

# Identification of novel genes associated with longevity in *Drosophila melanogaster* - a computational approach

Bethany S. Hall<sup>1</sup>, Yvonne A. Barnett<sup>1,2</sup>, Jonathan J. Crofts<sup>1</sup>, Nadia Chuzhanova<sup>1</sup>

<sup>1</sup>School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK

<sup>2</sup>Current address: Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, Cambridgeshire CB1 1PT, UK

**Correspondence to:** Nadia Chuzhanova; **email:** [nadia.chuzhanova@ntu.ac.uk](mailto:nadia.chuzhanova@ntu.ac.uk)

**Keywords:** single nucleotide polymorphisms, longevity, *Drosophila*, networks, target genes

**Received:** September 2, 2019

**Accepted:** November 18, 2019

**Published:** December 3, 2019

**Copyright:** Hall et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Despite a growing number of studies on longevity in *Drosophila*, genetic factors influencing lifespan are still poorly understood. In this paper we propose a conceptually new approach for the identification of novel longevity-associated genes and potential target genes for SNPs in non-coding regions by utilizing the knowledge of co-location of various loci, governed by the three-dimensional architecture of the *Drosophila* genome. Firstly, we created networks between genes/genomic regions harboring SNPs deemed to be significant in two longevity GWAS summary statistics datasets using intra- and inter-chromosomal interaction frequencies (Hi-C data) as a measure of co-location. These networks were further extended to include regions strongly interacting with previously selected regions. Using various network measures, literature search and additional bioinformatics resources, we investigated the plausibility of genes found to have genuine association with longevity. Several of the newly identified genes were common between the two GWAS datasets and these possessed human orthologs. We also found that the proportion of non-coding SNPs in borders between topologically associated domains is significantly higher than expected by chance. Assuming co-location, we investigated potential target genes for non-coding SNPs. This approach therefore offers a stepping stone to identification of novel genes and SNP targets linked to human longevity.

## INTRODUCTION

Despite a growing number of studies on survival into old ( $\geq 85$  years) and advanced ( $\geq 90$  years) age, factors influencing longevity (or lifespan) are still poorly understood. Human twin studies estimated that 20–30% of variation in survival into old and advanced age, besides maintaining a healthy life style, is determined by heritable genetic factors [1, 2].

In order to determine these genetic factors, several genome-wide scans for linkage, genome-wide association studies (GWAS) and genome-wide association meta-analyses have been carried out on panels of long-lived individuals. Variations in many loci, e.g. near the *D4S1564* [3], *MINPP1* [4], *HLA-DQA1/DRB1* and *LPA* [5] genes, have been identified as contributing to survival

into old age, but only single nucleotide polymorphisms (SNPs) in *TOMM40/APOE* and *FOXO3* loci were found to robustly associate with longevity [6–11]. In a whole-genome scan for genetic linkage performed by Kerber et al. [12] on individuals from the Utah Population Database, in which high levels of both familial longevity and individual longevity were exhibited, the strongest signal was observed in marker D3S3547 on chromosome 3p24.1. In addition, a locus on chromosome 3p24-22, previously identified in [13], was found to link to exceptional longevity [12], strengthening the case that genes found in these regions play a role in the regulation of human lifespan. Boyden and Kunkel [13] have identified several additional loci as having significant association with longevity, e.g. on chromosomes 9q31-34, 12q24 and 4q22-25. Recently, GWAS of parental longevity was performed on participants of European

descent available via the UK Biobank [14]. Several previously known variants have been confirmed in this study. In addition, other common variants previously found by disease-specific GWAS to associate with e.g. cellular senescence, inflammation, lipid metabolism and cardiovascular conditions were also found to associate with parental longevity [14]. Their results suggest that human longevity is a highly polygenic trait influenced by many variants with a small effect size [14].

Progress in studies of human longevity is being exacerbated by small sample sizes making model organisms, such as *Drosophila melanogaster*, increasingly important for studying and understanding genetic factors affecting longevity. The lifespan of *Drosophila* is affected by several factors including genetics, differences in environmental conditions, diet and overcrowding. In laboratory conditions under controlled environment the average lifespan is found to be 26 and 33 days for female and male *Drosophila*, respectively [15]. Mutations in several genes have been found to increase the lifespan of *Drosophila*. For example, a mutation in the *mth* (Methuselah) G protein-coupled receptor gene, which leads to the partial loss-of-function, has been found to extend the average lifespan by 35% [16]. Mutant versions of the *Indy* gene, which encodes an amino acid transporter, has been shown to double the average lifespan [17]. It was also shown that single gene mutations in the target of rapamycin (TOR) and the insulin/insulin-like growth factor (IIS) signaling pathways can slow down the aging process in model organisms including flies [18].

Up to date, *Drosophila* GWAS have identified millions of naturally occurring SNPs that potentially influence longevity. Burke et al. [19] compared allele frequencies in the oldest surviving *Drosophila* with the randomly selected individuals from the same “synthetic” populations, derived from eight inbred founders. Eight significantly divergent regions have been identified. A small proportion of genes, found in these regions, were enriched in Genome Ontology (GO) biological process terms ‘defense response’ and ‘glutathione metabolic process’ [19]. Ivanov et al. [20] used lines from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) to perform GWAS and identified ~2 M common SNPs. However, none of the SNPs found reached genome-wide significance level prompting the hypothesis of a possible combined effect of common SNPs on longevity. Gene-based analysis with either gene regions or gene regions extended into  $\pm 5$  Kb of flanking sequences had identified several top-ranked genes including the *CG11523* and *Neprilysin 1*. The former was found to have a *GSK3 $\beta$*  interaction domain that is a crucial component of the TOR pathway in human cell lines [20]; the latter could be essential for

female fitness [20]. Among the top-ranked 100 genes ( $p < 4.79 \times 10^{-6}$ ) found in this study were *Chrb*, *slif*, *mipp2*, *dredd*, *Rps9* and *dm* genes enriched in the ‘TOR pathway’ GO term [20]. Several of the longevity associated genes found are involved in processes which are known to impact aging (e.g. carbohydrate metabolism), however the function of others (although not known) provided opportunity for further, promising experimental examination. Polygenic score analysis was also used to find the additive effects of common SNPs [20]. In the absence of the second dataset, cross validation was performed. It was found that a small proportion of the observed lifespan variation (~4.7%) is explained by the additive effect of common SNPs. Despite the success in identification of variants, associated with longevity, the functional role of the majority of them – especially the variants residing outside the gene coding regions – remains to be determined.

In this paper we hypothesize that co-location of known longevity-associated genes with genes, not previously implicated in longevity, and their enrichment in the same biological function or pathway as known genes, make them novel candidate genes, potentially linked to longevity. We further hypothesize that both non-coding SNPs and their potential target genes also reside within co-located loci. To identify these novel genes/genomic regions we devised a computational approach based on analysis of networks of co-located loci, harboring both GWAS-identified variants and novel genes. Two datasets of SNPs generated by GWA studies [19–20] were used, comprising respectively ~1 million and ~2 million SNPs and sharing 2139 SNPs residing within 1515 (possibly overlapping) genes and 1044 non-coding SNPs.

As a measure of co-location (or proximity) of two distinct loci, not necessarily on the same chromosome, we used inter- and intra-chromosomal contacts generated by chromosome conformation capture Hi-C technique for the *Drosophila melanogaster* genome [21]. Studies of chromosome conformations have revealed that three-dimensional architecture of chromatin dictates the co-location of specific genes within the nucleus, thereby prompting the hypothesis of existence of common mechanisms controlling their transcription in a tissue-specific manner [22–23]. Recently, Won et al. [24] have demonstrated the advantages of using 3D chromatin maps for identifying target genes for schizophrenia-associated SNPs, residing within non-coding regions of the genome. The findings have shown that for many non-coding SNPs their target genes were neither adjacent to SNPs nor in linkage disequilibrium, proving the point that many regulatory interactions are not captured by linear chromosomal organization. Analysis of intra-chromosomal interactions showed more frequent and

stronger interactions within continuous genomic regions, called topologically associated domains (TADs), than with regions residing in other TADs [22–23]. TADs have been proven to play important roles in 3D organization of genomes and gene regulation and, when mutated, may lead to disease through disruption of gene regulatory pattern (reviewed in [25]).

A network of interactions was created from the inter- and intra-chromosomal contacts with nodes representing genomic regions, connected by edges, weighted by interaction frequencies. We calculated various network measures (e.g. degree [26]) and identified communities (i.e. densely connected subnetworks) existing within the network with the aim of revealing influential nodes/regions and densely connected communities (clusters) within networks. Candidate regions and communities were further explored using FlyBase (<http://flybase.org/>) and FlyMine (<http://www.flymine.org/>) resources, and GeneAge database (<http://genomics.senescence.info/genes/models.html>) to provide a body of evidence for genomic regions having genuine and/or previously unknown association with longevity.

To explore the role that SNPs occurring in TAD borders play in longevity, we analyzed genes residing in close proximity to TAD borders and sharing both ‘long-lived’ and ‘short-lived’ phenotypes. We hypothesized that a SNP(s) in nearby TAD borders may lead to a disruption of a regulatory pattern of a gene resulting in one of the phenotypes, ‘long-lived’ or ‘short-lived’, whereas the opposite phenotype could be a consequence of SNPs residing within genes themselves.

## RESULTS AND DISCUSSION

### Choice of interaction frequency thresholds and genome-wide significance level

To assess the strength of interactions between intra- and inter-chromosomal genomic regions, distributions of interacting frequencies were analyzed individually for each chromosome and between chromosomes. Only 1% of the strongest intra-chromosomal interactions corresponding to the tails of these distributions and resulting in frequencies greater than 247, 215, 1308 and 342 for chromosomes 2, 3, 4 and X, respectively, were considered. The threshold for inter-chromosomal interaction frequencies, corresponding to 1% of strongest interactions, was 10. We refer to interactions with frequencies exceeding these thresholds as “strong” interactions.

The genome-wide significance level, required for finding association between  $\sim 10^6$  SNPs, is usually set to

$p < 5 \times 10^{-8}$ . This value corresponds to 0.05 level of significance after correction for multiple testing. In our case, each SNP was binned into a 80 Kb region. There are 1503 distinct 80 Kb regions recorded in the *Drosophila* Hi-C data. Taking this into account, we corrected the required significance level to  $3.33 \times 10^{-5}$ . In the analysis of SNPs in non-coding regions the Hi-C data with finer resolution, 10 Kb, was used where interaction frequencies between 11,839 10 Kb bins were available [21]; in this case the genome-wide level of significance was set to  $0.05/11839 = 4.22 \times 10^{-6}$ . Following [19], SNPs with *D*-values exceeding 7.9 were deemed to be significant.

### Original networks of interaction based on Synthetic and DGRP GWAS data

The original network of interaction based on the Synthetic GWAS data consists of 279 nodes each representing a 80 Kb region harboring at least one SNP with *D* > 7.9. In turn, the original network of interaction based on the DGRP GWAS data consists of 80 nodes corresponding to regions harboring SNPs with *p*-values <  $3.33 \times 10^{-5}$ . The original networks share 14 common nodes covering 1.12 Mb of the *Drosophila* genome and harboring 168 genes. Only five genes – *Rim2* (replication in mitochondria 2), *GlyP* (glycogen phosphorylase), *aop* (anterior open), *HDAC1* (histone deacetylase 1) and *Tpi* (triose phosphate isomerase) – were found in FlyBase database as having “long-lived” phenotype. The number of SNPs residing within these common regions and satisfying chosen thresholds was 91 and 19 for Synthetic and DGRP GWAS-based data, respectively. Among the genes with the highest number of SNPs recorded in both GWAS datasets were *nmo*, *sima*, *axo*, *CG9967*, *eys*, *chinmo* and *dpr3* (for the full list of genes see Supplementary Table 1).

