreeding depressed lines results downregulated. This research outlines also a reduced expression for genes involved in protein targeting and localization, affirming that deregulation occurs for protein functional maturation. According to these results, observed protein overexpression in the inbred line and after cold shock could be considered as the result of the missing protein expression adjustment in post-translational mechanisms, causing deleterious effects in the observed line.

### 5.1.4 Stress response and heat shock proteins expression

Fundamental role in gene expression and protein correct folding is played by the heat shock proteins (HSP), a particular set of proteins preferentially expressed after stress: this extended protein family, of which the structure and domains are conserved among many animal species, regulates global cell homeostasis in normal conditions and at the same time balances deleterious effects under pathological conditions. HSP proteins family includes molecular chaperonines fundamental for protein stabilization and folding, and a broad group of gene expression activating factors, the heat shock factors (HSF), binding directly particular DNA loci and enhancing gene transcription (Morimoto, 1998). The latter mentioned family of proteins is involved in stress activation response accomplished through an inducible gene expression mechanism. The briefly described classes of regulatory HSP proteins could be potentially active in the lethal inbred line, even if not detected by the

proteome profile, and could fulfill a cardinal function under severe environmental conditions. The protein expression of the HSP proteins in *D. melanogaster* is mainly studied in larval stage (Burton et al., 1988): the HSP protein expression is enhanced 8 to 12 hours after a 0 °C brief cold treatment. Larvae with a higher expression of heat shock proteins showed an increased survival rate in comparison to individuals from the same population. Data from this study show an important molecular role for the heat shock proteins in the recovery from cold.

Here described candidate gene studies, carried out for the identification of characteristically altered genes in inbreeding depression, can't *per se* be indicative of the molecular function of the seven identified proteins: at this point of the research it's not possible to consider if they are a damaging effect of inbreeding under severe environment or the cellular response for contrasting the lower survival rate. Further investigations on their specific molecular function are necessary for the better comprehension of the deregulated patterns. Anyway a more general and globally assessing hypothesis of deregulation is here below proposed. Considering the two possible genetic mechanisms of inbreeding depression, the dominance and overdominance hypothesis, and the before described HSP homeostasis strategies devised by environmentally stressed organisms, it's now possible to speculate over the presence of deregulated patterns in *D. melanogaster* inbred lethal lines. The dominance hypothesis considers the damaging effect of rare recessive mutations present only in inbred lines. Mutations in coding regions could alter the protein structure or the oligomerization domains of regulatory proteins such the HSP and the global loss of molecular expression balance is then observable in the lower survival rate inbred lines. The overdominance hypothesis that affirms the advantageous effect of heterozygous combination of alleles could be also considered as one of the possible mechanisms affecting global organism homeostasis that leads to lethal effects in the inbred line. The overdominance inbreeding trend could affect at both protein and transcriptional level: a typical example explaining this disturbed balance is the case of transcriptional factors, as showed below in Figure 5.1. Homozygosis for one allele excludes the expression the other variant of transcriptional factor: only stable co-expression of both the factors induces optimal transcription level, condition reached exclusively in heterozygosis. Increased homozygosity affects products quantity through deregulated transcription, due to mutations in cis-regulatory elements, promoter of sequences or activity of enhancers and transcription factors. Theoretically, the optimum HSP levels could be damaged by genetic overdominant mechanism and the global gene and protein expression undergoes to deregulation.

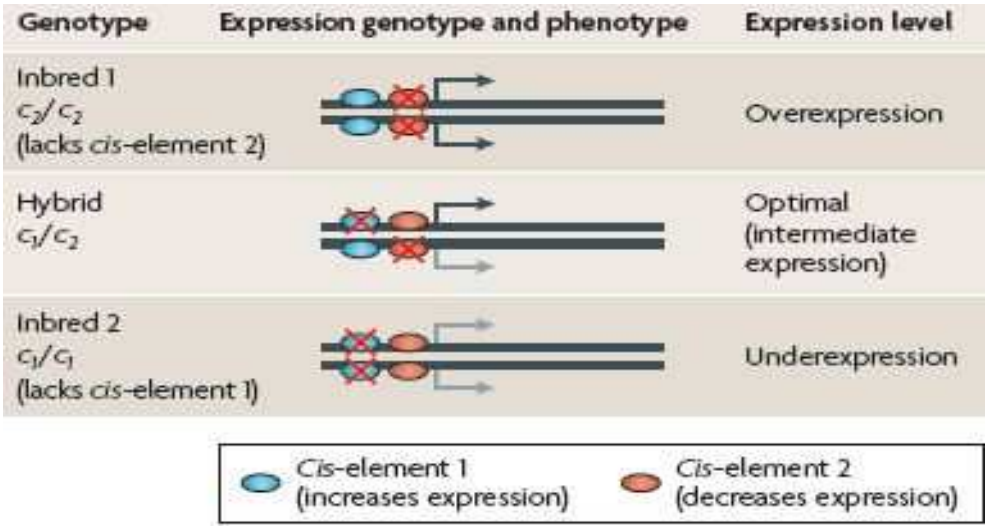



**Figure 5.1** *Example of ideal overdominance: balanced expression of heterogeneous transcriptional factors ensures optimal transcript production. Homozygosity related to transcriptional factors deregulates expression. This proposed mechanism could be hypothetically involved in global organism deregulation through unbalanced expression of HSP*

Effects of deregulation in gene expression are mostly observed at the 48 hours' time point, indicative of the fact that recorded expression modulation is mostly induced by slowly acting mechanisms. Gene expression for Prophenol oxidase, Transaldolase, Phosphogluconate mutase and β-Tubulin at 56 D follows a progressive downregulation in the considered time points. This aspect of the gene expression is indicative of a later response to cold shock, reaching the apex point 48 hours after the treatment. Hypothetically, expression for these genes is regulated by complex transcript inhibition factors or through the suppression of previously active transcriptional factors.

Previous microarrays studies (Ayroles et al., 2009) state that gene expression in *D. melanogaster* inbred lines is altered by additive or partially additive mechanisms of regulation. Differential gene expression in inbred individuals is driven through an epistatic effect or transactivation mechanisms. Considering this expression architecture involving few regulatory genes is probable that here analyzed candidate genes are not involved in the differential genes expression in inbred lines.

# 5.2 CANDIDATE GENES FROM MICROARRAYS

Oligonucleotide microarrays technique was here used for assessing different gene expression levels between *D. melanogaster* inbred and lethal lines, exposed to cold treatment comparing to expression at natural temperature of growth. The most differentially expressed results were validated in this project through RT-qPCR technique.

The six considered genes result highly deregulated only after cold shock and exclusively in the lethal inbred line. From these genes CG42369, Papilin and CG31606 follow negative gene expression and Turandot A, Turandot C and CG30151 an upregulation; expression observed in the control inbred line is maintained constant at the two considered treatments.

