

Effects of Life History on the Transcriptomic Phenotypes of Two Genotypes of *C. elegans*

David Angeles-Albores*, †, Daniel H Leighton*, ‡, Tiffany Tsou†, Tiffany K. Shaw†, Igor Antoshechkin§ and Paul W Sternberg†, 1

*Co-First Authors, †Department of Biology and Biological Engineering, and Howard Hughes Medical Institute, Caltech, Pasadena, CA, 91125, USA.,

‡Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA., §Department of Biology and Biological Engineering, Caltech, Pasadena, CA, 91125, USA

ABSTRACT Aging is a complex phenomenon that is likely to be affected by many different variables. Of key interest is the effect that life stresses may have in affecting the aging process. Here, we used *C. elegans* to study the effects of life history on global gene expression. We used a *fog-2* mutant, which is sperm-deficient and cannot self-fertilize as a result, to study the differential effects of reproduction on aging. We extracted RNA from wild-type hermaphrodites and *fog-2* females at the young and old adult stage and fit a generalized linear model to identify genes that are affected by life history. Using this method, we were able to define sets of genes that define the *fog-2* phenotype, genes that are commonly affected in aging between both genotypes and genes that change differently in either genotype. We performed three RNAi screens to identify novel genes associated with changes in brood size, lawn-leaving behaviour and ovulation. We identified XXX genes that led to altered brood size, 9 genes that altered lawn-leaving in N2 and 9 genes that increased lawn-leaving in *fog-2*.

KEYWORDS

C. elegans

fog-2

aging

transcriptome

INTRODUCTION

MATERIALS AND METHODS

RNA extraction

Synchronized worms were grown to either young adulthood or the 6th day of adulthood prior to RNA extraction. Synchronization and aging were carried out according to protocols described previously (1). 1000–5000 worms from each replicate were rinsed into a microcentrifuge tube in S basal (5.85g/L NaCl, 1g/L K₂HPO₄, 6g/L KH₂PO₄), and then spun at 14,000rpm for 30s. The supernatant was removed and 1mL of Trizol was added. Worms were lysed by vortexing for 30s at room temperature and then 20 minutes at 4°. The Trizol lysate was then spun at 14,000rpm for 10 minutes at 4° down to allow removal of insoluble materials. Thereafter the Ambion TRIzol protocol was followed to finish the RNA extraction (MAN0001271 Rev. Date: 13 Dec 2012).

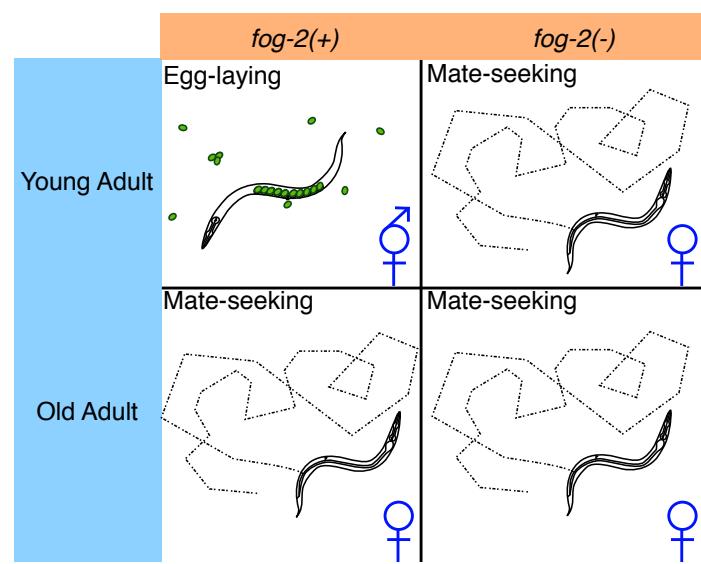


Figure 1 Experimental design to identify genes associated with life-history.

