The consequences of stochastic gene expression in the nematode *Caenorhabditis elegans*

Alejandro Raúl Burga Ramos

Tesi Doctoral UPF
Barcelona, 2012

Thesis Supervisor

Ben Lehner, PhD

Genetic Systems Laboratory
EMBL/CRG Systems Biology Research Unit,
Centre for Genomic Regulation (CRG)
Universitat Pompeu Fabra (UPF)

A mi familia y al Perú

Acknowledgements

Le debo tanto a tanta gente que me temo nunca podré saldar mi deuda. Hace ya 10 años que dejé mi país y con ello he dejado también a muchas personas queridas atrás quienes me han ayudado tanto que me cuesta enumerarlas, empezando por mi familia. Gracias a mis padres por su apoyo incondicional, por enseñarme con su ejemplo el valor del esfuerzo, la perseverancia y del trabajo duro, por enseñarme que siempre hay que dar sin esperar recibir nada a cambio. Sin aquellos valores, todo el conocimiento que uno pueda acumular no vale de nada. Gracias a mi hermana y mi abuela por su cariño y siento tanto no haber estado con todos ustedes todo este tiempo. Perdón por mi lejanía y ausencia, aunque poca justificación tengan.

Obviamente este trabajo no hubiera sido posible sin Ben Lehner. Ben, muchas gracias por darme la oportunidad de trabajar a tu lado, por las largas conversaciones y discusiones que dieron lugar a este trabajo. Por ser alguien cercano y compartir tu pasión por la ciencia.

Gracias a todos los miembros del laboratorio de Ben, en especial a Jenny por siempre ser un referente a la hora de hacer experimentos, a Laura por mantener todo en orden, a Adam y a Mirko por la mezcla de ciencia y risas (a

veces más risas que ciencia). A Mark Isalan y a los miembros de su laboratorio, por su ayuda y por hacer gratas las horas en el laboratorio.. A tantas personas del CRG que conocí a lo largo de los años en especial a Adam. Kiana, Camilla y Christian por su amistad. También quisiera agradecer a Timo y Raquel del ALMU por su ayuda. Tampoco puedo olvidar a mis amigos y profesores en Chile, con quienes aprendí y quienes me enseñaron a hacer ciencia.

Gracias Pancho Barriga, por ser como un hermano. Por todos esos viernes en el “Nevermind” o el “Gato Negro” con una piscola en la mano, hablando de nuestros sueños y temores. Gracias a ti y a la Javi por recordarme siempre (aunque muchas veces infructuosamente) que no es necesario renunciar a todo en la vida.

Table of Contents

Abstract…… 1

Preface…….. 5

Introduction……..7

1. Stochasticity as an inherent property of life……..7

2. Stochasticity in gene expression……11

2.1 Studies in prokaryotes…… 12
2.2 Studies in eukaryotes…….. 23

2.2.1 Stochastic gene expression in *S. cerevisiae*…… 23
2.2.2 Stochastic gene expression in metazoans……27

3. The genotype-phenotype map……33

3.1 The contribution of the genotype……34
3.2 The contribution of the environment……..40
3.3 Unexplained phenotypic variability……41

4. The interplay between mutations and stochastic gene expression……45

4.1 Phenotypic variability in model organisms……... 45
4.2 Variability in the outcome of mutations……. 50
4.3 Incomplete penetrance of mutations in *C. elegans*…… 55

Objectives……65

Results and Discussion……67

Predicting mutation outcome from early stochastic variation in genetic interaction partners……67

Discussion……... 105

Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans……...139

Discussion……... 175

Conclusions…….. 195

References……. 197

Abstract

Genetically identical cells and organisms growing in homogenous environmental conditions can show significant phenotypic variation. Furthermore, mutations often have consequences that vary among individuals (incomplete penetrance). Biochemical processes such as those involved in gene expression are subjected to fluctuations due to their inherent probabilistic nature. However, it is not clear how these fluctuations affect multicellular organisms carrying mutations and if stochastic variation in gene expression among individuals could confer any advantage to populations. We have investigated the consequences of stochastic gene expression using the nematode *Caenorhabditis elegans* as a model. Here we show that inter-individual stochastic variation in the induction of both specific and more general buffering systems combine to determine the outcome of inherited mutations in each individual. Also, we demonstrate that Individuals with higher induction of a stress response are more robust to the effects of diverse mutations, however they incur a fitness cost, thus suggesting that variation at the population level in stress resistance could be beneficial in unpredictable environments.

Resumen

Células y organismos genéticamente idénticos y creciendo en un ambiente homogéneo pueden mostrar diferencias en sus fenotipos. Además, una misma mutación puede afectar de un modo distinto a individuos de una misma población. Es sabido que los procesos bioquímicos responsables de la expresión de genes están sujetos a fluctuaciones debido a su inherentemente naturaleza probabilística. Sin embargo, el rol que juegan estas fluctuaciones en individuos portadores de mutaciones ha sido poco estudiado, así cómo si la expresión estocástica de genes puede conferir alguna ventaja al nivel poblacional. Para investigar las consecuencias de la expresión estocástica de genes usamos como modelo al nemátodo *Caenorhabditis elegans.* En este trabajo demostramos que existe variación estocástica entre individuos en la inducción de mecanismos (tanto gen específicos como globales) que confieren robustez al desarrollo. En consecuencia, diferencias fenotípicas entre mutantes están determinadas por la variación conjunta de dichos mecanismos. También, demostramos que aquellos Individuos que inducen estocásticamente una mayor respuesta a stress, están fenotípicamente mejor protegidos al efecto de diversas mutaciones pero incurren en un costo reproductivo importante. Eso sugiere, que la variación estocástica al nivel poblacional en la respuesta a stress podría ser

beneficiosa cuando las poblaciones afrontan ambientes impredecibles.

variation in the expression of genes could be exploited and beneficial under some circumstances.

Our work gave as a result two publications that I present here. The first publication (Burga et al.) had as a main objective to understand the phenomenon of incomplete penetrance. My supervisor Ben Lehner and I designed the experiments and conceived the model. I performed all the experiments and data analyses in this publication. Olivia Casanueva, a post-doc researcher in our lab, first showed that mutations in *C. elegans* could be rescued by chaperone overexpression and hormesis. The second manuscript (Casanueva et al.) studied the interplay between genetic and environmental robustness and the trade-off associated with it. This study was the main project of Olivia Casanueva. She performed all the experiments and analyses. My contribution to the second paper was the generation of a *daf-21* reporter transgenic line that was used by Olivia to show that variation in the levels of constitutively expressed chaperones can predict mutation outcome and its associated fitness trade-off. I also contributed to this work together with Ben Lehner with the conceptual framework of using stochastic variability in gene expression to explain the incomplete penetrance of mutations.

Preface

At the beginning of this PhD project it was already quite clear from studies starting in the early 2000's that microorganisms such as bacteria and yeast showed extensive variation in the expression of genes among individuals (noise). These differences in expression could not be attributed to genetic nor environmental variation. There was also evidence that mammalian cells in culture showed noisy gene expression. This stochastic scenario contrasted with the highly deterministic view of many biological processes. On the hand, it has been also known for many years that model organisms growing in the laboratory conditions show extensive phenotypic variation that could not be explained by neither genetics nor environment factors. Moreover, mutations often have consequences that vary among individuals, a phenomenon known as "incomplete penetrance". Incompletely penetrant mutations occur in organisms ranging from bacteria to humans, yet very little was known about its causes.

We decided to test whether the incomplete penetrance of mutations was related to the stochastic expression of genes using *Caenorhabditis elegans* as a model. Given that phenotypic variation was also evident in wild type worms, we were interested to investigate whether

Introduction

1. Stochasticity as an inherent property of life.

With the birth of life, the Earth was also witness to the transition from a world of possibilities to a world of probabilities and apparent certainties. Organisms have evolved from components governed by the laws of physics and chemistry and thus, have been and will continue to be influenced by their stochastic or probabilistic nature (see Table I for examples). On the other hand, biologists have typically been amazed by the reproducibility and efficiency in which diverse processes such as cell division, DNA replication and ultimately development are accomplished. This is particularly surprising given that these processes ultimately depend on the interaction of hundreds or even thousands of genes and their correct expression and function in both time and space. As a consequence of this apparent macroscopic determinism, a molecular view of the cell was born, in which every component in the cell had a particular role and function in close analogy to machines engineered by humans and any deviation from the expected outcome is usually called an “error”. But if cells are made of components that behave probabilistically, how is it possible that cells and organisms behave deterministically? How can then these two ideas be reconciled?

The physicist Erwin Schrödinger had already brought attention to this matter in his influential book *What is life?* (Schrödinger, 1945). He reasoned that one way in which cells could escape thermodynamic fluctuations due to the low copy number of their components was simply by having large copy number of those components. When Schrödinger wrote this book, virtually nothing was known about the molecular identity of the inheritance factors discovered by Mendel or how was this information decoded, as later described by the central dogma of molecular biology (Crick, 1970), yet his intuition regarding the low copy number or abundance of many key cellular components and its implications was right. Nowadays, we know that that the copy number of genes, mRNAs and even certain proteins can be extremely low (see Table II).

Conrad H. Waddington was one of the first biologists to consider the importance of errors and fluctuations in biological systems. Inspired by his studies in classical embryology and development, he concluded that organisms must have developed some way to guide and guarantee the reproducibility of development in the presence of perturbations. He termed such processes “canalization” (Waddington, 1942, 1957). In accordance with this view, mutants usually show increased phenotypic variance compared to wild-type individuals, suggesting that the normal canalization process has been perturbed.

Robustness, a concept used in physics and engineering, and analogous to biological canalization, refers to the correct function of a system in the presence of perturbations. The structure or topology of existing molecular networks have been shown to be an important factor conferring robustness to variation in biochemical parameters (Barkai and Leibler, 1997). For instance, simple negative feed-back mechanisms, prevalent in biological networks (Shen-Orr et al., 2002), could also have an important role minimizing these fluctuations (Becskei and Serrano, 2000).

Table I

Examples of stochastic phenomena leading to variability in diverse cellular processes.

Process	Component	Variation/Noise	Reference
Inheritance	DNA	Mutation	(Luria and Delbruck, 1943)
Gene expression	mRNA and protein	Variation in copy number	(Elowitz et al., 2002)
	mRNA and protein	Transcriptional or translational errors	(Meyerovich et al., 2010)
Gene expression/ Inheritance	DNA/histones	Variation in DNA methylation and histone pos-translational modifications	(Becker et al., 2011; Morgan et al., 1999)
Signaling	Protein	Promiscuous interactions	(Vavouri et al., 2009)
Metabolism	Protein	Promiscuous catalysis	(Aharoni et al., 2005)

2. Stochasticity in gene expression

Biochemists and geneticists studying the basic mechanisms of gene expression successfully identified, isolated and reconstituted *in vitro* the basic machinery responsible for transcription and translation. They relied on techniques that required large populations of cells as starting material. For instance, in order to detect or isolate a particular enzyme, sometimes liters of bacterial or cell culture are necessary. These strategies were incredibly successful in identifying the key molecules involved in gene expression, but did not allowed researchers to assess the variability present in single cell or individuals. For example, one can estimate the copy number of a particular protein in a cell to be 500 molecules from bulk measurements, but that doesn't mean that every cell has exactly 500 molecules; a bimodal distributions could also fit with this observation and that could have important implications (Huang, 2009). Novick and Weiner showed in 1957 that the expression of the ß-galactosidase enzyme was binary at the single bacterium level at low inducer levels (Novick and Weiner, 1957). Biochemical reactions that occur inside cells are expected to show intrinsic fluctuations determined by the structure, reactions rates and concentrations of the species in the biochemical network (Becskei and Serrano, 2000; Ozbudak et al., 2002; Thattai and van Oudenaarden, 2001).

Table II

Copy number of diverse molecules in *Escherichia coli*. (*) This number was measured just for genes whose protein copy number per cell was >100. Adapted and modified from (Xie et al., 2008)

Molecules	Copy number per cell
Gene	1-5
mRNA	0.05-5 (*)
Protein	0.1-10,000
Total RNAPs	1,500-10,000
Total Ribosomes	6,800-72,000

2.1 Studies in prokaryotes

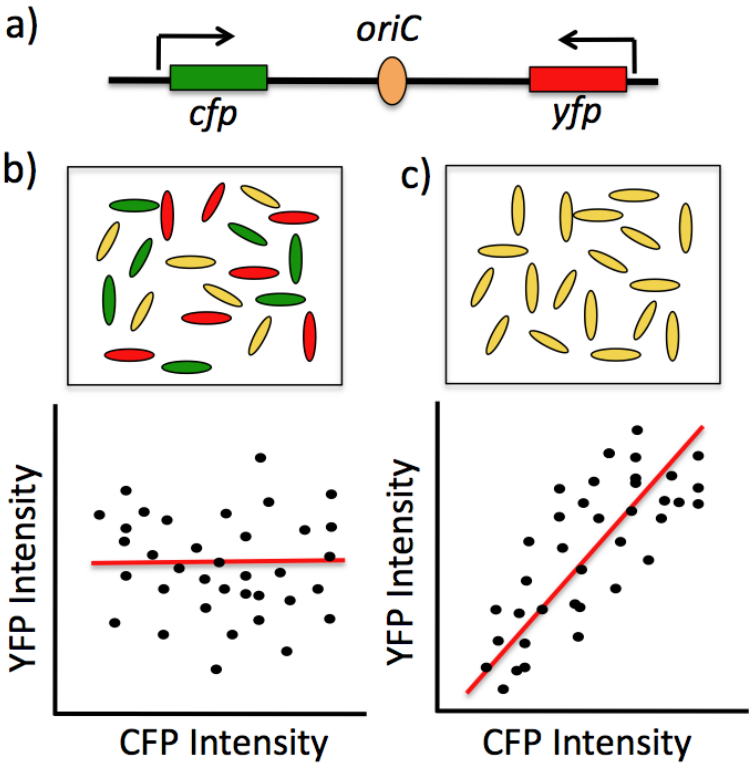
One of the first studies that took into consideration the stochastic expression of genes was performed by McAdams and colleagues (Arkin et al., 1998). The authors studied and modeled the phage λ lysis-lysogeny decision. Phage λ is a virus that infects *Escherichia coli*. Following bacterial infection two paths are possible. In the first possible path, more viral particles are produced leading to the lysis of the infected cell. Alternatively, phage DNA may integrate itself into the host cell chromosome (lysogenic pathway) and be continuously replicated together with the host genome. Environmental factors, such as stress, can lead to a switch from lysogenic to the lytic pathway. Both pathways are mutually exclusive and, interestingly, in a

population different individual infected bacterial cells follow different paths. This random developmental path choice between the lysogenic or lytic path was shown to result from the inevitable fluctuations in the temporal pattern of protein concentration caused by molecular-level thermal fluctuations in the rates of reactions (Arkin et al., 1998; McAdams and Arkin, 1997). However, the possibility always exists that part of the apparent stochasticity is due to some previously unidentified variable (Zeng et al., 2010). The fact that “switch” like behavior could be accomplished based on the inherent stochastic expression of genes has important implications for the understanding of differentiation process of more complex organisms. For instance, mammalian adult stem cells have the capacity to produce both stem cells and cells that are committed to terminal differentiation, ensuring their self-renewal and the homeostasis of tissues (Hall and Watt, 1989). There is growing evidence that in contrast to an asymmetric mechanism of division (one stem cells generates one stem cell and a committed cell), single stem cells can be heterogeneous in their decisions, being stochastic at the single cell level but generating the asymmetry at the population level (Simons and Clevers, 2011). Mechanistic insights gained in simpler systems, such as the choice between the lysogenic or lytic path in phage λ, may help us to understand how stochastic decisions at the single stem cell level impact tissue homeostasis.

Though experimental observations were consistent with the stochastic expression of genes, it was not until 2002 that a strategy designed by Elowitz and co-workers allowed the direct visualization of stochastic gene expression at the single cell level (Figure 1) (Elowitz et al., 2002).

Figure 1

Intrinsic and extrinsic noise can be measured and distinguished with two genes under the control of identical promoters in *E.coli* (a) Expression of both genes is uncorrelated when intrinsic noise is predominant (b) and correlated when extrinsic noise is predominant (c). Each dot in the plots represents a single bacterium. Adapted from (Elowitz et al., 2002)



A major advance in molecular biology that made this possible was the previous isolation, characterization and development of fluorescent protein reporters (Chalfie et al., 1994). Elowitz and co-workers inserted genes coding for cyan (cfp) and yellow (yfp) fluorescent proteins under the control of the same promoter in the genome of *E. coli*. Both genes were placed equidistant from the origin of replication to guarantee the absence of differences in gene copy number. Microscopy measurements of the simultaneous fluorescence of both proteins revealed that there was substantial variation in the expression of each gene among individual bacteria. The relative deviation from the average, measured as the standard deviation divided by the mean (sd/mean), is a common measure of gene expression noise and it's also known as the coefficient of variation (CV). The lack of correlation in the expression of both genes defines operationally the intrinsic noise due to biochemical fluctuations. Extrinsic sources of noise favor the correlation of both reporters. Dual reporter constructs driven by an artificial lac-repressible promoter we incorporated in different genetic backgrounds. In a strain where the *lacI* repressor was deleted, the total noise, measured as CV, was low (0.08) and the intrinsic noise was 0.05. However, in a wild type background (expressing *lacI*) where bacteria produced only 3-6% as much protein, both the intrinsic and extrinsic noise increased by a factor of 5. The authors thus

confirmed the existence of intrinsic noise or fluctuations, but also noticed that there was an extrinsic component that contributed substantially to the overall variation. This extrinsic noise is caused in principle by fluctuations of other cellular components, such as general transcriptional and translational machinery that should affect all genes equally.

That same year, studies performed by Van Oudenaarden and colleagues in *Bacillus subtilis* showed that stochastic variability could be regulated by genetic parameters. Stochastic modeling theory suggests that both the processes of transcription and translation can influence the variability in the expression of genes. The authors decided to test the relative contribution of these variables to the total expression variability. They used a model the expression of GFP reporter driven by an inducible promoter. They modulated the translational efficiency using different strains with mutations in the ribosomal binding site (RBS) and initiation codon of *gfp*. They could also vary the transcriptional efficiency by changing the concentration of inducer (IPTG) or by introducing point mutations in the promoter. The authors used noise strength, defined as the variance divided by the mean (sd^2/mean), as a measure of expression variability. This measure is used primarily to reveal trends otherwise obscured by the $1/\sqrt{N}$ (N=number of molecules) scaling of noise arising from low molecular abundance, circumventing the trivial effect of decreased

noise with increased mean. The authors found that noise strength showed a strong positive correlation with translational efficiency, in contrast to the correlation for transcriptional efficiency. Therefore, increased translational efficiency was shown to be the major source of increased inter-individual variability (Ozbudak et al., 2002). This is consistent with a model in which proteins are produced in random and sharp bursts, resulting in strong fluctuations. For genes expressed at similar levels, a combination of low transcription rate and high translation efficiency causes increased gene expression noise compared with a combination of high transcription rate and low translational efficiency (Kaern et al., 2005; Ozbudak et al., 2002). However, the authors measured the variability of using a single fluorescent reporter, so that it was not possible to differentiate the relative contribution of intrinsic and extrinsic sources of noise.

One major drawback of these early studies (Elowitz et al., 2002; Ozbudak et al., 2002) was that they focused on a very limited set of genes and many of those synthetic, so that no general conclusions about the sources and extent of this variability could be drawn. Xie and co-workers later overcame this issue (Taniguchi et al., 2010). They made use of an *E. coli* YFP tag reporter library to quantify with single molecule resolution the variation in the expression of ~1,000 endogenously tagged proteins. The abundance of

proteins ranged over five orders of magnitude from 10^{-1} to 10^4 copies per cell. Thus, at a given point some proteins are not even present in an individual bacterium. Further, the average protein in *E. coli* has ~10 copies per cell. They also found that below 10 copies per cell, protein noise is inversely proportional to abundance, indicative of a high intrinsic source of variation. Simple Poisson production and degradation of mRNA and protein, commonly termed intrinsic noise, are sufficient to account for the observed scaling. The inter-individual variation of higher expression levels proteins (>10 copies per cell) is mainly extrinsic (correlated between random pairs of abundant proteins) and reaches a plateau, so that proteins in *E. coli* have at least a 30% variation (measured as variance divided by $mean^2$) in their expression level (Taniguchi et al., 2010). The authors also measured the variability in mRNA levels for a set of abundant proteins (>100 copies per cell) and they found that variation ranged from 0.05 to 5 copies per cell (Taniguchi et al., 2010).