From these candidate genes only Papilin and both Turandot genes are well annotated and characterized also at protein level. Turandot is an extended protein family, secreted *in vivo* as a humoral factors in the cell fat bodies. These proteins are involved in the resistance against pathogen bacteria and environmental stress, inducing lethal effects at organism level. Turandot gene expression in the current study is highly upregulated after cold shock. Papilin is a component of the essential extracellular matrix, important element of the cell rearrangements at extracellular level. Papilin four isoforms are underexpressed after cold shock in the lethal line.

The two techniques used for the gene expression quantification are based on different detection approaches and, as a consequence, the statistics methods for data analysis are diverse. RT-qPCR is based on gene amplification reaction of RNA converted to cDNA and is highly dependent on efficiency. Calculated N0 values are afterwards normalized against constantly expressed reference genes. In the microarrays technique the quantification is carried out through RNA probe hybridization and resulted fluorescence assessment. Detected values are normalized against the background value and usually the fluorescence is biased by aspecific binding due to the high density probes of the microarrays. Even if biases and normalization methods of the two gene expression quantification techniques are based on different chemical and physical properties, validation led to a confirmation of the two data sets and the RT-qPCR technique here used for validating fluorescence data from microarrays gene expression analysis showed a positive correlation with the previously obtained results. Previous studies (Morey et al, 2006)

prove that positive correlation is only verified for genes that show at least 1,6 fold-change between treatments: in this way the results are affected at a lower rate by technical error, due to incompatibility between the two methods.

# APPENDIX TABLES

1)

### CG18067 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| Group | 3 | 28.515 | 9.505 | 20.56 | 9.79e-06 | *** |
| Residuals | 16 | 7.398 | 0.462 | | | |

| | diff | lwr | upr | p adj | |
|---|---|---|---|---|---|
| ICR-ICP | 0.2457715 | -0.9846682 | 1.476211 | 0.9391575 | |
| LP-ICP | 0.6028055 | -0.6276341 | 1.833245 | 0.5162274 | |
| LR-ICP | 2.9956081 | 1.7651684 | 4.226048 | 0.0000173 | *** |
| LP-ICR | 0.3570341 | -0.8734056 | 1.587474 | 0.8393989 | |
| LR-ICR | 2.7498366 | 1.5193969 | 3.980276 | 0.0000479 | *** |
| LR-LP | 2.3928025 | 1.1623629 | 3.623242 | 0.0002266 | *** |

### CG18067 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| Group | 3 | 39.20 | 13.066 | 18.64 | 1.8e-05 | *** |
| Residuals | 16 | 11.22 | 0.701 | | | |

| | diff | lwr | upr | p adj | |
|---|---|---|---|---|---|
| ICR-ICP | 0.4195814 | -1.0955824 | 1.934745 | 0.8568972 | |
| LP-ICP | 0.6045423 | -0.9106215 | 2.119706 | 0.6702504 | |
| LR-ICP | 3.5346911 | 2.0195273 | 5.049855 | 0.0000289 | *** |
| LP-ICR | 0.1849609 | -1.3302029 | 1.700125 | 0.9848329 | |
| LR-ICR | 3.1151097 | 1.5999459 | 4.630273 | 0.0001236 | *** |
| LR-LP | 2.9301488 | 1.4149851 | 4.445313 | 0.0002405 | *** |

### bTub56D 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| Group | 3 | 1.1077 | 0.3692 | 7.31 | 0.00264 | * |
| Residuals | 16 | 0.8081 | 0.0505 | | | |

| | diff | lwr | upr | p adj | |
|---|---|---|---|---|---|
| ICR-ICP | 0.02086212 | -0.3858016 | 0.4275259 | 0.9988234 | |
| LP-ICP | 0.19726345 | -0.2094003 | 0.6039272 | 0.5242660 | |
| LR-ICP | 0.58658091 | 0.1799172 | 0.9932447 | 0.003957 | *** |
| LP-ICR | 0.17640132 | -0.2302624 | 0.5830651 | 0.6111183 | |
| LR-ICR | 0.56571879 | 0.1590550 | 0.9723825 | 0.005335 | *** |
| LR-LP | 0.38931746 | -0.0173463 | 0.7959812 | 0.0630576 | |

### bTub56D 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| Group | 3 | 1.550 | 0.5167 | 4.647 | 0.0161 | * |
| Residuals | 16 | 1.779 | 0.1112 | | | |

| | diff | lwr | upr | p adj | |
|---|---|---|---|---|---|
| ICR-ICP | 0.139095404 | -0.4642339 | 0.7424247 | 0.9106167 | |
| LP-ICP | 0.137634612 | -0.4656947 | 0.740963940 | 0.9130742 | |
| LR-ICP | -0.537300200 | -1.1406295 | 0.0660291 | 0.0898836 | |
| LP-ICR | -0.001460792 | -0.6047901 | 0.60186853 | 0.999999 | |
| LR-ICR | -0.676395604 | -1.2797249 | -0.0730662 | 0.025408 | * |
| LR-LP | -0.674934812 | -1.2782641 | -0.0716054 | 0.025759 | * |

### CG11796 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| Group | 3 | 3.729 | 1.2429 | 8.655 | 0.00121 | *** |
| Residuals | 16 | 2.298 | 0.1436 | | | |

| | diff | lwr | upr | p adj | |
|---|---|---|---|---|---|
| ICR-ICP | 0.23967783 | -0.4460435 | 0.9253992 | 0.7515664 | |
| LP-ICP | 0.30655290 | -0.3791684 | 0.9922742 | 0.5884773 | |
| LR-ICP | 1.14388344 | 0.4581621 | 1.8296048 | 0.001072 | *** |
| LP-ICR | 0.06687506 | -0.6188463 | 0.7525964 | 0.9921151 | |
| LR-ICR | 0.90420561 | 0.2184843 | 1.5899269 | 0.008138 | *** |
| LR-LP | 0.83733054 | 0.1516092 | 1.5230519 | 0.0143269 | * |

### CG11796 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 0.8272 | 0.2757 | 2.714 | 0.0794 |
| Residuals | 16 | 1.6256 | 0.1016 | | |

### Angiotensin converting enzyme 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| Group | 3 | 0.6220 | 0.20732 | 4.062 | 0.0253 | * |
| Residuals | 16 | 0.8167 | 0.05104 | | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.20937202 | -0.6181829 | 0.19943889 | 0.4797623 |
| LP-ICP | 0.08729916 | -0.3215118 | 0.49611008 | 0.9271061 |
| LR-ICP | -0.36297841 | -0.7717893 | 0.04583250 | 0.0911436 |
| LP-ICR | 0.29667119 | -0.1121397 | 0.70548210 | 0.2027530 |
| LR-ICR | -0.15360639 | -0.5624173 | 0.25520453 | 0.7090917 |
| LR-LP | -0.45027757 | -0.8590885 | -0.04146666 | 0.0284043 |