Copyright © 2016 Angeles-Albores *et al.*

Manuscript compiled: Friday 30th September, 2016%

¹Contact: pws@caltech.edu, Address: Department of Biology and Biological Engineering, and Howard Hughes Medical Institute, Caltech, Pasadena, CA, 91106, USA

RNA-seq

RNA interference

RNAi was performed as described in previous protocols (). RNAi bacterial strains were grown in LB plus 100 μ g/mL ampicillin overnight. Fresh RNAi cultures were then plated onto NGM agar plates containing 25 μ g/mL carbenicillin and 1mM IPTG. N2 or *fog-2(q71)* hermaphrodites grown on *E. coli* OP50 were bleached onto sterile plates, and starved L1s transferred to recently seeded RNAi plates. All assays were performed on the offspring of these L1s. Worms grown on every strain were monitored for gross abnormalities, such as sterility, lethality and larval arrest. Control worms in all assays were fed with an anti-GFP RNAi strain. All RNAi worms were grown at 20° C.

Oocyte dropping assay

We performed a slightly modified oocyte dropping assay based on previously described protocols (). Feminized *fog-2(q71)* hermaphrodites were picked as virgin L4s the day before the assay to a fresh RNAi plate. The following day, the adult animals were placed on assay plates (NGM agar seeded with 15 μ L of *E. coli* OP50 four days earlier), twenty worms per assay, and allowed to incubate at room temperature. Laid oocytes were counted after two hours. Plates were then left at room temperature overnight to serve as lawn-leaving assays.

Lawn-leaving assay

Young adult N2 hermaphrodites were selected the day of the assay and placed on assay plates (same as the oocyte dropping plates), twenty worms per assay, and allowed to incubate at room temperature overnight. Assays for *fog-2(q71)* hermaphrodites were performed on the same worms as used in the oocyte dropping assays. The following morning, plates were scored for leaving, with any worm touching any part of the bacterial lawn with any part of its body deemed to be "on" the lawn, and all others deemed to be "off".

Brood size counting

Worms were selected for this assay as L4 hermaphrodites to ensure that all progeny could be counted. For each replicate of each assay, a single worm was placed on a fresh RNAi plate and incubated at 20°. Every 1–2 days, the test worm was moved to a fresh RNAi plate, until it stopped laying eggs. Progeny were counted on each plate before they reached adulthood to ensure that only a single generation was counted.

Statistical Analysis

RNA-Seq Analysis. RNA-seq alignment was performed using Kallisto () with 200 bootstraps. The commands used for read-alignment are in file . Differential expression analysis was performed using Sleuth (). The following Generalized Linear Model (GLM) was fit:

$$\log(y_i) = \beta_{0,i} + \beta_{G,i} G + \beta_{A,i} A + \beta_{A::G,i} A G, \quad (1)$$

where y_i are the TPM counts for the i th gene; $\beta_{0,i}$ is the intercept for the i th gene, and $\beta_{X,i}$ is the regression coefficient for variable X for the i th gene; A is a binary age variable indicating young adult (0) or old adult (1) and G is the genotype variable indicating wild-type (0) or *fog-2* (1); $\beta_{A::G,i}$ refers to the regression coefficient accounting for the interaction between the age and genotype variables in the i th gene. Genes were called significant if the FDR-adjusted q-value for any regression coefficient was less than 0.1. Our script for differential analysis is available on GitHub.

Regression coefficients and TPM counts were processed using Python. Data analysis was performed using the Pandas, NumPy and SciPy libraries (). Graphics were created using the Matplotlib and Seaborn libraries ().

Tissue Enrichment Analysis was performed using WormBase's TEA tool (). We also used the tissue annotation dictionary provided by WormBase to identify genes that were altered in specific tissues.

Brood Size Analysis. Brood size results were analyzed using Welch's t-test to identify genes that were significantly different from a GFP RNAi control. RNAi control results were pooled over multiple days because we could not detect systematic day-day variation. We did not apply FDR or Bonferroni correction because, at a p-value threshold for significance of 0.05, we expected 1 false positive on average per screen.

Lawn-leaving Analysis. lawn-leaving results were analyzed using a χ^2 test for categorical variables. Results were considered statistically significant if $p < 0.05$. No FDR or Bonferroni correction was applied because the size of the screen was too small, with 1 false positive expected on average per screen. However, the lawn-leaving results suffered from high variance, which can lead to false positive results. To safeguard against false positive discovery, we used a non-parametric bootstrap to estimate the true χ^2 value. Using a bootstrap does not lead to statistical acceptance of any gene that was not accepted without a bootstrap; however, applying a bootstrap does lead to statistical rejection of a large number of results.