Stochastic variability was initially studied as a snapshot, however important information was also hidden in the noise dynamics, as illustrated by noise propagation in transcriptional cascades (Hooshangi et al., 2005; Pedraza and van Oudenaarden, 2005; Rosenfeld et al., 2005). Dynamic noise studies of synthetic regulatory networks in *E. coli* revealed deviations of individual bacterial cells from

the mean behavior of the gene regulatory network inferred from population measurements. For instance, at a given transcription factor concentration different bacteria show different rates of a target gene transcription. The standard deviation of the production rates was ~55% of the mean. Interestingly, these deviations showed slow fluctuations, which could be as long as a cell cycle (Rosenfeld et al., 2005). Intrinsic noise was shown to decay quite rapidly (<10 min.) so that the observed slow fluctuations must be due to extrinsic factors. This work suggested the existence of a trade-off between accuracy and speed of responses in transcriptional cascades and also that, even in bacteria, extrinsic noise can play an important role, conferring a certain degree of cellular memory.

Is variability in gene expression of any use to bacteria? Both theoretical and experimental work has suggested an important role for positive feedback loops in the generation of gene expression bistability (Becskei et al., 2001; Ferrell and Machleder, 1998; Lisman, 1985; Xiong and Ferrell, 2003). Such “on” and “off” switches are a form of “cellular memory” necessary during developmental decisions. Phenotypic variability in such switches at the level of individual cells could also confer an advantage under fluctuating environmental conditions, especially in the case of unicellular organisms that reproduce mainly asexually (Acar et al., 2008). But what is the mechanism

behind these developmental switches? *B. subtilis* growing in adverse environmental conditions or entering stationary growth in the laboratory enter a “competence” state, which allows bacteria to uptake and incorporate foreign DNA from the environment, potentially increasing their fitness through the acquisition of novel traits. One key protein mediating competence in *B. subitlis* is ComK. The expression of *comK* is subject to a positive autoregulatory loop; ComK binds cooperatively to its own promoter. Also, *comK* expression is highly regulated at the mRNA and protein level. Once the stationary phase of growth is reached, cells randomly transition to competence during a period of just 2 hours. By the end of this period, approximately 15% of the cells have become competent and remained in that state (Maamar et al., 2007). Why do some cells differentiate into the competence state or others do not?

In the early stationary phase, variation among individual bacteria in *comK* mRNA levels was found to be largely intrinsic (average 1 molecule per cell), so that these differences could be responsible for the variability in the competence transition. In order to test this hypothesis, the authors engineered a *comK* low intrinsic noise strain. Rok is a transcriptional repressor of *comK*. Mutants for *rok*, show increased levels of *comK* transcription. Further mutating the initiation codon of c*omK* from ATG to GTG decreased *comK* translational efficiency. As a result, a strain carrying both

mutations had lower noise levels and slightly higher mean levels of ComK compared to the wild type. Interestingly, the number of competent cells was dramatically decreased from 15% to 1% in the low noise strain, thus suggesting that intrinsic fluctuations in *comK* expression were the responsible for the competence difference between cells (Maamar et al., 2007). Elowitz and colleagues reached a similar conclusion by different means. The authors used an inducible version of the *ftsW* gene, which is necessary for *B subtilis* septation. Upon inducer removal, the cells failed to septate resulting in long filaments of multiple cells units all sharing a common cytoplasm. Theoretically, a larger volume leads to a decrease in noise levels due to molecular averaging. The authors found that in this mutant background, the probability of competence initiation was decreased as cells elongated (decreased noise) (Suel et al., 2007). Thus, these studies suggest that microorganisms can tune noise in gene expression to genera potentially advantageous phenotypic variability in response to stress.

Another interesting point relates to the particular architecture or topology that characterizes some natural networks. Why is a circuit wired in a particular way and not in another? In order to address this issue, Suel and colleagues built a competence alternative circuit and implemented it into *B. subtilis*. In the native circuit, ComK represses its activator ComS, whereas in the alternative

circuit, ComK induces the expression of a repressor, MecA. Naively, a feedback characterized by the repression of an activator should be equivalent to the activation of a repressor. The alternative circuit indeed recapitulated the dynamics and frequency of the wild type competence in vivo, as suggested by simulations, however there was a key difference between the circuits. The alternative circuit showed a reduced level of variation in the duration of competence among individuals compared to the native one. The authors hypothesized that the inter-individual variation in the duration of competence could confer an advantage given the unpredictability in the concentration of DNA that bacteria could encounter in their natural environment. They tested this prediction by measuring the transformation efficiency of both circuits over a broad range of DNA concentrations. The native circuit was shown to confer a flatter response curve of normalized transformation efficiency compared to the alternative one. Thus, higher variation in duration of competence generated by the natural circuit guarantees an efficient response to a wide range of extracellular DNA concentrations. This suggest that native cells are better equipped to cope with variable environmental conditions and may explain the observed circuit choice (Cagatay et al., 2009).

2.2 Studies in Eukaryotes

2.2.1 Stochastic gene expression in *S. cerevisiae*

The first eukaryotic study on gene expression using the dual reporter technique was performed in the budding yeast *Saccharomyces cerevisiae* (Raser and O'Shea, 2004). In this case, the authors used a cyan (CFP) and yellow (YFP) fluorescent proteins as tag for each allele in the diploid cell. Looking at two yeast promoters of the *PHO5* and *GAL1* genes, they concluded that variability in yeast had a major extrinsic component. Deletion of UAS1 and UAS2 upstream activating sequences of the *PHO5* promoter, involved in the recruitment of chromatin remodeling complexes caused an increase in the intrinsic noise, as did deletion of SWI/SNF, INO80 and SAGA complexes members. This was the first indication that chromatin dynamics can affect the dynamics of stochastic gene expression *in vivo* (Raser and O'Shea, 2004). A previous study using a single reporter approach, demonstrated that, similar to the case in bacteria, noise can be propagated in a gene regulatory cascade in yeast (Blake et al., 2003). These studies were limited, however, to a few genes. A more global view of stochastic gene expression for budding yeast was accomplished by two studies (Bar-Even et al., 2006; Newman et al., 2006). Barkai and colleagues studied variation in the expression levels of 43

proteins (GFP tagged) in cells grown under 11 experimental conditions. Weissman and co-workers measured the variation of >2,500 proteins under 2 experimental conditions. Both studies reached the same basic conclusions. First, they observed a clear inverse correlation between protein noise (CV^2) and protein mean expression. What is the origin of this trend? Proteins are expressed is bursts both because each mRNA is typically translated many times (transcriptional bursting) and also because promoters can switch stochastically between “off” and “on” states (translation bursting). In prokaryotes, transcriptional initiation occurs frequently in comparison with synthesis and degradation events, so that translational bursting is more important (Ozbudak et al., 2002). By contrast, slow transition between open and closed chromatin states can be an important factor in eukaryotes (Kaern et al., 2005). The observed correlation between noise and protein abundance is characterized by a proportionality factor of ~1,200, which is very similar to the average number of proteins produced by mRNA. Therefore, translational bursting seems to be the mechanism behind this trend (Bar-Even et al., 2006; Newman et al., 2006). However, an alternative model based on the random activation and inactivation of promoters (slow promoter kinetics) can also explain the observed trend and cannot be entirely ruled out. Second, though a general trend is evident, there are deviations from this behavior for particular gene classes.

Stress related proteins (such as chaperones) are among the most variable independently of their mean expression level. Also, one of the most highly associated traits with high expression variability is the presence of a TATA box in the promoter region of genes. Indeed, TATA boxes are enriched in the promoter region of stress response genes. In addition, being regulated by a chromatin remodeling/histone modification complex also increases the chances of being highly variable, supporting previous findings. Essential genes are less variable than non-essential genes and there is a genome wide negative correlation between nucleosome occupancy and clusters of essential genes (Batada and Hurst, 2007). This suggests that genome wide evolution of chromosome organization could be driven in part by noise selection (Batada and Hurst, 2007). Third, for intermediate expression level genes (those that show the lineal dependence between noise and mean expression and thus the large majority of proteins), intrinsic noise is an important contributor to the total variability together with extrinsic noise. However, for highly expressed proteins, noise is mainly extrinsic.

Stochastic gene expression has been shown to be a complex genetic trait such as many other phenotypes. In particular, Yvert and colleagues mapped Quantitative Trait Loci (QTL) in *S. cerevisiae* for a *MET17* GFP transcriptional reporter. They could identify modifiers of gene expression

variability (both dependant and independent of the mean expression). This work also suggest, that stochastic variability in gene expression depends on the particular genetic background (Ansel et al., 2008). Similar conclusions have been reached mapping noise QTLs in the model plant *Arabidopsis thaliana* (Jimenez-Gomez et al., 2011).

Stochastic gene expression also influences the way cells respond to environmental information and how information is efficiently transmitted. One case study is the mating response of *S. cerevisiae*. Yeast has two haploid cell types: MATα and MATa and both secret peptide pheromones. The α-factor is secreted by the MATα type and stimulates a mating response in the MATa cells. This leads to the activation of a G protein and mitogen-activated protein (MAP) kinase-signaling pathway, which in turn activates the Ste12 transcription factor (Wang and Dohlman, 2004). Ste12 regulates the expression of many target genes, including PRM1. A PRM1 fluorescent transcriptional reporter (used as an output of this pathway) revealed that individual yeast cells vary as much as 35% in their transcriptional response to the α factor (Colman-Lerner et al., 2005). The authors were interested in understanding the origin of this variation. The dual reporter strategy revealed that intrinsic noise made a very small contribution, almost negligible. Furthermore, the differences in response among individuals were stable in time (lasting hours),

similar to the slow fluctuations previously described in bacteria (Rosenfeld et al., 2005). In addition, the authors measured the gene expression variability in a strain with YFP and CFP reporters driven by the α-factor-responsive PRIM1 promoter and by the α-factor-independent ACT1 promoter respectively, in this way they could decomposed the ruling extrinsic noise into pathway specific (PRIM1) and global noise (ACT1), common to all genes. At high pheromone concentration, variation was due to global regulators, probably an indication of the saturation of the system, whereas at low level of pheromone, the pathway noise became more important. Also, mutants in members of MAP kinase signaling pathway such as Fus3 and Kss1 (that phosphorylate Ste12) showed altered levels of variation in the mating response among individuals. This suggests that the autoregulatory feedback in which these proteins are involved may play a role in the modulation of the variation (Colman-Lerner et al., 2005).

2.2.2 Stochastic gene expression in metazoans

In addition to experiments performed in *S. cerevisiae*, stochastic gene expression has also been studied in metazoans. The groups of Alon and Tyagi performed the first studies of noise in mammalian cells in culture (Raj et al., 2006; Sigal et al., 2006). Alon and co-workers measured the variability of 20 endogenously YFP tagged human

proteins by time-lapse microscopy. They found that the coefficient of variation ranged from 0.12 to 0.28 depending on the protein. Interestingly, they found that the correlation in the variation of two proteins that belong to the same complex (such as the ribosome) was higher than the variation of two proteins that belong to different pathways (the correlation between different ribosomal subunits ranged from ~0.3-0.5, whereas the correlation among non related proteins ranged from ~0.05-0.1). This indicates that part of the noise in mammalian cells could be due to variation in upstream regulatory components, which are specific for each pathway or complex. However, other studies have not detected such a trend when studying variation at the mRNA level in *S. cerevisiae* (Gandhi et al., 2011). The differences could be explained by the role of post-transcriptional regulatory mechanisms. Another explanation is the disparate time scale of mRNA and proteins. In bugging yeast, the median half-life of mRNAs is 18 min (Munchel et al., 2011) and the one of proteins is 43 min (Belle et al., 2006). In humans, the median half-lives of mRNAs and proteins are 9 h and 46 h respectively, as measured in the NIH3T3 cell line (Schwanhausser et al., 2011). Also, the range of protein half-life is much wider than the one of mRNA in both organisms. Protein molecules degrade slower than mRNA molecules, thus filtering out fast fluctuations in their expression. As a consequence, dynamics of two mRNA species is dominated by

uncorrelated fast fluctuations, whereas the one of two proteins by their better correlated slow fluctuations. In this way, two mRNA species can be uncorrelated with each other, but produce protein in a coordinated manner (Munsky et al., 2012). Also, interestingly, the time that it took for a mammalian “high expressor” cell to turn into a “low expressor” cell for a given protein was longer than two cell generations. This memory could have important implications for cell differentiation and other processes. For instance, several cycles might be needed for some signals to affect populations of human cancer cells (Sigal et al., 2006).

Another example of the importance of stochastic gene expression in mammalian cells relates to cell death. Sorger and colleagues noticed that when treating a clonal culture of HeLa cells with TRAIL (tumour necrosis factor (TNF)-related apoptosis-inducing ligand), the apoptotic response of individual cells was highly variable. Some cells died within 45 min, others 8–12 h later, and yet others live indefinitely (Spencer et al., 2009). Time-lapse miscroscopy experiments revealed that recently divided sister cells behaved more similarly than a random pair of recently divided cells with respect to the time of cell death. This indicated that the variability in the time of death was due to pre-existing differences between cells and that these differences where transiently inheritable, in agreement with

previous studies (Sigal et al., 2006). These experiments were performed using an immortalized cell line and *in vitro*, future work should evaluate the role of stochastic gene expression in apoptosis in a more physiological context.

In addition to studies using fluorescent reporters, an alternative strategy makes use of the *in situ* hybridization technique. This method traditionally uses a labeled complementary DNA or RNA strand (i.e., probe) to localize a specific RNA sequence in a tissue. Recent advances in this technique allow the detection of single mRNA molecules. Tyagi and colleagues used a single molecule resolution Fluorescence In Situ Hybridization (FISH) technique to study gene expression noise in mammalian cells. They integrated an inducible t*etO* promoter and quantified the number of molecules under various conditions. They found great variation in the number of transcripts among cells. The mRNA distribution had long exponential tails, corresponding to a model in which mRNAs are produced in short and infrequent bursts and suggesting that the major contributor to this variation was intrinsic noise (Raj et al., 2006). One major drawback of this study is that they focused on a synthetic gene reporter (Tetracycline-Controlled Transcriptional Activation). This could explain why they didn't identify extrinsic sources of variation in contrast to previous studies (Sigal et al., 2006).

The importance of stochastic fluctuations in gene expression during normal development can be illustrated by the work of Huang and co-workers (Chang et al., 2008). The authors used the mouse hematopoietic differentiation system as a model to understand the role of stochastic gene expression in cell fate decisions. The expression of Sca-1, a stem cell-surface marker was found to be very variable (1,000-fold range) among individual EML cells derived from a clonal population. Moreover, “high” or “low” Sca-1 expressing cells showed differences in their differentiation rates into the erythroid and myeloid lineages, thus revealing that clonal heterogeneity governs differentiation potential. Interestingly, cells with “high” or “low” Sca-1 levels also showed transcriptome-wide differences in gene expression and the mixing (the time it takes for one population to convert into the other) of “high” and “low” populations was a slow process that required more than 12 cell doublings. This probably reflects the existence of slow transitions between metastable transcriptomes driven by stochastic dynamics. Thus, cells seem to have evolved to exploit spontaneous and transient gene expression differences due to stochastic processes to program cell-fate decisions. This concept was previously shown to apply for the development of the compound eye of *Drosophila melanogaster*. The authors showed that stochastic expression of a single transcription factor,

spineless, was sufficient to establish the mosaic pattern of photoreceptor's expression (Wernet et al., 2006).

There is also increasing evidence suggesting that early mammalian development relies on stochastic mechanisms. For instance, pluripotent embryonic stem (ES) cells derived from the inner cell mass (ICM) of mice blastocysts and cultivated *in vitro* show great variability in the expression of markers of pluripotency such as Nanog, Stella, Oct4, Cdx2, etc. Several studies have shown that differences in the expression of these makers, which show slow fluctuations, biases the cells towards particular fates when induced to differentiate (Balazsi et al., 2011; Canham et al., 2010; Chambers et al., 2007; Hayashi et al., 2008; Kalmar et al., 2009; Thomson et al., 2011). Also, variability in the expression of stem cell markers among blastomeres has been observed in the developing intact early embryo and it has been proposed that this variation may be necessary for normal development (Dietrich and Hiiragi, 2007). The first cell fate decision in early mouse development is whether to become the pluripotent inner cell mass (ICM) or an extra-embryonic tissue, the trophectoderm (TE). There are two key transcription factors involved in this process: CDX2, which promotes a TE fate and OCT4, which promotes an ICM fate. CDX2 expression is already heterogeneous among blastomeres at the eight-cell stage (cells that are thought to be equivalent) (Dietrich

and Hiiragi, 2007). Interestingly, experimental manipulation of the levels of Cdx2 in the early embryo showed that cells in which *Cdx2* levels are elevated prior to the generation of inside cells undertake more symmetric divisions and, consequently, contribute a greater proportion of their progeny to the trophectoderm than to the ICM. Conversely, the proportion of cells contributing to the trophectoderm is reduced following down-regulation of *Cdx2* (Jedrusik et al., 2008; Zernicka-Goetz et al., 2009). Also, recently, the variable kinetics of expression of an Oct4 reporter *in vivo* have been shown to predict cell lineage patterning in the early embryo (Plachta et al., 2011). Thus, the existence of differentiation biases seems to be also occurring in the early mammalian embryo, but it's still not clear how and when exactly these differences arise nor how they are reinforced by positional signaling between the cells in order to guarantee a robust development (Zernicka-Goetz and Huang, 2010).

3. The genotype-phenotype map

The Danish researcher Wilhelm Johannsen was the first person to clearly state that the environment, in addition to genes, influences the physical characteristic of organisms (Johannsen, 1909). Johannsen noticed that inbreed lines of beans (genetically identical) still showed

variability in traits such as seed size that followed a normal distribution. He coined the term "phenotype" to differentiate the actual expressed physical characteristics from the inherited characteristic (genotype). Thus the classical paradigm states that it is the joint contribution of genes and the environment that organisms experience what will determine their phenotype. But are these two variables enough to account for the total phenotypic variability among individuals?

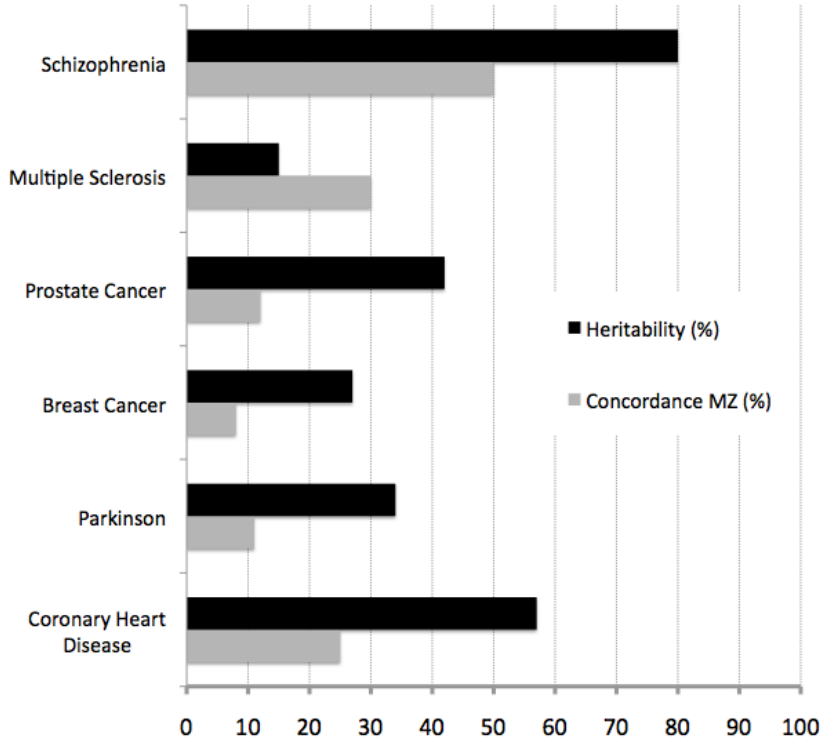
3.1 The contribution of the genotype

A lot of effort has been directed towards dissecting the contribution of natural genetic variation to phenotypic variation and establishing causality in model organisms (Rockman, 2008). Most phenotypic traits, including human diseases have a significant genetic component in all organisms studied to date (Visscher et al., 2008). Evidence for this is the success of artificial selection experiments performed in laboratories and by breeders (Visscher et al., 2008) and the estimated heritability values for most human traits and diseases (Smith, 2011). Heritability is a quantitative genetics measure, which quantifies what percentage of the total phenotypic variance is due to the underlying genetic variation present in a particular population. Indeed, for some human diseases heritability is

high. For instance, schizophrenia has a heritability of 0.8. However, we must bear in mind that heritability is a property of the population and that it doesn't inform us about the relative probability that an individual has a particular trait given a particular genome. In contrast to schizophrenia, most diseases, including different types of cancer have much lower heritabilities (Figure 2).

