



# HHS Public Access

Author manuscript

*Ann Gerontol Geriatr Res.* Author manuscript; available in PMC 2017 March 27.

Published in final edited form as:

*Ann Gerontol Geriatr Res.* 2016 ; 3(1): .

## Coordinated Action of Biological Processes during Embryogenesis Can Cause Genome-Wide Linkage Disequilibrium in the Human Genome and Influence Age-Related Phenotypes

Irina Culminkaya, Alexander M. Kulminski\*, and Anatoli I. Yashin

Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, USA

### Abstract

A role of non-Mendelian inheritance in genetics of complex, age-related traits is becoming increasingly recognized. Recently, we reported on two inheritable clusters of SNPs in extensive genome-wide linkage disequilibrium (LD) in the Framingham Heart Study (FHS), which were associated with the phenotype of premature death. Here we address biologically-related properties of these two clusters. These clusters have been unlikely selected randomly because they are functionally and structurally different from matched sets of randomly selected SNPs. For example, SNPs in LD from each cluster are highly significantly enriched in genes ( $p=7.1\times 10^{-22}$  and  $p=5.8\times 10^{-18}$ ), in general, and in short genes ( $p=1.4\times 10^{-47}$  and  $p=4.6\times 10^{-7}$ ), in particular. Mapping of SNPs in LD to genes resulted in two, partly overlapping, networks of 1764 and 4806 genes. Both these networks were gene enriched in developmental processes and in biological processes tightly linked with development including biological adhesion, cellular component organization, locomotion, localization, signaling, ( $p<10^{-4}$ ,  $q<10^{-4}$  for each category). Thorough analysis suggests connections of these genetic networks with different stages of embryogenesis and highlights biological interlink of *specific* processes enriched for genes from these networks. The results suggest that coordinated action of biological processes during embryogenesis may generate genome-wide networks of genetic variants, which may influence complex age-related phenotypes characterizing health span and lifespan.

### Keywords

Embryogenesis; Age-related traits; Healthspan; Linkage disequilibrium; Functional linkage

---

\*Corresponding author: Alexander Kulminski, Biodemography of Aging Research Unit, Duke University, Box 90420, Durham, NC, 27708, USA, Tel: 919-684-4962; Fax: 919-684-3861; Alexander.Kulminski@duke.edu.

### AUTHORS CONTRIBUTIONS

AMK and IC contributed to the study conception, design, interpretation of the results, and writing the manuscript. AMK performed statistical analysis. IC performed biological analysis. AIY contributed to discussion and interpretation of the results and drafting the manuscript.

## INTRODUCTION

Despite advance of genetic association studies, the optimism in the field is tempered because of so-called missing heritability problem, i.e., that proportions of phenotypic variance explained by genes detected in prior studies are much smaller than heritability estimates of those traits. This problem should be paralleled with two fundamental results of genome-wide association studies (GWAS). First, complex traits are likely influenced by a large number of alleles from genes spread through the entire genome. Second, individually these alleles can confer tiny risks of a complex trait. These results imply that in order to explain clinically meaningful genetic susceptibility for an individual, his/ her genome should include large number of the risk alleles. This is a major challenge.

Indeed, given clustering of complex human traits in families, heritability of the traits has to be explained in terms of transmittance of the risk alleles from parents to offspring. Originally, GWAS followed so-called common disease – common variant (CDCV) hypothesis assuming that a common trait could be influenced by a few common variants in *the person's genome*. Then transmittance of the risk alleles from parents to offspring follows Mendelians' logic. GWAS provided evidences that common traits are likely influenced by a large number of variants spread throughout the genome that strengthens Biometricians' logic. This logic poses theoretical challenges in terms of transmittance of the risk alleles to offspring. Fisher in his classical work [1] advocated a quantitative mechanism for explaining inheritance of polygenic traits. A key of this mechanism is the concept of allelic equivalence, i.e., those different alleles rather confer risk of, than cause, the same trait. Then, common traits are considered as non-Mendelian while the risk alleles are inherited following random (i.e., Mendelian) segregation.

If complex traits are influenced by a large number of risk alleles present in the genome of a given person, then, fundamental question is whether it is indeed always the case that these risk alleles are transmitted from parents to offspring randomly? A number of recent studies have documented co-adaptation of physically unlinked genes at loci on different chromosomes participating in the same biological processes (BP) in humans [2,3] and in mammals [4–6]. There are also increasing evidences that inter-chromosomal linkage may be much more common than previously believed [7–11]. Co-adaptation of alleles on the basis of biological mechanisms implies *functional linkage*. Although the mechanism of functional linkage is unclear yet, it can be associated with regulatory mechanisms controlling transcription of genes located on different autosomes [12,13]. Functionally linked alleles are phenotypically manifested and can be transmitted to offspring as a complex. This transmittance was observed in various studies [3,14–16]. These observations suggest that we can expect that the risk alleles from the person's genome could be transmitted to offspring non-randomly.

Theoretical basis of random mechanism of inheritance is the Mendel's Second Law which is believed to be proven by a number of cytological experiments. Much less expected was the fact that the very first cytological evidence was for non-random segregation of autosomes during meiosis [17]. Since that time non-random segregation was observed in various species [18–22]. Despite that, these evidences are not in mainstream of most genetic studies

[23]. These experimental evidences, however, may represent phenotypic manifestation of functional linkage among physically unlinked genes, implying that alleles in such genes can function as a complex and be transmitted to offspring together.