### Angiotensin converting enzyme 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 0.2169 | 0.07231 | 0.673 | 0.581 |
| Residuals | 16 | 1.7185 | 0.10741 | | |

### Pgm 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 0.6149 | 0.2050 | 1.363 | 0.29 |
| Residuals | 16 | 2.4057 | 0.1504 | | |

### Pgm 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 4.046 | 1.3487 | 35.86 | 2.48e-07 *** |
| Residuals | 16 | 0.602 | 0.0376 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.09425951 | -0.4451879 | 0.2566688 | 0.867425 |
| LP-ICP | 0.04360333 | -0.3073250 | 0.3945317 | 0.98403 |
| LR-ICP | -1.04920410 | -1.4001324 | -0.6982758 | 0.00000 *** |
| LP-ICR | 0.13786284 | -0.2130655 | 0.4887912 | 0.680589 |
| LR-ICR | -0.95494458 | -1.3058729 | -0.6040162 | 0.00000 *** |
| LR-LP | -1.09280743 | -1.4437358 | -0.7418791 | 0.00000 *** |

### Prophenol oxidase 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 4.315 | 1.4384 | 2.862 | 0.0696 |
| Residuals | 16 | 8.043 | 0.5027 | | |

### Prophenol oxidase 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 57.86 | 19.285 | 84.5 | 5.01e-10 *** |
| Residuals | 16 | 3.65 | 0.228 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | 0.1485945 | -0.7158272 | 1.0130163 | 0.9597916 |
| LP-ICP | -1.8012573 | -2.6656791 | -0.9368355 | 0.0001064 *** |
| LR-ICP | -4.0560996 | -4.9205214 | -3.1916778 | 0.000000 *** |
| LP-ICR | -1.9498519 | -2.8142737 | -1.0854301 | 0.000043 *** |
| LR-ICR | -4.2046941 | -5.0691159 | -3.3402723 | 0.000000 *** |
| LR-LP | -2.2548422 | -3.1192640 | -1.3904204 | 0.0000074 *** |

### Transaldolase 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 1.119 | 0.3731 | 7.98 | 0.00178 *** |
| Residuals | 16 | 0.748 | 0.0467 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.20995575 | -0.60119052 | 0.18127902 | 0.440795 |
| LP-ICP | 0.44329728 | 0.05206251 | 0.83453205 | 0.02373 * |
| LR-ICP | 0.03455545 | -0.35667933 | 0.4257902 | 0.9941066 |
| LP-ICR | 0.65325303 | 0.2620182 | 1.04448 | 0.0010627 *** |
| LR-ICR | 0.24451120 | -0.1467235 | 0.63574597 | 0.3144751 |
| LR-LP | -0.40874183 | -0.7999766 | -0.0175070 | 0.0390435 * |

### Transaldolase 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 16.069 | 5.356 | 48.51 | 2.95e-08 *** |
| Residuals | 16 | 1.767 | 0.110 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.09630383 | -0.6975693 | 0.5049617 | 0.9670256 |
| LP-ICP | 0.23406331 | -0.3672022 | 0.8353288 | 0.6865701 |
| LR-ICP | -2.00545345 | -2.6067189 | -1.40418 | 0.0000003 *** |
| LP-ICR | 0.33036714 | -0.2708983 | 0.9316326 | 0.4210255 |
| LR-ICR | -1.90914962 | -2.5104151 | -1.30788 | 0.0000006 *** |
| LR-LP | -2.23951676 | -2.8407822 | -1.638251 | 0.0000001 *** |

*Proteome profile, oneway ANOVA: comparison of log2 - relative expression (R.e) between four groups: ICR, ICP, LP, LR. Whether P value is significant, Tukey HSD at 95 % family-wise confidence is used for multiple comparisons. P-value < 0.05 (*), P-value < 0.01 (***)*

2)

| CG18067.24h | CG18067.48h |
|---|---|
| Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 14.050 14.050 30.39 4.73e-05 ***<br>Treatment 1 8.703 8.703 18.82 0.000508 ***<br>Line:Treatment 1 5.762 5.762 12.46 0.0027 ***<br>Residuals 16 7.398 0.462 | Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 17.295 17.295 24.67 0.000140 ***<br>Treatment 1 14.026 14.026 20.00 0.000385 ***<br>Line:Treatment 1 7.879 7.879 11.24 0.004049 ***<br>Residuals 16 11.219 0.701 |
| **bTub56D.24h** | **bTub56D.48h** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 0.7277 0.7277 14.407 0.00159 ***<br>Treatment 1 0.2103 0.2103 4.164 0.05816<br>Line:Treatment 1 0.1697 0.1697 3.360 0.08548<br>Residuals 16 0.8081 0.0505 | Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 0.3628 0.3628 3.264 0.0897<br>Treatment 1 0.3589 0.3589 3.228 0.0913<br>Line:Treatment 1 0.8283 0.8283 7.450 0.0148 *<br>Residuals 16 1.7788 0.1112 |
| **CG11796.24h** | **CG11796.48h** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 1.8324 1.8324 12.759 0.00255 ***<br>Treatment 1 1.4499 1.4499 10.09 0.00585 ***<br>Line:Treatment 1 0.4465 0.4465 3.109 0.09695<br>Residuals 16 2.2978 0.1436 | Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 0.3762 0.3762 3.702 0.0723<br>Treatment 1 0.4192 0.4192 4.126 0.0592<br>Line:Treatment 1 0.0318 0.0318 0.313 0.5834<br>Residuals 16 1.6256 0.1016 |
| **Ance.24h** | **Ance.48h** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 0.0055 0.0055 0.108 0.74707<br>Treatment 1 0.5439 0.5439 10.656 0.0048 ***<br>Line:Treatment 1 0.0725 0.0725 1.421 0.25059<br>Residuals 16 0.8167 0.0510 | Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 0.0788 0.07878 0.733 0.404<br>Treatment 1 0.0917 0.09174 0.854 0.369<br>Line:Treatment 1 0.0464 0.04641 0.432 0.520<br>Residuals 16 1.7185 0.10741 |
| **PPOA.24h** | **PPOA.48h** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 0.003 0.003 0.006 0.9415<br>Line 1 3.577 3.577 6.734 0.0189 *<br>Treatment:Line 1 0.144 0.144 0.272 0.6090<br>Residuals 17 9.031 0.531 | Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 5.55 5.55 24.30 0.000151 ***<br>Line 1 45.09 45.09 197.57 2.02e-10 ***<br>Treatment:Line 1 7.22 7.22 31.64 3.80e-05 ***<br>Residuals 16 3.65 0.23 |
| **Pgm.24h** | **Pgm.48h** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 0.4151 0.4151 2.761 0.116<br>Treatment 1 0.1996 0.1996 1.328 0.266<br>Line:Treatment 1 0.0002 0.0002 0.001 0.973<br>Residuals 16 2.4057 0.1504 | Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 1.0382 1.0382 27.60 7.87e-05 ***<br>Treatment 1 1.7614 1.7614 46.83 3.95e-06 ***<br>Line:Treatment 1 1.2464 1.2464 33.14 2.95e-05 ***<br>Residuals 16 0.6018 0.0376 |
| **Tal.24h** | **Tal.48h** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 0.5914 0.5914 12.649 0.00263 ***<br>Treatment 1 0.4785 0.4785 10.235 0.0055 ***<br>Line:Treatment 1 0.0494 0.0494 1.057 0.31928<br>Residuals 16 0.7480 0.0467 | Df Sum Sq Mean Sq F value Pr(>F)<br>Line 1 3.507 3.507 31.77 3.72e-05 ***<br>Treatment 1 6.820 6.820 61.77 6.98e-07 ***<br>Line:Treatment 1 5.742 5.742 52.00 2.08e-06 ***<br>Residuals 16 1.767 0.110 |