### Extended networks of interactions

Original networks were further expanded to create extended networks by adding extra nodes, corresponding to 80 Kb fragments that interact with frequencies meeting interaction frequency thresholds with the nodes, already present in the original networks. Together with regions that harbor SNPs recorded in the corresponding GWAS datasets, the extended networks contain novel regions that may not be covered by techniques used for SNP identification. We refer to these networks as Synthetic and DGRP GWAS-based (extended) networks.

The Synthetic GWAS-based extended network is fully connected and consists of 1099 nodes harboring  $\sim 75\%$  (69,951) of SNPs recorded in the Synthetic GWAS dataset with 2,409 SNPs residing within genes. Among

13,838 genes residing within the network nodes 217 genes were found to have “long-lived” phenotype as recorded in the FlyBase database. The node labelled 547 (corresponding to region Chr2R: 20800000-20880000) has the highest degree, 150.

The DGRP GWAS-based extended network has six disconnected components and consists of 671 nodes harboring ~50% (1,093,533) of SNPs recorded in the DGRP GWAS dataset with 114 SNPs residing within genes. Among 8,929 genes residing within the network nodes 145 genes were found to have “long-lived” phenotype according to the FlyBase database. The node labelled 1183 (region Chr3R: 25920000-26000000) has the highest degree of 68.

The extended networks share 527 common nodes covering 42.16 Mb of the *Drosophila* genome and harboring 7,413 genes among which 121 have “long-lived” phenotype. Fifteen common regions do not harbor any genes. For approximately 30% and 3% of genes residing within common regions no SNPs were recorded in the Synthetic and DGRP GWAS datasets, respectively. Among the genes with the highest number of SNPs recorded in both GWAS datasets were *Ptp61F*, *CG45186*, *kirre*, *Ptp99A* and *CG44153*. Only a small proportion of genes found in regions common for both datasets were harboring SNPs meeting our significance threshold – 717 and 57 in the Synthetic and DGRP GWAS-based networks, respectively.

Several novel regions with the highest degree were selected for further analysis and each of the subnetworks centered around these novel regions (i.e. together with all connected regions) were considered (Supplementary Table 2). Genes residing within these subnetworks were sought for enrichment in longevity-associated GO terms. The results are summarized in Table 1.

Genes residing within a subnetwork centered around node 928 (chr3R:5520000-5600000) in the extended Synthetic GWAS-based network were enriched in two GO terms, ‘apoptotic process’ and ‘nervous system development’ (Table 1). Among them the *trbd* and *CG8412* genes that have ‘short-lived’ phenotype according to in FlyBase resources. The loss of the *trbd* gene, a negative regulator of the *Drosophila* immune-deficiency pathway, has previously been observed to reduce lifespan [27]. A number of genes in this subnetwork, including *dmt*, *hyd*, *CG16908* and *CG9471*, were found to have phenotypes ‘increased mortality’ and ‘lethal’. The *MED6* gene was found to have a phenotype of ‘cell lethal’ and is known to be required for elevated expression of a distinct set of developmentally regulated genes. This gene is essential

for viability and/or proliferation of most cells and mutants of this gene have previously been observed to fail to pupate, dying in the third larval instar with severe proliferation defects in imaginal discs and other larval mitotic cells [28]. Finally, this subnetwork also contains the *FoxP* gene, a protein that encodes a transcription factor expressed in the nervous system. This gene has recently been shown to be important for regulating several neurodevelopmental processes and behaviors that are also related to human disease [29].

Many of the newly found genes (see Table 1) share the same biological function and co-locate with genes that have previously been reported to associate with longevity and/or aging, thus acting as a proof of concept. For example, the *sidpn*, *hook* and *CG12935* genes residing in subnetwork centered around bin 928 (chr3R:5520000-5600000) were reported to have a ‘short-lived’ phenotype. Loss-of-function mutation in the *hook* gene has been found to reduce maximum lifespan by up to 30% [30]. Mutant flies lacking mitochondrial *Top3alpha* gene have also been found to have decreased maximum lifespan by up to 25%, in which a premature aging phenotype was demonstrated and mobility defects were observed [31]. Several genes, e.g. *RpL30*, *Eps-15*, *Nipped-B* and *RPA2*, listed in Table 1 were also found to have an ‘increased mortality’ phenotype according to the FlyBase resources.

Five genes residing in a subnetwork centered around bin 1220, were enriched in the ‘DNA repair’ GO term. Interestingly, this novel region is located on chr4: 960000-1040000, a chromosome seen as an anomaly because of its small size in comparison to other chromosomes and its chromatin structure. Due to its size, this chromosome is often ignored, however it is known to harbor at least 16 genes where many of them are thought to have male-related functions [32]. Using a comprehensive database of *Drosophila* regulatory sequences available via RedFly database (<http://redfly.ccr.buffalo.edu>), several enhancers were found in this region that target lncRNA *sphinx* and the transcription factor *toy* residing within this novel region although for some enhancers their target genes are not known. One can speculate that these enhancers could target genes co-located in 3D, i.e. residing within the same subnetwork centered around bin 1220.

In the extended DGRP network two novel bins, 2 and 28, were found to have the highest degree. Seven and 17 genes residing in subnetworks centered around bin 28 (chr2L: 2160000-2240000) and bin 2 (chr2L:80000-160000) were enriched in the ‘immune system process’ and ‘cellular response to stress’ GO terms, respectively. Some of these genes have previously been implicated in aging or have phenotypes which could be linked to longevity. For example, flies heterozygous for the



**Table 1. Novel nodes with the highest degree in the Synthetic and DGRP GWAS-based networks harboring genes enriched in longevity-associated GO terms.**

Novel node	Network	GO term	P-value	Genes enriched in GO term	Number of nodes harboring genes enriched in GO term/ total number of nodes
928	Synthetic	Apoptotic process	2.27E-04	<i>E2f2</i> , <u><i>lola</i></u> *, <i>egr</i> , <i>Ret</i> , <i>Vps25</i> , <i>TER94</i> , <i>ptc</i> , <i>eEF5</i> (CG3186), <u><i>sname</i></u> , <i>ninaA</i> , <i>yki</i> , <u><i>sigmar</i></u> , <i>l(2)tid</i> , <i>Mcm10</i>	7/16
928	Synthetic	Nervous system development	4.37E-04	<i>CG10339</i> , <i>amos</i> , <i>CG10431</i> , <u><i>Sidpn</i></u> **, <i>RpL30</i> , <i>hook</i> , <i>Dap160</i> , <i>enok</i> , <i>lola</i> , <i>dgo</i> , <i>egr</i> , <i>CG12935</i> , <i>Ret</i> , <i>Pka-R2</i> , <i>Eps-15</i> , <i>Galphao</i>	9/16
1220	Synthetic	DNA repair	0.0294	<i>Top3alpha</i> , <i>PCNA2</i> (CG10262), <u><i>Nipped-B</i></u> , <i>CG9272</i> , <i>RPA2</i>	4/21
28	DGRP	immune system process	0.021515	<i>Vps16B</i> , <i>Cad99C</i> , <i>aop</i> , <i>DPCoAC</i> (CG4241), <i>Stat92E</i> , <i>Mtl</i> , <i>GlyP</i>	4/8
2	DGRP	cellular response to stress	0.006104	<i>CG11498</i> , <i>Clbn</i> , <i>CG13473</i> , <i>CG14130</i> , <i>Sld5</i> , <i>mu2</i> , <i>Atg16</i> , <i>kay</i> , <i>CG3448</i> , <i>Rad9</i> , <i>Mtl</i> , <i>Grx1</i> (CG6852), <i>Cat</i> , <i>HipHop</i> , <i>BI-1</i> , <i>Wdr24</i> (CG7609), <i>Drice</i>	13/15

\* Genes residing within original nodes, i.e. harboring SNPs with D > 7.9 are underlined.

\*\*Genes previously found to have association with longevity as recorded in FlyBase or GeneAge resources are shown in bold.

mutation in the *Stat92E* gene have been found to have maximum lifespan up to 30% shorter than those of wild-type control flies [33]. The mean lifespan of *Drosophila* was found to be increased through post developmental RNA interference of *GlyP* by up to 17.1% [34]. Another gene listed in Table 1 found to have a positive effect on lifespan is *Cat*, where an overexpression of this gene results in an increase in lifespan by up to a third [35]. Searches in the FlyBase database show that several other genes have phenotypes associated with aging, e.g. *Clbn* and *Atg16* genes have a ‘short-lived’ phenotype, the *BI-1* gene has both a ‘short-lived’ and ‘long-lived’ phenotype and genes *kay* and *HipHop* have phenotypes for increased mortality.

Using the RedFly database, we found that the novel region on chr2L:2160000-2240000 (bin 28), which was added to the original nodes of the DGRL GWAS-based network on the basis of its strong interactions with the original nodes, harbors several enhancers. Some of these enhancers target *CG34172*, *Uch* and the transcriptional-repressor protein *aop* genes. The latter strongly associates with longevity and is found to be central to lifespan extension caused by reduced IIS or Ras attenuation [36]. For some enhancers their target genes were not specified. One can speculate that these enhancers could target other co-located genes residing within the subnetwork centered around bin 28.

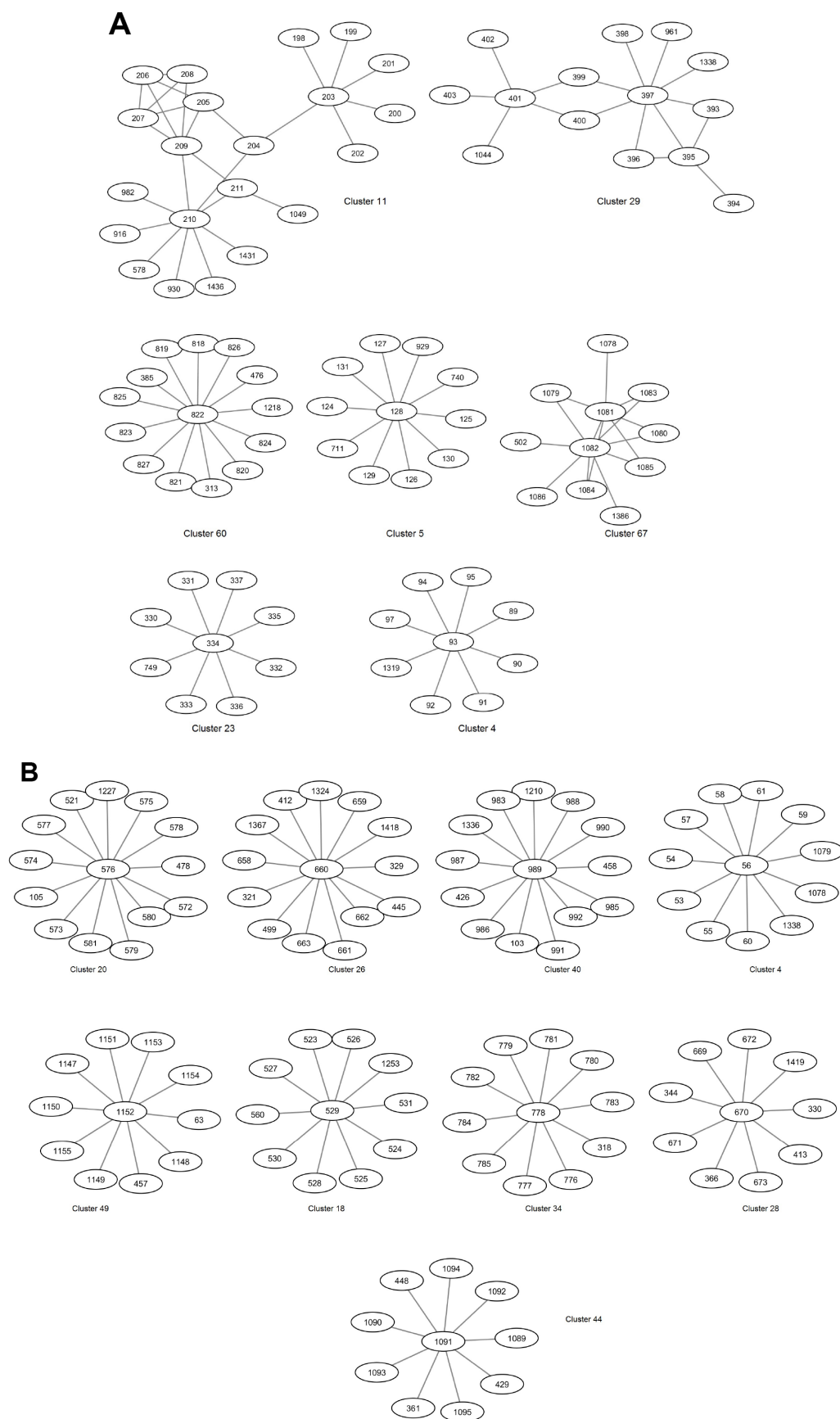
## Clusters in the extended GWAS-based networks

Community detection algorithm implemented in GEPHI which uses the Louvain modularity method [37] was performed to identify clusters in the Synthetic and DGRP GWAS-based networks. Selected clusters are shown in Figure 1. Complete sets of clusters for each network are shown in Supplementary Tables 3–4. A ‘resolution’ parameter was set to 0.1, enabling us to identify more communities/clusters as compared with the smaller number of communities that could be obtained by using a greater value for this parameter [38]. These clusters were further explored with the aim of identifying novel genes that co-locate with known longevity-associated genes and are enriched in the same biological function as known genes.