Genome-wide functional linkage of alleles implies that there should be fundamental bio-molecular mechanisms generating and supporting functional co-adaptation. A relevant mechanism could be associated with developmental processes [24] which affect the entire organism. Barker's hypothesis of developmental plasticity [25] also provides a link of the developmental processes with phenotypes characteristic for the late life including geriatric diseases and premature death.

Recently, we reported about inheritable genome-wide networks of SNPs in chromosome-wide and inter-chromosomal linkage disequilibrium (LD) which have been associated with premature death and major human diseases in the Framingham Heart Study (FHS) [26–28]. The analyses in [26–28] suggested that these networks were unlikely the matter of chance or technical artefacts. In this work we provide further evidence on non-random origin of the networks of SNPs in genome-wide LD. We also show that genes for these genome-wide SNP networks comprise a complex system of interlinked BPs which is essential for embryogenesis.

## METHODS

### Data

The FHS includes 14,428 participants comprising three cohorts, i.e., the FHS original (launched in 1948; 5,209 respondents), the FHS Offspring (FHSO, launched in 1971–1975; 5,124 offspring), and the 3<sup>rd</sup> Generation (launched in 2002; 4,095 grandchildren) cohorts. The FHS is a population-based longitudinal study following its participants for up to about 60 years. Selection criteria, study design and phenotypic data have been previously described [29–31]. SNP data are available for 9,274 participants for whom DNA samples have been drawn [32]. Genotyping in the FHS was done using the Affymetrix 500K (250K Nsp and 250K Sty) and an independent 50K Human Gene Focused arrays having no overlapping SNPs.

### Selection of sets of SNPs in inter-chromosomal linkage disequilibrium

In all analyses we used autosomal SNPs from the Affymetrix 500K and 50K arrays which passed quality control (QC) tests. QC was as in [26–28], i.e., SNPs/individuals were excluded if: call rate < 90%, Hardy-Weinberg equilibrium p-values < 10<sup>-2</sup>, and Mendel errors > 2%. The analyses in [26–28] identified two partly overlapping sets of SNPs in extensive intra- and inter-chromosomal LD among those showing associations with cardiovascular diseases, cancer, and premature death. This LD was observed among SNPs on *two independent* Affymetrix 500K and 50K arrays. In [28] we identified complexes of four (Y complex) and five (G complex) SNPs on Affymetrix 50K array which showed perfect intra-set LD ( $r^2 > 0.8$ ) and no inter-set LD. We evaluated LD of these nine SNPs with the other SNPs on both arrays (after QC) using *plink* [33]. We used empirically established in [26] cut off of  $r^2 = 0.02$  to define non-random LD and selected SNPs with minor allele

frequency (MAF) 0.009 MAF 0.499. The largest sets of SNPs in both nominal LD ( $r^2$  0.02) and moderately-strong LD ( $r^2$  0.50) were identified for rs1390694 from the Y complex and for rs9330200 from the G complex. The Y set included 3970 SNPs and the G set included 20254 SNPs called as the reference SNP sets hereafter.

### Mapping and functional annotation of genes

All SNPs were annotated and mapped to genes using the US FDA's Array Track [34]. Due to large number of genes for the reference SNP sets intergenic SNPs were not mapped to genes; annotation, thus, resulted in 1764 (Y set) and 4806 (G set) genes. We analyzed enrichment of genes in biological process Gene Ontology (GO) terms [35,36] in the Y and G sets separately. To ensure robustness of the results, three GO-based bioinformatics tools were used. For both the Y and G sets we used GOFFA [37] and high-throughput GoMiner [38]. Web-based tool DAVID [39] was aided in the analyses of the Y set alone because DAVID was limited to 3000 genes.

These tools evaluate enrichment of genes in GO terms relative to a list of all genes in the genome. Optionally, GoMiner also uses client supplied gene set. In our analysis we used both methods, i.e., GO database gene set (called "auto-generated" set) and set of genes for qualified SNPs on the arrays. The DAVID augments and integrates annotations from different databases [39].

All three tools provide Fisher's exact test p-values. Nominal level of significance according to this test (i.e.,  $p$  0.05) was used to pre-select GO terms. GoMiner and DAVID also provided false discovery rate (FDR) q-values which aid in offsetting the multiple comparisons problem in the analyses on enrichment of genes in specific GO terms [40,41]. As a rule, cut off for q-value at 0.1 at maximal number of randomizations (1000 in GoMiner) was used. Less stringent q-values were used in the analyses of enrichment of genes in small BPs. Although large high-level terms are more reliable [42] and small GO terms can be more readily enriched by chance, they are more specific and may be crucial for gaining insights into biological essence and for revealing interlink between the underlying BPs [35]. Interlink between such processes warrants using less stringent p- and q-values for the corresponding small GO terms because biologically-motivated interlink implies smaller chances of false findings.

The analyses using the bioinformatics tools were complemented by rigorous systemic analysis [43] of selected GO terms using manually curated information from relevant literature and public databases as advocated in [42,44]. Because GOFFA did not provide q-values, we validated our findings by contrasting them by 20 random gene sets (generated, using the set of genes for qualified SNPs on the 500K plus 50K arrays) for each reference gene set, i.e., by calculating and comparing enrichment statistics for GO terms in the reference and random gene sets (having the same number of genes as the reference sets). Given potential biases in the analyses of GO terms (e.g., gene length bias, large gene list size, gene enrichment in BP in genome, GO's structure [36,45]), we also verified in some analyses whether or not a complex of biologically interlinked GO terms enriched in the reference SNP sets is also enriched in 20 control random SNP sets (having the same number

of SNPs as the reference sets and generated, using the qualified SNPs on the 500K plus 50K arrays).