*Genes from proteome profile, twoway ANOVA: comparison of log2 - relative expression (R.e) between two groups: Line (IC, L) and Treatment (P,R). F value (1,16) for α - 0.05 = 4, 49. P-value < 0.05 (*), P-value < 0.01 (***)*

3)

| CG18067.L | CG18067.IC |
|---|---|
| Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 28.33 28.33 54.089 1.62e-06 ***<br>Threatment 1 47.88 47.88 91.412 5.11e-08 ***<br>Time:Threatment 1 1.37 1.37 2.611 0.126<br>Residuals 16 8.38 0.52 | Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 17.597 17.597 21.207 0.000293 ***<br>Threatment 1 0.795 0.795 0.958 0.342343<br>Time:Threatment 1 0.217 0.217 0.261 0.616129<br>Residuals 16 13.277 0.830 |
| bTub56D.L | bTub56D.IC |
| Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 0.1229 0.1229 0.839 0.373<br>Threatment 1 0.4228 0.4228 2.887 0.109<br>Time:Threatment 1 0.3805 0.3805 2.597 0.127<br>Residuals 16 2.3437 0.1465 | Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 0.1484 0.1484 1.156 0.298<br>Threatment 1 0.1070 0.1070 0.833 0.375<br>Time:Threatment 1 0.1602 0.1602 1.247 0.281<br>Residuals 16 2.0552 0.1285 |
| CG11796.L | CG11796.IC |
| Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 0.453 0.453 6.178 0.0244 *<br>Threatment 1 5.383 5.383 73.444 2.25e-07 ***<br>Time:Threatment 1 0.002 0.002 0.030 0.8638<br>Residuals 16 1.173 0.073 | Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 0.416 0.4162 1.762 0.203<br>Threatment 1 0.419 0.4191 1.775 0.201<br>Time:Threatment 1 0.055 0.0550 0.233 0.636<br>Residuals 16 3.779 0.2362 |
| Ance.L | Ance.IC |
| Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 5.225 5.225 61.706 7.02e-07 ***<br>Threatment 1 0.175 0.175 2.065 0.16997<br>Time:Threatment 1 1.059 1.059 12.501 0.00275 **<br>Residuals 16 1.355 0.085 | Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 11.269 11.269 128.799 4.6e-09 ***<br>Threatment 1 0.121 0.121 1.379 0.257<br>Time:Threatment 1 0.060 0.060 0.689 0.419<br>Residuals 16 1.400 0.087 |
| PPOA.L | PPOA.IC |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 12.520 12.520 59.51 8.86e-07 ***<br>Treatment 1 7.004 7.004 33.29 2.87e-05 ***<br>Tim: Treatment 1 5.739 5.739 27.28 8.38e-05 ***<br>Residuals 16 3.366 0.210 | Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 1.227 1.2268 2.357 0.144<br>Treatment 1 0.044 0.0439 0.084 0.775<br>Time:Treatment 1 0.015 0.0150 0.029 0.867<br>Residuals 16 8.328 0.5205 |
| Pgm.L | Pgm.IC |
| Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 2.3618 2.3618 29.917 5.14e-05 ***<br>Threatment 1 0.2225 0.2225 2.819 0.113<br>Time:Threatment 1 0.1942 0.1942 2.460 0.136<br>Residuals 16 1.2631 0.0789 | Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 0.3229 0.3229 2.660 0.122<br>Threatment 1 0.0293 0.0293 0.241 0.630<br>Time:Threatment 1 0.1401 0.1401 1.154 0.299<br>Residuals 16 1.9423 0.1214 |
| Tal.L | Tal.IC |
| Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 0.0024 0.00241 0.023 0.882<br>Threatment 1 0.0402 0.04020 0.383 0.545<br>Time:Threatment 1 0.1565 0.15653 1.491 0.240<br>Residuals 16 1.6794 0.10496 | Df Sum Sq Mean Sq F value Pr(F)<br>Time 1 3.866 3.866 86.80 7.29e-08 ***<br>Threatment 1 3.989 3.989 89.57 5.88e-08 ***<br>Time:Threatment 1 2.173 2.173 48.79 3.08e-06 ***<br>Residuals 16 0.713 0.045 |

*Genes from proteome profile, twoway ANOVA: comparison of log2 - relative expression (R.e) between two groups: Time point (24, 48) and Treatment (P, R). F value (1, 16) for α - 0.05 = 4, 49. P-value < 0.05 (*), P-value < 0.01 (***)*

4)

| Genes | Significant Fold Change | Direction |
|---|---|---|
| CG18067 | 24h : 4.3<br>48h : 7.7 | +<br>+ |
| B Tub 56D | 48 h: - 1.6 | - |
| CG11796 | 24h : 1.8<br>48h : 1.3 | +<br>+ |
| Ance | 24 h: - 1.4 | - |
| PPO-A | 48 h : -5 | - |
| Pgm | 48h : - 2.1 | - |
| Tal | 48 h: -4.6 | - |

*Significant fold change values for tested genes, chosen from the proteome profile. Direction of deregulation is showed besides (+ up and – down)*

5)

CG30151 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 2.148 | 0.716 | 10.69 | 0.000422 *** |
| Residuals | 16 | 1.072 | 0.067 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | 0.18120077 | -0.2871056 | 0.6495072 | 0.6905108 |
| LP-ICP | 0.13743256 | -0.3308738 | 0.6057389 | 0.8349057 |
| LR-ICP | 0.84710602 | 0.3787996 | 1.3154124 | 0.0004824 *** |
| LP-ICR | -0.04376821 | -0.5120746 | 0.4245382 | 0.9930410 |
| LR-ICR | 0.66590525 | 0.1975989 | 1.1342116 | 0.0044586 *** |
| LR-LP | 0.70967346 | 0.2413671 | 1.1779798 | 0.0025887 *** |