Oocyte Dropping Analysis. Oocyte dropping results were analyzed using a non-parametric bootstrapped Mann-Whitney U-test because the GFP control variance was very large relative to the variance of the RNAi treatments. Results were considered statistically significant if $p < 0.05$.

Data Availability

Strains are available upon request. File S1 XXXX. File S2 contains XXXX. File S3 contains XXXX. Sequence data are available at GenBank and the accession numbers are listed in File S3. Gene expression data are available at GEO with the accession number: XXXXXX. Code used to generate the simulated data is provided in file XXXX.

RESULTS

We used a linear generalized model (see Statistical Analysis) with interactions to identify a transcriptomic profile associated with the *fog-2* genotype, a transcriptomic profile of *C. elegans* aging common to both genotypes. By adding an interaction term to represent life history, we were also able to identify genes associated with life history. We identified an aging transcriptomic phenotypic consisting of 5,592 genes that were differentially expressed in 6 day old adults relative to young adults. This constitutes one quarter of the genes in *C. elegans*. Tissue Enrichment Analysis (TEA) showed that muscle-associated and hypodermis-associated genes are particularly enriched in this dataset (see Figure 2).

By extracting the regression coefficients associated with genotype change, we were also able to identify 1,881 genes associated with the *fog-2* genotype. Gonad-related tissues were enriched in this gene set, consistent with the function of *fog-2* as a sperm transcription factor. Of the 1,881 genes that we identified in the *fog-2* transcriptome, 1,040 genes were also identified in our aging set. Moreover, of these 1,040 genes, 905 genes changed in the same

G3 Genes | Genomes | Genetics Volume X September 2016 | Effects of Life History on the Transcriptomic Phenotypes of Two Genotypes of *C. elegans* | 3

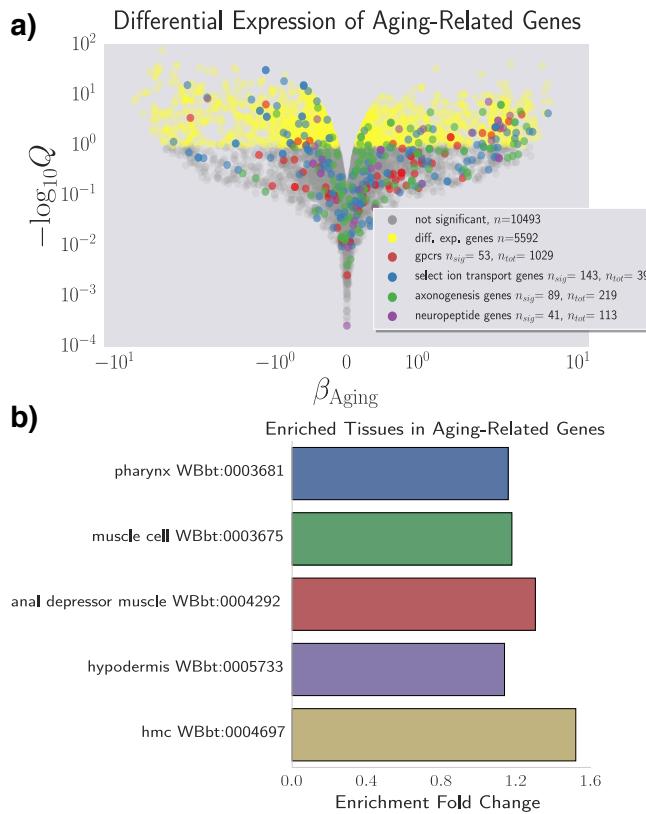


Figure 2 a) We identified a common aging transcriptome between N2 and *fog-2* animals, which showed more than 5,000 altered genes. The volcano plot was randomly down-sampled 30%. **b)** Tissue Enrichment Analysis showed that genes associated with muscle tissues and hypodermis are enriched in this dataset.

direction. The large overlap between these two sets could suggest that feminization is a developmental process in *C. elegans*.