BPs was prioritized according to the hierarchical structure of the GO tree of terms using GOFFA and the EBI's QuickGO tools (<http://www.ebi.ac.uk/QuickGO/>).

### Comparative and sensitivity analyses of BPs

Two reference sets of 1764 (Y set) and 4806 genes (G set) were linked to BP GO terms using GoMiner, GOFFA, and DAVID. Comparative analyses of the results using all three tools mostly show enrichment of genes in the same BPs regardless of the tool (see Supplementary Information, Table S1). Sensitivity analysis of the choice of the reference gene list was conducted using auto-generated (GO database) and an array-based gene lists in GoMiner for the Y and G sets. No qualitative differences were observed implying that genes were present on the arrays mostly in direct proportion to genes involved in BP in the genome (see Supplementary Information, Table S1 for the Y set). Accordingly, we discuss the results of gene enrichment analysis for the Y and G sets based on the auto-generated gene list in GoMiner (see Supplementary Information, Tables S2 and S3).

## RESULTS AND DISCUSSION

### Non-random selection of SNPs and genes in the Y and G reference sets

Genome-wide LD implies coherent behavior of SNPs seen as statistical clustering; this property is not typical for traditional GWAS. This property can be efficiently used to ensure that random origin of the observed LD is implausible. We prove this by providing three evidences highlighting biologically-motivated functional and structural differences among the two reference sets of SNPs in LD and the randomly selected SNP sets.

First, mapping the reference SNPs to genes resulted in 2318 gene entries (including repeats) and 1764 unique genes in the Y set and 10026 gene entries and 4806 unique genes in the G set. Annotation of the same number of SNPs to genes in the 20 control random sets for each reference set shows that genes in the random sets are significantly underrepresented compared to the reference sets. Highly significant enrichment in the Y and G reference sets is observed for both unique genes (Figure 1) and all gene entries (Supplementary Information, Figure S1). This finding ensures functional difference between the random SNP sets and the sets of SNPs in LD.

Second, because genes with large number of SNPs have higher chances to be selected by chance, we evaluated the difference in the number of SNPs in genes (that is a proxy for gene length) identified in the reference and random sets. Contrary to the expectation by chance, this analysis shows significant enrichment of shorter genes in the Y and G reference sets compared to the random sets. Specifically, mean number of SNPs in genes identified in the Y (G) set was 28.3 (25.8) whereas mean number of SNPs in the Y-(G-) related random sets was 49.2 (29.4). This difference was highly significant with  $p=1.4\times 10^{-47}$  for the Y set and  $p=4.6\times 10^{-7}$  for the G set (Figure 2). This finding provides another robust support for functional difference between the random and the reference SNP sets. Significantly larger

number of shorter genes in the reference sets partly explains larger number of genes in these sets.

Third, SNPs from the Y and G sets were associated with the same phenotype of premature death, although of different severity [27], whereas SNPs from the random sets were generated regardless of their connection (except by chance) with any health- and/or lifespan-related phenotype. Accordingly, if the associations of SNPs from the Y and G sets with premature death are plausible, then SNPs from these sets should overlap non-randomly. We found that 1541 SNPs overlap between the Y (N=3970) and G (N=20254) reference sets. Analysis of pairs of SNP sets from the Y- and G-related control random sets shows that stochastic overlap between them is on average of only 385 SNPs. The difference in these overlaps of SNPs is highly significant,  $p=8.6 \times 10^{-35}$  (Figure 3). This result ensures that the overlap of SNPs from the Y and G sets is unlikely by chance. It highlights structural difference between the reference and random SNP sets.

These evidences, along with those on enrichment of non-synonymous coding SNPs in the reference sets [28], suggest that genes for SNPs in genome-wide LD selected against the phenotype of premature death [27] should have relevant biological functions to work in coordinate fashion. Accordingly, it is imperative to gain profound insights on this role.

#### **Enrichment of genes for SNPs from the Y and G reference sets in BP GO terms**

The most significant ( $p < 10^{-4}$ ) enrichment of genes from the Y and G sets in general GO categories was observed in developmental process terms and in a number of other complex, biologically important processes that drive many aspects of development (Figure 4). The full list of enriched BPs is given in Supplementary Information, Tables S2 and S3. These tables show significant enrichment of genes from the reference sets in higher-level GO terms as well as in a variety of lower-level categories, partly overlapping through GO hierarchy. Larger size of the G set results in longer list of significant BPs compared to the Y set at different hierarchical levels. Importantly, higher-level categories related to growth (e.g., developmental cell growth) and cell cycle (e.g., cell cycle process, mitotic cell cycle, cell cycle phase) mostly characterized the G set, i.e., they were neither significant in the Y set nor in the majority of the G-set-related control random SNP sets.

Although high-level GO terms per se provide limited information for biological interpretation of the results (e.g., due to GO structure [35,36] and biases of pathway analyses [35,44]), enrichment of genes in the aforementioned categories suggests that specifics of the Y and G sets is an aggregation genes in BPs essential for development (Figure 4). This general architecture of biological specifics of the reference sets is supported by enrichment of genes in a variety of biologically interrelated specific processes which are characteristic for these sets but, as a rule, not for the random sets (selected list of such processes is shown in Table 1). Next sections provide the results of detail biological analyses of developmental processes enriched in the Y and G sets (see “Developmental processes”) and interlinks between the selected Y set specific developmental and signaling processes (see “Functional crosstalk: the Wnt signaling”). The analyses of other major processes sketched in Figure 4 are detailed in Supplementary Information, Text “Biological analyses”. These analyses

highlight a systemic structure of the Y and G sets which is lacking in the control random SNP sets.