CG30151 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 7.801 | 2.6004 | 19.39 | 1.41e-05 *** |
| Residuals | 16 | 2.146 | 0.1341 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.171192 | -0.8338606 | 0.4914765 | 0.8798947 |
| LP-ICP | -0.502387 | -1.1650555 | 0.1602816 | 0.1742091 |
| LR-ICP | 1.156183 | 0.4935149 | 1.8188519 | 0.0006930 *** |
| LP-ICR | -0.331195 | -0.9938635 | 0.3314736 | 0.4999153 |
| LR-ICR | 1.327375 | 0.6647069 | 1.9900440 | 0.0001646 *** |
| LR-LP | 1.658570 | 0.9959018 | 2.3212389 | 0.0000123 *** |

CG31606 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 21.604 | 7.201 | 47.58 | 3.39e-08 *** |
| Residuals | 16 | 2.422 | 0.151 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | 0.2944210 | -0.4095248 | 0.99836691 | 0.6376067 |
| LP-ICP | -0.4330501 | -1.1369960 | 0.27089576 | 0.3272432 |
| LR-ICP | -2.3708322 | -3.0747780 | -1.66688629 | 0.0000003 *** |
| LP-ICR | -0.7274711 | -1.4314170 | -0.02352527 | 0.0415810 * |
| LR-ICR | -2.6652532 | -3.3691991 | -1.96130732 | 0.0000000 *** |
| LR-LP | -1.9377820 | -2.6417279 | -1.23383617 | 0.0000037 *** |

CG31606 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 21.935 | 7.312 | 34.83 | 3.03e-07 *** |
| Residuals | 16 | 3.358 | 0.210 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.1718233 | -1.000835 | 0.6571882 | 0.9327403 |
| LP-ICP | -0.6987944 | -1.527806 | 0.1302172 | 0.1148411 |
| LR-ICP | -2.6344982 | -3.463510 | -1.8054867 | 0.0000006 *** |
| LP-ICR | -0.5269710 | -1.355983 | 0.3020405 | 0.3009256 |
| LR-ICR | -2.4626749 | -3.291686 | -1.6336633 | 0.0000014 *** |
| LR-LP | -1.9357038 | -2.764715 | -1.1066923 | 0.0000286*** |

### CG42369 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 9.653 | 3.218 | 10.13 | 0.000557 *** |
| Residuals | 16 | 5.084 | 0.318 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.22992303 | -1.2499137 | 0.7900676 | 0.9157804 |
| LP-ICP | -0.14437761 | -1.1643683 | 0.8756131 | 0.9767833 |
| LR-ICP | -1.71793518 | -2.7379259 | -0.6979445 | 0.0009779 *** |
| LP-ICR | 0.08554542 | -0.9344453 | 1.1055361 | 0.9949402 |
| LR-ICR | -1.48801214 | -2.5080028 | -0.4680215 | 0.0035964 *** |
| LR-LP | -1.57355756 | -2.5935482 | -0.5535669 | 0.0022095 *** |

### CG42369 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 8.173 | 2.7243 | 6.236 | 0.00523 *** |
| Residuals | 16 | 6.990 | 0.4369 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.1964093 | -1.392430 | 0.9996110 | 0.9646245 |
| LP-ICP | -0.8740670 | -2.070087 | 0.3219533 | 0.1980156 |
| LR-ICP | -1.6290898 | -2.825110 | -0.4330695 | 0.0063189 *** |
| LP-ICR | -0.6776577 | -1.873678 | 0.5183627 | 0.3952742 |
| LR-ICR | -1.4326805 | -2.628701 | -0.2366602 | 0.0163788 * |
| LR-LP | -0.7550228 | -1.951043 | 0.4409975 | 0.3064329 |

### Papilin 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 8.324 | 2.7748 | 28.47 | 1.19e-06 *** |
| Residuals | 16 | 1.559 | 0.0975 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.05476114 | -0.6196647 | 0.5101425 | 0.9922532 |
| LP-ICP | -0.40105824 | -0.9659618 | 0.1638454 | 0.2178685 |
| LR-ICP | -1.59891489 | -2.1638185 | -1.0340113 | 0.0000026 *** |
| LP-ICR | -0.34629710 | -0.9112007 | 0.2186065 | 0.3300964 |
| LR-ICR | -1.54415376 | -2.1090573 | -0.9792502 | 0.0000041 *** |
| LR-LP | -1.19785666 | -1.7627602 | -0.6329531 | 0.0000875 *** |

### Papilin 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 14.880 | 4.960 | 31.17 | 6.46e-07 *** |
| Residuals | 16 | 2.546 | 0.159 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | -0.3434199 | -1.0652159 | 0.3783761 | 0.5398156 |
| LP-ICP | -0.6178488 | -1.3396448 | 0.1039472 | 0.1074541 |
| LR-ICP | -2.2472283 | -2.9690243 | -1.5254323 | 0.0000007 *** |
| LP-ICR | -0.2744289 | -0.9962249 | 0.4473671 | 0.7017061 |
| LR-ICR | -1.9038084 | -2.6256044 | -1.1820124 | 0.0000064 *** |
| LR-LP | -1.6293795 | -2.3511755 | -0.9075835 | 0.0000426 *** |

### TurandotA 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 6.863 | 2.2877 | 4.342 | 0.0203 * |
| Residuals | 16 | 8.430 | 0.5269 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | 0.9309302 | -0.3825036 | 2.2443639 | 0.2190392 |
| LP-ICP | 0.2576490 | -1.0557847 | 1.5710828 | 0.9420850 |
| LR-ICP | 1.4979996 | 0.1845658 | 2.8114333 | 0.0227505 * |
| LP-ICR | -0.6732811 | -1.9867149 | 0.6401526 | 0.4790140 |
| LR-ICR | 0.5670694 | -0.7463643 | 1.8805032 | 0.6145844 |
| LR-LP | 1.2403505 | -0.0730832 | 2.5537843 | 0.0676181 |

### TurandotA 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 31.32 | 10.441 | 14.02 | 9.66e-05 *** |
| Residuals | 16 | 11.92 | 0.745 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | 0.8456703 | -0.7160541 | 2.407395 | 0.4332573 |
| LP-ICP | 0.4375877 | -1.1241368 | 1.999312 | 0.8526553 |
| LR-ICP | 3.2342110 | 1.6724866 | 4.795935 | 0.0001140 *** |
| LP-ICR | -0.4080827 | -1.9698071 | 1.153642 | 0.8763521 |
| LR-ICR | 2.3885407 | 0.8268163 | 3.950265 | 0.0023865 *** |
| LR-LP | 2.7966234 | 1.2348989 | 4.358348 | 0.0005343 *** |