Finally, we extracted the genes that were associated with life history. As expected, this is the smallest gene set, consisting of 1,243 genes that are mainly associated with muscle-tissue(see Figure 3). 457 genes in this set were also present in the aging and genotype transcriptomes. In particular,

Next, we were interested in testing the transcriptomic signatures we identified. We designed three screens to test our gene sets. We selected gene targets by identifying genes that belonged exclusively to one of the three sets defined previously. We selected two assays that could be performed to test genes associated with genotype change. We reasoned that genes that are negatively correlated with the *fog-2* genotype could be associated with fertility, whereas genes that are positively correlated with the *fog-2* genotype could be associated with ovulation rate. Moreover, lawn leaving is known to be associated with sperm status (), so we decided to perform a second screen to study lawn-leaving behaviour.

We performed a brood size screen, selecting as targets genes that were upregulated in N2 but downregulated in *fog-2*, and we also included some genes from alternative categories if we visually noticed serious decreases in brood size. We identified 9 genes that altered brood size (see Figure 4). Of these 7 genes, XXX were previously known. XXX/7 were genes that were associated specifically with the *fog-2* phenotype. XXX/7 were associated with aging or life history.

We reasoned that some of the genes in our genotype dataset might also be associated with ovulation rate in *fog-2*. We selected genes that showed upregulation in *fog-2* animals and placed them on a lawn for two hours. We observed large variation in the ovulation rate for the control RNAi, and as a result no genes were associated with alterations in ovulation rate. However, the pooled variance across the screen was very similar to the control variance, suggesting that our failure to identify genes was not a result of poor conditions or experimental variance. Our screen identified a single hit: C23H4.4, a carboxy ester lipase with unknown function.

We also performed lawn-leaving assays because previous research has reported differences in N2 and *fog-2* leaving rates (). We tested genes that increased in *fog-2* animals in a *fog-2* background, expecting that decreasing these genes should lead to a reversal of the lawn-leaving assay. Likewise, we tested genes that were decreased in *fog-2* animals in an N2 background, expecting that knockdown of these genes would cause lawn-leaving. We identified 9 genes that had an altered lawn-leaving profile in an N2 background. We also identified 9 genes that had an altered lawn-leaving profile in a *fog-2* background. Oddly, both screens identified genes that mainly stimulate lawn-leaving. Initially, we had expected that RNAi knockdown would allow us to identify genes that inhibit lawn-leaving behaviour in an N2 background; whereas RNAi knockdown in a *fog-2* background would identify genes that promote lawn-leaving. However, our screen only identified genes that suppress lawn-leaving in an N2 background. All other hits stimulated lawn-leaving.