Figure 2

Heritability and concordance rate of monozygotic twins for several human diseases. Data from (Cardno and Gottesman, 2000; Kuusisto et al., 2008; Lichtenstein et al., 2000; Soffritti et al., 2006; Wirdefeldt et al., 2011)



Despite the fact that most traits have a clear genetic component, it has proven difficult in practice to identify the genetic variants underlying this heritability (Manolio et al., 2009). With the exception of some diseases caused just by single gene mutations such as phenylketonuria or Huntington disease, most diseases and importantly those more prevalent worldwide can be defined as “complex”. These complex diseases, such as schizophrenia, Alzheimer’s disease, multiple sclerosis, autism, and diabetes are thought to arise from the combination of several genetic variants, each of them giving a small contribution to the phenotype and they can also be modified by the environment. Genome-wide association studies (GWAS), in which several hundred thousand to more than a million single nucleotide polymorphisms (SNPs) are assayed in thousands of individuals, represent a powerful new tool for investigating the genetic architecture of complex diseases (Manolio et al., 2009). Initial GWAS studies made substantial findings by studying only a few hundred individuals (Maller et al., 2006). However, more often the small effect-size of associated SNPs requires the genotyping of thousands of individuals when studying complex diseases across a population. Myocardial infarction is a leading cause of death and disability worldwide, and its early form is characterized by a high heritability (56%) (Nora et al., 1980). A GWAS studying early onset of myocardial infarction in 2,967 cases and

3,075 controls found 9 loci, which explained just 2.8% of the heritability (Kathiresan et al., 2009). A meta-analysis of more than 100,000 individuals identified 95 (59 new) associated loci with cholesterol and blood lipid levels. These loci could explain 12% of the Low-density lipoprotein (LDL) level's total phenotypic variance and ~30% of the heritability (Teslovich et al., 2010). Other GWAS have been more successful. For example, in the case of schizophrenia, recent studies estimated that ~26-24% of the heritability could be explained by common variants (Lee et al., 2012; Purcell et al., 2009). Lack of statistical power due to small sample size makes detection of these variants difficult by GWAS, but there is evidence that this limitation may be overcome (Ehrenreich et al., 2010). Also, it has been suggested that the main contributors to human disease are rare and severe effect alleles in populations (McClellan and King, 2010).

One reason why it may be difficult to estimate the contribution of genetic variability to the phenotype determination is the fact that genes do not work in isolation in cells but the numerous interactions between them define a complex genetic network (Costanzo et al., 2010; Lehner et al., 2006). Synthetic lethality screens have been used in diverse organisms to map these interactions (Costanzo et al., 2010; Lehner et al., 2006). These studies have revealed that most genes have few interactions, but a small number

of genes (termed “hubs”) have many. As a matter of fact, heritability estimates in the literature refer most of the time to narrow-sense heritability, which take into consideration just additive effects of genes and neglects any epistatic or dominance interactions (Manolio et al., 2009). In addition, genetic variability can also affect these interactions. For instance, it has been shown that the set of yeast essential genes is dependant on the particular genetic background (Dowell et al., 2010). Also, these interactions are not static but rather dynamic. Ideker and colleagues studied how genetic interactions change in response to an environmental stress in *S. cerevisiae* (Bandyopadhyay et al., 2010). They selected a subset of 418 yeast genes and obtained genetic networks for cells growing with and without a DNA-damaging agent. Comparison of the genetic networks across conditions revealed large differences, with more interactions unique to each map than in common. For instance, more than 70% of positive interactions identified under DNA damage stress were not identified in the untreated sample, which reflects widespread DNA damage–induced epistasis. Knowledge gained by studying gene interaction networks is giving us clues on how complexes diseases originate (Lehner, 2007). But the fact that genetic interactions are complex and disease may arise from tens or hundreds of small contributing variants does not imply that these will explain all the phenotypic variability among individuals.

In summary, most traits have a clear genetic component. Heritability, the proportion of phenotypic variation that can be explained by genetic variation, is a good estimate of the contribution of genes. Nevertheless, for the large majority of traits and diseases, most of the contributing alleles found in GWAS often explain just a small proportion of this heritability. Human height is a multigenic complex trait characterized by a high heritability ~0.8. However, a set of the 50 most significantly associated variants identified by GWAS could explain just 5% of the total phenotypic variation. A recent study suggested that as much as 45% of variance could be explained taking into consideration the contribution of 294,831 single nucleotide polymorphisms (SNPs) (Yang et al., 2010). But even if these estimations are correct, the total phenotypic variance explained for one of the most highly studied traits and with an important genetic component would be 0.45 times 0.8, just 36% of the total phenotypic variation. There have also been recent claims that heritability inferred from populations may be overestimated, so that epistastic interaction between already identified alleles could explain most of the heritability for some traits (Carlborg and Haley, 2004; Zuk et al., 2012). Yet, heritability is a property of populations and that we can explain most of it does not imply that a good predictive power can be attained at the level of the

individual, which is one of the aims of personalized medicine (Ng et al., 2009).

3.2 The contribution of the environment

If genes and environment are the only or major contributors to phenotypic variation, then for a particular trait whose heritability is 0.3, it must follow that the environment is responsible for 0.7 of the variance. Is that really the case? The environment in humans can refer to our diet, exposure to chemicals or pathogens, life habits such as frequency of exercise and exposure to stress. No single pair of individuals has experienced the same environment, so it's complicated to directly estimate the environmental contribution. However several studies performed with twins growing up in the same or different families have tried to estimate the contribution of "shared-environment". Shared-environment makes reference to the environmental factors that two siblings growing in the same house have in common. The surprising result of these studies is that the proportion of the variance explained for most traits and diseases is quite small, around 0.1. In other words, genetics apart, two siblings growing in the same house are not much more similar than any two random kids taken from the same population (Plomin and Daniels, 2011; Smith, 2011). If we add the heritability and the shared-environment contribution of most cancers, then that

explains just 40% of the total phenotypic variance, leaving the large majority of the variance unexplained (Lichtenstein et al., 2000). Of course, this could indicate that we are missing key environmental factors or triggers for most diseases, which is certainly a possibility. However, this may also indicate that relatively small or even insignificant events that happen through our life could have a major impact in our phenotype. This will indeed be bad news for epidemiology (Smith, 2011). A third, non-mutually exclusive alternative is that stochastic variation not in the environment we experience, but happening inside our own cells, in processes such gene expression could play an important role in phenotypic determination.

3.3 Unexplained phenotypic variability

Predicting an individual's phenotype when genetic variation is present is very challenging. Populations in the wild are characterized by a large number of genetic variants. Moreover, the number of genetics differences is almost insignificant compared to the number of possible genetic interactions among them. But how easy would it be to predict a phenotype in the absence of genetic variability? Monozygotic twins (who originate from a single zygote) are assumed to be essentially isogenic, so that any phenotypic difference that exists between them is usually attributed to the environment. Twins have offered researchers the

possibility to investigate the contribution of genes and environment to disease phenotypes. These studies have revealed for instance that if a woman suffers breast cancer, the chance that her monozygotic twin sister also suffers from this condition is just ~0.2 (Hamilton and Mack, 2003). If a man suffers from schizophrenia, the chance of his twin brother suffering the same disease is ~0.5 (Cardno and Gottesman, 2000), i.e. the genetic equivalent of tossing a coin. Remarkably, those numbers represent the best we could aim for when predicting the disease phenotype of an individual and not controlling for the environment (normal scenario for humans). Phenotypic discordance among twins is the rule rather than the exception (Wong et al., 2005). Examples are provided in Figure 2.

Genes and environment apart, what other factors could possibly cause this phenotypic discordance? It has been shown that even monozygotic twins show unique somatic mutations (Bruder et al., 2008) and differences in epigenetic marks appear during their lifetime (Fraga et al., 2005). Several studies have evaluated the role of these somatic single-nucleotide polymorphisms (SNPs,) copy number variants (CNVs) and epigenetic marks in the genome, but have failed so far to explain the dissimilar phenotypic outcome of mutations in monozygotic twins (Baranzini et al., 2010; Kimani et al., 2009; Lasa et al., 2010). However, somatic mutations are known to be

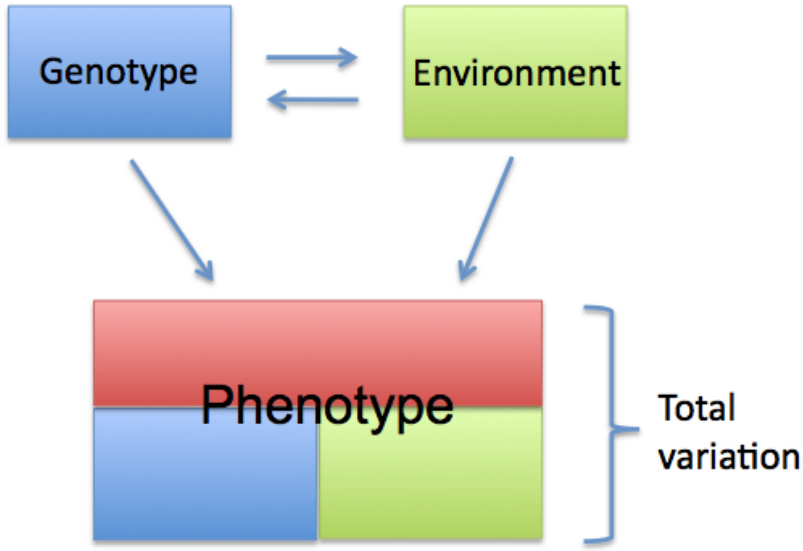
important in diseases such as cancer. The “two-hit hypothesis” of recessive tumor suppressor genes is the current paradigm of cancer initiation (Knudson, 1971). The development of restriction fragment length polymorphism (RFLP) mapping technology made possible the confirmation of this hypothesis in the 80s (Berger et al., 2011). It was shown that many families with familial retinoblastoma (an hereditary cancer that develops in the cells of retina) harbored mutations or deletions of the *RB1* gene in the germ line (the first hit). In addition, the tumors from retinoblastoma patients nearly always contained mutation or loss of the other *RB1* allele (the second hit) (Cavenee et al., 1985; Friend et al., 1986; Lee et al., 1987). This latter event most often results from somatic mitotic recombination, whereas in a minority of cases, the normal allele acquires a distinct mutation. Therefore, we can expect that at least some cases of phenotypic discordance between twins are due to somatic mutations.

Another factor that could play a role is obviously the environment, which varies between twins. It is quite likely that some unknown environmental variable, such as pathogens, is influencing disease phenotypes to some degree. But, to determine the sole contribution of the environment is difficult in humans, because there is no environmental equivalent of monozygotic twins. However, studies performed in model organisms where the

contribution of genes and the environment can be carefully controlled have shown that even when these two factors are kept constant, phenotypic differences among individuals are still evident (Figure 3) (Gartner, 1990; Horvitz and Sulston, 1980; Pereira et al., 1994).

Figure 3

The colored areas represent the fraction of the total phenotypic variance explained by genetic (blue) and environmental (green) variance. The red are represents the unexplained phenotypic variance.



4. The interplay between mutations and stochastic gene expression

4.1 Phenotypic variability in model organisms

Early observations performed more than 30 years ago already suggested a role for stochastic processes in phenotypic variability. In a pioneering study, Spudich and Koshland showed that behavioral differences between isogenic bacteria persisted over generations, demonstrating a non-genetic source of individuality in microorganisms (Spudich and Koshland, 1976). They also hypothesized that the differences were caused by fluctuations in the level of key molecules due to its low copy number in the bacterial cell (Spudich and Koshland, 1976). But this phenotypic variation is not evident just in microorganisms. For instance, the life span of *Caenorhabditis elegans* is also considerably variable among isogenic individuals growing under the same environment. This variability in time of death is similar to the one of human populations (Rea et al., 2005). Isogenic lines of rats growing in laboratory controlled conditions also show a large range of variability in their life spans (ranging from 60-140 weeks) (Soffritti et al., 2006). In addition, electron-microscopic studies have revealed stochastic variation in cellular demise within the same cell type of individual *C. elegans* worms (Herndon et al., 2002).

Johnson and colleagues noticed that transgenic worms carrying a fluorescent (GFP) transcriptional reporter for the heat shock response gene *hsp-16.2* showed high inter-individual variation in the induction of the transgene after a heat stress. The level of expression of this reporter in adults partially predicted the remaining life span of individual worms. Those worms with higher induction of the *hsp-16.2* reporter typically live longer than their lower induction counterparts. Overexpression of *hsp-16.2* resulted in a very modest lifespan extension, so the authors hypothesized that the reporter conveys information about a general and variable physiological state of the worm. However since the heat shock treatment alone is sufficient to increase the mean life span of the population, it is not clear if differences in *hsp-16.2* reporter expression levels are mainly due to stochasticity in the heat shock response or whether these differences are somehow already present in the unstressed population.

Two recent studies investigated the predictability of individuals' life spans in the absence of a heat stress. Surprisingly, part of the variation could be due variability in pathogenicity among individuals (Sanchez-Blanco and Kim, 2011). The *E. coli* OP50 strain usually used to feed worms is slightly pathogenic for *C. elegans*. A transcriptional reporter for the gene *sod-3* (superoxide dismutase-3) measured during mid-life partially predicted the remaining

life span, and this correlation was dependant on *daf-16* activity. Comparison of the predictive power and expression levels of *sod-3* between growth in *E. coli* and in the non-pathogenic *B. subtilis* bacteria suggested that variable pathogenicity from individual to individual may cause variation in abundance of *sod-3* and lifespan variability (*sod-3* predictive power was lost in worms fed with *B. subtilis*). However, whether the differences in pathogenicity are intrinsic to worms or extrinsic (coming from the bacteria itself) remains as an open question (Sanchez-Blanco and Kim, 2011).

The second study focused on the variable expression of microRNAs previously identified as regulators of the insulin pathway in *C. elegans*. *mir-71* and *mir-246* mutants were long lived, whereas *mir-239* mutants were short lived compared to the wild type N2 strain (de Lencastre et al., 2010). A very elegant experimental set up (single-animal culture apparatus) allowed the authors to measure the fluorescence of transcriptional reporters and individual size continuously from embryonic development until the time of death. These measurements revealed that inter-individual variability in *mir-71* and *mir-246* expression was positively correlated with life span, and *mir-239* expression was negatively correlated (Pincus et al., 2011). However, one drawback of this study is that the imaging settings required genetically or chemically germline-ablated

worms, and it is known that the germline has an important influence on lifespan. When the germline of *C. elegans* is removed either by laser microsurgery or by mutations that block germ cell proliferation, animals live up to 60% longer than control animals (Hsin and Kenyon, 1999). In animals lacking germ cells, DAF-16 accumulates specifically in the intestinal nuclei and activates the transcription of stress-related and metabolic genes (Lin et al., 2001). It was recently described that *mir-71* was required for the life-span expansion caused by germ line removal. Experiments suggest that *mir-71* functions in the nervous system to facilitate the localization and transcriptional activity of DAF-16 in the intestine (Boulias and Horvitz, 2012).

Another example of a stochastic variable phenotype in wild type strain takes place in the vulva development. Though, the embryonic and post-embryonic somatic development of *C. elegans* is virtually invariable, there are a couple of exceptions to this rule. The vulva is specified during the third larval stage of postembryonic development from a row of six vulva precursor cells (VPCs), named P3.p to P8.p. The P3.p cell may either divide in a similar way to P4.p and P8.p, or fuse to the syncytial epidermis hyp7 at the L2 stage with a 0.5 probability (Felix and Barkoulas, 2012). The stochastic lineage of the anchor cell (AC) is another example. The development pattern varies randomly in different animals between two distinct lineages. In half

the animals, by chance, the AC is derived from the Z1 gonadal precursor cell; in the other half, it is derived from the Z4 gonadal precursor. Ablation of either cells results in the other becoming the AC. The final regulatory outcome, however, is not affected because a regulatory feedback via intercellular signaling involving the receptor LIN-12 and its ligand LAG-2 causes the alternative cell to become a ventral uterus cell (Wilkinson et al., 1994). Thus, in this case, random developmental lineage variation is dynamically compensated by cellular level feedback (McAdams and Arkin, 1999). Interestingly, different species of related nematodes have evolved different preferences. In some species, the decision is stochastic but biased towards Z4 being the AC (*Acrobeloides)* and in others just the Z1 cell adopts the fate (*Panalograimus*) (Sternberg and Felix, 1997).

Is this unexplained variability in traits between individuals just a particularity of lower metazoans and microorganisms? Studies performed in mice, suggest that that is not the case. In a seminal study by Gärtner and colleagues, it was shown how mice strains subject to inbreeding for many years and growing in controlled laboratory conditions still show phenotypic variability (Gartner, 1990). For instance, individuals' body weight was found to follow a normal distribution. Remarkably, the authors estimated that just 20-30% of the body weight

variability could be attributed to differences in the environment, thus leaving 70-80% of the phenotypic variability unexplained.

4.2 Variability in the outcome of mutations

Unexplained phenotypic variability has also been reported for disease models in mice. Osteogenesis imperfecta (OI) is a human genetic disorder, which affects the connective tissues of individuals. OI is incompletely penetrant and variable in human families and has been described to be caused in most cases by of a mutation in the gene coding for Type-I collagen (Sippola et al., 1984). An inbreed mouse model for this disease carrying a dominant negative form of COL1A1 was generated and showed incomplete penetrance and variability of the disease phenotype. ~22% of individuals were severely affected (fractured long bones), ~51% showed mild defects and ~27% had no fractures at all (Pereira et al., 1994). The mean level of expression of the gene was not correlated with the penetrance and no environmental factor was found that could explain the differences. More recently, a knock out of the mouse *mir-290–295* cluster was shown to have incompletely penetrant phenotypes affecting embryonic survival and germ-line development (Medeiros et al., 2011). These are just a few examples, but many mutants show some degree of variability between individuals. Altogether,

these studies suggest that genes and environment cannot explain all the phenotypic variability between individuals, also that this phenomena is widely conserved in evolution and there is no reason to think that this may not be also the case for humans.

The penetrance of a mutation describes the percentage of individuals in a population that manifest phenotypically the effect of a certain genetic variant. For example, a penetrance of 30% means that 70 out of 100 individuals carrying a particular mutation do not show any phenotypic effect. For instance, not all women carrying a mutation in the BRCA1 gene actually develop breast cancer, and the penetrance has been estimated to be 68% by the age of 80 (Chen and Parmigiani, 2007). This reported number is variable in different studies, ranging from 30-90% (Offit, 2006). It has been suggested that the penetrance of BRCA1 and BRCA2 mutations is usually overestimated, since there is a clinical bias towards genotyping or studying families with an increased risk due to familial history of breast cancer (Offit, 2006). However, independent of the penetrance value, it is clear, that some women do not develop breast cancer despite carrying one of the most paradigmatic cancer causing mutations. Why is this the case? There are three obvious explanations. First, although two unrelated individuals may share the same mutation in a particular gene, there is natural genetic

variation in the rest of their genomes. Genetic interactions between background variants and the mutation could be responsible for the dissimilar phenotypic outcome. Second, the environment experienced by different individuals, such as exposure to chemicals, diet, infections, etc. could cause differences in disease susceptibility. Third, somatic mutations occur during the lifetime of person, and they can also have and important impact. However, as mentioned in previous sections, even model organisms, where genetic and environmental variation is minimized or practically non-existent can show extensive phenotypic variability. So, it seems reasonable that to stochastic processes, such as gene expression can explain part of the incomplete penetrance cases.

The first study that established a link between the incomplete penetrance of mutations and stochastic variation in gene expression was performed in the prokaryote *Bacillus subtilis* (Eldar et al., 2009). *B. subtilis* cells growing under nutrient-limited conditions develop into dormant spores. These spores are highly resistant to stress conditions that would normally kill normal cells, such as high temperatures, ionizing radiation, detergents, etc. and they can remain inert for large number of years. Even though development is usually conceptualized in a multi-cellular context, the process of spore formation can be seen as a differentiation process involving a cell-fate decision.