### Developmental processes

General GO term developmental process included a variety of significant sub-categories, typically, in strong hierarchic sequence. Highly significant enrichment of genes from the Y and G sets ( $p < 10^{-4}$ ) was observed in large BPs related to organismal development, morphogenesis, and cell differentiation. BPs related to developmental growth were the G set-specific (e.g., developmental growth involved in morphogenesis,  $p_G = 0.008$  vs.  $p_Y = 0.358$ ).

Morphogenesis, differentiation and growth are BPs controlling fundamental aspects of the developmental program. Accordingly, enrichment of genes in these processes suggests that the Y and G sets are likely linked to the program of embryonic development. This is also supported by enrichment of genes in more specific large GO term in utero embryonic development ( $p_Y = 0.024$  and  $p_G = 0.0009$ ) and in a number of specific BPs related to gastrulation.

### Gastrulation

Gastrulation, an early stage of embryogenesis [46] determining the basic body plan, includes the Y set specific terms placenta development ( $p_Y = 0.032$ ), branching involved in labyrinthine layer morphogenesis ( $p_Y = 0.021$ ), gastrulation with mouth forming second ( $p_Y = 0.026$ ), formation of primary germ layer ( $p_Y = 0.053$ ). The development of the labyrinthine layer is among the earliest steps of placenta formation [47] and patterning along the primary anterior–posterior (A-P) axis [48] occurs first in development [48]. These specific processes suggest relevance of the Y set to earlier stage of the embryogenesis.

Paraxial mesoderm formation ( $p_Y = 0.013$ ) specifies mesoderm formation in the Y set whereas axial mesoderm morphogenesis ( $p_G = 0.016$ ) is the G set-specific term. The unsegmented cranial paraxial mesoderm particularly contributes to the neurocranium (the part of the skull that protects the brain) [49]. Accordingly, small process embryonic neurocranium morphogenesis ( $p_Y = 0.023$ ) characterizes the Y set. The paraxial mesoderm (but not the axial) has been suggested as a source of neural crest (neural crest cell development ( $p_Y = 0.071$  vs.  $p_G = 0.296$ )) inductive signals [50,51]. The cranial paraxial mesoderm together with cranial neural crest cells from the hindbrain migrate into developing pharyngeal arches [51,52]. The latter is supported by gene enrichment in pharyngeal system development ( $p_Y = 0.003$ ).

Also, the neural plate, the central nervous system (CNS) primordium, can be patterned by signals from the paraxial mesoderm [53,54]. Specific terms neural plate regionalization ( $p_Y = 0.007$ ) and neural plate anterior/posterior regionalization ( $p_Y = 0.003$ ) are neither significant in the G set nor in each control random set; they logically complement the discussed above earlier developmental processes characteristic for the Y set. Regionalization of neural plate is the first step in neural patterning with establishment of the initial polarity of the neural plate along the A-P axis.

## Nervous system development

The development of the nervous systems is one of the most important processes in embryo that allows to even pursue so-called neural default model [55]. Our results show enrichment of genes involved in nervous system development, neurogenesis (all  $p_{Y,G} < 10^{-4}$ ), and central nervous system development ( $p_Y < 10^{-4}$  and  $p_G = 0.002$ ) along with variety of related sub-categories (Supplementary Information, Tables S2 and S3).

**Brain development** ( $p_Y = 0.003$  and  $p_G = 0.028$ ) in both sets is mostly represented by sub-categories of forebrain development which are more pronounced in the G set. Cerebral cortex development ( $p_G = 0.0003$  vs.  $p_Y = 0.014$ ) is specified in the G set by child term superior temporal gyrus development ( $p_G = 0.001$ ). The superior temporal gyrus contains the primary auditory cortex [56]. The G set genes are enriched in this process in parallel with enrichment in other terms linked to developmental processes of the auditory pathway (e.g., inner ear receptor cell differentiation, auditory receptor cell differentiation; see “Sensory organ development”). Further, the superior temporal gyrus is a part of brain network which includes the orbitofrontal cortex [57,58] that is also represented by orbitofrontal cortex development ( $p_G = 0.014$ ). The orbitofrontal cortex integrates the auditory, visual and somatosensory inputs, participates in learning, prediction and decision making, and controls behaviors [59,60]. Again, this is in line with gene enrichment in large terms learning or memory, learning, locomotory behavior, visual behavior, startle response, and sensory perception of mechanical stimulus in the G set (Supplementary Information, Table S3).

Development of the hindbrain, the second of the three primary regions of brain, is specified by child terms cell migration in hindbrain ( $p_Y = 0.006$ ) and cerebellum development ( $p_Y = 0.025$ ) along with cerebellum morphogenesis and cerebellar cortex morphogenesis in the Y set. The hindbrain, a key source of patterning information in the developing head [61], controls basic life-supporting functions (e.g., breathing and heart rate) and reflexes. Also, the motor innervation is supplied by the hindbrain’s neurons [62]. Activity of motor neuron directly leads to muscle contraction. The cerebellum coordinates and controls motor activity (e.g., body movement, balance, coordination) through the cerebro-cerebellar network [63]. As discussed below, other processes associated with motor control and activity, e.g., motor axon guidance (see “Neurogenesis”), neuromuscular process controlling balance (see “Sensory organ development), that are early events of prenatal development [64,65], are consistently specific for the Y set.

**Neurogenesis**—Enrichment of genes from the Y and G sets in neurogenesis and the lower-level sub-categories in hierarchic sequence (Figure 5) is primarily related to axonal outgrowth during neuron development. Genes are also enriched in central nervous system neuron axonogenesis ( $p_Y = 0.003$  and  $p_G = 0.005$ ) and other related terms.