### Clusters in the extended Synthetic GWAS-based network

The Synthetic GWAS-based network was found to have 81 communities/clusters with the smallest consisting of three nodes and the largest of 72 nodes (see Supplementary Table 3). Selected clusters with the most significant enrichment in longevity-associated GO terms are summarized in Table 2. Nodes constituting these clusters are listed in Supplementary Table 5.

Six genes residing within five nodes of cluster 11 were enriched in the ‘DNA repair’ GO term (p-value = 0.022).



**Figure 1.** Selected clusters in the Synthetic (A) and DGRP (B) GWAS-based extended networks of interactions.

**Table 2. Number of nodes in the selected clusters in the Synthetic GWAS-base network and genes enriched in longevity-associated GO terms.**

Cluster	GO term	P-value	Genes enriched in GO term	Number of nodes harboring genes enriched in GO term/ total number of nodes
4	cellular response to stimulus	0.004503	<i>Rab39, Tom40, santa-maria, <b>Mnn1</b>*, sem1, Pvf2, <b>Gr28b</b>, Pvf3, Ziz, RapGAP1, Wnt4, wg, Wnt6, <u>Wnt10**</u>, ninaC, <u>CG5160</u>, CG5181, mir-305</i>	6/9
4	localization	0.007119	<i>Rab39, Tom40, Sem1, Pvf2, CG13793, CG13794, CG13795, CG13796, CG31904, CG31907, CG33296, Pvf3, Ndae1, Wnt4, ninaC, Ntl, ATPsynGL, Nuf2</i>	5/9
4	cell communication	0.023993	<i>Rab39, santa-maria, <b>Mnn1</b>, Pvf2, Gr28b, Pvf3, Ziz, RapGAP, Wnt4, wg, Wnt6, <u>Wnt10</u>, ninaC, <u>CG5160</u>, mir-305</i>	7/9
5	macromolecule modification	0.003413	<i><b>Atg1</b>, Ptp69D, Cnot4, RluA-I, CG32847, CG33303, CG34183, CG42366, Fkbp59, CG4839, Ror, CG4968, Sps2, gny, STUB1, Sp27A, LManI, Bug22, Cdk1, Cand1, Usp14, <b>CYLD</b>, Utx, <b>Pten</b>, <b>bsk</b>, Dref, RluA-2, LMannII, FBXO11</i>	9/11
5	cellular catabolic process	0.020971	<i><b>Atg1</b>, lft, CG32847, CG4592, CG4594, CG4598, yip2, Prosalpha6, RpS27A, CG5367, Usp14, Utx, <b>Pten</b>, CG5676, <b>bsk</b>, <b>chico</b>, CG5731, CG8526, FBXO11</i>	9/11
11	DNA repair	0.021953	<i>CG17329, <b>ku80</b>, CG31807, CG33552, EndoGI, CG5316</i>	5/21
11	developmental process	0.010492	<i><b>cact</b>, <u>Cas</u>, <u>chif</u>, <u>cni</u>, <u>crp</u>, <u>dac</u>, <u>foxo</u>, <u>fzy</u>, <u>glu</u>, <u>goe</u>, <u>grp</u>, <u>heix</u>, <u>her</u>, <u>mdy</u>, <u>mir-9b</u>, <u>mir-9c</u>, <u>sing</u>, <u>squ</u>, <u>twe</u>, <u>wek</u>, <u>yellow-b</u>, <u>BicC</u>, <u>BuGZ</u>, CG17328, CG32572, CG4793, <u>CG5953</u>, <u><b>Ca-alpha1D</b></u>, Cyp303a1, Cyt-c-d, EndoGI, GMF, Idgfl, Idgf2, Idgf3, <u>Mhc</u>, Npc2b, Syx5, TwdIX, TwdIY, TwdIZ, TwdIapIha, <u><b>VhaSFD</b></u>, beat-Ia, beat-Ib, beat-Ic</i>	14/21
23	apoptotic process	0.033954	<i><b>azot</b>, tor, <b>cathD</b>, Cull1, fwe, mir-263b</i>	4/9
23	positive regulation of gene expression	0.028920	<i>CG12769, Rpt1, Kdm4A, udd, <b>Nup50</b>, nito, CG6244, Lpin, lig</i>	6/9
29	negative regulation of transcription, DNA-templated	0.024472	<i>CG10038, spt4, wuc, Iz, seq, Kdm4B, <b>sug</b>, <b>Psc</b>, <b>Su(z)2</b>, Iswi</i>	7/14
60	gene expression	$5.6 \times 10^{-4}$	<i>CG10474, Rpb8, sa, CG11906, mip40, Pc, croc, barc, CRIF, Hr78, wbl, rib, Tsr1, eg, CycH, CG7414, Nopp140, <b>mub</b>, RpLPO, Cdk12, TfAP-2, <b>rho-7</b></i>	9/14

\* Genes previously found to have association with longevity as recorded in FlyBase or GeneAge resources are shown in bold.

\*\*Genes residing within original nodes, i.e. harboring SNPs with D > 7.9 are underlined.

DNA integrity and stability depend upon the ability of DNA repair mechanisms to detect and repair damaged DNA. A DNA repair gene *ku80* is involved in repair of double-stranded DNA breaks [39] and was found to have

a ‘short-lived’ phenotype. The *EndoGI* gene is involved in positive regulation of the Notch signaling pathway and associated with an ‘increased mortality’ phenotype. Notch signaling is important for cell-cell communication and

plays an important role in processes such as neuronal function and development (reviewed in [40]). The other four genes residing within this cluster, *CG17329*, *CG31807*, *CG33552* and *CG5316*, are currently not fully characterized. One can speculate that close proximity of these genes within the cell nucleus and shared biological function with the *ku80* and *EndoG1* genes, make them potential candidate genes, linked to longevity. In fact, the human ortholog of the *CG5316* gene identified via the Integrative Ortholog Prediction Tool available at [https://www.flyrnai.org/cgi-bin/DRSC\\_orthologs.pl](https://www.flyrnai.org/cgi-bin/DRSC_orthologs.pl), the *APTX* gene, which encodes the DNA strand-break repair protein aprataxin was found to have a broader role in DNA single-strand break repair in neurodegenerative disease (reviewed in [41]) that shortens lifespan. In a longitudinal study with 11 years of follow-up on survival in the oldest-old Danes, Soerensen et al. [42] have shown that rs705649 SNP in DNA repair protein XRCC5, which is the human ortholog of the *ku80* gene, is associated with mortality in late life.

Six genes, *azot*, *tor*, *cathD*, *Cull1*, *fwe* and *mir-263b*, residing in cluster 23 were enriched in the ‘apoptotic process’ GO term (p-value = 0.034). The apoptotic process has almost an opposite role to the previously discussed GO term ‘DNA repair’, whereby when DNA is damaged, the checkpoint protein p53 is activated and the decision is made as to whether replication should be stopped and the DNA repaired, or the cell made to die by apoptosis (reviewed in [43]). Studies have found that in mammals, at least in part, apoptosis plays an important role in the process of aging and tumorigenesis and that age-enhanced apoptosis may work as a protective mechanism against age-associated tumorigenesis [44]. The *ahuizotl* gene, *azot*, which encodes a calcium dependent protein responsible for the elimination of less fit cells, is known to play a role in delaying aging and extending lifespan. This gene has the ‘long-lived’ phenotype and was previously found to increase lifespan [45]. Another gene in this group, *cathD*, a gene with phenotypes that include those that associate with apoptosis such as ‘increased cell death’ as well as longevity associated phenotype ‘short-lived’. The human ortholog of this gene, *CTSD*, encoding cathepsin D was found to associate with cognitive abilities in both demented and non-demented individuals [46] and was also implicated in increasing the risk of developing Alzheimer's disease [47]. Another gene in this group, *Cull1*, belongs to the cullin family and has phenotypes of ‘increased mortality’ and ‘neuroanatomy defective’. The *fwe* gene encodes a transmembrane protein that mediates win/lose decisions in cell competition and neuronal culling during development and aging; this gene has longevity related phenotypes - ‘increased mortality’ and ‘lethal’. Given the longevity association that these genes in this cluster have, through both phenotypes and

biological functions, one can speculate that other genes that are found to reside within this cluster 23 may also influence longevity in the same way as the genes discussed above, due to their close proximity and strong interaction in the genome.

Four out of 19 genes enriched in the ‘cellular catabolic process’ GO term (cluster 5; Table 2) have been previously shown to have association with longevity or display phenotypes which associate with aging and, in most cases, with increased lifespan. This includes *chico*, a gene encoding an insulin receptor substrate that belongs to an insulin/insulin-like growth factor (IGF) signaling pathway and found to increase lifespan by up to 48% [48]. Koohy et al. [49] have identified transcriptional downregulation of components of the insulin-like growth factor signaling pathway in mouse, in particular downregulation of the mouse homolog of *chico* gene, *IRS1*, as a signature of aging in developing B cells. The overexpression of the *Pten* gene was found to delay the process of proteostasis and therefore resulted in a decrease in the loss of muscle strength during muscle aging, increasing maximum lifespan in *Drosophila* by up to 7.7% in comparison with matched controls [50]. Interestingly, the human *PTEN* (phosphatase and tensin homolog) gene was found to encode upstream regulators for the *FOXO3* gene [51], one of the few loci robustly associated with longevity in humans [10], stressing that longevity-associated SNPs may reside in regulatory regions as well as in protein-coding genes [52]. The *Drosophila* ortholog of the *FOXO3* gene, *foxo*, a transcription factor involved in the regulation of the insulin signaling pathway, is a commonly known longevity gene [53–56]. The 80 Kb region harboring this gene interacts with the *Pten* region although the interaction frequency is below the threshold chosen in this study. The neuronal-specific upregulation of the *Atg1* gene was found to result in increased median lifespan of *Drosophila* by up to 25% [57]. The human ortholog of this gene, *ULK1*, involved in longevity-regulating pathways identified by the KEGG database [58]. Salas-Pérez et al. [59] have shown that methylation level of the CpG region residing within this gene strongly associates with age-related obesity and metabolic syndrome traits, suggesting a role for DNA methylation in aging-related metabolic alterations. Another gene found to be enriched in the ‘cellular catabolic process’ GO term was the *bsk* gene which is involved in RNA interference. Such interference in intestinal stem cells results in short life due to impaired intestinal homeostasis and tissue regeneration and has been found to reduce mean lifespan by 16.4% and 10.2% in males and females, respectively [60].