The regulatory network underlying this process has been studied genetically and molecularly in detail (Errington, 2003; Losick et al., 1986). The endospore is formed by a mechanism involving asymmetric cell division, followed by engulfment of the smaller cell (forespore) by its larger sibling (mother cell).

Normal vegetative cell division in *B. subtilis* (and most bacteria) depends on the assembly of the Z-ring in the mid-part of the cell. This ring is made largely of polymerized FtsZ protein, a bacterial homologue of tubulin. Sporulation also uses the Z-ring but expression of *SpoIIE* causes a repositioning of FtsZ into two polar rings, through a helical intermediate. For unknown reasons, one ring develops faster than the other, establishing the asymmetric septum (Errington, 2003). Septation leads to activation of the transcriptional activator σ^F in the forespore. σ^F activates the expression of *SpoIIR*, which in turn activates σ^E in the mother cell. The transcriptional program of the mother cell abolishes the formation of the second polar septum. Another key process in sporulation regards chromosome segregation. A mechanism that specifically anchors the origin of replication (*oriC*) region of the genome to the bacterial poles ensures the rapid segregation of the chromosomes. Interestingly, both the σ^F -› *SpoIIR* -› σ^F transcriptional cascade and forespore septation start before the future spore chromosome has fully segregated into the

forespore compartment. The *SpoIIR* locus is located close to the *oriC* region, which guarantees its early transcription and proper function. Placing the *SpoIIR* locus opposite the oriC in the genome its sufficient to cause developmental problems, because even if σ^F is present in the forespore comportment, the *SpoIIR* locus is lags behind and is not available to be transcribed (Khvorova et al., 2000). These "positional" mutants are incompletely penetrant. Some individuals of the same isogenic population manage to develop the forespore and mother cell correctly. However, other cells suffer problems and develop into abortive dispores. Intriguingly, other cells follow an 'escape" pathway which leads them to develop into a polyploidy mother or twins cell, producing two mature viable spores. Single cell dynamic measurements of *SpoIIR* mutant fluorescent reporters showed that *SpoIIR* expression was reduced and delayed 10 min in the mutants compared to the wild type. Also, there was ~5 min. inter-individual variation in the timing and ~56% in rate of *SpoIIR* expression (Eldar et al., 2009). Stochastic variation in the expression rate of *SpoIIR* explained approximately 15% of the decision between sporulation and the other alternative fates. Thus, for this apparently simple case, most of the phenotypic variation remains still unexplained and variation in one or more unidentified factors could play an important role.

4.3. Incomplete penetrance of mutations in *C. elegans*

C. elegans is a small free-living nematode that survives feeding on microbes. In the laboratory, it is usually grown on agar plates covered with a lawn of *E. coli*. Among its many advantages for study are its short life cycle, compact genome, stereotypical development, ease of propagation and small size. The adult body plan is anatomically simple with 959 somatic nuclei and over 1000 germline nuclei in the adult. *C. elegans* has two sexes: hermaphrodite and male. Hermaphrodites have 2 copies of the X chromosome (XX), whereas males have just one (XO). Hermaphrodites produce both oocytes and sperm and reproduce by selfing, thus given rise to large isogenic populations. In benign conditions, males arise infrequently (0.1%) by spontaneous non-disjunction in the hermaphrodite germ line and at higher frequency (up to 50%) through mating.

C. elegans is amenable to genetic crosses and produces a large number of progeny per adult (around 300 in 4-5 days). It reproduces with a life cycle of about 3 days under optimal conditions and their average life span is 3 weeks at 20°C. Embryogenesis in *C. elegans* is roughly divided into two stages: (1) proliferation and (2) organogenesis/ morphogenesis. Proliferation (0 to 330-350 min. post-fertilization at 22°C) includes cell divisions from a

single cell to about 550 largely undifferentiated cells. This stage is traditionally further subdivided into two phases: The first phase (0-150 min) spans the time between zygote formation to generation of embryonic founder cells, and the second phase (150-350 min) covers the bulk of cell divisions and gastrulation until the beginning of organogenesis. The initial 150 minutes of proliferation takes place within the mother's uterus, and the embryo is laid outside when it reaches the approximate 30-cell stage (at gastrulation) (Altun and Hall, 2008; Sulston et al., 1983). The comma stage occurs 430 minutes after fertilization. The embryo starts pharyngeal pumping at 800 min after the first cell cleavage and hatches at 840 min (Figure 4). Of the 671 nuclei generated in the embryo, 113 undergo programmed death in the course of embryogenesis.

Figure 4

Embryonic stages of development. Numbers below the horizontal axis show approximate time in minutes after fertilization at 20-22°C. Image credit Wormatlas (www.wormatlas.org)

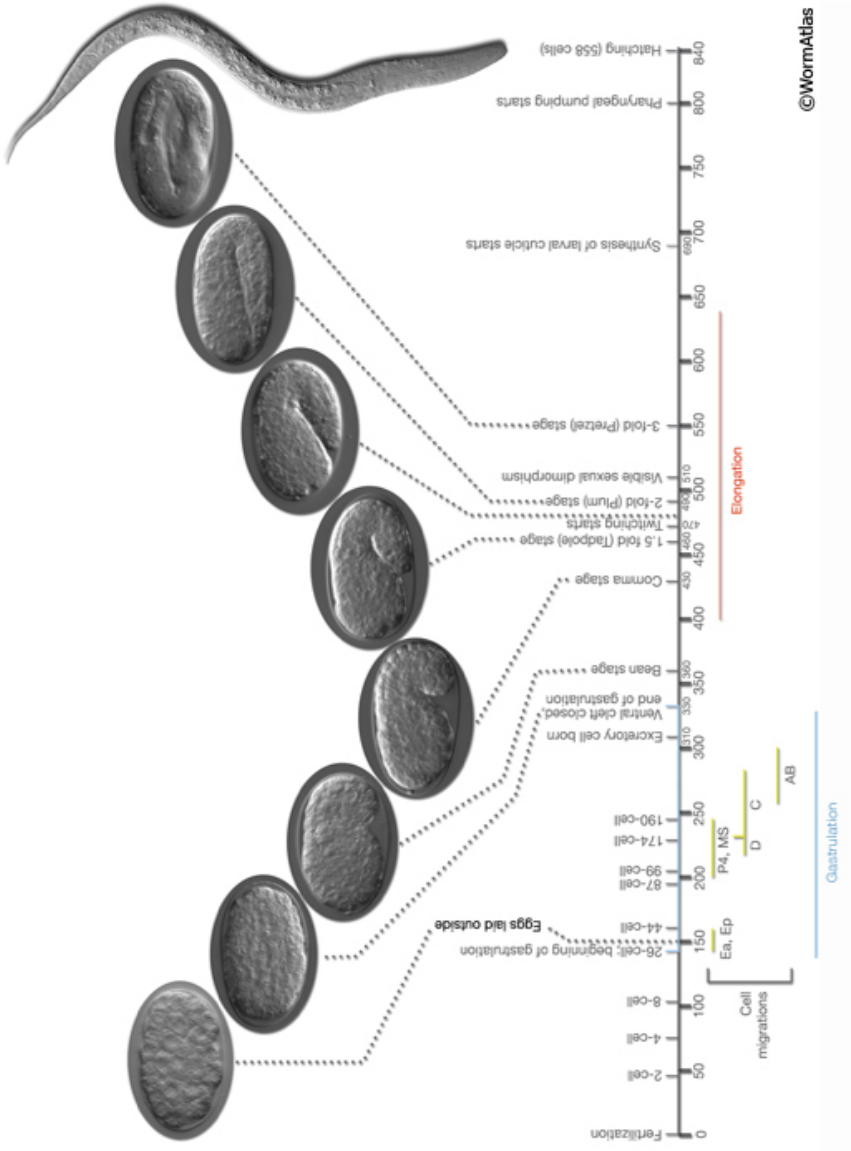
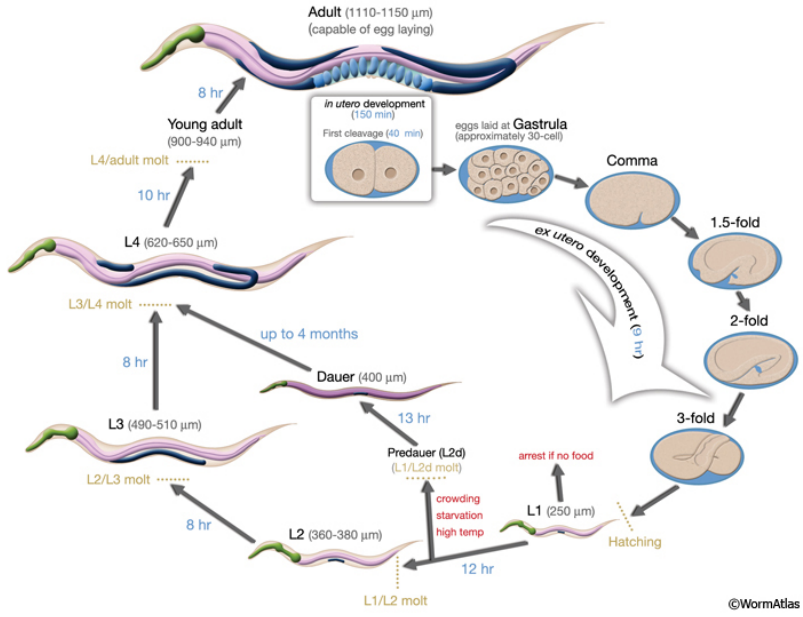


Figure 5

Life cycle of *C. elegans* at 22°C. 0 min is fertilization. Numbers in blue along the arrows indicate the length of time the animal spends at a certain stage. First cleavage occurs at about 40 min. post-fertilization. Eggs are laid outside at about 150 min. post-fertilization and during the gastrula stage. Image credit Wormatlas (www.wormatlas.org)



Feeding after hatching triggers post-embryonic development. In the presence of food, cell divisions resume and the post-embryonic developmental program begins 3 hours after hatching. If the embryos hatch in the absence of food, however, they arrest development until food becomes available. Once food becomes available, these arrested L1 stage larvae progress through normal molting and development (Altun and Hall, 2008; Sulston and Horvitz,

1977). The animal normally passes through four larval (L1-L4) stages to reach adulthood (Figure 5). Numerous blast cells set aside at the end of embryogenesis divide in nearly invariant temporal and spatial patterns through the four larval stages and give rise to a fixed number of cells with determined fates.

In 1980, Horwitz and Sulston reported the isolation of many mutants that altered the otherwise invariant pattern of post-embryonic cell divisions in *C. elegans* (Horvitz and Sulston, 1980). Interestingly, all of the 24 mutants found (including the mutant *lin-4,* that would later lead to the discovery of microRNAs (Lee et al., 1993)) displayed an incompletely penetrant phenotype and/or variable expressivity, even when grown in a controlled environment. Given that *C. elegans* hermaphrodites reproduce by selfing, populations of worms are assumed to be largely isogenic. However, spontaneous mutations do occur in a population. The measured mutation rate in *C. elegans* is 2.1×10^{-8} mutations per site per generation (~2.1 point mutations per genome per generation) (Denver et al., 2004). The duplication rate of genes in C. elegans is two orders of magnitude greater than the spontaneous rate of point mutation per nucleotide site, on the order of 10^{-7} duplications per gene per generation (Lipinski et al., 2011). In addition, if a mutant worm with an abnormal phenotype is allowed to self fertilize, the resulting F1 progeny shows both

phenotypically wild type (WT) and abnormal worms recapitulating the parental phenotypic distribution. Therefore, it is extremely unlikely that novel mutations explain this phenotypic variability. If not genetic or environmental, then what is the difference between two worms with the same mutation such that one hatches to be phenotypically WT and the other phenotypically abnormal? The authors hypothesized that this variation may be due to "leaky" expression of alleles (hypomorphic forms) whose fluctuations would lead to all-or-none variation in phenotypes, but they also made the important observation that this phenomena is also observed for putative null mutations alleles, implying that fluctuations in another component (related gene or pathway) is the origin of the phenotypic variation leading to the incomplete penetrance. It was not until almost 30 years later that studies have provided molecular insights into the phenotypic inter-individual variability caused by mutations in *C. elegans* (Burga et al., 2011; Casanueva et al., 2012; Raj et al., 2010; Topalidou et al., 2011).

Van Oudenaarden and colleagues were the first to study how noise in gene expression propagates in a transcriptional network in *C. elegans* (Raj et al., 2010). The authors used as a model the transcriptional network involved in the specification of the worm endoderm (Maduro, 2009). Previous genetic screens for maternal

embryonic lethal genes lead to the identification of the gene *skn-1* (Bowerman et al., 1992). *skn-1* null mutants arrest during embryonic development and fail to develop a pharynx 100% of the time. Endoderm development is also affected in these mutants, being absent in 70% of the embryos. *skn-1* transcripts and proteins are maternally deposited in the oocyte and this gene is thought to start the transcriptional cascade that leads to the activation of *elt-2*, the master regulator of gut development. This transcriptional cascade is characterized by a high degree of genetic redundancy. *skn-1* first activates two genes *med-1* and *med-2*, which in turn activate *end-1* and *end-3*. Independent deletion of each of these genes has none or very little consequences for the development of the endoderm, however, simultaneous deletion of *med-1/med-2* or *end-1/end-3* gene pairs result in severe defects compromising endoderm development (Maduro et al., 2005; Maduro et al., 2001). The authors used a modified Fluorescence *In Situ* Hybridization (FISH) technique that allowed them to count the number of single mRNA molecules present in embryos. They first observed that *elt-2* expression was bimodal in the *skn-1* mutant background, exhibiting an ON/OFF expression pattern. Analysis of other genes downstream to *skn-1* (but upstream of *elt-2*) revealed that *med-1/med-2* and *end-3* expression was almost abolished. However, end-1 expression was still detectable. *End-1* mean transcript expression values were lower and

more variable in *skn-1* mutants than in wild type embryos (WT). Given that *end-1* was the only transcription factor in this network whose expression was detectable, the authors hypothesized that its expression was responsible for the bimodal distribution of *elt-2*. The authors found that *elt-2* expression was only found in embryos with high levels of *end-1* expression between the 65-cell and 120-cell stage, whereas both genes were highly expressed in wild-type embryos during this time, suggesting that *end-1* expression needed to reach a threshold level during a critical developmental time window to activate *elt-2*. In addition, lower penetrance mutants of skn-1 showed a lower threshold of elt-2 expression compared to those with a higher penetrance (Raj et al., 2010). However, it is evident that some embryos with *end-1* expression above the threshold of *elt-2* activation still fail to activate *elt-2* expression, indicating there are still unknown determinants of variation.

In a similar study, Chalfie and colleagues showed that *alr-1* mutants (*alr-1* codes for a transcription factor) had increased expression variation of its downstream target *mec-3*. Individuals carrying a mutation in *mec-3* fail in the final step of differentiation of the six touch receptor neurons. Interestingly, *alr-1* mutants showed increased inter-individual variation in touch insensitivity assays. The authors proposed that this behavioral variability could be

linked to the increased inter-individual variability in *mec-3* transcripts. Though this idea is very appealing stronger evidence is needed to support such claims, given that *mec-3* transcripts are quantified in fixed samples so that no direct correlation between behavior and expression variability could be tested (Topalidou et al., 2011). In summary, recent studies have illustrated how noise in gene expression can propagate in a transcriptional network, but it is not really clear what are the determinants that lead to the phenotypic variation and whether it could be possible to predict the outcome of such incompletely penetrant mutations in single individuals based on pre-existing variation.

Objectives

Here I present the main objectives of this thesis

1. To determine if there is stochastic variation in the expression of genes in *C. elegans.*

   1.1 To develop a methodology which allow us to reliably quantify the expression of fluorescent reporters in *C. elegans*.
   1.2 To determine if there is inter-individual variation in the expression of fluorescent reporters and gain insights into its origin.

2. To determine what the phenotypic consequences of stochastic gene expression are in *C. elegans*.

   2.1 To test if the incomplete penetrance of mutations in *C. elegans* is due to the inherent stochastic expression of genes.

   2.2 To evaluate whether robustness to environmental and genetic perturbations are coupled in *C. elegans* and if there is any benefit associated with the stochastic response to stress experienced by individual worms.

Burga A, Casanueva MO, Lehner B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. Nature. 2011 Dec 7;480 (7376):250-253.

Burga A, Casanueva MO, Lehner B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. Supplementary information. Nature. 2011 Dec 7;480(7376):250-253.

Results and Discussion

Predicting mutation outcome from early stochastic variation in genetic interaction partners.

Alejandro Burga, M. Olivia Casanueva & Ben Lehner
Nature 480, 250-253 (2011)

On genetic redundancy and incomplete penetrance

In order to explain the incomplete penetrance of mutations in the absence of genetic and environmental variation, we proposed a model based on genetic interactions. Two genes, A and B, are said to be synthetic lethal if the effect of deleting both A and B is more detrimental to the fitness of an organism than the deletion of just A or B (Figure 1b). It has been proposed that genetic interactions, such as those described in synthetic lethality screens, underlie the buffering capacity of organisms (Hartman et al., 2001). Deletion of single genes in *S. Cerevesiae* often has a small effect on growth rate (Giaever et al., 2002). Such robustness to mutations could be due in principle to two mechanisms. First, compensation can arise from gene duplication, such that the loss of one gene can be compensated by the presence of the other copy (Wagner, 1996). Second, compensation can originate from different regulatory or metabolic pathways (Wagner, 2000a).Genome wide studies have revealed a higher probability of functional compensation for a duplicate gene

than for a singleton in *S. cerevisiae* (Gu et al., 2003). In addition, theoretical (Bergman and Siegal, 2003; Waddington, 1957) as well as experimental (Landry et al., 2007; Lehner, 2010; Queitsch et al., 2002) work has suggested that buffering mechanisms that confer robustness to genetic, environmental and stochastic perturbations are coupled. Thus if genetic interactions confer robustness to mutations, it is likely that they also confer robustness to stochastic events. Our model states that in the wild type (WT) situation, gene A normally buffers the effects of the inherent variability in the expression of gene B and vice versa. But if gene A is absent or its function highly compromised, then variability in the expression of gene B becomes a significant factor influencing the phenotypic outcome of the mutation (Figure 1d). Since gene duplication is the simplest and more intuitive origin of mutational robustness, we searched for cases of redundant genes originated by gene duplication events and characterized by incomplete penetrance in the literature.

tbx-8 and *tbx-9* are two genes that belong to the T-box family of transcription factors of C. elegans. The first member of this family to be identified was the gene *Brachyury* in mice (also known as T-gene) (Herrmann et al., 1990). Mutations in this gene cause truncated tails in heterozygous mice and homozygous mutants fail to form

the notochord, the entire posterior region and the allantois, and die at about 10 days of gestation (Herrmann et al., 1990). The T-box family is present in all metazoans studied so far and is defined by the presence of a conserved DNA binding domain, the T-domain. Inactivation and mutations in diverse member of this family have shown that T-box genes have diverse roles in patterning and morphogenesis. In addition, human developmental abnormalities such as: Holt-Oram syndrome, Ulnar-Mammary syndrome, X linked and cleft palate are caused by mutations in T-box genes (Naiche et al., 2005). The T-domain is about 180 amino-acid residues. Crystallographic studies of Xenopus *Brachyury* have revealed that this protein binds as a dimer, interacting with the major and minor grooves of DNA (Muller and Herrmann, 1997). In vitro binding assays have shown that *Brachyury* recognizes a palindromic DNA sequence. Each monomer of *Brachyury* binds the consensus sequence 5'-AGGTGTGAAATT-3' (Naiche et al., 2005).

tbx-8 and *tbx-9* map to *C. elegans* chromosome III, they are located only 3.75 kb apart and are divergently transcribed (Pocock et al., 2004). Both proteins are 59% identical in their T-box domain. Inactivation of each of these genes results in morphogenetic problems affecting a very small proportion of the population at 20°C (<5%). However, inactivation of both *tbx-8* and *tbx-9* simultaneously results in 100% embryonic lethality. Tbx-8 and tbx-9 have a

redundant role in posterior patterning of *C. elegans*, and double mutants show clear disorganization of the hypodermis and muscle, failing to elongate properly, but the cell number of these tissues is unaltered (Andachi, 2004; Pocock et al., 2004). Thus, the genomic location, conservation, phylogeny (Reece-Hoyes et al., 2007) and function of these two genes indicate that they were originated by a gene duplication event and that both proteins still retain redundant functions. Interestingly, both genes are also present in the *Caenorhabditis briggsae* (separated from *C. elegans* by 80–110 million years) and they also interact genetically (Baugh et al., 2005). This indicates that the duplication event that originated *tbx-8* and *tbx-9* occurred before the speciation event of *C. elegans* and *C. briggsae,* and that functional redundancy has been retained since then. We also studied *C. elegans' flh-1* and *flh-2*, another pair of functionally redundant transcription factors. These two genes belong to the class of Zn-finger FLYWCH transcription factors. Single *flh-1(bc374)* and *flh-2(bc375)* mutants are characterized by very low penetrance (~3%) of abnormal morphologies at 20°C, but 100% embryonic lethality is detected when both genes are deleted (Ow et al., 2008). *Flh-1* mutants have increased levels of miRNAs: *lin-4, mir-241* and *mir-48* compared to the wild type. However, the lethality of *flh-1; flh-2* double mutants is not primarily a consequence of excessive levels those miRNAs (Ow et al., 2008).