Directed axon outgrowth is an essential process for the formation of neuronal network during embryogenesis [66]. Axon guidance in both sets is specified by child terms dorsal/ventral axon guidance ( $p_Y = 0.023$  and  $p_G = 0.014$ ) and retinal ganglion cell axon guidance ( $p_Y = 0.008$  and  $p_G = 0.003$ ) which are paralleled with axial patterning and development of the optic nerve containing the ganglion cell axons running to the brain (see “Sensory organ development”).

The Y set is characterized by a small term motor axon guidance ( $p_Y=0.0005$  vs.  $p_G=0.187$ ) that is in line with a number of the Y set-specific terms discussed in “Gastrulation”. The hindbrain, or cranial motor neurons [62], project their axons to the skeletal (striated) muscles which, in turn, are derived from paraxial mesoderm [67]. The face and neck muscles developing from pharyngeal arches are striated muscles as well. Skeletal muscle organ development ( $p_Y=0.038$ ), skeletal muscle tissue development ( $p_Y=0.037$ ) and more specific process myotube differentiation ( $p_Y=0.024$ ) characterize the Y set along with the other related terms observed in both sets (e.g., muscle structure development, muscle cell differentiation, myoblast fusion, and neuromuscular junction development) (Supplementary Information, Tables S2 and S3). A skeletal muscle fiber formed by the fusion of myoblasts during a developmental stage is myotube. Motor axons contact myotubes in skeletal muscles [68,69]. Besides, genes are enriched in the Y set-specific child terms astrocyte development and astrocyte activation ( $p_Y=0.003$ ). The astrocytes are essential for axon guidance, synapse formation, neuronal homeostasis and myelination; they are also implicated in cerebellar development and motor control [70]. Thus, this system of specialized processes (see also Table 1) in the Y set highlights specific stage related to earlier development of motor control system.

Interestingly, leading SNP rs1390694 in the Y set (see “Methods”) is located in *FGGY* gene which is highly expressed in fetal brain and is implicated in neuron homeostasis. *FGGY* may confer susceptibility to sporadic amyotrophic lateral sclerosis (ALS) [71,72]. ALS is the most common motor neuron disease with astrocyte-mediated toxicity on motor neurons [73]. Also, the involvement of the cerebellum in ALS is suggested [74] that again is in line with development of this part of brain and basic motor control system specifically in the Y set. Moreover, demyelination may be a secondary symptom for ALS patients arriving after muscle atrophy [75,76]. *FGGY* encodes a protein that phosphorylates carbohydrates such as ribulose, ribitol and L-arabinitol. Specifically, ribitol is an essential component of the riboflavin (vitamin B-2) that is important for fetal development and implicated in the riboflavin/free radical-induced axonal degeneration and in demyelination as well [77,78]. Accordingly, more complex interplay of ALS determinants can be expected.

After neuronal differentiation axons extend towards their target cells. Axon extension ( $p_G=0.009$ ) is significant in the G set in agreement with gene enrichment in the related term developmental cell growth ( $p_G=0.008$ ).

Axons and dendrites are two parts of nerve cells involved in the conduction of nerve impulses in CNS [79]. Axons extend rapidly and begin to develop first. Dendrite-related terms (e.g., dendrite development ( $p_G=0.019$  vs.  $p_Y=0.13$ )) are specific for the G set. Neuron recognition ( $p_G=0.008$ ) is necessary in determining synaptic connectivity to generate early behaviors [80, 81], though, the first functional networks may develop without recognition of “correct” target neurons [80].

Like other developmental processes, generation of the nervous system occurs in stages. Because terms dendrite development and neuron recognition along with behavior, learning, and startle response are not enriched in the Y set, this set likely specifies the earlier stage of the development. Importantly, the lack of enrichment of these processes is contrasted by

their enrichment (at  $p < 0.05$ ) in majority ( $> 18$ ) of control Y-set-related random SNP sets that is largely due to gene length bias.

**Sensory organ development**—The development of normal brain organization requires input via all the major sensory systems [82]. Development of the sensory organs is characterized in both sets by specific terms related to the development of the olfactory bulb and nose, e.g., olfactory bulb interneuron development ( $p_Y = 0.002$  and  $p_G = 0.001$ ), chemorepulsion involved in embryonic olfactory bulb interneuron migration ( $p_Y = 0.0007$  and  $p_G = 0.014$ ), nose development ( $p_Y = 0.018$ ), the eyes, e.g., optic nerve morphogenesis ( $p_Y = 0.043$  and  $p_G = 0.004$ ) and the inner ear, e.g., inner ear receptor cell differentiation ( $p_Y = 0.011$  and  $p_G = 0.002$ ), inner ear receptor stereocilium organization ( $p_Y = 0.019$ ), auditory receptor cell differentiation ( $p_G = 0.013$ ).

The olfactory system is one of the earliest sensory systems to develop in the human embryo [83]. The optic nerve [84] is derived from the retinal ganglion cell and contains the ganglion cell axons running to the brain (see “Neurogenesis”). The inner ear contains the sensory organs for balance and hearing [85]. The hearing function is paralleled with neurological system process sensory perception of sound ( $p_Y = 0.001$  and  $p_G = 0.002$ ) in both sets. The inner ear’s balance function is in line with cerebellum development (cerebellum controls balance) and neuromuscular process controlling balance ( $p_Y = 0.018$ ) in the Y set.