Several genes residing within other clusters and enriched in longevity-related GO terms have been previously implicated in longevity. The overexpression of the



*VhaSFD* gene that encodes a regulatory subunit of the vacuolar ATPase proton pump (H<sup>+</sup>-ATPase) and Sugar baby (*Sug*) gene related to a maltose permease from *Bacillus* result in an increase in mean life span by 5–10% [61]. The *mushroom-body expressed (mub)* gene has previously been found to have an association with longevity when mutated; the insertion of a p-element in the gene resulted in an increased lifespan up to 21.4% [62]. On the contrary, the *rho-7* gene was found to decrease lifespan; the knocked-out study showed that flies develop severe neurological defects as well as a greatly reduced lifespan [63]. Several genes in Table 1 were found to display the phenotype ‘increased mortality’ according to FlyBase resources. These includes the *CYLD* gene, a cancer consensus gene responsible for tightly limiting the immune response duration [64]. A mutant of this gene, *dCYLD*, was proven to be essential for JNK (Jun-N-terminal Kinase)-dependent oxidative stress resistance and normal lifespan and has also been indicated to play a critical role in modulating TNF-JNK-mediated cell death [65]. The *Mnn1* gene that also play a role in the regulation of stress response in *Drosophila* [66] displays this phenotype. The association between stress and lifespan has often been made, and previous studies have observed differences in gene expression when comparing normal and stress conditions which has resulted in the identification of aging genes in *Drosophila*. The genes found to reside in the same clusters as the genes previously shown to play roles in biological processes associated with longevity were found to harbor a number of SNPs. Although not all SNPs residing within genes enriched in the same GO term had a significant *D*-value (*D* > 7.9), one can speculate that SNPs in one or several functionally-related gene(s) co-located within the cell nucleus may contribute collectively to the longevity phenotype.

#### Clusters in the DGRP GWAS-based network

The DGRP GWAS-based network comprised 61 communities, where the smallest consisted of three nodes and the largest of 42 nodes (see Supplementary Table 4). Selected clusters with the most significant enrichment in longevity-associated GO terms are summarized in Table 3. Nodes constituting these clusters are listed in Supplementary Table 6.

Thirteen genes in cluster 20 and eight genes in cluster 26 were enriched in the ‘immune system process’ (p-value = 0.036) and ‘regulation of immune system process’ (p-value = 0.0195), respectively. Immune senescence is the deterioration of immune function with age. As well as resistance to infection, immunosenescence may also reduce resistance to cancer and chronic activation of the immune system, usually as a result of autoimmune diseases, cancer, HIV infection and other chronic infections. The changes in immune response were found

to be very similar to the changes that occur in elderly individuals [67]. In response to aging most physiological functions are altered, e.g. the declination in cellular and humoral immunity. The most sensitive immune cells to aging appeared to be T cells, and the most critical component of immunological aging is known to be changes in the T lymphocyte compartment, concluded by studies on aging in humans [68], documenting significant changes in the functional and phenotypic profiles of T cells. Further analysis of literature has also suggested that the inability of the innate immune system to work efficiently is a contributing factor to the development of many diseases observed in the elderly [69].

Several genes shown in Table 3 have been found previously to have association with longevity, with many of them being associated with a decrease in life span. It has been found that *Drosophila*, heterozygous for the tumor suppressor gene *ft*, had a shorter lifespan, where it was suggested that this mortality effect was associated with the interaction between this *ft* tumour suppressor and signal transduction pathways mediated by the Hippo pathway [70]. Phenotype searches for genes in this table found *grim*, *Btk29A* and *tko* to express the phenotype ‘increased mortality’ whereas *Chmp1* was found to express the phenotype ‘short-lived’. An increase in the proapoptotic protein *grim* has been shown to significantly reduce lifespan in female *drosophila* by up to 34% in median lifespan and 25% in maximum lifespan [71]. The *Btk29A* and *Traf6* genes are FOXO targets in the JNK signaling pathway. This signaling pathway is stress-activated and involved in developmental and metabolic regulation, immune responses and lifespan extension [72–73].

The *Sod2* gene has been observed, in separate studies, to have both a positive and negative effect on lifespan in *Drosophila*. When overexpressed, the gene was found to result in a 20% increase in both mean and maximum lifespan [74] whereas RNA interference-mediated silencing of the *Sod2* gene caused an increase in oxidative stress leading to early-onset mortality in young adults [75]. The *PGRP-SA* gene has also been observed as one of few genes to show age-related changes in expression without being affected by diet, allowing this gene to be considered a candidate marker of aging [76].

#### SNPs in non-coding regions

Total of 26,499 and 653,030 non-coding SNPs were recorded in the Synthetic and DGRP GWAS datasets, respectively. First, we explored whether these SNPs tend to occur within border regions separating adjacent topologically associated domains (TADs). Second, using intra-chromosomal Hi-C data with finer resolution

**Table 3. Number of nodes in the selected clusters in the DGRP GWAS-base network and genes enriched in longevity-associated GO terms.**

Cluster	GO term	P-value	Genes enriched in GO term	Number of nodes harboring genes enriched in GO term/ total number of nodes
4	development growth	0.030823	<i>Elp3, ine, bdl, <b>ft*</b>, CASK, tsl</i>	4/12
18	nervous system process	0.033071	<i>Or59c, <u>bw**</u>, Gr59c, Gr59a, Gr59b, Gr59d, Gr59e, Gr59f, Or59b, Or59a, tko</i>	4/11
20	organelle assembly	0.016230	<i>Oseg2, Pp2A-29B, Rcd4, sls, Oseg4, CG42787, hts, Cnb, RpL11, Ar16, mtsh, RpL23A</i>	9/14
20	immune system process	0.035555	<i>CG10764, asrij, HBS1, sls, Rap1, ac, ecd, cnk, Ostgamma, Bgb, Bro, <b>Btk29A</b>, par-1</i>	9/14
26	regulation of immune system process	0.019532	<i><b>Traf6, PGRP-SA</b>, CG1572, Cyt-b5, GGBP3, GstO2, <b>Sod2</b>, Spn42Dd</i>	8/14
34	response to stimulus	0.031506	<i>geko, skl, AstC-R2, Adf1, Dic4, Trap1, geminin, Bap170, <b>Debcl</b>, <b>Chmp1</b>, GGBP2, not, CG4306, <b>rpr</b>, <b>grim</b>, hid, CG6893, GGBP1</i>	8/11
40	open tracheal system development	0.001555	<i>stumps, <u>Cad88C</u>, cv-c, grh, btsz, thr, put, scb</i>	5/14

\* Genes previously found to have association with longevity as recorded in FlyBase or GeneAge resources are shown in bold.

\*\*Genes residing within original nodes, i.e. harboring SNPs with  $p < 3.33 \times 10^{-5}$  are underlined.

we explore potential target genes for SNPs residing in non-coding regions utilizing co-location of SNP- and gene-harboring loci.

#### **SNPs in Topologically Associated Domain (TAD) boundary regions**

Approximately 2% (11,982) of all SNPs recorded in non-coding regions in the DGRP GWAS dataset were found in TAD boundary regions as compared to 9,321 SNPs in controls. Fisher's exact test shows that TAD boundary regions are enriched in SNPs ( $p = 1.0376 \times 10^{-75}$ ). These SNPs were found in 998 (~35%) of all TAD boundary regions. On the contrary, just a small proportion of SNPs from the Synthetic GWAS dataset were found within TAD boundary regions.

In the absence of individual genotype data, it is extremely difficult to assess the effect that SNPs in TAD borders may have on the genes residing within a given TAD. We assumed that one of the observed manifestations of latent changes in patterns of interactions between genomic regions could be in longevity-associated genes known to share both 'long-lived' and 'short-lived' phenotypes. We hypothesized that a SNP(s) in nearby TAD borders may lead to a disruption of a regulatory pattern of these genes resulting in one of the phenotypes, either 'long-lived' or

'short-lived', whereas the opposite phenotype could be caused by SNPs residing within genes themselves. Genomic positions were available for 124 out of 131 genes recorded in FlyBase resources as sharing both 'long-lived' and 'short-lived' phenotypes. We found that the majority of these genes, 106, were residing within 30 Kb regions spanning bins harboring a TAD border and including  $\pm 10$  Kb of flanking regions (i.e. two adjacent bins). From these genes, 43 were found to reside in the vicinity of 51 TAD borders that harbor SNPs; 89 genes were found to reside in the vicinity of 120 TAD borders that don't harbor SNPs. Thirty of these genes were found in the vicinity of both mutated and non-mutated TAD borders. (Note that the length of the topologically associated domains in our dataset for *Drosophila* varies between ~2 and 436 Kb that leads to the same gene being in the vicinity of two or more borders depending upon its length). Seventeen genes including *Charon*, *foxo*, *Jafrac1*, *mei-9* and *sun* occurred exclusively in the vicinity of mutated TAD borders (see Supplementary Table 7). Based on these observations one can speculate that SNPs residing in border regions of TADs may disrupt regulatory pattern of longevity related genes in the corresponding TADs by forming looping interactions with regulatory elements residing in the adjacent TADs potentially leading to the change of function, e.g. phenotype.

### Target genes for SNPs in non-coding regions

It is often assumed that a SNP residing in non-coding regions could potentially occur within a regulatory region(s) for a nearby gene(s). In many cases, if this nearby gene carries out biological functions which are related to the disease being studied, the SNPs found are automatically become a subject of further investigations. However, various looping interactions could happen between seemingly remote DNA fragments. For example, Sahlén et al. [77] have observed looping interactions between different promoters and postulated that promoters can also have enhancer activity influencing the expression of other genes not necessarily the nearest ones [78].

Analysis of the intra-chromosomal Hi-C data at 10 Kb resolution shows that in most cases interactions between adjacent bins are the strongest. This observation justifies the extension of gene boundaries to include SNPs residing within  $\pm 10$  Kb of flanking non-coding regions as it is often done in analyses of GWAS data. However, there are regions harboring non-coding SNPs recorded in the Synthetic GWAS dataset and DGRP GWAS datasets for which the strongest interacting regions were as distant as 50 Kb and 100 Kb, respectively. Thirty of these top long-range interacting pairs in both datasets were selected for further investigation. The summary of these regions is given in Supplementary Tables 8–9.

The calculated *D*-value for only one non-coding SNPs from the Synthetic GWAS dataset in the selected regions residing in bin 2006 (chr2L:20170000-20180000; Supplementary Table 8) exceeded the significance level ( $D=12.009 > 7.9$ ). The strongest interacting region for this bin was found 30 Kb upstream of the SNP. Not a single SNP chosen from the DGRP GWAS dataset residing within the selected regions meets the genome-wide significance level which in this case was set to  $4.22 \times 10^{-6}$ .

A total of 73 and 59 genes were found in the top 30 regions selected for SNPs from the Synthetic and DGRP GWAS datasets, respectively. Several bins, 1165, 4366, 7816 and 7887, which correspond to regions chr2L:11,680,000-11,690,000, chr3L:380,000-390,000, chr3R:10,760,000-10,770,000 and chr3R:11,470,000-11,480,000, respectively, were gene-less. Further analysis of potential target genes residing within long-range interacting regions using FlyBase resources have found many genes that share phenotypes that could be associated with longevity, e.g. ‘increased mortality’, ‘lethal’ and ‘immune response defective’ (Tables 4 and 5). Only one gene, *AttC*, encoding an immune inducible peptide homologous to antibacterial peptides having activity

against Gram-negative bacteria was previously considered to be a candidate marker of aging [76]. For the interacting bins containing more than one gene with longevity related phenotypes, we can speculate that non-coding SNPs could reside in an enhancer and this single enhancer may target all these genes, influencing their expressions and phenotypes. Phenotypes of potential target genes and their human orthologs are summarized in Tables 4–5.