### TurandotC 24h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 17.38 | 5.794 | 5.547 | 0.00835 *** |
| Residuals | 16 | 16.71 | 1.044 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | 1.6867046 | -0.1625514 | 3.535961 | 0.0802771 |
| LP-ICP | 0.1074305 | -1.7418255 | 1.956687 | 0.9982960 |
| LR-ICP | 2.1004488 | 0.2511928 | 3.949705 | 0.0233664 * |
| LP-ICR | -1.5792741 | -3.4285301 | 0.269982 | 0.1085480 |
| LR-ICR | 0.4137442 | -1.4355118 | 2.263000 | 0.9174376 |
| LR-LP | 1.9930183 | 0.1437623 | 3.842274 | 0.0324628 * |

### TurandotC 48h

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Group | 3 | 60.83 | 20.276 | 15.71 | 5e-05 *** |
| Residuals | 16 | 20.65 | 1.291 | | |

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| ICR-ICP | 1.2321776 | -0.8236019 | 3.287957 | 0.3485263 |
| LP-ICP | 0.3273859 | -1.7283936 | 2.383165 | 0.9675552 |
| LR-ICP | 4.4101870 | 2.3544075 | 6.465966 | 0.0000767 *** |
| LP-ICR | -0.9047917 | -2.9605713 | 1.150988 | 0.6002933 |
| LR-ICR | 3.1780093 | 1.1222298 | 5.233789 | 0.0021692 *** |
| LR-LP | 4.0828011 | 2.0270215 | 6.138581 | 0.0001807 *** |

*Microarray validation, oneway ANOVA: comparison of log2 - relative expression (R.e) between four groups: ICR, ICP, LP, LR. Whether P value is significant, Tukey HSD at 95 % family-wise confidence is used for multiple comparisons. P-value < 0.05 (*), P-value < 0.01 (***)*

6)

| CG30151.24h | CG30151.48h |
|---|---|
| Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 0.9921 0.9921 14.811 0.0014 ***<br>Line 1 0.8067 0.8067 12.043 0.0031 ***<br>Treatment:Line 1 0.3491 0.3491 5.212 0.03644 *<br>Residuals 16 1.0717 0.0670 | Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 2.765 2.765 20.619 0.000334 ***<br>Line 1 0.851 0.851 6.343 0.022804 *<br>Treatment:Line 1 4.185 4.185 31.204 4.1e-05 ***<br>Residuals 16 2.146 0.134 |
| CG31606.24h | CG31606.48h |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 3.376 3.376 22.30 0.00023 ***<br>Line 1 11.999 11.999 79.28 1.35e-07 ***<br>Treatment:Line 1 6.22 6.22 41.15 8.55e-06 ***<br>Residuals 16 2.422 0.151 | Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 5.552 5.552 26.45 9.82e-05 ***<br>Line 1 12.494 12.494 59.52 8.85e-07 ***<br>Treatment:Line 1 3.889 3.889 18.53 0.000546 ***<br>Residuals 16 3.358 0.210 |
| CG42369.24h | CG42369.48h |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 4.066 4.066 12.795 0.00252 ***<br>Line 1 3.331 3.331 10.483 0.00515 ***<br>Treatment:Line 1 2.257 2.257 7.102 0.01694 *<br>Residuals 16 5.084 0.318 | Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 1.132 1.132 2.590 0.12709<br>Line 1 6.651 6.651 15.224 0.00127 ***<br>Treatment:Line 1 0.390 0.390 0.893 0.35877<br>Residuals 16 6.990 0.437 |
| Papilin.24h | Papiln.48h |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 1.961 1.961 20.12 0.000374 ***<br>Line 1 4.730 4.730 48.53 3.18e-06 ***<br>Treatment:Line 1 1.633 1.633 16.76 0.000848 ***<br>Residuals 16 1.559 0.097 | Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 4.865 4.865 30.57 4.58e-05 ***<br>Line 1 7.948 7.948 49.95 2.66e-06 ***<br>Treatment:Line 1 2.067 2.067 12.99 0.00238 ***<br>Residuals 16 2.546 0.159 |
| TurandotA.24h | TurandotA.48h |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 5.893 5.893 11.185 0.00412 ***<br>Line 1 0.850 0.850 1.614 0.22214<br>Treatment:Line 1 0.120 0.120 0.227 0.64010<br>Residuals 16 8.430 0.527 | Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 16.583 16.583 22.261 0.000232 ***<br>Line 1 9.984 9.984 13.403 0.002109 ***<br>Treatment:Line 1 4.758 4.758 6.387 0.022410 *<br>Residuals 16 11.919 0.745 |
| TurandotC.24h | TurandotC.48h |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 16.925 16.925 16.205 0.000978 ***<br>Line 1 0.340 0.340 0.325 0.576493<br>Treatment:Line 1 0.117 0.117 0.112 0.741902<br>Residuals 16 16.711 1.044 | Df Sum Sq Mean Sq F value Pr(>F)<br>Treatment 1 33.54 33.54 24.387 0.000179 ***<br>Line 1 16.20 16.20 11.777 0.003708 **<br>Treatment:Line 1 9.34 9.34 6.791 0.019864 *<br>Residuals 15 20.63 1.38 |

*Microarray validation, twoway ANOVA: comparison of log2 - relative expression (R.e) between groups: Line (IC, L) and Treatment (P, R). F value (1, 16) for α - 0.05 = 4, 49. P-value < 0.05 (*), P-value < 0.01 (***)*

7)