The sensory organs, such the nose, the ear, and the eye appear very early in development as specialized surface regions (placode) [86]. Enrichment of the Y set genes in olfactory placode formation ( $p_Y = 0.043$  vs.  $p_G = 0.671$ ) and optic placode formation ( $p_Y = 0.023$  vs.  $p_G = 0.566$ ) supports connection of the Y set with earlier developmental stage of these sensory organs. Meanwhile, progress in the development of the inner ear and the eyes is in line with gene enrichment in processes related to development of the superior temporal gyrus and the orbitofrontal cortex in the G set as well as in the related neurophysiological processes and behaviors (see “Brain development”). This connection supports more pronounced role of the G set at later stage of development of the sensory pathways. This functional difference between the Y and G sets suggests that embryonic induction of the inner ear, particularly, may be paralleled with development of the responding organs and tissue and be at different stages controlled by distinct sets of genes.

Thus, the Y set emphasizes the earlier processes of embryogenesis related to the formation of the head and the establishment of the network of neuromuscular connection in head region.

### **Cardiovascular system (CV) development**

The development of the nervous system is paralleled with the development of the CV system and the CV system is the first major system to function in the human embryo. Among processes related to earlier events of cardiogenesis and development of blood vessels characteristic for the Y set, we observe cardiac atrium development ( $p_Y = 0.031$ ) along with venous blood vessel morphogenesis ( $p_Y = 0.023$ ).

Formation of the de novo blood vessels (vasculogenesis) begins from endothelial cell precursors (angioblasts). Endothelial progenitor cells from paraxial mesoderm, that are characteristic for the Y set, have angioblastic capacity [87]. Endothelial cell differentiation ( $p_Y=0.035$ ) together with endothelium development ( $p_Y=0.041$ ) and regulation of vasoconstriction ( $p_Y=0.049$ ) are characteristic for the Y set as well. In addition, specific process lymphangiogenesis ( $p_Y=0.021$ ) characterizes early development of the lymph vessels [88]. Lymphatic vessels arise via transdifferentiation of venous endothelial cells. The endothelial cells line vessels and regulate the flow of body fluid. The blood and lymphatic vascular systems function in concert that is characterized by regulation of body fluid levels ( $p_Y=0.026$  and  $p_G=0.0008$ ).

BPs aorta smooth muscle tissue morphogenesis ( $p_Y=0.008$ ) and vein smooth muscle contraction ( $p_Y=0.045$ ) along with smooth muscle cell proliferation ( $p_G=0.009$ ) and smooth muscle tissue development ( $p_G=0.005$ ) are plausible in connection with developmental processes of circulatory system because involuntary smooth muscles form the muscle layers in the walls of blood vessels [89].

Genes from the G set, in contrast to the Y set, are predominantly enriched in cardiovascular system development ( $p_G<10^{-4}$  vs.  $p_Y=0.038$ ) and, specifically, in heart development ( $p_G<10^{-4}$  vs.  $p_Y=0.028$ ), blood vessel development ( $p_G=0.008$  vs.  $p_Y=0.118$ ) and sprouting angiogenesis ( $p_G=0.007$ ). The latter is a specific process crucial for the development of new blood vessels from pre-existing blood vessels [90]. Heart development is paralleled with cardiac muscle fiber development ( $p_G=0.034$ ) in the G set. Gene enrichment in these BPs supports connection of the G set with later stage of embryogenesis when development and functioning of the CV system are in progress.

**Anatomical structure development in the G set**—Highly significant higher-level term anatomical structure development ( $p_{Y,G}<10^{-4}$ ) is specified by a substantially wider array of subcategories in the G set than in the Y set. This differentiation is logical given potential link of the G set with later stage of embryogenesis and with some initial developmental processes. In particular, enrichment of genes in head development ( $p_G=0.004$ ) and face morphogenesis ( $p_G=0.006$ ) is paralleled with prevalence of BPs related to the forebrain (see “Brain development”) that acts as a structural support for facial development [91] as well as the axial mesoderm (see “Gastrulation”) that comprises the head process [92]. The development of complex asymmetric organs is represented in the G set not only by heart development but also by liver development ( $p_G=0.027$ ) along with hepaticobiliary system development ( $p_G=0.012$ ). The lungs (respiratory system development ( $p_G=0.049$ )) are markers of left-right (L-R) asymmetry as well [93]. Renal system development ( $p_G=0.046$ ) also requires both the L-R and A-P patterning signals to determine the kidney field [94].

Female sex differentiation ( $p_G=0.007$ ) along with development of primary female sexual characteristics, female gonad development, and apoptosis involved in luteolysis represent developmental process involved in reproduction.

Mesenchymal cell development ( $p_G=0.043$ ), morphogenesis of an epithelium ( $p_G=0.023$ ), and columnar/cuboidal epithelial cell differentiation ( $p_G=0.012$ ) characterize progress in development of the connective and the epithelial tissues in addition to the development of two other basic tissue types (neuronal and muscular). The mammary gland [95] and the salivary gland [96] develop as a result of mesenchymal-epithelial interactions. Accordingly, specific processes salivary gland development together with child term dichotomous subdivision of terminal units involved in salivary gland branching ( $p_G=0.0008$ ) and mammary gland development together with mammary gland duct branching involved in pregnancy ( $p_G=0.034$ ) characterize gland morphogenesis ( $p_G=0.009$ ). Epithelio-mesenchymal interactions are also essential for series of other events [97] including G-set-specific GO terms (e.g., cell-cell junction organization, cell-matrix adhesion ; see “Cellular component organization” and “Biological adhesion” in Supplementary Information, Text “Biological analyses”).