Several genes with longevity-associated phenotypes were common between two datasets: *CG45186*, *CG4611*, *jing*, *Ca-alpha1D*, *Hml*, *CG32298*, *SNCF*, *CG14107*, *AttC*, *CG4597* and *CG43335*. The first three genes in this list have matched human orthologs *SVIL*, *PTCD1* and *AEBP2*, respectively. The *Ca-alpha1D* gene was found to match two human genes, *CACNA1D* and *CACNA1S*, whereas the *Hml* gene has four human orthologs: *SSPO*, *VWF*, *OTOG* and *MUC5B*. The human *SSPO* gene is involved in the modulation of neuronal aggregation and was suggested to be involved in developmental events during the formation of the central nervous system (<https://www.uniprot.org/uniprot/A2VEC9>). Dysregulation of the *CACNA1D* gene and loss-of-function mutations in the *SSPO* gene were found to associate with age-related diseases such as Alzheimer’s [79] and Parkinson’s [80]. Although no other human orthologs have been previously implicated in longevity, one can speculate that SNPs in non-coding regions may target these genes remotely in a similar way as was found in *Drosophila* and play a role in longevity.

## CONCLUSIONS

In this study we applied a conceptually new approach for identification of novel genes associated with longevity in *Drosophila* and provided the evidence for using co-location of genes/genomic regions governed by the 3D architecture of the *Drosophila* genome for predicting these novel genes. First, we created networks of interactions between genes and genomic regions harboring SNPs that meet a predefined level of significance for each GWAS dataset by using intra- and inter-chromosomal interaction frequencies (Hi-C data) as a measure of co-location. Then each of these networks was extended by adding regions that co-locate with the existing regions. We identified several genes residing within these newly added regions both known to associate with longevity and the novel ones that were not originally included in the analysis. Community detection algorithm identified several tightly-knit clusters in both networks. Genes residing within the same clusters were found to be enriched in longevity-related GO terms including ‘DNA repair’, ‘apoptotic process’, ‘nervous system process’ and ‘immune system

**Table 4. Phenotypes of genes, found in regions most strongly interacting with regions containing non-coding SNPs from the Synthetic GWAS dataset, and their human orthologs.**

SNP-harbours bin	Possible target gene	Longevity related phenotypes	Human ortholog
4566	<i>CG45186</i>	lethal; increased mortality during development; increased mortality	<i>SVIL</i>
	<i>CG32298</i>	partially lethal - majority die; flightless	
5660	<i>SNCF</i>	lethal - all die during P-stage	
	<i>CG14107</i>	partially lethal - majority die; some die during pupal stage; lethal - all die during P-stage	
1608	<b><i>Ca-alpha1D*</i></b>	increased mortality during development; lethal - all die before end of P-stage	<i>CACNA1D</i> , <i>CACNA1S</i>
2464	<i>jing 16</i>	locomotor behavior defective; cell death defective	<i>AEBP2</i>
3149	<i>AttC 3</i>	partially lethal; some die during pupal stage; neuroanatomy defective	
4840	<i>CG4597</i>	some die during pupal stage; partially lethal - majority die	
	<i>CG4611</i>	lethal - all die during P-stage	<i>PTCD1</i>
5705	<i>Hml</i>	immune response defective	<i>SSPO</i> , <i>VWF</i> , <i>OTOG</i> , <i>MUC5B</i>
7886	<i>CG43335</i>	partially lethal - majority die; some die during pupal stage; partially lethal	
554	<i>GluRIIA</i>	locomotor behavior defective; neurophysiology defective; neuroanatomy defective; lethal	
554	<i>GluRIIB</i>	neuroanatomy defective; neurophysiology defective	
943	<i>numb</i>	decreased cell number; some die during embryonic stage; increased mortality; increased cell number; lethal - all die before end of prepupal stage; flight defective; tumorigenic	<i>NUMBL</i>
996	<i>bib</i>	lethal - all die before end of pupal stage	
1176	<b><i>crol</i></b>	locomotor behavior defective; increased occurrence of cell division; increased mortality; cell death defective	<i>ZNF569</i> , <i>ZNF99</i> , <i>ZNF841</i> , <i>ZNF814</i>

\* Genes previously found to have association with longevity as recorded in FlyBase or GeneAge resources are shown in bold.

process'. Using literature search and additional bioinformatics resources we investigated the plausibility of genes found to have genuine association with longevity. Our network approach identified several novel genes (see Tables 1–3) with no prior known associations with longevity as well as genes with prior reported associations with longevity, acting as a proof of concept. Among these genes are the *Vps16B*, *Cad99C*, *DPCoAC* (*CG4241*) and *Mtl* genes residing within important nodes in the DGRP GWAS-based network and, together with the known longevity-associated genes *aop*, *Stat92E*, *GlyP*, being enriched in 'immune system process' GO term. The *Cad99C* gene, which encodes a member of the cadherin superfamily of transmembrane proteins, harbors an SNP (chr3R:25674492) which is present in both GWAS datasets. Another novel gene, *CG5316*, co-located with genes *CG17329*, *CG31807*, *CG33552*, *EndoGI* and the longevity-associated gene *ku80* in the Synthetic

GWAS-based network, was found to be enriched in 'DNA repair' GO term. Although the function of this gene is unknown, its human ortholog – the *APTX* gene – was implicated in longevity [41]. A group of genes – *CG1572*, *Cyt-b5*, *GNBP3*, *GstO2* and *Spn42Dd* – residing within the cluster 26 in the DGRP GWAS-based network together with the known longevity associated genes *Traf6*, *PGRP-SA* and *Sod2* are also strong candidates for novel longevity genes. None of these genes harbor SNPs that reach genome-wide significance level in the DGRP GWAS dataset. One gene, the glutathione S transferase O2 gene *GstO2*, harbors one SNP with genome-wide significant D-value 8.38 in the Synthetic GWAS dataset. Interestingly, all these genes reside in close proximity (i.e. within 30 Kb region) to TAD borders. Three of the nearest TAD border regions harbor SNPs prompting speculations that these SNPs could influence longevity either separately or together with SNPs residing within the



**Table 5. Phenotypes of genes, found in regions most strongly interacting with regions containing non-coding SNPs from the DGRP GWAS dataset, and their human orthologs.**

SNP-harboured bin	Possible target gene	Longevity related phenotypes	Human ortholog
<b>2962</b>	<i>en</i>	lethal - all die during embryonic stage; size defective; planar polarity defective; increased cell death; some die during pupal stage; partially lethal - majority die	
<b>2453</b>	<i>Pld</i>	developmental rate defective; partially lethal - majority die; some die during embryonic stage; neurophysiology defective; lethal - all die before end of embryonic stage	<i>PLD2</i>
<b>2463-2464</b>	<i>jing</i>	locomotor behavior defective; cell death defective	<i>AEBP2</i>
<b>4366</b>	<i>trh</i>	neuroanatomy defective; partially lethal - majority die; lethal; some die during embryonic stage; lethal - all die before end of embryonic stage	<i>NPASI</i>
<b>4566</b>	<i>CG45186</i>	lethal; increased mortality during development; increased mortality	<i>SVIL</i>
	<i>CG32298</i>	some die during pupal stage; partially lethal - majority die; flightless	
<b>5660</b>	<i>SNCF</i>	lethal - all die during P-stage	
	<i>CG14107</i>	partially lethal - majority die; some die during pupal stage; lethal - all die during P-stage	
<b>1608</b>	<b><i>Ca-alpha1D</i>*</b>	increased mortality during development; lethal - all die before end of P-stage	<i>CACNAID</i> , <i>CACNAIS</i>
<b>2286</b>	<i>RpL38</i>	increased mortality; increased mortality during development; developmental rate defective	<i>RPL38</i>
<b>2375</b>	<i>laccase2</i>	lethal; partially lethal; lethal - all die during embryonic stage;	<i>HEPHL1</i> , <i>CP</i> , <i>HEPH</i>
<b>3149</b>	<i>AttC</i>	partially lethal; some die during pupal stage; neuroanatomy defective	
<b>4840</b>	<i>CG4597</i>	some die during pupal stage; partially lethal - majority die	<i>PTCD1</i>
	<i>CG4611</i>	lethal - all die during P-stage	
<b>5705</b>	<i>Hml</i>	immune response defective	<i>SSPO</i> , <i>VWF</i> , <i>OTOG</i> , <i>MUC5B</i>
<b>7636</b>	<i>timeout 1</i>	increased mortality; lethal - all die before end of P-stage; some die during P-stage	<i>TIMELESS</i>
<b>7886</b>	<i>CG43335</i>	partially lethal - majority die; some die during pupal stage; partially lethal	
<b>8955</b>	<i>CG33970</i>	lethal; sleep defective; flightless	

\* Genes previously found to have association with longevity as recorded in FlyBase or GeneAge resources are shown in bold.

genes. This observation could explain the enrichment of TAD borders in SNPs. Many of these newly found genes harbor SNPs that do not reach a predefined genome-wide significance level, leading to speculation that SNPs residing within genes enriched in the same GO term may influence longevity collectively (when one or several SNPs in these functionally-related genes occur in the same fly to cause a phenotype) rather than individually when a single SNP in one of these genes could cause a

phenotype. Further, we explored potential target genes for SNPs in non-coding regions also assuming co-location of SNP-harboring loci and target genes within cell nucleus. Several novel target genes for non-coding SNPs have been identified using our network approach (see Tables 4–5) including genes such as *CG45186*, *CG4611*, *Ca-alpha1D*, *Hml*, and *AttC* that are common between two datasets and have human orthologs associated with age-related diseases.

Further experimental validation is required in order to establish the functional significance of the genes and SNP-target gene pairs found. This computational approach promises to be a stepping stone to identification of novel genes and SNP targets linked to longevity in humans and further understanding of genetic factors associated with this phenotype.

## MATERIALS AND METHODS

### *Drosophila* GWAS data used in this study

Two datasets of SNPs summary statistics generated by GWA studies reported in Burke et al. [19] and Ivanov et al. [20], containing respectively ~1 million and ~2 million SNPs, were used in this study. The first dataset was obtained from a “synthetic” population of *Drosophila* derived from a small number of inbred founders. Two independent sets of seven inbred *Drosophila* lines with another founding line added to both sets, were crossed to initiate two synthetic recombinant populations, A and B. Populations A and B were then maintained as four independent large populations (A1/A2, B1/B2). Next-generation sequencing was used to identify allele frequencies in the ‘young’ control group, comprising 120 14-day-old females, and the last surviving ~2% of females from the remaining cohort (an ‘old’ group). The occurrence of SNP in each of the eight ‘old’ populations and eight ‘young’ control populations was recorded, resulting in ~1.2M SNPs in the A populations and ~1.1M SNPs in the B populations (see [19] for details). The SNPs for both populations were combined; duplicated SNPs were recorded only once with haplotype allele frequencies combined. Henceforth, we will refer to this data set as the Synthetic GWAS data.

The second dataset was obtained by GWAS performed on The *Drosophila* Genetic Reference Panel (DRGP), Freeze 2.0 [20], which comprises 205 *D. melanogaster* lines derived from 20 generations of full-sib mating from inseminated wild-type caught females from Raleigh, North Carolina. Lifespan data was available for virgin females for 197 DGRP lines, with ~25 females per line. A total of 2,193,745 SNPs was recorded together with the corresponding p-values, quantifying association with lifespan. P-values were calculated using linear regression under an additive model with four first principal components and the presence of *Wolbachia pipientis* included as a covariate (see [20] for details). Henceforth, we will refer to this data set as the DRGP GWAS data.