| CG30151.L | CG30151.IC |
|---|---|
| Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 1.347 1.347 14.59 0.00151***<br>Treatment 1 7.011 7.011 75.94 1.8e-07 ***<br>Time:Treatment 1 1.126 1.126 12.19 0.00302 ***<br>Residuals 16 1.477 0.092 | Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 1.2917 1.2917 11.875 0.00332***<br>Treatment 1 0.0001 0.0001 0.001 0.97335<br>Time:Treatment 1 0.1552 0.1552 1.427 0.24967<br>Residuals 16 1.7405 0.1088 |
| **CG31606.L** | **CG31606.IC** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 0.769 0.769 5.205 0.0366 *<br>Treatment 1 18.755 18.755 126.987 5.09e-09 ***<br>Time:Treatment 1 0.000 0.000 0.000 0.9953<br>Residuals 16 2.363 0.148 | Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 0.650 0.6499 3.043 0.100<br>Treatment 1 0.019 0.0188 0.088 0.771<br>Time:Treatment 1 0.272 0.2717 1.272 0.276<br>Residuals 16 3.417 0.2136 |
| **CG42369.L** | **CG42369.IC** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 5.887 5.887 12.454 0.00279 ***<br>Treatment 1 6.778 6.778 14.337 0.00162***<br>Time:Treatment 1 0.837 0.837 1.772 0.20185<br>Residuals 16 7.564 0.473 | Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 10.115 10.115 35.881 1.89e-05 ***<br>Treatment 1 0.227 0.227 0.806 0.383<br>Time:Treatment 1 0.001 0.001 0.005 0.945<br>Residuals 16 4.510 0.282 |
| **Papilin.L** | **PapiIn.IC** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 0.017 0.017 0.148 0.706<br>Treatment 1 9.992 9.992 86.105 7.7e-08 ***<br>Time:Treatment 1 0.233 0.233 2.006 0.176<br>Residuals 16 1.857 0.116 | Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 0.6013 0.6013 4.279 0.0552<br>Treatment 1 0.1982 0.1982 1.410 0.2524<br>Time:Treatment 1 0.1042 0.1042 0.741 0.4020<br>Residuals 16 2.2487 0.1405 |
| **TurandotA.L** | **TurandotA.IC** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 26.108 26.108 60.641 7.85e-07 ***<br>Treatment 1 20.371 20.371 47.317 3.71e-06 ***<br>Time:Treatment 1 3.027 3.027 7.032 0.0174 *<br>Residuals 16 6.889 0.431 | Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 8.248 8.248 9.804 0.00644 ***<br>Treatment 1 3.945 3.945 4.690 0.04579 *<br>Time:Treatment 1 0.009 0.009 0.011 0.91852<br>Residuals 16 13.460 0.841 |
| **TurandotC.L** | **TurandotC.IC** |
| Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 28.40 28.40 33.037 3.0e-05 ***<br>Treatment 1 46.14 46.14 53.670 1.7e-06 ***<br>Time:Treatment 1 5.46 5.46 6.349 0.0227 *<br>Residuals 16 13.76 0.86 | Df Sum Sq Mean Sq F value Pr(>F)<br>Time 1 3.972 3.972 2.692 0.1203<br>Treatment 1 10.650 10.650 7.218 0.0162 *<br>Time:Treatment 1 0.258 0.258 0.175 0.6812<br>Residuals 16 23.607 1.475 |

*Genes from micrarrays, twoway ANOVA: comparison of log2 - relative expression (R.e) between two groups: Time point (24, 48) and Treatment (P,R). F value (1,16) for α - 0.05 = 4, 49. P-value < 0.05 (*), P-value < 0.01 (***)*

8)

| Genes | Significant Fold Change | Direction |
|---|---|---|
| CG42368 | 24h : - 2.8 | - |
| Ppn | 24h: - 2.3<br>48 h: - 2.9 | -<br>- |
| CG31606 | 24h : - 3.8<br>48h : - 3.7 | -<br>- |
| CG30151 | 24 h: 1.6<br>48 h: 3.3 | +<br>+ |
| TotA | 24h : 2.4<br>48 h : 7.3 | +<br>+ |
| TotC | 24h : 4<br>48h : 16.5 | +<br>+ |

*Significant fold change values for tested genes, chosen for microarrays validation. Direction of deregulation is showed besides (+ up and – down)*

# APPENDIX PROTOCOLS

## 1. RNA extraction using TRIzol® reagent

1. Preparing samples:

i. Add 100 µL Trizol to each sample
ii. Homogenizing for 30-60 seconds
iii. Add other 900 µL Trizol to each sample

2. RNA extraction

i. When all your samples are finished, let your samples incubate at room temperature for 5 minutes.
ii. Add 200 µL chloroform to the samples
iii. Close the tubes, cover the rack and shake as hard as possible for 15 seconds
iv. Incubate 2-3 minutes at room temperature.
v. Centrifuge the samples at 12.000x *g* for 15 minutes at 2°C
vi. Add 500 µL isopropanol and mix your sample.
vii. Incubate at room temperature for 10 minutes
viii. Centrifuge at 12.000 x *g* for 10 minutes at 2°C.
ix. Discard supernatant carefully not to discard the (blue) pellet.
x. Add 1mL 75% EtOH (RNAse free) and vortex
xi. Centrifuge at 7.500 x *g* for 5 minutes at 2°C and discard supernatant.
xii. Airdry the samples for ±15 minutes
xiii. Add 26 µL $H_2O$/DEPC or MilliQ from the kit, mix and leave at room temperature for ±15 minutes.

## 2. DNA – free Kit Applied Biosystems, DNA removal after TRIzoL Extraction Routine DNase digestion ≤200 µg DNA

i. Transfer RNA volumes to 0.5 mL eppendorf.
ii. Add 2.6 µl 10X DNase I Buffer and 1 µl rDNase I to the DNA and mix gently.
iii. Incubate at 37 ° for 20-30 min.
iv. Add 3 µl resuspended DNase Inactivation Reagent.
v. Incubate 2 min at room temperature, mixing occasionally by tapping gently the eppendorf.
vi. Centrifuge at 10 000 x g for 1.5 min and transfer the RNA to a fresh tube.

1. Armbruster, Reed: Inbreeding depression in benign and stressful environments, Heredity (Edinb).;95(3):235-42, 2005

2. Bakker , van Rijswijk, Weissing, Bijlsma: Consequences of fragmentation for the ability to adapt to novel environments in experimental Drosophila metapopulations, Conserv Genet 11:435–448, DOI 10.1007/s10592-010-0052-5, 2010

3. Bell, Gonzalez: Evolutionary rescue can prevent extinction following environmental change, Ecology Letters, 12: 942–948 doi: 10.1111/j.1461-0248.2009.01350.x, 2009

4. Bijsma, Bundgaard, Boerema: Does inbreeding affect the extinction risk of small populations? :predictions from Drosophila, EVOL. BIOL. 13,502±514, 2000

5. Bio-Rad Laboratories: Real-Time PCR Applications Guide, 2006

6. Biron, Loxdale, Ponton, Moura, Marché, Brugidou, Thomas: Population proteomics: An emerging discipline to study metapopulation ecology, PROTEOMICS Volume 6, Issue 6, pages 1712–1715, 2006

7. Bougdour , Lelong , Geiselmann: Crl, a low temperature-induced protein in Escherichia coli that binds directly to the stationary phase sigma subunit of RNA polymerase, J Biol Chem. 2004 May 7;279(19):19540-50. Epub 2004

8. Brenda Thornton, Chhandak Basu: Real-time PCR (qPCR) primer design using free online software, Biochemistry and Molecular Biology Education, Volume 39, Issue 2, pages 145–154, 2011

9. Burton, Mitchell, Young, Petersen: Heat shock protection against cold stress of Drosophila melanogaster, doi: 10.1128/MCB.8.8.3550Mol. Cell. Biol. vol. 8 no. 8 3550-3552, 1988

10. Charlesworth, Charlesworth: Inbreeding Depression and its Evolutionary Consequences, Annual Review of Ecology and Systematics, Vol. 18: 237-268, DOI:10.1146/annurev.es.18.110187.001321, 1987

11. Charlesworth, Willis: The genetics of inbreeding depression, Nature Reviews Genetics 10, 783-79 | doi:10.1038/nrg2664, 2009