The *tbx-8*/*tbx-9* pair provided a good model to study whether stochastic gene expression had a role in the incomplete penetrance of mutations, because of their very well described redundant roles and the easily recognizable incompletely penetrant phenotype of their mutants. We reasoned that in the absence of a functional TBX-9, the most likely candidate to account for the phenotypic variance of the mutant was TBX-8. We obtained a strain carrying a deletion in the *tbx-9 locus* generated by the C. elegans Knock out Consortium. The exact location and extent of the deletion was not known. We sequenced the mutant allele *ok2473* and determined the boundaries of the deletion (Sup. Figure 1). Just the first out of four exons is still present and importantly, half of the T-box domain is missing, including conserved residues necessary of the protein-DNA interaction. We conclude that this is a null allele of *tbx-9*. The mutant *tbx-8(ok656)* was previously characterized and the deletion affects the four exons, also removing virtually entirely the T-box domain (Andachi, 2004). The use of null mutants facilitates the test of our hypothesis, since if some residual activity were still present in the mutated gene, then this would add another variable in addition to the variation in the expression of the paralog. The penetrance of *tbx-8/9* and *flh-1/2* mutants was very low (<5%) when worms were grown at 20°C, which is the standard temperature to grow worms in the lab. Given that

our set up allows one to quantify tens but nor hundreds of embryos simultaneously, it would have been very difficult to perform phenotypic predictions with such low expected number of abnormal phenotypes. We found that growing these mutants at higher temperatures (>25°C) increased the penetrance of these mutations, reaching around 40-50% (except for the *flh-2(bc375)*). Thus, the use of a temperature-controlled (both chilling and heating) chamber set up as part of the fluorescence microscope was of great importance for the feasibility of our experimental approach.

Measuring inter-individual gene expression variation in *C. elegans*

One of the requisites of our model is the pre-existence of variation in the expression of *tbx-8* and *tbx-9* in the wild type strain N2. We decided to use fluorescent reporters, as used by many others previous studies (Elowitz et al., 2002; Newman et al., 2006; Rea et al., 2005) to study the extent of inter-individual variation between embryos. We first focused on a transcriptional reporter for *tbx-8*, GFP (Green Fluorescent Protein) under the control of the promoter region of *tbx-8*. There are two alternatives to measure fluorescence in biological samples: fluorescence microscopy and fluorescence flow cytometry. Fluorescence microscopy revealed that the expression pattern of the tbx-8 reporter is dynamic during development. As expected, for a

zygotically expressed gene, no expression is detected in the early embryo and expression increases in time and peaks approximately during the 1.5-fold-stage (Figure 2a). Given this temporal dynamics in expression, it is important to compare developmentally synchronized embryos, otherwise difference in expression among embryos could be simply due to temporal differences. There is no easy way to obtain the large amount of living highly synchronized embryos necessary for flow cytometry measurements, so we discarded this option. We dissected young adult hermaphrodites and manually sorted early 4-8 cell stage embryos. This strategy allowed us to obtain a set of highly synchronized and unperturbed embryos, however a major limitation of this set up is that it has to be performed fast and is difficult to collect more than 30 embryos, especially when worms are grown at 25°C and the development proceeds faster.

This experimental set up revealed that individual embryos carrying the *tbx-8* reporter vary in their total fluorescence expression. Importantly, this was not due to developmental differences. The embryos were obtained from an isogenic population and environmental conditions were kept homogenous, therefore we refer to this inter-individual variation in gene expression as stochastic. Differences among embryos were relatively stable in time (Figure 2a). However, caution must be taken when

interpreting the dynamics of expression of any fluorescence transcriptional reporter with respect to the endogenous protein. The half-life of proteins varies considerably, and currently there is no data available on the half-life of proteins (including GFP) in the developing *C. elegans* embryo. The half-life of GFP in mammalian cells is approximately 26 hours (Corish and Tyler-Smith, 1999), but this value is cell line dependent. *C. elegans* embryonic development lasts ~10 hours at 25°C, so the GFP half-life could mask rapid fluctuations of the endogenous protein. Future work could make use of GFP fused to PEST domains, which have been shown to decrease significantly the half-life of proteins (Corish and Tyler-Smith, 1999).

Technical sources of noise could theoretically explain the differences in fluorescence between embryos. Our experiments suggest that this was not the case. We used a relatively low magnification (10X objective) so that typically around 10 embryos can perfectly fit in one field of view. Expression variability is already evident in one field of view and not just between two different fields. We manually adjusted the focus for each field of vision to compensate for small differences in the focal plane of the plate. The variation in fluorescence of the same embryo measured in different field of views is minimal compared to the embryo-to-embryo variation. Embryo orientation is also not a major determinant of signal differences in our settings. After

morphogenesis, embryos move very actively inside their eggshell, but those rapid movements do no affect the inter-individual variation measurements (Figure 2c). In addition, we optimized exposure times and we used a high sensitivity CCD camera to obtain a good signal/background noise ratio. Non-homogenous lamp illumination can create a gradient of intensity that can affect measurements. We observed this effect was important when measuring the *elt-5* reporter (due to the low fluorescence and the particular gfp/rfp dual filter used for this experiment). For this case, a flat field correction was applied that compensated for the technical differences. In summary, we minimized, controlled or compensated for several possible source of technical variation and still observed differences in fluorescence between embryos.

Validation of C. elegans transgenic reporters

One limitation of transgenesis in *C. elegans* is the lack of an efficient homologous recombination technology as the one available for *S. cerevisiae* and many other organisms (Capecchi, 1989; Petes et al., 1989). Homologous recombination allows the addition of a fluorescent tag to the protein by targeting its endogenous locus. Reporter genes generated in this fashion are ideal, since they should resemble the endogenous gene transcriptional regulation and variation. In *C. elegans* for

reasons not fully understood, the efficiency of homologous recombination is extremely low, therefore the community lacks this powerful tool. Traditionally, most transgenic lines are generated by microinjection of the transgene construct into the hermaphrodite germline, which is incorporated by some embryos as a repetitive extrachromosomal array (Mello et al., 1991). This array can later be incorporated spontaneously or by UV irradiation into the genome. As a result, typically hundreds of copies of the transgene are incorporated at some random location in the genome. Alternatively, bombardment of worms with gold particles coated with the transgene construct can also lead to direct integration of the transgene into the genome and typically in lower copies than with the microinjection technique (Wilm et al., 1999). It is known that transgenes can show position and copy number dependant effects. In particular, high copy transgenes are known to be silenced in the germline of *C. elegans* (Kelly et al., 1997), and some degree of somatic silencing has also been described in the worm soma (Grishok et al., 2005). In our work, we assumed that inter-individual variation of transgenes is actually reporting variation of the endogenous gene. Since no previous study has dealt with this issue, we decided to study if these factors affected our measurements of inter-individual variability in gene expression. Here, we discuss evidence that transgene variation correlates to a good extent with that of endogenous genes. One piece of evidence comes from a

double fluorescent reporter strategy, comparing the correlation of two independent transcriptional reporters for the same gene with that of two no-related genes. We generated a mCherry fluorescent transcriptional reporter strain for *tbx-8 (ptbx-8::mCherry)* and crossed it with the *ptbx-8::GFP* reporter. We found that the inter-individual variation of both independent transgenes was positively correlated (Sup. Figure 3a, r=0.71). In contrast, we found no correlation between variation of *ptbx-8::GFP* and other mCherry transgenes (such as *pdaf-21::mCherry* and *phsp-4::mCherry*) (Sup. Figure 3d). The positive correlation between dual reporters for the same gene was also observed for *daf-21* and *hsp-4* reporters (Figure 3b and Sup. Figure 3c and 3d. These results suggest that inter-individual variation in *tbx-8* and other reporters had an important extrinsic component and consequently, one could expect a correlation between the transgene and the endogenous gene. In contrast to studies in yeast and bacteria (Elowitz et al., 2002; Raser and O'Shea, 2004), in this case, it was not possible to differentiate between the contribution of intrinsic noise and transgene-related effects. Ideally, we could have tested the direct correlation between transgene reporter and endogenous gene by means of single molecule FISH (Fluorescence *In Situ* Hybridization). Unfortunately, we could not obtain a good signal for the GFP mRNA, possibly due to the small size of the GFP Open Reading Frame (ORF) and the low amount of probes

that could be target against it (32 probes, instead of the 48 probes recommended). Antibodies against TBX-8 or TBX-9 are not available, so we couldn't test the correlation of the transgene with those of the endogenous protein. However, the fact that *daf-21* reporter is constitutively expressed through all of *C. elegans* life span and the availability of an anti-DAF-21 antibody offered us the possibility to test the correlation between the variability of the transgene and the endogenous gene. As previously mentioned, multi-copy transgenes are silenced in the germline, however, DAF-21 is primarily expressed in the germline of developing and adults worms (Inoue et al., 2003). In order to circumvent this problem, we crossed the *pdaf-21::mCherry* reporter strain with *glp-1(e2141)* mutants that do not develop germline at 25°C (restrictive temperature). In this strain, all the DAF-21 protein is contributed by the soma, just like the expression of the *pdaf-21::mCherry* transgene. Sorting worms manually into High and Low groups based on their *daf-21* reporter fluorescence levels, followed by western blot, showed that there was a positive correlation between inter-individual transgene and endogenous protein variation (Figure 3a). Recently, a novel transgenesis technology based on Mos1 transposon was developed which allows the insertion of a single copy of the transgene of interest in *C. elegans* (Frokjaer-Jensen et al., 2008). Single-copy insertion allows also the expression of the transgene in the germline. We recently obtained a single copy *pdaf-21::GFP*

reporter. As expected expression levels are lower than those of the multi-copy transgene, yet quantifiable and the transgene is expressed strongly in the germline. We generated a dual reporter crossing the strains carrying the multi-copy *pdaf-21::mCherry* and the single copy *pdaf-21::GFP* and found that inter-individual variation was positively correlated (Additional Figure 1). The extent to which single copy transgene reporters can be used depends on the particular abundance of the studied gene. *daf-21* is one of the most highly expressed genes in *C. elegans*, and thus, fluorescence from the single copy can be reliably measured. But for genes expressed in lower levels, such as most transcription factors, the specific fluorescent signal may not be higher than the auto-fluorescence of the worms. This is a problem that also affects measurements in yeast and human cells (Li and Xie, 2011). In summary, the correlation between independent multi-copy transgene reporters; between multi-copy and single-copy reporters and the correlation between *daf-21* reporter and endogenous DAF-21 protein levels argue that multi-copy transgenes can be used as a proxy to the inter-individual stochastic variation of endogenous genes.

A previous report studying the propagation of noise in the mesoderm specification network of *C. elegans* made use of single molecule FISH to study the incomplete penetrance *skn-1* mutants (Raj et al., 2010). One of the

disadvantages of using fluorescent reporters is that inter-individual variation in gene expression cannot be directly linked to the appearance of phenotypes. FISH techniques rely on fixed samples, so that it is not possible to know if differences in gene expression are the cause or the consequences of the phenotypic differences. However, the great resolution of single molecule FISH makes it a very valuable tool and together with fluorescent reporters, provide a powerful toolkit for the study of stochastic variation.

Variable levels of paralog induction among individuals predicts the outcome of mutations

Given that *tbx-8* expression is variable among individuals, we tested whether this variation could be responsible for the incomplete penetrance of *tbx-9* mutants. We generated a strain carrying the *ptbx-8::GFP* and the *tbx-9(ok2473)* mutation. First, we compared the expression levels of *ptbx-8::GFP* reporter in the WT to those in the *tbx-9* mutant. Interestingly, the expression of the reporter was up-regulated in the mutant background. This up-regulation was confirmed for endogenous gene using single molecule FISH (Sup. Figure 5) , thus confirming that fluorescence of the transgene is a good indicator of the endogenous levels of expression. We also observed up-regulation of *ptbx-9::GFP* in a *tbx-8(ok656)* background and *pflh-2::GFP* in a

flh-1(bc374). We also performed control experiments and showed that the up-regulation is paralog specific. This back-up mechanism, in which upon deletion of a gene its paralog is transcriptionaly reprogrammed to compensate for its lost, was first proposed and described by Pilpel and colleagues (Kafri et al., 2005). The up-regulation of genes to compensate for deletion of their paralog has been described in the literature in prokaryotes, yeast and mammals (Kafri et al., 2005; Rudnicki et al., 1992). Probably the most exhaustive study designed to quantify the prevalence of this paralog responsiveness was performed in yeast by Kishony and colleagues in *S. cerevisiae* (DeLuna et al., 2010). They found that only a modest fraction of proteins (22 out of 202) showed a significant up-regulation upon deletion of their duplicate genes in mid-log phase in rich medium (YPD). However, these paralog-responsive proteins matched almost exclusively duplicate pairs whose overlapping function is required for growth (as defined by a synthetic interaction). Moreover, many paralogs were up-regulated just when the mutant strain was grown in an environmental condition that made the gene necessary for growth (DeLuna et al., 2010). We have tested just three pairs of genes combinations for paralog up-regulation and those three showed this behavior. The function of all these genes is essential for normal development; therefore our results are consistent with previous observations. The observed paralog up-

regulation is relatively small in the cases we have studied, 1.2-4.3 fold increase at the transcriptional level. Kishony and colleagues also observed a modest up-regulation in their systematic survey in yeast (1.13-fold to over 20-fold; median value 1.7-fold at the protein level) (DeLuna et al., 2010). This indicates that modest changes at the transcriptome or proteome level could have an important impact at the phenotypic level, especially in mutants. Given that the transcriptional back up-mechanism of a given “gene A” in a “gene B” mutant background seems to be a good predictor of genetic interaction between genes A and B, maybe global gene expression measurements in different genetic backgrounds could be used to predict genetic interactions. This strategy could be of special interest for model organisms in which genetic interactions are difficult to test or in which advanced genetics tools have not been developed. A previous study investigated this matter, using expression data from a collection of knock-out mutants (Hughes et al., 2000) and genome wide genetic interaction studies in S. *Cerevisiae* (Tong et al., 2004). The authors found just very few examples of transcriptional response of a gene upon deletion of its interacting partner. Among 872 studied pairs, just 1.6% (14) were significantly changed of which 93% (13/14) exhibited significant up-regulation (Wong and Roth, 2005). However, one caveat of this study is that they used data generated growing yeast in different conditions. The expression experiments were conducted in

rich medium, while the genetic interactions were assessed in near minimal medium. Also, gene expression measurements were performed with microarrays and the sensitivity may not have been high enough to detect small changes in expression. Therefore, more careful studies could reveal a stronger link between transcriptional compensation and genetic interactions.

In order to test our model for incomplete penetrance, we followed by fluorescence and bright field time-lapse microscopy the expression of the *ptbx-8::GFP* reporter in *tbx-9* mutant embryos. The characteristic anterior morphogenetic defects of *tbx-9* mutants allowed us to classify larvae at the moment of hatching into phenotypically wild type or phenotypically abnormal, if a defect was detected. Then, we asked if those *tbx-9* mutant embryos that hatched into phenotypically wild type had higher expression levels of the *tbx-8* reporter. That was indeed the case (Figure 2c and 2d). Expression levels of the *tbx-8* reporter measured at the comma stage of embryogenesis were predictive of the phenotypic outcome of *tbx-9* mutation. We also quantified the predictive power of total expression and rate of induction at other developmental time points, but predictions at the comma stage were better (Sup. Table 3). Given that we are quantifying the abundance of GFP and not the one of the endogenous TBX-8 (the half-life of this proteins could be

different) and GFP has a known fluorescence maturation time, we cannot conclude that variation of TBX-8 at the comma stage is the best predictor, since it could actually take place earlier.

We performed control experiments in order to test if the predictive power of the *tbx-8* reporter was due to general transgene effects. We performed the same experiment but using a reporter for the constitutively expressed gene *plet858::GFP* and we found no correlation between the variation in the expression of this reporter and the phenotype of *tbx-9* embryos (Sup. Figure 7a,b). We performed an additional set of experiments using a *pelt-5::mCherry* reporter, *elt-5* codes for a transcription factor expressed in the hypodermis of the embryo. This reporter was also not predictive of the penetrance of the *tbx-9* mutation (Sup. Figure 7c,d). Altogether, these experiments suggest that in the absence of *tbx-9* function, a feed-back mechanism is responsible for the up-regulation of its paralog, *tbx-8.* However, the level of induction of *tbx-8* is variable among individual embryos, so that embryos that express lower levels of *tbx-8* have a greater chance of developing anterior morphogenetic problems compared to high *tbx-8* expressor embryos.

To explore the generality of such mechanism, we also generated a *tbx-9* reporter (*ptbx-9::GFP*) and crossed it

with a *tbx-8(ok656)* mutant strain. As previously mentioned, we observed the up-regulation of the *tbx-9* reporter in the *tbx-8* mutant background and also that variation in *tbx-9* reporter´s expression levels was predictive of the phenotypic outcome of *tbx-8(ok656)* mutants (Figure 2). Furthermore, we looked at an additional pair of redundant genes, *flh-1* and *flh-2*, also originating from a gene duplication event. An *flh-2* reporter was also up-regulated in an *flh-1* mutant background and variation in *flh-2* reporter´s expression levels was predictive of the phenotypic outcome of *flh-1(bc374)* mutants (Figure 2). We performed an additional set of controls and showed by measuring *tbx-8* reporter expression in *flh-1* mutants and *flh-2* reporter expression in *tbx-9* mutant background that the phenotypic predictions were paralog specific (Sup. Figure 7). *Flh-2* expression pattern differs from that of *tbx-8* and *tbx-9* in that is not exclusively embryonic, but its expression if also observed in all larval developmental stages and also in the adult worm. We based our *flh-1* phenotypic predictions on a very spatially restricted *flh-2* expression pattern in the pre-comma stage embryo. Quantification of *flh-2* reporter was not measured as the total fluorescence of the embryo as for the other reporters. The auto-fluorescence background of the green channel (which itself varies between embryos) was high compared to the specific *flh-2* expression signal in the early embryo, therefore we measured the expression of the small cluster of cells expressing *flh-2* by using a region

of interest (ROI) mask of the same size for all embryos. More ubiquitous *flh-2* expression in the late embryo did not predict the *flh-1* phenotype. This could be due to the apparent pleiotropic role of *flh-2*. *Flh-2* is involved in morphogenesis in the early embryo, but its strong and constitutive expression pattern in the anterior nervous system from the late embryo to adult stages suggests that it could have additional roles which are not redundant to *flh-1* function (Ow et al., 2008).

Global inter-individual variation of chaperones in developing embryos

We found in our studies that in general, transcriptional reporters for different genes showed no correlation in their inter-individual variation, but there were some exceptions. Interestingly, what these correlated genes had in common was they were all genes involved in stress-response. For example, we found a positive correlation, in the absence of stress, between the variation of *daf-21* and *hsp-4* (r=0.63, Figure 3) and between *hsp-4 and hsp-3* (r=0.73, Sup. Figure 14). All of these are constitutively expressed chaperones. However, not all stress-response genes were correlated in their variation, as exemplified by the lack of correlation between *sod-3* and *daf-21* reporters (r=0.10, Sup. Figure 14). So, it is unlikely that this applies to all stress genes. These results suggest that not all genes

vary independently, but some genes vary in concert. We have just tested correlations in developing embryos, so it could be possible that this concerted variation among genes is also stage specific. For the particular case of chaperone co-variation, it is likely that variation in an upstream factor that regulates the expression of many chaperones, is the responsible for this observation. We think that a good candidate for this upstream factor is the transcription factor HSF-1, the master regulator of the heat shock response.