SNPs were considered to be in coding regions of certain genes if they resided between gene start and gene end

positions as defined by BDGP Release 6/dm6 assembly [81] and recorded in FlyBase database (<http://flybase.org/>). All other SNPs were considered to reside within non-coding regions.

### Intra- and inter-chromosomal interaction (Hi-C) data

A dataset of intra- and inter-chromosomal normalized contacts (interaction frequencies) between 1503 80 Kb regions (bins) obtained by Sexton et al. [21] was downloaded from GEO database (accession number GSM849422). In this dataset bins 1-287 correspond to Chromosome 2L, bins 288-551 to Chromosome 2R, bins 552-858 to Chromosome 3L, bins 859-1207 to Chromosome 3R, bins 1208-1223 to Chromosome 4 and bins 1224-1503 to Chromosome X. In addition, a dataset of intra-chromosomal interaction between 11,839 10Kb regions was downloaded from the same GEO database.

### Dataset of Topologically Associated Domain (TAD) boundary regions

For each of 2,847 TAD borders compiled by [82] 100 bp of flanking sequences were added from both sides to create a dataset of TAD border regions. Using SNP position data, non-coding SNPs residing in each TAD border region in the GWAS datasets were counted. Matched control dataset was generated as follows. For each TAD border a random border was generated by randomly selecting a position on the matching chromosome and adding ±100 bp of flanking sequences not overlapping with any “real” border region. This process was repeated 100 times. The number of SNPs residing within generated control datasets were counted and averaged across 100 control datasets. Fisher’s Exact Test was then used to assess the overrepresentations of SNPs within TAD borders.

### Pre-processing of Synthetic GWAS data

To identify positions with divergent haplotype frequencies in the young (control) and old groups in this dataset, Euclidean distances between the control and old groups were calculated for haplotype data for populations A1/A2 and B1/B2 combined. All duplicates were removed. The distance,  $D$ , for a given SNP was calculated as suggested in Burke et al. [19]:

$$D = 100 \cdot \sqrt{\frac{\sum_{j=1}^n (h_{O,j} - h_{Y,j})^2}{n}}$$

where  $h_{oj}$  is the haplotype frequency of the  $j^{\text{th}}$  founder in the old samples,  $h_{Yj}$  is the haplotype frequency of the  $j^{\text{th}}$  founder in the young control sample,  $n$  is the number of haplotypes found at that position. For SNP positions with combined haplotype frequencies, the  $n$  in the equation used was altered accordingly. SNP positions with the largest calculated  $D$  values were those showing the largest differences between haplotype frequencies in the control and old groups, and it was therefore these SNPs that were indicated as most likely to have association with longevity. Following Burke et al. [19],  $D > 7.9$  was considered to correspond to genome-wide significance  $p\text{-value} < 0.05$ .

### Identification of candidate longevity genes and potential regulatory regions

To align SNP positions with Hi-C data, SNPs were binned into 80 Kb regions. Start and end positions of each bin and corresponding chromosomes are given in Supplementary Table 10. Each region, harboring SNP(s) meeting the  $D > 7.9$  threshold in the Synthetic GWAS data or a predefined  $p\text{-value}$  threshold (see in the Results and Discussion section) in the DRGP GWAS data, was identified and considered as a node of an *original* network of interactions. Links between the nodes were added to create a network of interactions when an intra- or inter-chromosomal interaction between two nodes was recorded in Hi-C data and the frequency of interaction exceeded a certain threshold. Further, the resulting network was expanded to create an *extended network* by adding extra nodes, corresponding to 80 Kb fragments that interact with the nodes, already present in the original network, with frequencies exceeding a predefined threshold. For each node in networks its degree, i.e. the number of connections a given node has with other nodes, was calculated with the aim of finding influential nodes/regions and novel genes not necessarily covered by GWAS SNP array. In addition, the Louvain modularity method [33] was used to detect communities within the resulting networks, i.e. groups of nodes/regions that are densely connected to each other within a given community but sparsely connected to nodes in other communities of the network. All measures were calculated using GEPHI software tool available at <https://gephi.org/>.

Genes residing within important nodes/regions of interest were identified using genomic coordinates corresponding to the BDGP Release 6/dm6 assembly [81] downloaded from the FlyBase database (<http://flybase.org/>). To align the Hi-C data and GWAS SNP positions, all gene positions were lifted over to BDGP Release 5/dm3. This was done using a

LiftOver tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>).

### Gene ontology enrichment analysis

The FlyMine software (<http://www.flymine.org/>) was used to analyze the enrichment of the set of genes, residing within important nodes/clusters, in Gene Ontology (GO) terms for cellular component, biological process and molecular function. Each gene was also sought in the GeneAge database (<http://genomics.senescence.info/genes/models.html>) of longevity genes and FlyBase resources (<http://flybase.org/>) as having longevity-related phenotype.

### ACKNOWLEDGMENTS

The authors are grateful to Dr MK Burke and Dr DK Ivanov for making their GWAS summary statistics available for analysis.

### CONFLICTS OF INTEREST

All authors declare no conflicts of interest.

### FUNDING

The work of BSH was supported by a Vice-Chancellor PhD bursary at Nottingham Trent University (UK).

### REFERENCES

1. Herskind AM, McGue M, Holm NV, Sørensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet.* 1996; 97:319–23. <https://doi.org/10.1007/BF02185763> PMID:8786073
2. Skytthe A, Pedersen NL, Kaprio J, Stazi MA, Hjelmberg JV, Iachine I, Vaupel JW, Christensen K. Longevity studies in GenomEUtwin. *Twin Res.* 2003; 6:448–54. <https://doi.org/10.1375/136905203770326457> PMID:14624729
3. Puca AA, Daly MJ, Brewster SJ, Matise TC, Barrett J, Shea-Drinkwater M, Kang S, Joyce E, Nicoli J, Benson E, Kunkel LM, Perls T. A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proc Natl Acad Sci USA.* 2001; 98:10505–08. <https://doi.org/10.1073/pnas.181337598> PMID:11526246
4. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, Arnold AM, Aspelund T,

- Aulchenko YS, Benjamin EJ, Christiansen L, D'Agostino RB Sr, Fitzpatrick AL, et al. A meta-analysis of four genome-wide association studies of survival to age 90 years or older: the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. *J Gerontol A Biol Sci Med Sci*. 2010; 65:478–87. <https://doi.org/10.1093/gerona/glq028> PMID:20304771
5. Joshi PK, Pirastu N, Kentistou KA, Fischer K, Hofer E, Schraut KE, Clark DW, Nutile T, Barnes CL, Timmers PR, Shen X, Gandin I, McDaid AF, et al. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat Commun*. 2017; 8:910. <https://doi.org/10.1038/s41467-017-00934-5> PMID:29030599
  6. Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, Kremer D, van der Breggen R, Suchiman HE, Lakenberg N, van den Akker EB, Passtoors WM, Tiemeier H, et al. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell*. 2011; 10:686–98. <https://doi.org/10.1111/j.1474-9726.2011.00705.x> PMID:21418511
  7. Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB, Myers RH, Steinberg MH, Montano M, et al. Genetic signatures of exceptional longevity in humans. *PLoS One*. 2012; 7:e29848. <https://doi.org/10.1371/journal.pone.0029848> PMID:22279548
  8. Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, Tamm R, Trompet S, Guðbjartsson DF, Flachsbarth F, Rose G, et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet*. 2014; 23:4420–32. <https://doi.org/10.1093/hmg/ddu139> PMID:24688116
  9. Lu F, Guan H, Gong B, Liu X, Zhu R, Wang Y, Qian J, Zhou T, Lan X, Wang P, Lin Y, Ma S, Lin H, et al. Genetic variants in PVRL2-TOMM40-APOE region are associated with human longevity in a Han Chinese population. *PLoS One*. 2014; 9:e99580. <https://doi.org/10.1371/journal.pone.0099580> PMID:24924924
  10. Broer L, Buchman AS, Deelen J, Evans DS, Faul JD, Lunetta KL, Sebastiani P, Smith JA, Smith AV, Tanaka T, Yu L, Arnold AM, Aspelund T, et al. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J Gerontol A Biol Sci Med Sci*. 2015; 70:110–18. <https://doi.org/10.1093/gerona/glu166> PMID:25199915
  11. Zeng Y, Nie C, Min J, Liu X, Li M, Chen H, Xu H, Wang M, Ni T, Li Y, Yan H, Zhang JP, Song C, et al. Novel loci and pathways significantly associated with longevity. *Sci Rep*. 2016; 6:21243–58. <https://doi.org/10.1038/srep21243> PMID:26912274
  12. Kerber RA, O'Brien E, Boucher KM, Smith KR, Cawthon RM. A genome-wide study replicates linkage of 3p22-24 to extreme longevity in humans and identifies possible additional loci. *PLoS One*. 2012; 7:e34746. <https://doi.org/10.1371/journal.pone.0034746> PMID:22506048
  13. Boyden SE, Kunkel LM. High-density genomewide linkage analysis of exceptional human longevity identifies multiple novel loci. *PLoS One*. 2010; 5:e12432. <https://doi.org/10.1371/journal.pone.0012432> PMID:20824210
  14. Pilling LC, Kuo CL, Sicinski K, Tamosauskaite J, Kuchel GA, Harries LW, Herd P, Wallace R, Ferrucci L, Melzer D. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)*. 2017; 9:2504–20. <https://doi.org/10.18632/aging.101334> PMID:29227965
  15. Linford NJ, Bilgir C, Ro J, Pletcher SD. Measurement of lifespan in *Drosophila melanogaster*. *J Vis Exp*. 2013; 7:50068. <https://doi.org/10.3791/50068> PMID:23328955
  16. Lin YJ, Seroude L, Benzer S. Extended life-span and stress resistance in the *Drosophila* mutant methuselah. *Science*. 1998; 282:943–46. <https://doi.org/10.1126/science.282.5390.943> PMID:9794765
  17. Rogina B, Reenan RA, Nilsen SP, Helfand SL. Extended life-span conferred by cotransporter gene mutations in *Drosophila*. *Science*. 2000; 290:2137–40. <https://doi.org/10.1126/science.290.5499.2137> PMID:11118146
  18. Fontana L, Partridge L, Longo VD. Extending healthy life span—from yeast to humans. *Science*. 2010; 328:321–26. <https://doi.org/10.1126/science.1172539> PMID:20395504
  19. Burke MK, King EG, Shahrestani P, Rose MR, Long AD. Genome-wide association study of extreme longevity in *Drosophila melanogaster*. *Genome Biol Evol*. 2014; 6:1–11. <https://doi.org/10.1093/gbe/evt180> PMID:24259311
  20. Ivanov DK, Escott-Price V, Ziehm M, Magwire MM, Mackay TF, Partridge L, Thornton JM. Longevity GWAS using the *Drosophila* genetic reference panel. *J*