Thus, the systems of the interlinked developmental BPs enriched in the Y and G sets suggest involvement of genes from these sets in regulation of processes characteristic for embryonic development. The Y set of genes appears to be more characteristic for the early stages whereas the G set of genes is more characteristic for the later stage. Importantly, epigenetic processes, that are crucial for normal early development in mammals [98], genetic imprinting and regulation of gene expression by genetic imprinting ( $p_Y=0.003$ ,  $q_Y=0.05$  and  $p_G=0.029$ ,  $q_G=0.21$ ), are characteristic for both reference sets in contrast to the control random sets (also, see Supplementary Information, Text “Biological analyses”).

As an example, next section details the results of biological analyses of interlink between the selected Y set specific developmental and signaling processes.

### Functional crosstalk: the Wnt signaling

Wnt signaling plays a critical role in the embryonic development and the maintenance of adult tissues [99,100]. Wnt receptor signaling pathway is more pronounced in the Y set ( $p_Y=0.005$  vs.  $p_G=0.064$ ). It is specified by regulation of Wnt receptor signaling pathway ( $p_Y=0.009$  and  $p_G=0.023$ ) along with child terms negative regulation of Wnt receptor signaling pathway ( $p_Y=0.016$ ). Gene enrichment in the Wnt signaling is consistent with enrichment of a number of the Y set-specific interlinked processes in which Wnt signaling is implicated, including (among others) the early placental differentiation [101], the development of neural crest [102] and pharyngeal system [103], the paraxial mesoderm formation [104], specification of skeletal muscle cells [105], and cranial skeleton development [106]. The key role of Wnt family proteins in A-P patterning has been confirmed [48, 107] including establishing the oral-aboral axis [108] and the initial A-P regionalization of the neural plate [48]. Wnt signaling in growth cone is a general and evolutionary conserved mechanism of axon guidance [109]. Specifically, Wnts mediate A-P axon guidance of motor neurons [110, 111]. Also, the Wnt signaling plays an essential role in establishment of the embryonic cerebellum [112], in formation of synapse [113] and neuromuscular junction [105], and regulates astrocyte development [114]. These developmental events characteristic for the Y set in which Wnt signaling is implicated are sketched in Figure 6 (see also “Developmental process”).

The most studied canonical Wnt/ $\beta$ -catenin pathway (regulation of canonical Wnt receptor signaling pathway ( $p_Y=0.016$  and  $p_G=0.017$ )) controls diverse BPs [99]. Among others, Wnt/ $\beta$ -catenin signaling regulates the formation of axes and dorsal structures, the stem cells maintenance, the skin and cardiac development, and angiogenesis [99]. This is in line with a number of the G set specific developmental processes as well (see “Developmental process”).

Moreover, Wnt and BMP (member of the TGF $\beta$  family) signals pattern the dorsal spinal cord [115] whereas the G set-specific Sonic hedgehog signaling is implicated in the ventral region (see Supplementary Information, Text “Biological analyses”). Interaction between Wnt and BMP signaling is further consistently supported by the Y set-specific small terms dorsal spinal cord development and BMP signaling pathway involved in spinal cord dorsal/ventral patterning. Also, regarding motor control and the Y set *FGGY* gene (see “Neurogenesis”), it has been demonstrated that Wnt- and BMP-dependent signaling could play relevant roles in the neurodegeneration of motor neurons in the context of ALS [116].

Wnt, BMP/TGF $\beta$ , and Sonic hedgehog function as morphogens early in development and also mediate axon guidance [117, 118]. Given the observed differences between the Y and G sets in signaling of these morphogens and the related developmental processes as well as apparent interlink between temporally coordinated processes regulating the A-P and dorsal-ventral (D-V) patterning [119] and guidance of axons [118], we suggest that at least part of overlapping SNPs in the Y and G reference sets may be linked to genes which are associated at different time points with specific processes of early development involved in A-P, D-V and L-R patterning pathways.

Thus, this section highlights crosstalk of specific signaling events with processes related to development in a time and context specific manner. Further support to our conclusions on connections of the Y and G sets of genes with embryogenesis and to the interlink of developmental processes with other groups of BPs ensuring proper development and related to different levels of GO hierarchy is detailed in Supplementary Information, Text “Biological analyses”.

## CONCLUSIONS AND IMPLICATIONS

This work focuses on comprehensive analyses of biologically-related specifics of two inheritable extensive clusters of SNPs in genome-wide LD which have been revealed as correlates of premature death in the FHS [27]. The results of our analyses support three important findings. First, they show that SNPs in genome-wide LD were unlikely selected as a matter of chance. Second, they document the enrichment of genes for these SNPs in BPs which are remarkably important for an organism development in utero. Third, they highlight striking importance of systemic analyses of biological machinery involved in regulation of complex phenotypes.

### Non-random nature of SNPs in genome-wide LD

A striking result was apparent functional and structural difference between SNPs in LD and those in matching random sets. That is, the Y and G sets of SNPs in genome-wide LD have

substantially different functions and structure compared to the matching sets of randomly selected SNPs. These are the differences which are not driven by genotyping errors; a problem which is of inherent concern in genome-wide association studies. The evidences are: (i) SNPs from the reference sets are highly significantly enriched in genes compared to SNPs from the matching random control sets (Figure 1), (ii) SNPs from the reference sets are highly significantly enriched in shorter genes in contrast to SNPs from the random sets (Figure 2); (iii) the difference between the overlap of SNPs in the two reference sets and the overlaps of SNPs in pairs of random sets is highly significant (Figure 3). These evidences are supported by our prior finding of enrichment of non-synonymous coding SNPs in the reference sets [28]. These functional and structural dissimilarities suggest that networks of genes from the Y and G sets should play profound biological role.