Stress-induced transcription requires activation of Heat Shock Factor (HSF) that binds to the heat shock promoter element (HSE), characterized as multiple adjacent and inverse iterations of the pentanucleotide motif 5′-nGAAn-3′. Yeast, *D. melanogaster* and *C. elegans* have one HSF protein (HSF-1); whereas vertebrates have several paralogs (Morimoto, 1998). It is thought that in the absence of stress, intra-molecular interactions between the HSF amino- and carboxyl-terminal coiled coil domains sequester the protein in an inactive form. However, in addition to mediating the induction of many chaperones exclusively under heat shock treatment, HSF also mediates the constitutive expression of many genes. Indeed, Hsf1 is an essential gene for both S. cerevisiae and *C. elegans* in the absence of stress (Sorger and Pelham, 1988; Sugi et al., 2011). It would be interesting to test if variation in *C.*

elegans hsf-1, *daf-16, skn-1 or* other stress response genes explains the variation of chaperones in the WT.

Extrinsic noise can be due to variation in global factors affecting all genes and to variation in factors that are pathway specific (Colman-Lerner et al., 2005). However, their relative contribution is not well understood. A recent study by El-Samad and colleagues has shed light into this matter (Stewart-Ornstein et al., 2012). The authors performed a study in a broad array of proteins (456) in C*. cerevisiae* and found, as previously reported, that all genes experience intrinsic noise that dominates at low levels and a modest amount of floor extrinsic noise (CV=9%). However a substantial subset of genes experience high levels extrinsic noise. Examination of the proteins that exhibited the top 10% of highest extrinsic noise revealed that their respective promoters were enriched for several transcription factor -binding motifs. Thus the authors hypothesized that some of the extrinsic noise could be produced by pathway-specific fluctuations (Stewart-Ornstein et al., 2012). The authors name *"noise regulons"* as those genes that correlate in their variation because of pathway-specific fluctuations. For instance, yeast general stress response depends on Msn2/4 two partially redundant transcription factors. Target genes of these TFs showed more co-variation with Pgm2 (a known target of Msn2/4) than with Rpl17b (ribosomal subunit). Other identified *noise regulons*

correspond to amino-acid biosynthesis and mitochondrial maintenance. In summary, the co-variation of stress response genes that we described in *C. elegans* has been confirmed in S. *cerevisiae* and this has opened immediate question such as what's the global picture of pathway co-variation in *C. elegans* and how does it vary during development and evolution?

Variability in paralog and chaperone induction among embryos explains incomplete penetrance of mutations.

We have found significant differences in the expression level of the *tbx-8* reporter between phenotypically wild type and abnormal *tbx-9* mutant worms, however there was not a clear threshold and the predictive power was not so high as indicated by Receiver operating characteristic (ROC) analysis (Sup. Figure 17, Area under the Curve (AUC)=0.67). There are several possible explanations for this. First, even though we have minimized sources of technical noise, there probably are still limitations in this respect, especially for lowly expressed transgenes. Second, though we have provided extensive evidence that our transcriptional reporters are a good proxy to study the inter-individual variation in the endogenous genes, this may not be the case. Even if the source of noise was entirely (or largely) extrinsic and the correlation

between the transgene and the endogenous gene was perfect, we must bear in mind that the actual function is performed at the protein level (in the genes we have studied). Thus, we are assuming a good correlation between mRNA and protein level. Genome wide studies have revealed that overall mRNA and protein abundance are clearly positively correlated; nevertheless there is clear variation, in a gene-to-gene basis (Maier et al., 2009). Deviation from linearity in the mRNA and protein abundance relationship could add noise to our predictions. Also, our transgenes may be lacking important information at the level of transcription, such as long distance enhancers. The mRNA 3' UTR (un-translated region) of genes also encodes important regulatory information. All our transgenic lines use the same 3' UTR from the unc-54 gene. This particular 3'UTR is largely used in the *C. elegans* community and thought not to be target of any particular regulatory mechanism (such as miRNA regulation). It is possible that the use of the endogenous 3'UTR of the studied genes could have given better predictions. Histone modifications and nucleosomes positioning of the transgene could also have important differences compared to the endogenous locus. However, as previously discussed, currently, it's not possible to tag genes at their endogenous locus in *C. elegans*. A third and more appealing possibility is that the penetrance of mutations does not depend on the variation of just one

particular gene, but could be the result of fluctuations in multiple buffering pathways.

We decided to test if variation in chaperone buffering could explain part of the missing phenotypic variance. We had several reasons to think that variation in chaperone expresison could play a role. First, previous research focused on the chaperone HSP90 in models such as bacteria, yeast, plant and flies has shown that chaperones can act as promiscuous buffers of genetic variation (Queitsch et al., 2002; Rutherford and Lindquist, 1998; Tokuriki and Tawfik, 2009). Work performed in our lab has shown that increased chaperone dosage, by means of HSF-1 overexpression or heat shock, can decrease the penetrance of many mutations in *C. elegans* (Casanueva et al., 2012). Indeed, HSF-1 overexpression could reduce the penetrance of *tbx-9* mutants. Though it has been previously shown that the level of induction of the *hsp-16.2* chaperone upon a heat stress is variable among individuals (Rea et al., 2005), no previous work had looked at the variation of constitutively expressed chaperones. *daf-21* is the worm ortholog of HSP90 and is one of the most abundant chaperones in the developing embryo. It is expressed both maternally and zygotically and it's essential for embryonic development. We created a transcriptional reporter for *daf-21*, *pdaf-21::mCherry*. Expression of this transgene was highly variable among individual developing embryos and

during later larval stages. Interestingly, inter-individual variation in *daf-21* expression was predictive of *tbx-9* mutation outcome. Thus, *tbx-9* mutant embryos expressing higher levels of *daf-21* reporter early in life had higher chances of developing into phenotypically wild type larvae. It is not entirely clear how higher levels of DAF-21 or other chaperones could buffer the effects null genetic mutations. Studies in tyrosine kinase and steroid hormone receptor maturation suggest that HSP90 binds to “clients” that are already substantially folded and facilitates their conformational remodeling (Pearl and Prodromou, 2006). In addition, genome wide epistatic studies performed in yeast revealed that HSP90 interacts with hundreds of genes in an environment dependant manner (Pearl and Prodromou, 2006). The authors used macbecin II to chemically inhibit HSP90 in the yeast single-gene knockout collection. Interestingly, these knockouts remove the genes ORF, thus showing that Hsp90 can interact with null mutations, as our experiments also suggest. This chemical-genetic screen revealed that under normal growth conditions (30°C), HSP90 plays a major role in various aspects of the secretory pathway and cellular transport, among others. However, during environmental stress (37°C), HSP90 is required for the cell cycle, meiosis, and cytokinesis. Among *Hsp90* interacting genes there was a significant enrichment in genes encoding subunits of oligomeric complexes, which suggest that HSP90 could be regulating the stability of

proteins prior to the assembly of large complexes or facilitatating their conformational transitions during either binding or removal of these complexes (Pearl and Prodromou, 2006).

We found no correlation in inter-individual levels of variation between *daf-21* and *tbx-8* reporters, indicating that these pathways vary independently (Figure 4). In fact, considering the variation in expression of both reporters simultaneously increased the accuracy of phenotypic predictions for *tbx-9* mutants based on ROC (Sup. Figure 17) analysis (AUC=0.68, 0.69, 0.77 for *tbx-8*, *daf-21* and the joint contribution, respectively). Predictions are still not perfect, and the explanations previously enumerated could still be valid for this situation, not discarding even fluctuations in a third unknown component. A genome wide RNAi genetic interaction screen could be performed in the single mutant background in order to find new interacting genes and potential buffers. Our results indicate that variation of two independent buffering pathways, one that originated through a gene duplication event (gene specific) and another related to global chaperone variation (promiscuous) determines the variable phenotypic outcome of a mutation.

Although we have focused on this work on redundant genes originating by gene duplication, the large majority of

the redundancy in genomes comes from interactions between different pathways not necessarily sharing a common phylogenetic origin. (Costanzo et al., 2010). Our model for incomplete penetrance is perfectly compatible with such redundancy and we expect it apply for such cases as well. Moreover, it could also apply to cases of positive epistasis (Lehner, 2011), where the double mutant has a less severe phenotype than either single mutant. In this situation, lower (instead of higher) levels of the interacting partner could be predictive of the better-fit phenotype. Also, we speculate that stochastic variability in the activity rather than expression levels of genes could predict cases of incomplete penetrance. This could be of importance when looking at enzymes involved in metabolism or epigenetic modification of DNA and chromatin.

Implications for evolution

Both small-scale and large-scale genetic studies in diverse models have revealed that genetic redundancy is a common feature of organisms (Pickett and Meeks-Wagner, 1995; Thomas, 1993). Gene duplications events, in addition to being a source of evolutionary novelty (Ohno, 1970), can also confer such redundancy (Li et al., 2010; Wagner, 1996). However, it has also been proposed that genetic redundancy should be an evolutionary unstable state. Given complete redundancy between two genes, there would be

no selective constraints preventing the appearance of deleterious mutations in a second copy. This second copy would accumulate mutations, drift and eventually decay into a pseudogene (Brookfield, 1992). Phylogenetic analyses have shown that in multiple examples, genes have retained redundant functions since the divergence of the animal, plant and fungi kingdoms over a billion years ago (Vavouri et al., 2008). So, what are the mechanisms behind the evolutionary stability of redundant genes? To date, many mechanistic models have been proposed. Population genetics simulations indicate that high mutation rates or populations sizes can lead to the maintenance of redundancy (Wagner, 2000b). Redundancy can be stable under certain combinations of mutation rates, efficacy of function and pleiotropy (Nowak et al., 1997). In addition, a mechanism has been proposed which is based on “developmental errors”. This term refers to the inaccuracy of information transmission in biological systems. The authors found that in order for redundancy to be stable, the mutation rate in each gene has to be smaller than the developmental error rate in the other gene (Nowak et al., 1997). Stochastic variability in gene expression can be understood as an example of development error, since it can lead to inaccuracy in the transmission of information in transcriptional and signaling networks. As we have previously mentioned (see Table I in Introduction), errors or noise are far more frequent than genetic mutations. On a

more macroscopic level, there are few studies that have quantified developmental errors, partly because measuring them is a complicated task. The development of the vulva in *C. elegans* offers a system where these errors can be quantified. Braendle and Felix found that although the vulval pattern is robustly generated in the wild-type, it does fail (Braendle and Felix, 2008). Error rates for diverse deviant morphologies depend on the particular environmental conditions. For instance, error rates in the division of P4.p and P8.p cells ranges from ~0.1% in dish culture to ~1.2% in liquid culture. Although the first impression is that these values are low and not so relevant, they are orders of magnitude higher than estimated mutation rates. The measured mutation rate in *C. elegans* is 2.1×10^{-8} mutations per site per generation (Denver et al., 2004), very similar to the one of humans, 2.5×10^{-8} mutations per site per generation (Nachman and Crowell, 2000). Thus errors are much more frequent and could indeed act as a selective force in addition to mutations. Genetic background also influenced failure rates and they were consistently found to be increased in mutation accumulation lines, thus suggesting that robustness of vulval development is maintained by selection (Braendle and Felix, 2008). Thus the assumption of Nowak and colleagues regarding the frequency of developmental errors seems more than plausible. Our work provides experimental evidence that supports such a model, since we have shown that

stochastic variability (developmental errors) in the expression of paralogs has important phenotypic consequences and could provide enough pressure to maintain redundancy. However, we do not argue that the "developmental error" model is the sole contributor to the maintenance of the *tbx-8/9* pair. Interestingly, this gene pair shows divergence at the protein and regulatory level (expression patterns are highly similar, yet different). Selection on their differential pattern of expression could be also a contributor. What's the relative importance of all these factors is a matter of debate and an interesting direction for future research.

It has also been previously suggested that the incomplete penetrance of mutations could have an important role in evolution (Eldar et al., 2009). The evolution of complex novel traits that rely on the independent change of several components has puzzled evolutionary biologists. In particular, if intermediate state mutations show a decrease in fitness. Low penetrance of mutations affecting a small subset of individuals could allow organisms to accumulate different mutations without affecting the large majority of the population, yet "testing" their impact at the phenotypic level. If one particular combination of mutations confers a fitness advantage, then secondary mutations could increase the penetrance, so that now most of the individuals of the population will be fitter (Eldar et al., 2009).

It would be interesting to study how genetically interacting partners and their stochastic buffering constrain the evolutionary landscape of organisms and if this mechanism could also be of importance in non-isogenic populations.

Implications for personalized medicine

In the last years, with the advent and massification of high throughput sequencing technologies, the idea of a personalized medicine has gained adherents in the scientific and medical community. In the practical sense, the full sequence of a given individual will provide information that could influence the diagnosis, prognosis and treatment of diseases (Chen et al., 2012). Given that an important part of the phenotypic variation in human disease is non heritable, together with the fact that concordance rate between monozygotic twins for most diseases is below 50%, especial caution must be taken regarding such claims. Probably personal genome information will be useful when it comes to predict the “typical” outcome of a mutation (Jelier et al., 2011), but since “personalized” medicine is about *individuals,* the biggest unknown is to what extent it will help us increase the success of predictions for a each patient. If the principles we have uncovered in *C. elegans* are conserved in humans, an important part of the development of personalized medicine should be directed

into understanding the role of stochastic processes and how to measure such variables in humans in order to predict disease.

Casanueva MO, Burga A, Lehner B. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. Science. 2012 Jan 6;335(6064):82-85.

Casanueva MO, Burga A, Lehner B. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. Elegans. Supplementary Online Material. Science. 2012 Jan 6;335(6064):82-85.

Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans.

M. Olivia Casanueva, Alejandro Burga & Ben Lehner
Science 335, 82-85 (2012)

Correlation between environmental and mutational robustness

The environment has a fundamental role in the determination of phenotypes and organisms have generated and evolved ways to sense and respond to it. Some environmental conditions can be potentially very harmful to organisms such as stress caused by a fast increase in temperature (heat shock). Proteins need to adopt particular tertiary structures to fulfill their biological functions. Pioneering studies performed by Anfinsen showed that the primary structure (linear sequence of aminoacids) of certain proteins contains all the necessary information for proteins to fold *in vitro* (Anfinsen, 1973). However, a large fraction of proteins in the cell require assistance to reach their folded states efficiently and on a biologically relevant timescale (Hartl, 1996). High temperatures among other kind of stresses can alter the folding path or destabilize the native state of proteins causing the accumulation of misfolded protein states. Misfolded proteins have a reduced activity and are also prone to experience aggregation (the association of two or more non-native protein molecules). For a subset of

proteins, one of the outcomes of protein aggregation is the formation of highly ordered, fibrillar aggregates called amyloid that are usually toxic to cells and may give rise to some of the most debilitating neurodegenerative diseases (Hartl and Hayer-Hartl, 2009). Under heat stress, cells and organisms respond inducing the expression of hundreds of genes, among those, a large group of proteins called chaperones. Chaperones are a very conserved group of proteins (from bacteria to mammals) and their canonical function is to help other proteins to fold. They protect stressed cells by their ability to recognize nascent polypeptides, unstructured regions of proteins and exposed hydrophobic stretches of amino acids. In doing so, chaperones hold, translocate or refold stress-denatured proteins and prevent their irreversible aggregation with other proteins in the cell (Nollen and Morimoto, 2002). As previously mentioned, some proteins need help to fold even in the absence of a heat-stress and therefore chaperones have a constitutive role in protein homeostasis. As a matter of fact some chaperones are essential for viability. Chaperones, therefore, are fundamental players mediating the environmental robustness of cells. However, studies initially performed in *Drosophila melanogaster* (Rutherford and Lindquist, 1998) and later confirmed in many other species suggested that chaperones also provide a link between environmental and genetic robustness (Jarosz and Lindquist, 2010; Queitsch et al., 2002; Tokuriki and Tawfik,

2009). HSP90 is essential for Drosophila embryonic development. Heterozygous mutants (reduced HSP90 activity) are viable, but interestingly, a small proportion of them show a myriad of developmental abnormalities. Moreover, these abnormal phenotypes are specific to different genetic backgrounds. These observations suggested that reduction of HSP90 function didn't just affect the stability of development, but that it revealed previous cryptic genetic variation present in the population that was being buffered, acting like a capacitor. (Rutherford and Lindquist, 1998). A similar conclusion was reached when studying HSP90 function in the mustard plant *Arabidopsis thaliana,* by pharmologically inhibiting its seven HSP90 proteins with geldanamycin (GDA) (Queitsch et al., 2002). The authors studied recombinant inbred lines (RI lines) originating from crosses between two accessions followed by single-seed self-propagation for eight generations. In contrast to Drosophila, which is an obligatorily outcrossing species, the Arabidopsis IR lines were largely isogenic. Surprisingly, most abnormal phenotypes were incompletely penetrant, affecting a small percentage of the population. In some cases, partial penetrance was not due to segregating genetic variation (low but present). These results suggested that Hsp90 normally acts to reduce the likelihood that stochastic events will alter the deterministic outcome of developmental pathways (Queitsch et al., 2002). However, given that in this case HSP90 perturbation is not genetic,

but pharmacological, is not clear if there is variation in the effective inhibition *per se* among individuals. Nevertheless, work on the chaperone HSP90 supports the idea of a strong correlation between robustness to environmental, genetic and even stochastic variation. Furthermore, both *in silico* studies of genetic network evolution (Bergman and Siegal, 2003; Kaneko, 2007) and genome-wide analysis in *C. cerevisiae* (Landry et al., 2007; Lehner, 2010; Levy and Siegal, 2008) have consistently found, a coupling between environmental, genetic and stochastic buffering, as initially suggested by Waddington (Waddington, 1957). Thus, even though HSP90 could be a major player in these phenomena, evidence indicates that many genes can act as capacitors, especially those which are hubs or highly connected nodes in genetic networks (Lehner et al., 2006; Levy and Siegal, 2008).

The induction a stress response decreases the penetrance of mutations

HSF-1 is transcription factor, known to be the master regulator of the heat shock stress response from yeast to mammals. Previous studies have shown that overexpression of HSF-1 in *C. elegans* ubiquitously in somatic cells under the control of the *let-858* promoter resulted in a 22% increase in life span, whereas expression of a mutant HSF-1 lacking the amino terminal DNA binding

domain had no effect. In addition, overexpression of *Hsp16* or *Hsp70* in *Drosophila* and *C. elegans* can increase life span (Tatar et al., 1997; Yokoyama et al., 2002). We decided to test if HSF-1 overexpression could affect the penetrance of several incompletely penetrant mutations in *C. elegans*. We found that penetrance was decreased in eight out of eleven (8/11) mutations tested carrying the HSF-1 transgene. Why are some mutations rescued and not others? Interestingly, those mutations that could be rescued showed a temperature dependant effect on the penetrance (Sup. Table 5). Many of the mutants studied here are likely not null alleles, and maybe direct physical interactions of chaperones with the mutated protein could alleviate the effect of the mutations. Also, interaction of chaperones with members of the affected pathway, like downstream targets, could be involved; we currently cannot discriminate between these options. The fact that the penetrance of the mutants is reduced is also interesting, given that all worms carry the same transgene overexpressing HSF-1. Why are some worms rescued and some not? If HSF-1 dosage is important, and if HSF-1 expression from the transgene is noisy then that could explain this observation. Alternatively, independent of HSF-1 dosage there could be intrinsic limitations in the heat shock response among individuals and also other buffering mechanism not involved in stress response.

Inter-individual variation in environmentally induced mutational robustness

Worms growing at 20°C and then changed to an incubator with a 35°C temperature will trigger the heat shock response will eventually die because of this stress. However, a short heat shock (2-3 hours at 35°C) is sufficient to trigger the heat shock response and moreover worms subjected to this mild stress live longer than their non-stressed counterparts (Cypser and Johnson, 2002). This phenomenon is termed “hormesis”. Also, worms subjected to the hormetic treatment are more resistant to subsequent stresses. The most likely mechanism is that the short heat shock induces chaperones that can act more promptly when the second stress arrives. As previously mentioned, overexpression of HSF-1 increases life span of worms, and this may explain the increased life span of worms subjected to a mild heat shock. In concordance with our previous observation, hormetic stimulation of worms by a mild heat shock early in life, protected worms phenotypically against mutations that acted in later development (Figure 1C). For instance, *lin-31(n1053)* mutants show an incompletely penetrant phenotype characterized by the induction of multiple or no vulva in contrast to the single vulva present in wild type worms. The induction of the vulva occurs at the L2 stage of postembryonic development. Performing the hormesis

treatment in L1 larvae diminished the penetrance of *lin-31(n1053)* mutants. But again, like in the case of HSF-1 overexpression, not all individuals were equally protected by the mild heat shock treatment. Could these observations be an indication of a intrinsic variability of the heat shock response among individuals?