- Gerontol A Biol Sci Med Sci. 2015; 70:1470–78.  
<https://doi.org/10.1093/gerona/glv047>  
 PMID:25922346
21. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*. 2012; 148:458–72. <https://doi.org/10.1016/j.cell.2012.01.010>  
 PMID:22265598
  22. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009; 326:289–93.  
<https://doi.org/10.1126/science.1181369>  
 PMID:19815776
  23. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159:1665–80.  
<https://doi.org/10.1016/j.cell.2014.11.021>  
 PMID:25497547
  24. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, Gandal MJ, Sutton GJ, Hormozdiari F, Lu D, Lee C, Eskin E, Voineagu I, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016; 538:523–27.  
<https://doi.org/10.1038/nature19847>  
 PMID:27760116
  25. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet*. 2016; 32:225–37.  
<https://doi.org/10.1016/j.tig.2016.01.003>  
 PMID:26862051
  26. Estrada E. The structure of complex networks: Theory and applications. Oxford University Press; 2011.  
<https://doi.org/10.1093/acprof:oso/9780199591756.01.0001>
  27. Fernando MD, Kounatidis I, Ligoxygakis P. Loss of Trabad, a new negative regulator of the drosophila immune-deficiency pathway at the level of TAK1, reduces life span. *PLoS Genet*. 2014; 10:e1004117.  
<https://doi.org/10.1371/journal.pgen.1004117>  
 PMID:24586180
  28. Gim BS, Park JM, Yoon JH, Kang C, Kim YJ. Drosophila Med6 is required for elevated expression of a large but distinct set of developmentally regulated genes. *Mol Cell Biol*. 2001; 21:5242–55.  
<https://doi.org/10.1128/MCB.21.15.5242-5255.2001>  
 PMID:11438678
  29. Castells-Nobau A, Eidhof I, Fencikova M, Brenman-Suttner DB, Scheffer-de Gooyert JM, Christine S, Schellevis RL, van der Laan K, Quentin C, van Nijhuijs L, Hofmann F, Ejsmont R, Fisher SE, et al. Conserved regulation of neurodevelopmental processes and behavior by FoxP in Drosophila. *PLoS One*. 2019; 14:e0211652.  
<https://doi.org/10.1371/journal.pone.0211652>  
 PMID:30753188
  30. Simonsen A, Cumming RC, Lindmo K, Galaviz V, Cheng S, Rusten TE, Finley KD. Genetic modifiers of the Drosophila blue cheese gene link defects in lysosomal transport with decreased life span and altered ubiquitinated-protein profiles. *Genetics*. 2007; 176:1283–97.  
<https://doi.org/10.1534/genetics.106.065011>  
 PMID:17435236
  31. Tsai HZ, Lin RK, Hsieh TS. Drosophila mitochondrial topoisomerase III alpha affects the aging process via maintenance of mitochondrial function and genome integrity. *J Biomed Sci*. 2016; 23:38.  
<https://doi.org/10.1186/s12929-016-0255-2>  
 PMID:27067525
  32. Carvalho AB. Origin and evolution of the Drosophila Y chromosome. *Curr Opin Genet Dev*. 2002; 12:664–68.  
[https://doi.org/10.1016/S0959-437X\(02\)00356-8](https://doi.org/10.1016/S0959-437X(02)00356-8)  
 PMID:12433579
  33. Larson K, Yan SJ, Tsurumi A, Liu J, Zhou J, Gaur K, Guo D, Eickbush TH, Li WX. Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. *PLoS Genet*. 2012; 8:e1002473.  
<https://doi.org/10.1371/journal.pgen.1002473>  
 PMID:22291607
  34. Bai H, Kang P, Hernandez AM, Tatar M. Activin signaling targeted by insulin/dFOXO regulates aging and muscle proteostasis in Drosophila. *PLoS Genet*. 2013; 9:e1003941.  
<https://doi.org/10.1371/journal.pgen.1003941>  
 PMID:24244197
  35. Orr WC, Sohal RS. Extension of life-span by overexpression of superoxide dismutase and catalase in Drosophila melanogaster. *Science*. 1994; 263:1128–30.  
<https://doi.org/10.1126/science.8108730>  
 PMID:8108730
  36. Slack C, Alic N, Foley A, Cabecinha M, Hoddinott MP, Partridge L. The Ras-Erk-ETS-signaling pathway is a drug target for longevity. *Cell*. 2015; 162:72–83.  
<https://doi.org/10.1016/j.cell.2015.06.023>  
 PMID:26119340
  37. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech*. 2008; 10:10008.

<https://doi.org/10.1088/1742-5468/2008/10/P10008>

38. Lambiotte R, Delvenne JC, Barahona M. Laplacian dynamics and multiscale modular structure in networks. 2008; arXiv preprint. <https://arxiv.org/abs/0812.1770>
39. Shaposhnikov M, Proshkina E, Shilova L, Zhavoronkov A, Moskalev A. Lifespan and stress resistance in *Drosophila* with overexpressed DNA repair genes. *Sci Rep*. 2015; 5:15299. <https://doi.org/10.1038/srep15299> PMID:26477511
40. Gaiano N, Fishell G. The role of notch in promoting glial and neural stem cell fates. *Annu Rev Neurosci*. 2002; 25:471–90. <https://doi.org/10.1146/annurev.neuro.25.030702.130823> PMID:12052917
41. Kao AW, Eisenhut RJ, Martens LH, Nakamura A, Huang A, Bagley JA, Zhou P, de Luis A, Neukomm LJ, Cabello J, Farese RV Jr, Kenyon C. A neurodegenerative disease mutation that accelerates the clearance of apoptotic cells. *Proc Natl Acad Sci USA*. 2011; 108:4441–6. <https://doi.org/10.1073/pnas.1100650108> PMID:21368173
42. Soerensen M, Dato S, Tan Q, Thinggaard M, Kleindorp R, Beekman M, Jacobsen R, Suchiman HE, de Craen AJ, Westendorp RG, Schreiber S, Stevnsner T, Bohr VA, et al. Human longevity and variation in GH/IGF-1/insulin signaling, DNA damage signaling and repair and pro/antioxidant pathway genes: cross sectional and longitudinal studies. *Exp Gerontol*. 2012; 47:379–87. <https://doi.org/10.1016/j.exger.2012.02.010> PMID:22406557
43. Warner HR. Apoptosis: a two-edged sword in aging. *Ann N Y Acad Sci*. 1999; 887:1–11. <https://doi.org/10.1111/j.1749-6632.1999.tb07917.x> PMID:10668459
44. Higami Y, Shimokawa I. Apoptosis in the aging process. *Cell Tissue Res*. 2000; 301:125–32. <https://doi.org/10.1007/s004419900156> PMID:10928285
45. Proshkina EN, Shaposhnikov MV, Sadritdinova AF, Kudryavtseva AV, Moskalev AA. Basic mechanisms of longevity: A case study of *Drosophila* pro-longevity genes. *Ageing Res Rev*. 2015; 24:218–31. <https://doi.org/10.1016/j.arr.2015.08.005> PMID:26318059
46. Payton A, van den Boogerd E, Davidson Y, Gibbons L, Ollier W, Rabbitt P, Worthington J, Horan M, Pendleton N. Influence and interactions of cathepsin D, HLA-DRB1 and APOE C. *Genes Brain Behav*. 2006 (Suppl 1); 5:23–31. <https://doi.org/10.1111/j.1601-183X.2006.00191.x> PMID:16417614
47. Schuur M, Ikram MA, van Swieten JC, Isaacs A, Vergeer-Drop JM, Hofman A, Oostra BA, Breteler MM, van Duijn CM. Cathepsin D gene and the risk of Alzheimer's disease: a population-based study and meta-analysis. *Neurobiol Aging*. 2011; 32:1607–14. <https://doi.org/10.1016/j.neurobiolaging.2009.10.011> PMID:19926167
48. Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafen E, Leivers SJ, Partridge L. Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science*. 2001; 292:104–06. <https://doi.org/10.1126/science.1057991> PMID:11292874
49. Koohy H, Bolland DJ, Matheson LS, Schoenfelder S, Stellato C, Dimond A, Várnai C, Chovanec P, Chessa T, Denizot J, Manzano Garcia R, Wingett SW, Freire-Pritchett P, et al. Genome organization and chromatin analysis identify transcriptional downregulation of insulin-like growth factor signaling as a hallmark of aging in developing B cells. *Genome Biol*. 2018; 19:126. <https://doi.org/10.1186/s13059-018-1489-y> PMID:30180872
50. Demontis F, Perrimon N. FOXO/4E-BP signaling in *Drosophila* muscles regulates organism-wide proteostasis during aging. *Cell*. 2010; 143:813–25. <https://doi.org/10.1016/j.cell.2010.10.007> PMID:21111239
51. Chiacchiera F, Simone C. The AMPK-FoxO3A axis as a target for cancer treatment. *Cell Cycle*. 2010; 9:1091–96. <https://doi.org/10.4161/cc.9.6.11035> PMID:20190568
52. Sanese P, Forte G, Disciglio V, Grossi V, Simone C. FOXO3 on the road to longevity: lessons from SNPs and chromatin hubs. *Comput Struct Biotechnol J*. 2019; 17:737–45. <https://doi.org/10.1016/j.csbj.2019.06.011> PMID:31303978
53. Giannakou ME, Goss M, Jünger MA, Hafen E, Leivers SJ, Partridge L. Long-lived *Drosophila* with overexpressed dFOXO in adult fat body. *Science*. 2004; 305:361–361. <https://doi.org/10.1126/science.1098219> PMID:15192154
54. Giannakou ME, Partridge L. Role of insulin-like signalling in *Drosophila* lifespan. *Trends Biochem Sci*. 2007; 32:180–88. <https://doi.org/10.1016/j.tibs.2007.02.007> PMID:17412594
55. Giannakou ME, Goss M, Jacobson J, Vinti G, Leivers SJ, Partridge L. Dynamics of the action of dFOXO on adult mortality in *Drosophila*. *Aging Cell*. 2007; 6:429–38. <https://doi.org/10.1111/j.1474-9726.2007.00290.x> PMID:17465980