DISCUSSION

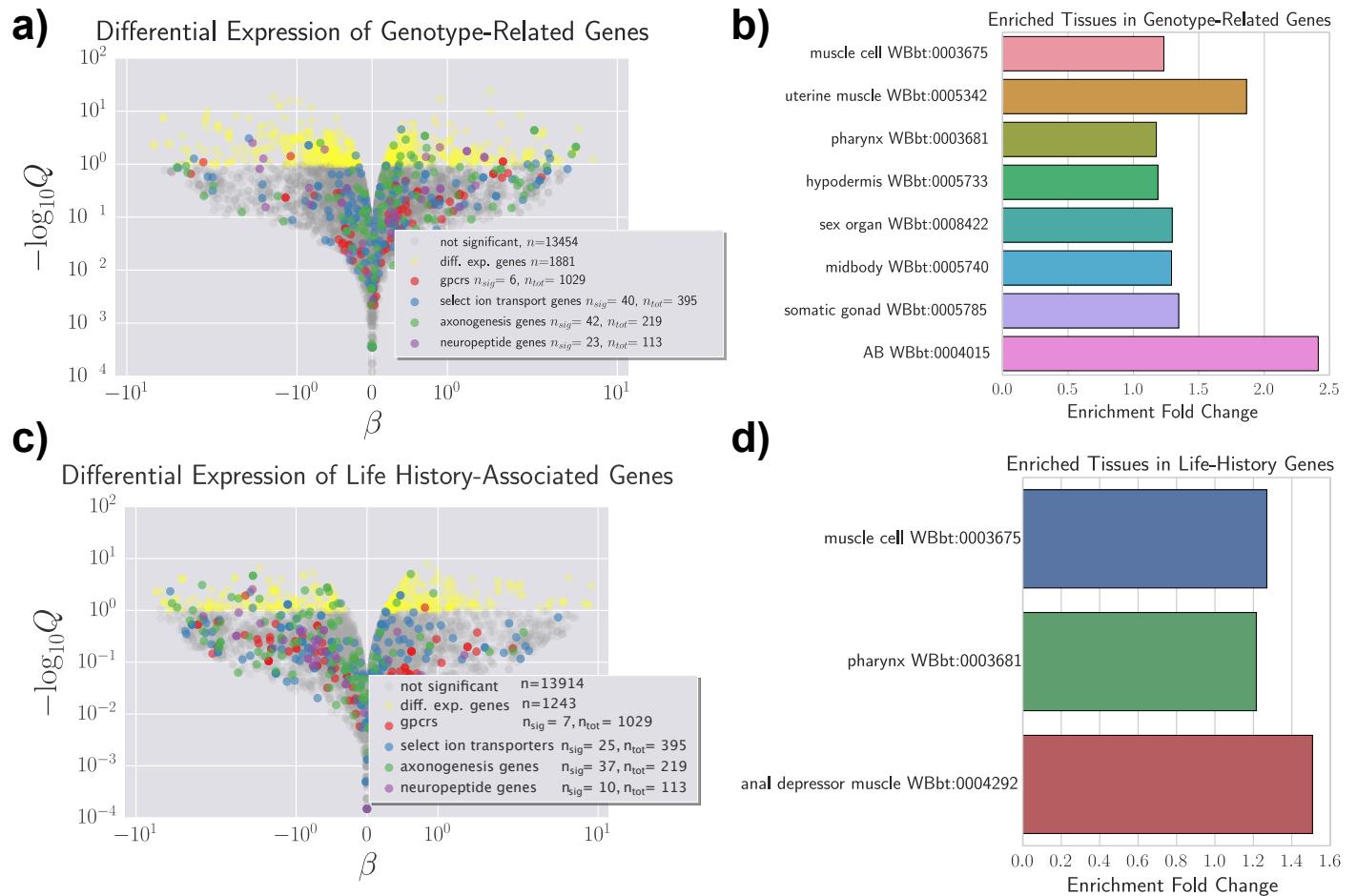


Figure 3 a) Genotype transcriptome volcano plot. **b)** TEA results. **c)** Life history transcriptome volcano plot. **d)** TEA results for life history genes. Volcano plots were randomly down-sampled 30%.

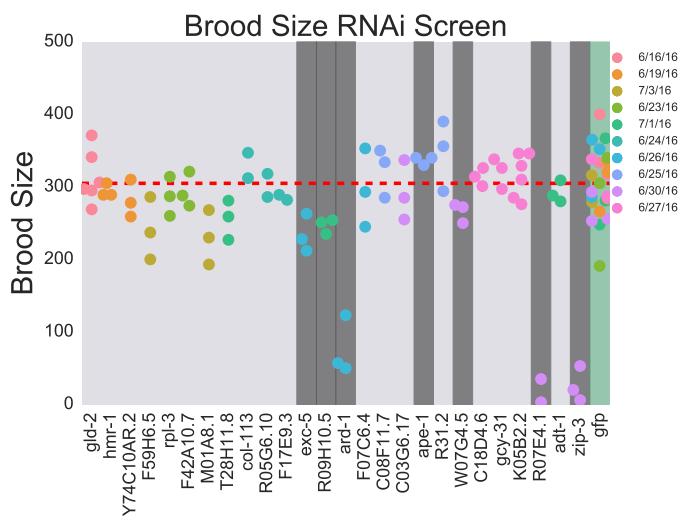


Figure 4 a) Brood size screen results. GFP control RNAi is shaded in green. Results that are statistically significantly different from GFP are shaded in dark grey. Dotted red line shows the mean of the pooled GFP controls. Points are colored by the date the assay was started on.