A pioneering study by Johnson and colleagues had previously shown that the heat shock response is variable among individual worms (Rea et al., 2005). They used a transcriptional reporter of the *hsp-16.2* gene, fusing the promoter of this gene to the Green Fluorescent Protein (GFP). *hsp-16.2* is not detectably expressed under normal conditions but is induced after a heat shock treatment and this induction is at least partially dependant on HSF-1 (Hajdu-Cronin et al., 2004). Worms carrying this transgene were subjected to a mild heat shock, such as the hormesis treatment previously described. Heat shocked worms induced the expression of GFP in a ubiquitous fashion. A later follow study by the same group later showed Interestingly, even though the population of worms is isogenic and experienced the same environmental insult, the induction of GFP levels was highly variable among individual worms. The authors made use of a COPAS Biosorter machine, a continuous flow system capable of analyzing, sorting and dispensing objects ranging in size from 10 to 1,500 µm (an adult *C. elegans* is approximately 1

mm. long). The fluorescence distribution of the population had a Gaussian shape. Sorting of the individuals at the extremes of the distribution into “High” and “Low” *hsp-16.2* fluorescence groups (from now on referred as High and Low groups) showed that worms with higher *hsp-16.2* reporter induction live longer and where more resistant to stress than the “Low” induction group (Rea et al., 2005). This observation was later confirmed by the same group using a single copy insertion of the *hsp-16.2* reporter strain, ruling out the possibility that the phenotypic or gene induction differences are an artifact from the multi-copy transgene (Mendenhall et al., 2012). Also, later studies have shown that High and Low induction worms are not only different quantitatively but that qualitative differences in the cells that express *hsp-16.2* are evident (Seewald et al., 2010). Most of the expression signal originated in the intestinal cells and that is true for both the High and Low groups. However, a deeper analysis revealed two patterns in the Low induction group (A and B). In the High animals, most of *hsp-16.2* reporter expression emanated from cells comprising the first intestinal ring (cells intDL, intDR, intVL, and intVR) and the fifth intestinal ring (int5L and int5R); the same pattern is apparent in LowA clustered animals. However, the expression pattern of LowB animals showed more prominent expression in the fifth and ninth (int9L and int9R) intestinal rings. Thus varying tissue specificity of *hsp-16.2* expression and/or quantitative differences could

underlie some or all of the variation in longevity that distinguishes High and Low worms (Seewald et al., 2010). Given that our previous HSF-1 overexpression and hormesis experiments suggested some level of inter-individual variability in genetic buffering we decided to test that directly by sorting High and Low inducers of *hsp-16.2* reporter after a heat shock and then quantifying the penetrance of mutations in both groups. The penetrance of mutations was found to be lower in the High *hsp-16.2* reporter group compared to the Low inducers (Figure 2A and 2B). This was true for the four post–embryonic mutations (*lin-31(n1053)*, *lin-29(ga94)*, *mab-19(bx83)* and *vab-9(ju6)*) that we have previously shown were rescued by HSF-1 overexpression and hormesis treatment. Furthermore, the same trend was apparent when the fluorescence was measured in an individual basis (Figure 2C and Sup. Table 6). So, based on our results and work of others we can conclude that upon a heat shock treatment, there are inter-individual differences in the induction of the heat shock response as indicated by a *hsp-16.2* reporter. Those worms that induce higher levels of the response are longer lived, more resistant to stress and more resistant to genetic perturbations than worms that induce lower levels. However, given that the hormesis treatment *per se* is sufficient to increase the life span, thermotolerance and decrease the penetrance of mutations, it is not clear whether variability of these traits in un-stressed animals has

the same origin. Some experiments trying to address this issue will be discussed later. Another important issue relates to the reason for this variability. If those worms that induce higher levels of *hsp-16.2* reporter live longer, are more thermotolerant and more resistant to certain mutations, why are all not worms high inducers? Is there any price that the high *hsp-16.2* inducers are paying?

The insulin signaling pathway and phenotypic trade-offs

daf-2 mutants are long lived and resist stress better than their wild type counterparts (Kenyon et al., 1993). DAF-2 is the sole receptor ortholog in *C. elegans* of the Insulin/IGF-1-like (IIS) pathway. The Insulin pathway is well conserved in worms, flies and mammals. For instance, an inverse correlation between IGF-1 levels and lifespan has been described among inbred mice strains, implicating IGF-1 in lifespan regulation in mammals (Yuan et al., 2009). In addition, mutations that inhibit the insulin receptor (specifically in adipose tissue)(Bluher et al., 2003) and the IGF-1 receptor (Kappeler et al., 2008) can extend mice lifespan (Kenyon, 2010). The life span extension phenotype of *daf-2* mutants is completely dependant on the gene *daf-16* (Kenyon et al., 1993; Ogg et al., 1997). DAF-16 is a transcription factor member of the FOXO family and is closely related to the mammalian FOX3a. Worms growing

in the presence of food and without any kind of stress activate the DAF-2 receptor by binding of an insulin-like agonist ligand. This initiates a phosphorylation cascade which ultimate leads to the inhibition of DAF-16 activity. Activated DAF-2 phosphorylates AGE-1, a phosphoinositide 3-kinase, which recruits the AKT-1, AKT-2 and SGK-1 kinases to the plasma membrane where PDK-1 phosphorylates AKT and SGK-1. The activated AKT-1/AKT-2/SGK-1 complex phosphorylates DAF-16. Phosphorylated DAF-16 is sequestered in the cytoplasm and prevented from activating the transcription of its target genes (Landis and Murphy, 2010). *daf-2* mutants show nuclear localization of DAF-16 under normal growing conditions in contrast to their wild type counterparts. In addition to receiving signals from the IIS pathway, there is increasing evidence that DAF-16 also integrates multiple other pathways. DAF-16's activity is crucial for many processes including lifespan, development, stress resistance, thermotolerance, pathogen response, metabolism and autophagy, probably through the direct and indirect transcriptional control of hundreds of genes (Kenyon, 2010).

In addition to the benefits that mutations in the insulin pathway confer on individuals, such as increased lifespan and stress resistance, some studies have provided evidence that there is an inherent fitness cost in these mutations. The first evidence came from studies on *age-1*

mutants. AGE-1 is a phosphatidylinositol 3-OH kinase catalytic subunit component of the insulin-like signalling pathway and mutants on this gene can extend adult lifespan by up to 80% (Friedman and Johnson, 1988). If equal number of wild type and *age-1(hx546)* mutants worms are allowed to grow in the same plates and the populations go through several cycles of starvation, the mutant *age-1* worms are outcompeted (Walker et al., 2000). This competition experiment suggests the existence of a fitness cost in long-lived worms. Under normal growth conditions (without starvation cycles), no differences in fitness were detected. Interestingly, similar competition experiments showed that longed lived *daf-2(e1368)* mutants had an associated fitness cost, both in regimes of constant food and starvation cycles (Jenkins et al., 2004). Though the developmental time of *daf-2(e1368)* mutants was similar to that of the wild type, fertility was found to be affected. Under conditions rich in food, wild-type *C. elegans* hermaphrodites begin to produce progeny soon after reaching adulthood and deplete endogenous sperm reserves within 3–4 days. *daf-2* mutants have a reduced early fertility compared to the wild type (adults lay fewer eggs during the first days) (Jenkins et al., 2004). Also, increased dosage of DAF-16 that makes worms longer-lived and stress resistant causes worms to reproduce slower. This phenotype, can be reversed by inhibition of *daf-16* activity using RNAi (Henderson and Johnson, 2001).

Furthermore, DAF-16 depletion from wild type by means of *daf-16* RNAi causes worms to reproduce faster, suggesting that DAF-16 activity is influencing the rate of development in the wild type (Henderson and Johnson, 2001). Altogether, these studies suggest that there is a trade-off between lifespan/stress resistance and the rate of reproduction in *C. elegans* and that the Insulin/IGF-1-like (IIS) pathway is directly involved, though the mechanism behind this trade-off is not yet clear. Is this trade-off a consequence of the regulatory wiring of the network regulating these processes? Or is it an unavoidable dichotomy of resource allocation between the soma and the germline?

We decided to study if high induction of *hsp-16.2* has an associated fitness cost, like the one of mutants in the insulin pathway. High inducers of *hsp-16.2* reporter after a heat shock a reduced early fecundity compared to Low inducers (Sup. Figure 3D and Sup. Table 9). Thus, we describe a fitness cost associated to the higher induction of the heat shock response pathway, which is stochastically variable in an isogenic population. This diversification of phenotypes in an unpredictable environment could be beneficial for the population. In evolutionary ecology, risk spreading or “bet-hedging” is the idea that unpredictably variable environments favor genotypes with lower variance in fitness (between generations) at the cost of lower

arithmetic mean fitness (Cohen, 1968; Hopper, 1999). Bet-hedging can be achieved by a single phenotype that avoid risks or by phenotypic variation expressed by a single genotype. Most of studies about bet-hedging have been performed in insects but in those cases it has proven very difficult to distinguish between the effects of genetic and non-genetic variation (Hopper, 1999). It has been previously shown that the induction levels of *hsp-16.2* reporter are not inheritable (Rea et al., 2005) and given the isogenicity of C. elegans worm strains, we can almost certainly discard genetic differences. How could this bet-hedging strategy work? For instance, following a first heat stress, a subpopulation will be more resistant to a second stress (high *hsp-16.2* inducers) and fitter if such event occurs. But if the second stress does not occur, the second subpopulation (low *hsp-16.2* inducers) will be fitter, reproducing faster and thus making more likely the survival of the population as a whole. We have not directly shown that this is a bet-hedging strategy, but we have described a stochastic phenotypic trade-off upon environmental stress that is consistent with such a mechanism. If natural selection has ultimately selected for this variation is still matter of debate. In order to gain more insights into this proposed strategy, we would need to find a genotype whose inter-individual variation in the heat shock response is much lower than the current wild type thus characterized by a lower geometric mean fitness and a higher arithmetic

mean fitness. Direct competition between the non-risky and the bet-hedging strain should give as a winner the non-risky under a predictable environment and the bet-hedging strain in a fluctuating environment. Also, the bet-hedging mechanism should evolve from the non-risky under a selection regime of unpredictable stress, in analogy to recent experiments performed in bacteria (Beaumont et al., 2009).

What could be mediating such differences in stress response among individuals? In order to gain some mechanistic insight, we made use of a transgenic line developed by Johnson and colleagues, which carries extra copies of a DAF-16::GFP fusion protein under the control of *daf-16* promoter to study its dynamics in individuals (Henderson and Johnson, 2001). After a heat shock treatment, it is possible to follow the transient translocation of DAF-16::GFP into the nucleus and its later export into the cytoplasm. In addition to heat shock, several other environmental stresses, such as starvation and oxidative stress, have been shown to trigger rapid nuclear localization of DAF-16 (Herrmann et al., 1990). DAF-16 translocation into the nucleus after a heat shock is rapid, but we could observe large variation in the nuclear residency time of DAF-16 among individual worms. Interestingly, sorting worms into “long” and “short” groups relative to their DAF-16 nuclear residency time revealed physiological

differences between the two groups. “Long residence“ DAF-16 worms had increased thermotolerance and also showed reduced early fecundity compared to the “short” residency group, supporting the existence of a trade-off (Fig 3). Further quantitative real time PCR confirmed that there were differences in the induction of heat shock response genes between the two groups (Sup. Figure S4). The strain used in these experiments has extra functional copies of DAF-16 integrated in its genome, so it doesn’t resemble exactly the wild type physiology (Henderson and Johnson, 2001). Unfortunately, there is currently no experimental option what would allow the differences between long and short DAF-16 residence groups in the wild type to be tested. However, the fact that this trade-off is also observed in the *hsp-16.2* reporter strain after a heat shock indicates that the observed differences are not likely to be a DAF-16 overexpression artifact. We don’t know what determines the inter-individual variation in the DAF-16 nuclear time of residence. Is this variation correlated with the *hsp-16.2* reporter variation? A worm strain carrying the *DAF-16::GFP* transgene and *hsp-16.2::mCherry* reporter could be used to test this hypothesis. However, given that DAF-16 has been shown to cooperate with HSF-1 in the transcription of several heat shock proteins, a positive correlation between DAF-16 long residence and higher hsp-16.2 expression could be an artifact of the DAF-16 overexpression strain and not necessarily represent the wild type situation (Hsu et

al., 2003). It has been described that *hsf-1* RNAi affects the export of DAF-16 from the nucleus after a heat shock treatment, delaying it considerably. This effect seems to be largely dependant on the expression of *hsp-1* (Singh and Aballay, 2009). *Hsp-1* is expressed constitutively, but up-regulated after a heat shock, thus implying that some feed-back mechanism may be acting to avoid the over-activation of the DAF-16 response.

Intrinsic variation of chaperones affects the outcome of mutations

The fact that the hormetic heat shock treatment necessary to visualize the inter-individual variability in the heat shock response also causes a decrease in the mean penetrance made difficult to infer if chaperones had a role in the penetrance of mutations in unstressed conditions. In order to test such a possibility, we used a bacterial feeding RNAi screen to test if down-regulation of chaperones had an effect on the penetrance of mutations (Sup. Figure 5). We tested the major transcription factors involved in the stress response (*hsf-1, daf-16, skn-1*), the major chaperones (*hsp-1, daf-21, hsp-6, hsp-12, hsp-43, hsp-16.1*) and the members of Dnj family of chaperones (*dnj-7, dnj-12, dnj-13, dnj-15, dnj-19, dnj-28, dnj-29 and rme-8*). The screened revealed that the different mutants tested (*vab-1(e2)*, *ifb-1(ju71)* , *lin-29(ga94)*, *lin-31(n1053)* , *vab-*

9(ju6) and *vab-3(ju468)*) varied in their dependence on chaperone activity (Sup Table 13-20). For instance, *lin-31(n1053)* penetrance was increased in *daf-21* RNAi worms but was non affected by *dnj-12* RNAi. In contrast, *lin-29(ga94)* mutant's penetrance was increase by dnj-12 RNAi and not by daf-21 RNAi treatment (Sup. Figure 5). These results indicate that chaperones are buffering mutations in unstressed worms and moreover, that different mutations are buffered by different chaperones. Future work could try to understand the reasons for these differences, but factors such as chaperone-client specificity and expression patterns are likely playing a role.

We generated a transcriptional reporter for *daf-21*, the worm ortholog of HSP90 (see Discussion of Burga et al. for strain characterization) and we could also show that stochastic variation in the *daf-21* reporter predicted the penetrance of *lin-31(n1053)* mutants as suggested by the RNAi screen. *lin-31(n1053)* worms with higher expression of the *daf-21* reporter had a lower penetrance compared to the group of lower expressors. In contrast, the penetrance of neither *vab-9(ju6)* nor *lin-29(ga94)* mutants was correlated with the levels of the *daf-21* reporter, in accordance with the RNAi screen results. So, together with other results (see Burga et al.), these results strongly suggest that chaperones levels are variable among non-stressed individuals and that this variation influences the

outcome of mutations in chaperone-mutation specific way. Individuals with higher levels of *daf-21* reporter expression before a hormesis treatment, where later more resist to a second heat shock treatment than the low expressors and also had a reduced early fecundity (Sup. Figure 5). This supports the idea that there is some level of intrinsic variability in chaperones with important phenotypic consequences. We could not directly test if the expression levels of the *daf-21* reporter were predictive of the variability in the *hsp-16.2* reporter, because unfortunately both transgenes are integrated in a very close region of chromosome IV (A.B and O.C unpublished observations). We cannot, therefore, still rule out the possibility that the trade-off incurred after the hormesis is predicted by *daf-21* and *hsp-16.2* independently.

Conclusions

The main conclusions of our work that I present in this thesis are:

1. We have detected stochastic variability in the expression of genes among isogenic individual *C. elegans* worms growing in a homogenous environment
2. Variability in the induction of redundant gene pairs originating from gene duplication events accounts for part of the phenotypic variability among individuals carrying null mutations in the other pair (incomplete penetrance).
3. Variability in the expression of constitutively expressed chaperones such as DAF-21 (HSP90) can also account for part of this phenotypic variability.
4. Variability in both buffering pathways, one gene-specific (gene duplication) and the other more promiscuous (chaperones), independently predicts the variable phenotypic outcome of null mutants.
5. Our conceptual model based on the stochastic gene expression of genetically interacting partners has proven useful to understand the causes of the incomplete penetrance of mutations.

6. Robustness to genetic and environmental perturbations are coupled in *C. elegans*. Those individuals that stochastically induce a higher response are more protected to genetic mutations and subsequent environmental stresses.
7. We have found a trade-off associated to this robustness, such that individuals that are more protected against mutations and stress incur a fitness cost characterized by delay in early fecundity.
8. Mutations interact genetically with several constitutively expressed stress-response genes, however the nature of these interactions are mutation and chaperone specific.

Bibliography

Acar, M., Mettetal, J.T., and van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. Nat Genet *40*, 471-475.

Andachi, Y. (2004). Caenorhabditis elegans T-box genes tbx-9 and tbx-8 are required for formation of hypodermis and body-wall muscle in embryogenesis. Genes Cells *9*, 331-344.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223-230.

Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., Francois, J., and Yvert, G. (2008). Cell-to-cell stochastic variation in gene expression is a complex genetic trait. PLoS Genet *4*, e1000049.

Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics *149*, 1633-1648.

Balazsi, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. Cell *144*, 910-925.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., and Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. Nat Genet *38*, 636-643.

Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks. Nature *387*, 913-917.

Batada, N.N., and Hurst, L.D. (2007). Evolution of chromosome organization driven by selection for reduced gene expression noise. Nat Genet *39*, 945-949.

Baugh, L.R., Wen, J.C., Hill, A.A., Slonim, D.K., Brown, E.L., and Hunter, C.P. (2005). Synthetic lethal analysis of Caenorhabditis elegans posterior embryonic patterning genes identifies conserved genetic interactions. Genome Biol *6*, R45.

Beaumont, H.J., Gallie, J., Kost, C., Ferguson, G.C., and Rainey, P.B. (2009). Experimental evolution of bet hedging. Nature *462*, 90-93.

Becskei, A., Seraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J *20*, 2528-2535.

Becskei, A., and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. Nature *405*, 590-593.

Belle, A., Tanay, A., Bitincka, L., Shamir, R., and O'Shea, E.K. (2006). Quantification of protein half-lives in the budding yeast proteome. Proc Natl Acad Sci U S A *103*, 13004-13009.

Berger, A.H., Knudson, A.G., and Pandolfi, P.P. (2011). A continuum model for tumour suppression. Nature *476*, 163-169.

Blake, W.J., M, K.A., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. Nature *422*, 633-637.

Bluher, M., Kahn, B.B., and Kahn, C.R. (2003). Extended longevity in mice lacking the insulin receptor in adipose tissue. Science *299*, 572-574.

Boulias, K., and Horvitz, H.R. (2012). The C. elegans MicroRNA mir-71 Acts in Neurons to Promote Germline-Mediated Longevity through Regulation of DAF-16/FOXO. Cell Metab *15*, 439-450.

Bowerman, B., Eaton, B.A., and Priess, J.R. (1992). skn-1, a maternally expressed gene required to specify the fate of

ventral blastomeres in the early C. elegans embryo. Cell *68*, 1061-1075.

Braendle, C., and Felix, M.A. (2008). Plasticity and errors of a robust developmental system in different environments. Dev Cell *15*, 714-724.

Brookfield, J. (1992). Can genes be truly redundant? Curr Biol *2*, 553-554.

Cagatay, T., Turcotte, M., Elowitz, M.B., Garcia-Ojalvo, J., and Suel, G.M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell *139*, 512-522.

Canham, M.A., Sharov, A.A., Ko, M.S., and Brickman, J.M. (2010). Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biol *8*, e1000379.

Cardno, A.G., and Gottesman, II (2000). Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. Am J Med Genet *97*, 12-17.