56. Hwangbo DS, Gershman B, Tu MP, Palmer M, Tatar M. Drosophila dFOXO controls lifespan and regulates insulin signalling in brain and fat body. *Nature*. 2004; 429:562–66. <https://doi.org/10.1038/nature02549> PMID:15175753
57. Ulgherait M, Rana A, Rera M, Graniel J, Walker DW. AMPK modulates tissue and organismal aging in a non-cell-autonomous manner. *Cell Rep*. 2014; 8:1767–80. <https://doi.org/10.1016/j.celrep.2014.08.006> PMID:25199830
58. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017; 45:D353–61. <https://doi.org/10.1093/nar/gkw1092> PMID:27899662
59. Salas-Pérez F, Ramos-Lopez O, Mansego ML, Milagro FI, Santos JL, Riezu-Boj JL, Martínez JA. DNA methylation in genes of longevity-regulating pathways: association with obesity and metabolic complications. *Aging (Albany NY)*. 2019; 11:1874–99. <https://doi.org/10.18632/aging.101882> PMID:30926763
60. Biteau B, Karpac J, Supoyo S, Degennaro M, Lehmann R, Jasper H. Lifespan extension by preserving proliferative homeostasis in Drosophila. *PLoS Genet*. 2010; 6:e1001159. <https://doi.org/10.1371/journal.pgen.1001159> PMID:20976250
61. Landis GN, Bhole D, Tower J. A search for doxycycline-dependent mutations that increase Drosophila melanogaster life span identifies the VhaSFD, Sugar baby, filamin, fwd and Cctl genes. *Genome Biol*. 2003; 4:R8. <https://doi.org/10.1186/gb-2003-4-2-r8> PMID:12620118
62. Magwire MM, Yamamoto A, Carbone MA, Roshina NV, Symonenko AV, Pasyukova EG, Morozova TV, Mackay TF. Quantitative and molecular genetic analyses of mutations increasing Drosophila life span. *PLoS Genet*. 2010; 6:e1001037. <https://doi.org/10.1371/journal.pgen.1001037> PMID:20686706
63. McQuibban GA, Lee JR, Zheng L, Juusola M, Freeman M. Normal mitochondrial dynamics requires rhomboid-7 and affects Drosophila lifespan and neuronal function. *Curr Biol*. 2006; 16:982–89. <https://doi.org/10.1016/j.cub.2006.03.062> PMID:16713954
64. Zerofsky M, Harel E, Silverman N, Tatar M. Aging of the innate immune response in Drosophila melanogaster. *Aging Cell*. 2005; 4:103–08. <https://doi.org/10.1111/j.1474-9728.2005.00147.x> PMID:15771614
65. Xue L, Igaki T, Kuranaga E, Kanda H, Miura M, Xu T. Tumor suppressor CYLD regulates JNK-induced cell death in Drosophila. *Dev Cell*. 2007; 13:446–54. <https://doi.org/10.1016/j.devcel.2007.07.012> PMID:17765686
66. Papaconstantinou M, Wu Y, Pretorius HN, Singh N, Gianfelice G, Tanguay RM, Campos AR, Bédard PA. Menin is a regulator of the stress response in Drosophila melanogaster. *Mol Cell Biol*. 2005; 25:9960–72. <https://doi.org/10.1128/MCB.25.22.9960-9972.2005> PMID:16260610
67. Tarazona R, Solana R, Ouyang Q, Pawelec G. Basic biology and clinical impact of immunosenescence. *Exp Gerontol*. 2002; 37:183–89. [https://doi.org/10.1016/S0531-5565\(01\)00182-6](https://doi.org/10.1016/S0531-5565(01)00182-6) PMID:11772502
68. Goronzy JJ, Fang F, Cavanagh MM, Qi Q, Weyand CM. Naive T cell maintenance and function in human aging. *J Immunol*. 2015; 194:4073–80. <https://doi.org/10.4049/jimmunol.1500046> PMID:25888703
69. Gomez CR, Nomellini V, Faunce DE, Kovacs EJ. Innate immunity and aging. *Exp Gerontol*. 2008; 43:718–28. <https://doi.org/10.1016/j.exger.2008.05.016> PMID:18586079
70. Kopyl SA, Omelyanchuk LV, Shaposhnikov MV, Moskalev AA. Role of tumor suppressor genes in aging and longevity mechanisms in Drosophila melanogaster. *Russian Journal of Genetics: Applied Research*. 2014; 4:8–14. <https://doi.org/10.1134/S2079059714010043>
71. Bauer JH, Poon PC, Glatt-Deeley H, Abrams JM, Helfand SL. Neuronal expression of p53 dominant-negative proteins in adult Drosophila melanogaster extends life span. *Curr Biol*. 2005; 15:2063–68. <https://doi.org/10.1016/j.cub.2005.10.051> PMID:16303568
72. Biteau B, Karpac J, Hwangbo D, Jasper H. Regulation of Drosophila lifespan by JNK signaling. *Exp Gerontol*. 2011; 46:349–54. <https://doi.org/10.1016/j.exger.2010.11.003> PMID:21111799
73. Karpac J, Hull-Thompson J, Falleur M, Jasper H. JNK signaling in insulin-producing cells is required for adaptive responses to stress in Drosophila. *Aging Cell*. 2009; 8:288–95. <https://doi.org/10.1111/j.1474-9726.2009.00476.x> PMID:19627268
74. Curtis C, Landis GN, Folk D, Wehr NB, Hoe N, Waskar M, Abdueva D, Skvortsov D, Ford D, Luu A, Badrinath A, Levine RL, Bradley TJ, et al. Transcriptional

- profiling of MnSOD-mediated lifespan extension in *Drosophila* reveals a species-general network of aging and metabolic genes. *Genome Biol.* 2007; 8:R262.  
<https://doi.org/10.1186/gb-2007-8-12-r262>  
PMID:18067683
75. Kirby K, Hu J, Hilliker AJ, Phillips JP. RNA interference-mediated silencing of Sod2 in *Drosophila* leads to early adult-onset mortality and elevated endogenous oxidative stress. *Proc Natl Acad Sci USA.* 2002; 99:16162–67.  
<https://doi.org/10.1073/pnas.252342899>  
PMID:12456885
  76. Doroszuk A, Jonker MJ, Pul N, Breit TM, Zwaan BJ. Transcriptome analysis of a long-lived natural *Drosophila* variant: a prominent role of stress- and reproduction-genes in lifespan extension. *BMC Genomics.* 2012; 13:167.  
<https://doi.org/10.1186/1471-2164-13-167>  
PMID:22559237
  77. Sahlén P, Abdullayev I, Ramsköld D, Matskova L, Rilakovic N, Lötstedt B, Albert TJ, Lundeberg J, Sandberg R. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.* 2015; 16:156.  
<https://doi.org/10.1186/s13059-015-0727-9>  
PMID:26313521
  78. Andersson R, Sandelin A, Danko CG. A unified architecture of transcriptional regulatory elements. *Trends Genet.* 2015; 31:426–33.  
<https://doi.org/10.1016/j.tig.2015.05.007>  
PMID:26073855
  79. Villela D, Suemoto CK, Pasqualucci CA, Grinberg LT, Rosenberg C. Do copy number changes in CACNA2D2, CACNA2D3, and CACNA1D constitute a predisposing risk factor for Alzheimer's disease? *Front Genet.* 2016; 7:107.  
<https://doi.org/10.3389/fgene.2016.00107>  
PMID:27379157
  80. Chew EG, Liany H, Tan LC, Au WL, Prakash KM, Annur AA, Chan AY, Lim SY, Mok V, Chung SJ, Song K, Liu J, Foo JN, Tan EK. Evaluation of novel Parkinson's disease candidate genes in the Chinese population. *Neurobiol Aging.* 2019; 74:235.e1–4.  
<https://doi.org/10.1016/j.neurobiolaging.2018.09.013>  
PMID:30337193
  81. dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, and FlyBase Consortium. FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* 2015; 43:D690–97.  
<https://doi.org/10.1093/nar/gku1099> PMID:25398896
  82. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun.* 2018; 9:189.  
<https://doi.org/10.1038/s41467-017-02525-w>  
PMID:29335486



## SUPPLEMENTARY MATERIALS

### Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 1.

**Supplementary Table 1. Number of SNPs residing in regions common between the original GWAS-based networks. Genes with ‘long-lived’ phenotype are shown in bold.**

**Supplementary Table 2. Selected subnetworks centered around novel nodes with the highest degree in the extended Synthetic and DGRP GWAS-based networks.**

Novel node	Network	All other nodes in subnetwork	Degree	Total number of genes
1220	Synthetic	244, 245, 255, 262, 270, 271, 272, 273, 275, 276, 277, 295, 302, 305, 334, 359, 799, 848, 920, 923	20	188
928	Synthetic	11, 233, 234, 238, 265, 343, 360, 361, 366, 370, 409, 456, 536, 531, 545	15	290
28	DGRP	27, 29, 30, 1063, 1124, 1152, 1179	7	114
2	DGRP	576, 655, 660, 670, 699, 736, 787, 1120, 1131, 1132, 1152, 1178, 1179, 1183	14	255

Please browse Full Text version to see the data of Supplementary Tables 3 and 4.

**Supplementary Table 3. Nodes and interactions comprising each cluster of the extended Synthetic GWAS-based networks.**

**Supplementary Table 4. Nodes and interactions comprising each cluster of the extended DGRP GWAS-based networks.**

**Supplementary Table 5. List of nodes constituting selected clusters in the Synthetic GWAS-based network.**

Cluster	Nodes in cluster	Total number of	
		nodes in cluster	genes in cluster
4	89, 90, 91, 92, 93, 94, 95, 97, 1319	9	92
5	124, 125, 126, 127, 128, 129, 130, 131, 711, 740, 929	11	229
11	198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 578, 916, 930, 982, 1049, 1431, 1436	21	249
23	330, 331, 332, 333, 334, 335, 336, 337, 749	9	141
29	393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 961, 1044, 1338	14	222
60	313, 385, 476, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 1218	14	174
67	502, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1386	11	150

**Supplementary Table 6. List of nodes constituting selected clusters in the DGRP GWAS-based network.**

Cluster	Nodes in cluster	Total number of	
		nodes in cluster	genes in cluster
4	53, 54, 55, 56, 57, 58, 59, 60, 61, 1078, 1079, 1338	12	125
18	523, 524, 525, 526, 527, 528, 529, 530, 531, 560, 1253	11	155
20	105, 478, 521, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 1227	14	256
26	321, 329, 412, 445, 499, 658, 659, 660, 661, 662, 663, 1324, 1367, 1418	14	223
28	330, 344, 366, 413, 669, 670, 671, 672, 673, 1419	10	177
34	318, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785	11	103
40	103, 426, 458, 983, 985, 986, 987, 988, 989, 990, 991, 992, 1210, 1336	14	156
44	361, 429, 448, 1089, 1090, 1091, 1092, 1093, 1094, 1095	10	156
49	63, 457, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155	11	142

Please browse Full Text version to see the data of Supplementary Table 7.

**Supplementary Table 7. Genes sharing ‘long-lived’ and ‘short-lived’ phenotypes found in close proximity to mutated and non-mutated TAD border regions.**

**Supplementary Table 8. Top 30 long-range interacting regions for non-coding SNPs from the Synthetic GWAS dataset.**

Non-coding SNP region			Interacting region		
Bin number	Number of SNPs in bin	Highest recorded SNP's D-value	Strongest interacting bin	Distance (Kb)	Number of genes in bin
2367	3	2.727	2362	50	2
3982	7	5.421	3987	50	16
5470	9	5.765	5465	50	13
2459	4	5.170	2463	40	6
4370	12	7.189	4366	40	10
4560	4	3.814	4566	40	10
5656	9	6.284	5660	40	16
11643	2	5.029	11647	40	6
1605	7	7.657	1608	30	9
2006	3	12.009	2003	30	7
2461	1	4.592	2464	30	6
3146	13	3.468	3149	30	13
3447	1	2.160	3444	30	10
4837	8	6.213	4840	30	15
5702	1	3.582	5705	30	8
6981	6	4.577	6978	30	4
7819	4	3.189	7816	30	6
7883	6	5.029	7886	30	11
7884	4	5.989	7887	30	12
8958	11	4.396	8995	30	11
155	14	7.658	153	20	6
229	7	7.550	227	20	10
324	3	5.987	326	20	15
556	11	6.606	554	20	10
592	1	2.066	590	20	12
945	6	4.168	943	20	10
994	5	5.459	996	20	3
1167	10	6.455	1165	20	13
1173	7	6.222	1171	20	10
1174	11	5.694	1176	20	9

**Supplementary Table 9. Top 30 long-range interacting regions for non-coding SNPs from the DGRP GWAS dataset.**

Non-coding SNP region			Interacting region		
Bin number	Number of SNPs in	Lowest recorded SNP P-value	Strongest interacting bin	Distance (Kb)	Number of SNPs
2388	8	0.3858	2378	100	168
3644	26	0.1303	3637	70	190
2458	34	0.06431	2464	60	171
2367	53	0.05865	2362	50	74
2957	16	0.06397	2962	50	228
3982	256	0.007077	3987	50	205
5470	280	0.003503	5465	50	420
2457	28	0.08089	2453	40	81
2459	73	0.5731	2463	40	120
4370	297	0.008632	4366	40	192
4560	88	0.02264	4566	40	280
5656	222	0.002052	5660	40	307
11643	51	0.05711	11647	40	138
1605	95	0.0236	1608	30	230
2006	64	0.01037	2003	30	153
2124	14	0.0354	2121	30	22
2283	17	0.1617	2286	30	62
2372	68	0.01846	2375	30	128
2456	44	0.01432	2453	30	81
2461	32	0.04341	2464	30	171
3146	377	0.006937	3149	30	295
3447	44	0.13	3444	30	285
4837	181	0.007141	4840	30	359
5702	27	0.01703	5705	30	322
6981	87	0.00301	6978	30	83
7639	11	0.09939	7636	30	126
7819	238	0.001871	7816	30	203
7883	108	0.007967	7886	30	234
7884	120	0.006706	7887	30	255
8958	222	0.001062	8955	30	197

Please browse Full Text version to see the data of Supplementary Table 10.

**Supplementary Table 10. Chromosome, start and end positions for each 80 Kb bin.**