12. Chomczynski , Sacchi: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, Anal Biochem;162(1):156-9, 1987

13. Crow: Minor viability mutants in Drosophila, Genetics [ 92(1 Pt 1 Suppl):s165-72], 1979

14. East: Studies on Size Inheritance in Nicotiana, Genetics; 1(2): 164–176, 1916

15. Frankham, Conservation genetics, AnnlL Rev. Genetics. 29:305-27, 1995

16. Fleige , Pfaffl: RNA integrity and the effect on the real-time qRT-PCR performance, Mol Aspects Med. 2006 Apr-Jun;27(2-3):126-39. Epub 2006

17. García, Ávila, Quesada, Caballero: Gene expression changes caused by inbreeding protect against inbreeding depression in Drosophila, Genetics,10.1534/genetics.112.142687, 2012

18. Gomez , Gomez : Statistical Procedures in Agricultural Research, By K. A. New York, Chichester, etc.: Wiley, 2nd edition, paperback, pp. 680, price not stated, 1984

19. Hästbacka, de la Chapelle, Kaitila, Sistonen, Weaver, Lander: Linkage disequilbrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nature Genetics 2:204-21, 1992
20. Hedrick: Genetics of Populations, 2005

21. Hellemans, Mortier, De Paepe, Speleman, Vandesompele: qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data, Genome Biol.;8(2):R19, 2007

22. Issaq, Veenstra:Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives, Biotechniques;44(5):697-8, 700. doi: 10.2144/000112823, 2008

23. Jonge, Fehrmann, de Bont, Hofstra, Gerbens, Kamps, de Vries, van der Zee, Meerman, Elst :Evidence Based Selection of Housekeeping Genes, PLoS ONE 2(9): e898. doi:10.1371/journal.pone.0000898, 2007

24. Karlen, McNair, Perseguers, Mazza, Mermod: Statistical significance of quantitative PCR, BMC Bioinformatics, 8:131 doi:10.1186/1471-2105-8-131, 2007

25. Keightley, Trivedi, Thomson, et al.: Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines, Genome Res. 2009 19: 1195-1201, 2009

26. Keller, Waller: Inbreeding effects in wild populations, TRENDS in Ecology & Evolution Vol.17 No.5, 2002

27. Kristensen, Pedersen, Vermeulen, Loeschcke: Research on inbreeding in the 'omic' era, Trends Ecol Evol. Jan;25(1):44-52. doi: 10.1016/j.tree.2009.06.014. 2009

28. Liberg, Andre´, Pedersen, Sand, Sejberg, Wabakken, Kesson, Bensch: Severe inbreeding depression in a wild wolf (Canis lupus) population, Biol. Lett. (2005) 1, 17–20, doi:10.1098/rsbl.2004.0266, 2005

29. Madsen, Stille c, Shine b: Inbreeding depression in an isolated population of adders Vipera Berus, Biological Conservation 75 (1996) 113-118, 1995

30. McCune, Houle, McMillan, Annable, Kondrashov: Two classes of deleterious recessive alleles in a natural population of zebrafish, Danio rerio, Proc. R. Soc. Lond. B 271, 2025–2033, doi:10.1098/rspb.2004.2787, 2004

31. Morey, Ryan, Van Dolah: Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR, Biol Proced Online; 8:175-93. 2006.

32. Morimoto: Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators, 0.1101/gad.12.24.3788Genes & Dev. 12:3788-3796, 1998

33. Morton, Crow, Muller. An estimate of the mutational damage in man from data on consanguineous marriage, Proc Natl Acad Sci USA. 1956

34. Nolan, Hands, Bustin: Quantification of mRNA using real-time RT-PCR, Nat Protoc; 1(3):1559-82. 2006

35. Pardo, Piras, Asproni, van der Gaag, Gabbas, Ruiz-Linares, de Knijff, Monne, Rizzu, Heutink: Dissecting the genetic make-up of North-East Sardinia using a large set of haploid and autosomal markers, European Journal of Human Genetics 20, 956–964, 2012

36. Ponton, Chapuis, Pernice, Sword , Simpson: Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in Drosophila melanogaster, J Insect Physiol. 2011 Jun;57(6):840-50. doi: 10.1016/j.jinsphys.2011.03.014. Epub 2011

37. Rieu, Powers: Real-Time Quantitative RT-PCR: Design, Calculations and Statistics, The Plant Cell, Vol. 21: 1031–1033, 2009

38. Ruijter, Ramakers, Hoogaars, Karlen, Bakker , van den Hoff, Moorman: Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data, Nucleic Acids Res. 2009 Apr;37(6):e45. doi: 10.1093/nar/gkp045. Epub 2009

39. Sarup, Pedersen, Nielsen, Malmendal, Loeschcke: The metabolic profile of long-lived Drosophila melanogaster, PLoS One.;7(10):e47461. doi: 10.1371/journal.pone.0047461. 2012

40. Shul : Hybrid Seed Corn, Science;103(2679):547-50, 1946

41. Slate, Kruuk, Marshall, Pemberton, Clutton-Brock: Inbreeding depression influences lifetime breeding success in a wild population of red deer (Cervus elaphus), Proc. R. Soc. Lond. B 2000 267, doi: 10.1098/rspb.2000.1192, 2000

42. Tan: Experimental validation of microarrays data

43. Taylor, Wakem, Dijkman, Alsarraj, Nguyen: A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines, Methods;50(4):S1-5. doi: 10.1016/j.ymeth.2010.01.005, 2010

44. Vandesompele, De Preter, Pattyn, Poppe, Van Roy, De Paepe, Speleman: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 2002 Jun 18;3(7):RESEARCH0034. Epub 2002

45. VanGuilder , Vrana, Freeman : Twenty-five years of quantitative PCR for gene expression analysis, Biotechniques;44(5):619-26. doi: 10.2144/000112776, 2008

46. Vermeulen, Bijlsma: Characterization of conditionally expressed mutants affecting age-specific survival in inbred lines of Drosophila melanogaster: lethal conditions and temperature-sensitive periods, Genetics.;167(3):1241-8, 2004

47. Vermeulen, Bijlsma: Changes in mortality patterns and temperature dependence of lifespan in Drosophila melanogaster caused by inbreeding, Heredity (Edinb).;92(4):275-81, 2004

48. Ye, Coulouris , Zaretskaya, Cutcutache, Rozen , Madden: Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinformatics ;13:134. doi: 10.1186/1471-2105-13-134,2012

FlyBase: http://flybase.org/

FlyAtlas: http://www.flyatlas.org/

Gene Ontology: http://www.geneontology.org/

Uniprot: http://www.uniprot.org/

NCBI: http://www.ncbi.nlm.nih.gov/tools/primer-blast/

Primer3: http://frodo.wi.mit.edu/

Beacon Desiner free online software: http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1

mFold free online software: http://eu.idtdna.com/UNAFold?