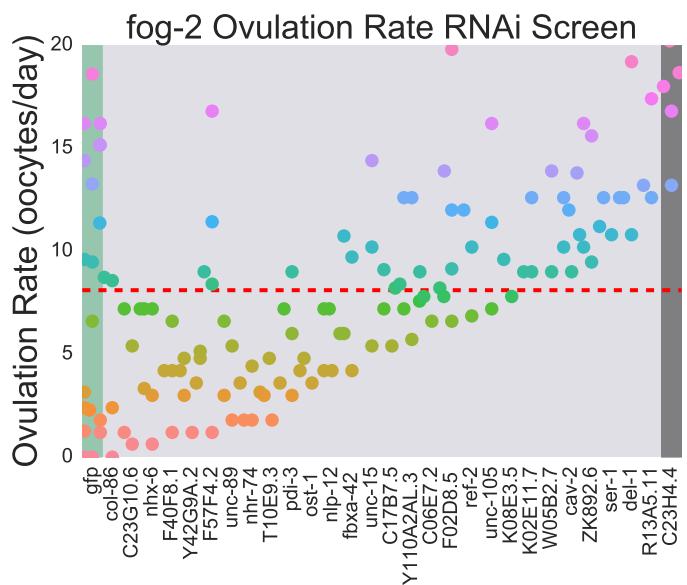


Figure 5 a) Ovulation Rate Assay. GFP control RNAi is shaded in green. Results that are statistically significantly different from GFP are shaded in dark grey. Dotted red line shows the mean of the pooled GFP controls. Points are colored proportionally to their y-coordinate

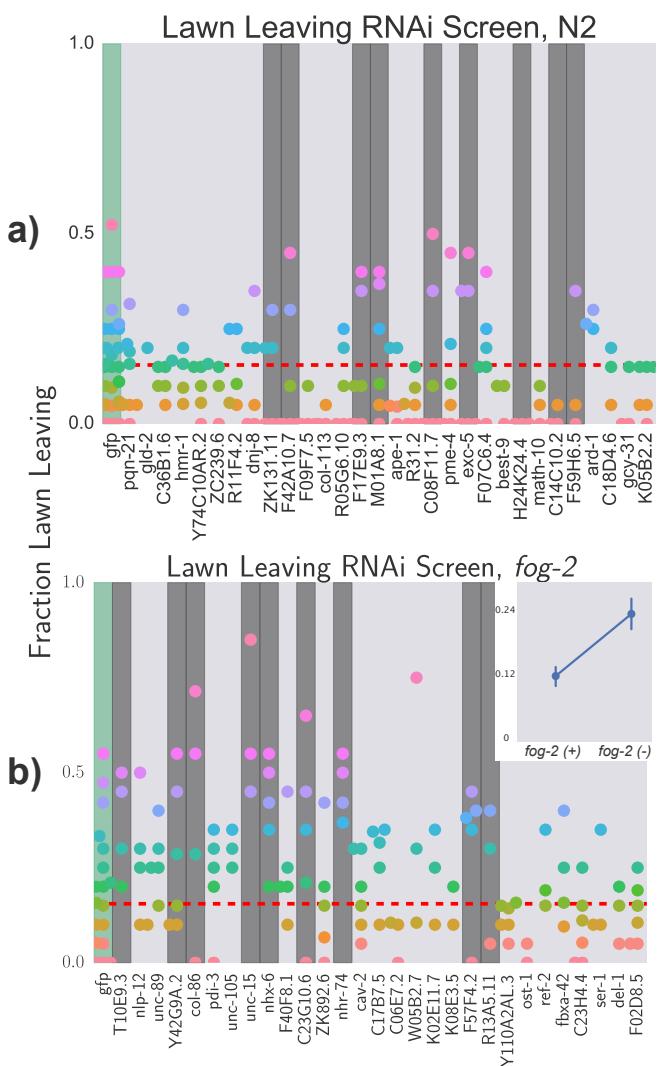


Figure 6 lawn-leaving Screen Results. GFP control RNAi is shaded in green. Results that are statistically significantly different from GFP are shaded in dark grey. Dotted red line shows the mean of the pooled GFP controls. Points are colored proportionally to their y-coordinate. **a)** N2 lawn-leaving assay. **b)** *fog-2* lawn-leaving assay. Inset shows the screen-wide average lawn-leaving rate of N2 and *fog-2*, which shows that the screen-wide lawn-leaving rate of *fog-2* worms is twice the rate of lawn-leaving of N2 worms, in agreement with previous literature ()�.