Casanueva, M.O., Burga, A., and Lehner, B. (2012). Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. Science *335*, 82-85.

Cavenee, W.K., Hansen, M.F., Nordenskjold, M., Kock, E., Maumenee, I., Squire, J.A., Phillips, R.A., and Gallie, B.L. (1985). Genetic origin of mutations predisposing to retinoblastoma. Science *228*, 501-503.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544-547.

Chen, S., and Parmigiani, G. (2007). Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol *25*, 1329-1333.

Cohen, D. (1968). A General Model of Optimal Reproduction in a Randomly Varying Environment. Journal of Ecology *56*, 219-228.

Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C.G., and Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. Nature *437*, 699-706.

Corish, P., and Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng *12*, 1035-1040.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S.*, et al.* (2010). The genetic landscape of a cell. Science *327*, 425-431.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Cypser, J.R., and Johnson, T.E. (2002). Multiple stressors in Caenorhabditis elegans induce stress hormesis and extended longevity. J Gerontol A Biol Sci Med Sci *57*, B109-114.

de Lencastre, A., Pincus, Z., Zhou, K., Kato, M., Lee, S.S., and Slack, F.J. (2010). MicroRNAs both promote and antagonize longevity in C. elegans. Curr Biol *20*, 2159-2168.

DeLuna, A., Springer, M., Kirschner, M.W., and Kishony, R. (2010). Need-based up-regulation of protein levels in

response to deletion of their duplicate genes. PLoS Biol *8*, e1000347.

Denver, D.R., Morris, K., Lynch, M., and Thomas, W.K. (2004). High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Nature *430*, 679-682.

Dietrich, J.E., and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. Development *134*, 4219-4231.

Ehrenreich, I.M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J.A., Gresham, D., Caudy, A.A., and Kruglyak, L. (2010). Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature *464*, 1039-1042.

Eldar, A., Chary, V.K., Xenopoulos, P., Fontes, M.E., Loson, O.C., Dworkin, J., Piggot, P.J., and Elowitz, M.B. (2009). Partial penetrance facilitates developmental evolution in bacteria. Nature *460*, 510-514.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Errington, J. (2003). Regulation of endospore formation in Bacillus subtilis. Nat Rev Microbiol *1*, 117-126.

Felix, M.A., and Barkoulas, M. (2012). Robustness and flexibility in nematode vulva development. Trends Genet *28*, 185-195.

Ferrell, J.E., Jr., and Machleder, E.M. (1998). The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. Science *280*, 895-898.

Friedman, D.B., and Johnson, T.E. (1988). Three mutants that extend both mean and maximum life span of the nematode, Caenorhabditis elegans, define the age-1 gene. J Gerontol *43*, B102-109.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M., and Dryja, T.P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature *323*, 643-646.

Frokjaer-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nat Genet *40*, 1375-1383.

Gandhi, S.J., Zenklusen, D., Lionnet, T., and Singer, R.H. (2011). Transcription of functionally related constitutive genes is not coordinated. Nat Struct Mol Biol *18*, 27-34.

Gartner, K. (1990). A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? Lab Anim *24*, 71-77.

Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W., and Li, W.H. (2003). Role of duplicate genes in genetic robustness against null mutations. Nature *421*, 63-66.

Hall, P.A., and Watt, F.M. (1989). Stem cells: the generation and maintenance of cellular diversity. Development *106*, 619-633.

Hamilton, A.S., and Mack, T.M. (2003). Puberty and genetic susceptibility to breast cancer in a case-control study in twins. N Engl J Med *348*, 2313-2322.

Hartl, F.U. (1996). Molecular chaperones in cellular protein folding. Nature *381*, 571-579.

Hartl, F.U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol *16*, 574-581.

Hartman, J.L.t., Garvik, B., and Hartwell, L. (2001). Principles for the buffering of genetic variation. Science *291*, 1001-1004.

Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell *3*, 391-401.

Henderson, S.T., and Johnson, T.E. (2001). daf-16 integrates developmental and environmental inputs to mediate aging in the nematode Caenorhabditis elegans. Curr Biol *11*, 1975-1980.

Herrmann, B.G., Labeit, S., Poustka, A., King, T.R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. Nature *343*, 617-622.

Hooshangi, S., Thiberge, S., and Weiss, R. (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. Proc Natl Acad Sci U S A *102*, 3581-3586.

Hopper, K.R. (1999). Risk-spreading and bet-hedging in insect population biology. Annu Rev Entomol *44*, 535-560.

Horvitz, H.R., and Sulston, J.E. (1980). Isolation and genetic characterization of cell-lineage mutants of the nematode Caenorhabditis elegans. Genetics *96*, 435-454.

Hsin, H., and Kenyon, C. (1999). Signals from the reproductive system regulate the lifespan of C. elegans. Nature *399*, 362-366.

Hsu, A.L., Murphy, C.T., and Kenyon, C. (2003). Regulation of aging and age-related disease by DAF-16 and heat-shock factor. Science *300*, 1142-1145.

Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. Development *136*, 3853-3862.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., *et al.* (2000). Functional discovery via a compendium of expression profiles. Cell *102*, 109-126.

Inoue, T., Takamura, K., Yamae, H., Ise, N., Kawakami, M., Tabuse, Y., Miwa, J., and Yamaguchi, Y. (2003). Caenorhabditis elegans DAF-21 (HSP90) is characteristically and predominantly expressed in germline cells: spatial and temporal analysis. Dev Growth Differ *45*, 369-376.

Jedrusik, A., Parfitt, D.E., Guo, G., Skamagki, M., Grabarek, J.B., Johnson, M.H., Robson, P., and Zernicka-Goetz, M. (2008). Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. Genes Dev *22*, 2692-2706.

Jelier, R., Semple, J.I., Garcia-Verdugo, R., and Lehner, B. (2011). Predicting phenotypic variation in yeast from individual genome sequences. Nat Genet *43*, 1270-1274.

Jenkins, N.L., McColl, G., and Lithgow, G.J. (2004). Fitness cost of extended lifespan in Caenorhabditis elegans. Proc Biol Sci *271*, 2523-2526.

Jimenez-Gomez, J.M., Corwin, J.A., Joseph, B., Maloof, J.N., and Kliebenstein, D.J. (2011). Genomic analysis of QTLs and genes altering natural variation in stochastic noise. PLoS Genet *7*, e1002295.

Johannsen, W. (1909). Elemente der exakten Erblichkeitslehre (Jena, Gustav Fischer.).

Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet *6*, 451-464.

Kafri, R., Bar-Even, A., and Pilpel, Y. (2005). Transcription control reprogramming in genetic backup circuits. Nat Genet *37*, 295-299.

Kalmar, T., Lim, C., Hayward, P., Munoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol *7*, e1000149.

Kappeler, L., De Magalhaes Filho, C., Dupont, J., Leneuve, P., Cervera, P., Perin, L., Loudes, C., Blaise, A., Klein, R., Epelbaum, J.*, et al.* (2008). Brain IGF-1 receptors control mammalian growth and lifespan through a neuroendocrine mechanism. PLoS Biol *6*, e254.

Kathiresan, S., Voight, B.F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P.M., Anand, S., Engert, J.C., Samani, N.J., Schunkert, H.*, et al.* (2009). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet *41*, 334-341.

Kelly, W.G., Xu, S., Montgomery, M.K., and Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics *146*, 227-238.

Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A C. elegans mutant that lives twice as long as wild type. Nature *366*, 461-464.

Kenyon, C.J. (2010). The genetics of ageing. Nature *464*, 504-512.

Khvorova, A., Chary, V.K., Hilbert, D.W., and Piggot, P.J. (2000). The chromosomal location of the Bacillus subtilis sporulation gene spoIIR is important for its function. J Bacteriol *182*, 4425-4429.

Knudson, A.G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A *68*, 820-823.

Kuusisto, H., Kaprio, J., Kinnunen, E., Luukkaala, T., Koskenvuo, M., and Elovaara, I. (2008). Concordance and heritability of multiple sclerosis in Finland: study on a nationwide series of twins. Eur J Neurol *15*, 1106-1110.

Landis, J.N., and Murphy, C.T. (2010). Integration of diverse inputs in the regulation of Caenorhabditis elegans DAF-16/FOXO. Dev Dyn *239*, 1405-1412.

Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J., and Hartl, D.L. (2007). Genetic properties influencing the evolvability of gene expression. Science *317*, 118-121.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet *44*, 247-250.

Lee, W.H., Bookstein, R., Hong, F., Young, L.J., Shew, J.Y., and Lee, E.Y. (1987). Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science *235*, 1394-1399.

Lehner, B. (2007). Modelling genotype-phenotype relationships and human disease with genetic interaction networks. J Exp Biol *210*, 1559-1566.

Lehner, B. (2010). Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. PLoS One *5*, e9035.

Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. Trends Genet *27*, 323-331.

Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A.G. (2006). Systematic mapping of genetic

interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. Nat Genet *38*, 896-903.

Li, G.W., and Xie, X.S. (2011). Central dogma at the single-molecule level in living cells. Nature *475*, 308-315.

Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med *343*, 78-85.

Lipinski, K.J., Farslow, J.C., Fitzpatrick, K.A., Lynch, M., Katju, V., and Bergthorsson, U. (2011). High spontaneous rate of gene duplication in Caenorhabditis elegans. Curr Biol *21*, 306-310.

Lisman, J.E. (1985). A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. Proc Natl Acad Sci U S A *82*, 3055-3057.

Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in Bacillus subtilis. Science *317*, 526-529.

Maier, T., Guell, M., and Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. FEBS Lett *583*, 3966-3973.

Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M.J., and Seddon, J.M. (2006). Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nat Genet *38*, 1055-1059.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A.*, et al.* (2009). Finding the

missing heritability of complex diseases. Nature *461*, 747-753.

McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A *94*, 814-819.

McAdams, H.H., and Arkin, A. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet *15*, 65-69.

McClellan, J., and King, M.C. (2010). Genetic heterogeneity in human disease. Cell *141*, 210-217.

Mello, C.C., Kramer, J.M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. EMBO J *10*, 3959-3970.

Mendenhall, A.R., Tedesco, P.M., Taylor, L.D., Lowe, A., Cypser, J.R., and Johnson, T.E. (2012). Expression of a Single-Copy hsp-16.2 Reporter Predicts Life span. J Gerontol A Biol Sci Med Sci.

Morimoto, R.I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev *12*, 3788-3796.

Muller, C.W., and Herrmann, B.G. (1997). Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. Nature *389*, 884-888.

Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol Biol Cell *22*, 2787-2795.

Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. Science *336*, 183-187.

Nachman, M.W., and Crowell, S.L. (2000). Estimate of the mutation rate per nucleotide in humans. Genetics *156*, 297-304.

Naiche, L.A., Harrelson, Z., Kelly, R.G., and Papaioannou, V.E. (2005). T-box genes in vertebrate development. Annu Rev Genet *39*, 219-239.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840-846.

Nollen, E.A., and Morimoto, R.I. (2002). Chaperoning signaling pathways: molecular chaperones as stress-sensing 'heat shock' proteins. J Cell Sci *115*, 2809-2816.

Nora, J.J., Lortscher, R.H., Spangler, R.D., Nora, A.H., and Kimberling, W.J. (1980). Genetic--epidemiologic study of early-onset ischemic heart disease. Circulation *61*, 503-508.

Novick, A., and Weiner, M. (1957). Enzyme Induction as an All-or-None Phenomenon. Proc Natl Acad Sci U S A *43*, 553-566.

Nowak, M.A., Boerlijst, M.C., Cooke, J., and Smith, J.M. (1997). Evolution of genetic redundancy. Nature *388*, 167-171.

Offit, K. (2006). BRCA mutation frequency and penetrance: new data, old debate. J Natl Cancer Inst *98*, 1675-1677.

Ogg, S., Paradis, S., Gottlieb, S., Patterson, G.I., Lee, L., Tissenbaum, H.A., and Ruvkun, G. (1997). The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. Nature *389*, 994-999.

Ohno, S. (1970). Evolution by gene duplication (New York, Springer Verlag).

Ow, M.C., Martinez, N.J., Olsen, P.H., Silverman, H.S., Barrasa, M.I., Conradt, B., Walhout, A.J., and Ambros, V. (2008). The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microRNA genes in C. elegans. Genes Dev *22*, 2520-2534.

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. Nat Genet *31*, 69-73.

Pearl, L.H., and Prodromou, C. (2006). Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu Rev Biochem *75*, 271-294.

Pedraza, J.M., and van Oudenaarden, A. (2005). Noise propagation in gene networks. Science *307*, 1965-1969.

Pereira, R., Halford, K., Sokolov, B.P., Khillan, J.S., and Prockop, D.J. (1994). Phenotypic variability and incomplete penetrance of spontaneous fractures in an inbred strain of transgenic mice expressing a mutated collagen gene (COL1A1). J Clin Invest *93*, 1765-1769.

Pincus, Z., Smith-Vikos, T., and Slack, F.J. (2011). MicroRNA predictors of longevity in Caenorhabditis elegans. PLoS Genet *7*, e1002306.

Plachta, N., Bollenbach, T., Pease, S., Fraser, S.E., and Pantazis, P. (2011). Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. Nat Cell Biol *13*, 117-123.

Pocock, R., Ahringer, J., Mitsch, M., Maxwell, S., and Woollard, A. (2004). A regulatory network of T-box genes and the even-skipped homologue vab-7 controls patterning and morphogenesis in C. elegans. Development *131*, 2373-2385.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009).

Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748-752.

Queitsch, C., Sangster, T.A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. Nature *417*, 618-624.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol *4*, e309.

Raj, A., Rifkin, S.A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. Nature *463*, 913-918.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811-1814.

Rea, S.L., Wu, D., Cypser, J.R., Vaupel, J.W., and Johnson, T.E. (2005). A stress-sensitive reporter predicts longevity in isogenic populations of Caenorhabditis elegans. Nat Genet *37*, 894-898.

Rockman, M.V. (2008). Reverse engineering the genotype-phenotype map with natural genetic variation. Nature *456*, 738-744.

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., and Elowitz, M.B. (2005). Gene regulation at the single-cell level. Science *307*, 1962-1965.

Rutherford, S.L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. Nature *396*, 336-342.

Sanchez-Blanco, A., and Kim, S.K. (2011). Variable pathogenicity determines individual lifespan in Caenorhabditis elegans. PLoS Genet *7*, e1002047.

Schrödinger, E. (1945). What is life? : the physical aspect of the living cell (Cambridge [Eng], University Press).

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Seewald, A.K., Cypser, J., Mendenhall, A., and Johnson, T. (2010). Quantifying phenotypic variation in isogenic Caenorhabditis elegans expressing Phsp-16.2::gfp by clustering 2D expression patterns. PLoS One *5*, e11426.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet *31*, 64-68.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. (2006). Variability and memory of protein levels in human cells. Nature *444*, 643-646.

Simons, B.D., and Clevers, H. (2011). Strategies for homeostatic stem cell self-renewal in adult tissues. Cell *145*, 851-862.

Sippola, M., Kaffe, S., and Prockop, D.J. (1984). A heterozygous defect for structurally altered pro-alpha 2 chain of type I procollagen in a mild variant of osteogenesis imperfecta. The altered structure decreases the thermal stability of procollagen and makes it resistant to procollagen N-proteinase. J Biol Chem *259*, 14094-14100.

Smith, G.D. (2011). Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. Int J Epidemiol *40*, 537-562.

Soffritti, M., Belpoggi, F., Degli Esposti, D., Lambertini, L., Tibaldi, E., and Rigano, A. (2006). First experimental demonstration of the multipotential carcinogenic effects of aspartame administered in the feed to Sprague-Dawley rats. Environ Health Perspect *114*, 379-385.

Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428-432.

Spudich, J.L., and Koshland, D.E., Jr. (1976). Non-genetic individuality: chance in the single cell. Nature *262*, 467-471.

Sternberg, P.W., and Felix, M.A. (1997). Evolution of cell lineage. Curr Opin Genet Dev *7*, 543-550.

Stewart-Ornstein, J., Weissman, J.S., and El-Samad, H. (2012). Cellular Noise Regulons Underlie Fluctuations in Saccharomyces cerevisiae. Mol Cell *45*, 483-493.

Suel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J., and Elowitz, M.B. (2007). Tunability and noise dependence in differentiation dynamics. Science *315*, 1716-1719.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533-538.

Tatar, M., Khazaeli, A.A., and Curtsinger, J.W. (1997). Chaperoning extended life. Nature *390*, 30.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J.*, et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707-713.

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A *98*, 8614-8619.

Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875-889.

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M.*, et al.* (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808-813.

Topalidou, I., van Oudenaarden, A., and Chalfie, M. (2011). Caenorhabditis elegans aristaless/Arx gene alr-1 restricts variable gene expression. Proc Natl Acad Sci U S A *108*, 4063-4068.

Vavouri, T., Semple, J.I., and Lehner, B. (2008). Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. Trends Genet *24*, 485-488.

Visscher, P.M., Hill, W.G., and Wray, N.R. (2008). Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet *9*, 255-266.

Waddington, C.H. (1942). Canalization of Development and the Inheritance of Acquired Characters. Nature *150*, 563-565.

Waddington, C.H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology (London, Allen & Unwin).

Wagner, A. (1996). Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. Biol Cybern *74*, 557-567.

Wagner, A. (2000a). Robustness against mutations in genetic networks of yeast. Nat Genet *24*, 355-361.

Wagner, A. (2000b). The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. Genetics *154*, 1389-1401.

Walker, D.W., McColl, G., Jenkins, N.L., Harris, J., and Lithgow, G.J. (2000). Evolution of lifespan in C. elegans. Nature *405*, 296-297.

Wang, Y., and Dohlman, H.G. (2004). Pheromone signaling mechanisms in yeast: a prototypical sex machine. Science *306*, 1508-1509.

Wernet, M.F., Mazzoni, E.O., Celik, A., Duncan, D.M., Duncan, I., and Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. Nature *440*, 174-180.

Wilkinson, H.A., Fitzgerald, K., and Greenwald, I. (1994). Reciprocal changes in expression of the receptor lin-12 and its ligand lag-2 prior to commitment in a C. elegans cell fate decision. Cell *79*, 1187-1198.

Wilm, T., Demel, P., Koop, H.U., Schnabel, H., and Schnabel, R. (1999). Ballistic transformation of Caenorhabditis elegans. Gene *229*, 31-35.

Wirdefeldt, K., Gatz, M., Reynolds, C.A., Prescott, C.A., and Pedersen, N.L. (2011). Heritability of Parkinson disease in Swedish twins: a longitudinal study. Neurobiol Aging *32*, 1923 e1921-1928.

Wong, A.H., Gottesman, II, and Petronis, A. (2005). Phenotypic differences in genetically identical organisms: the epigenetic perspective. Hum Mol Genet *14 Spec No 1*, R11-18.

Wong, S.L., and Roth, F.P. (2005). Transcriptional compensation for gene loss plays a minor role in maintaining genetic robustness in Saccharomyces cerevisiae. Genetics *171*, 829-833.

Xie, X.S., Choi, P.J., Li, G.W., Lee, N.K., and Lia, G. (2008). Single-molecule approach to molecular biology in living bacterial cells. Annu Rev Biophys *37*, 417-444.

Xiong, W., and Ferrell, J.E., Jr. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature *426*, 460-465.

Yokoyama, K., Fukumoto, K., Murakami, T., Harada, S., Hosono, R., Wadhwa, R., Mitsui, Y., and Ohkuma, S. (2002). Extended longevity of Caenorhabditis elegans by knocking in extra copies of hsp70F, a homolog of mot-2 (mortalin)/mthsp70/Grp75. FEBS Lett *516*, 53-57.

Yuan, R., Tsaih, S.W., Petkova, S.B., Marin de Evsikova, C., Xing, S., Marion, M.A., Bogue, M.A., Mills, K.D., Peters, L.L., Bult, C.J.*, et al.* (2009). Aging in inbred strains of mice: study design and interim report on median lifespans and circulating IGF1 levels. Aging Cell *8*, 277-287.

Zeng, L., Skinner, S.O., Zong, C., Sippy, J., Feiss, M., and Golding, I. (2010). Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell *141*, 682-691.

Zernicka-Goetz, M., and Huang, S. (2010). Stochasticity versus determinism in development: a false dichotomy? Nat Rev Genet *11*, 743-744.

Zernicka-Goetz, M., Morris, S.A., and Bruce, A.W. (2009). Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. Nat Rev Genet *10*, 467-477.