### Genes for SNPs in genome-wide LD are implicated in embryonic development

The analyses using bioinformatics tools highlighted highly significant enrichment of genes in developmental process and major BP GO terms tightly linked with development. High-levels BPs provide, however, limited information for biological interpretation of the results and for gaining insights into specifics of networks of interlinked BPs generated during embryogenesis. Given that we deal with genes for SNPs in genome-wide LD, we should expect a system of more informative *specific* BPs with a *well-defined* cross-talk. To gain farther insights into biological role of networks of genes for SNPs from the Y and G sets, we conducted in depth manual targeted analysis of literature and public databases. The focus was on revealing well-documented interlinks between specific BPs on lower levels in GO hierarchy. These analyses focused on nominally significant functional terms because interlink per se implies smaller chances of false enrichment.

In depth targeted biological analyses provide a striking result on enrichment of genes from the reference sets in variety of interlinked BPs which represent a system of time-synchronized events characterizing different stages of embryonic development of specific organs/tissue and systems which is a hallmark of both reference sets.

In both reference sets, contrary to the control random SNP sets, genes are enriched in BPs related to genetic imprinting (an epigenetic mechanism of inheritance that exhibit a parent-of-origin specific pattern of gene expression) and to carbohydrate metabolism (in support of glucose as the principal metabolic fuel for the growing embryo).

Our analyses suggest that genes from the Y set likely control earlier events in embryogenesis and neurulation with emphasis on A-P patterning and on organization in the head region. The analyses suggests coherent enrichment of the Y set genes in processes related to initial development of the hindbrain (i.e., cerebellum), sensory organs, skeletal muscles, astrocyte, and axonal outgrowth of motor neurons. These evidences, among the others, support and emphasize formation of basic apparatus for motor control that is an early event of embryonic and fetal development in accordance with priority of the organization and production of movement for primary organism functioning and survival, and for development of simple learning and behavioral processes [64,120].

Genes in the G set emphasize cell cycle and developmental cell growth along with specific processes related to the cytoskeleton. The G set genes are involved in a system of BPs highlighting progressive development of nervous and cardiovascular systems. This set was also characterized by progress in body patterning and axis formation along with initial development of other organs and systems as well as by progressing motor control, neurophysiological functions and the early behaviors along with accumulation of the relevant signaling and metabolic events.

Our results also suggest consistent biological role of two leading genes (*FGGY* and *TUBB4B*) with the key groups of specific BPs characterizing the Y and the G sets (as partly summarized in Table 1). Given these encouraging results, more detail analysis of the functional role of other specific genes from these sets is underway.

Our analyses indicate that different stages in embryonic induction of individual organs (e.g., inner ear) can be paralleled with development of responding organs and tissues and be regulated by the Y and G sets independently. In this regard, it appears that the systems of interlinked BPs may pinpoint different sets of biologically relevant genes from the total list of genes defining these BPs at different stages of an organism development and growth. This implies that specific sets of interlinked genes from different functional pathways may work together to ensure sequential development of specific organs and systems. Alternations in functions of these specific sets of genes during embryogenesis may lead to development of diseases in late life (supported, particularly, by findings on connections of fetal development with neurological processes in later life [121]). For example, alterations in the Y set specific synchronous development of early cerebellum and skeletal muscles and axonal outgrowth of motor neurons linked with development and activation of astrocytes may result in developing ALS in late life (see “Nervous system development”).

### The need for targeted analyses

The analyses using bioinformatics tools are often limited to functional terms with larger number of genes. The argument is that statistics on enrichment of terms with small number of genes may be unreliable. However, the larger GO terms are, the less specific biological information is. In depth analyses of cross-talk of biological processes conducted in our paper suggests that specific BPs, in general, are more informative for biological analysis than the higher-level terms. Indeed, small child terms characteristic for the Y and G sets highlight *systems of coherent biological processes* which characterize coordinated events during embryonic development. These BPs also pinpoint differences in the systems of such embryonic events characteristic for the Y and G sets. It is important to emphasize that in contrast to the Y and G sets, no such systems of coherent BPs were observed in the matching control random sets. As a result, confining bioinformatics analyses to higher levels in the GO hierarchy may lead to loss of biological information. Therefore, analysis of small GO terms is needed to gain more specific insights into underlying biology which has high potential to be translated into clinical interventions. Accordingly, analyses using bioinformatics tools should be considered as a preliminary step with the goal to pinpoint potentially relevant specific processes (e.g., at a nominal level of significance). Then, these

results should be complemented by targeted systemic biological analysis highlighting functional interlinks between the selected processes.

Taken together, our results suggest that coordinated action of BPs during embryogenesis may generate genome-wide LD. These genome-wide networks of genes and genetic variants may control complex phenotypes in late (post-reproductive) life. These results challenge our understanding of the role of genes in complex phenotypes and suggest an important role of genomics in health span and lifespan.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The research reported in this paper was supported by Grants No 1P01 AG043352-01 (AIY, AMK) and 1R01 AG047310 (AMK) from the National Institute on Aging. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI. Funding for SHARe Affymetrix genotyping was provided by NHLBI Contract N02-HL-64278. SHARe Illumina genotyping was provided under an agreement between Illumina and Boston University. This manuscript was prepared using a limited access dataset obtained from the NHLBI and the Framingham SHARe data obtained through dbGaP (accession number phs000007.v14.p5).

## References

1. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*. 1918; 52:399–433.
2. Rohlfes RV, Swanson WJ, Weir BS. Detecting coevolution through allelic association between physically unlinked loci. *Am J Hum Genet*. 2010; 86:674–685. [PubMed: 20381007]
3. Sivagnanasundaram S, Broman KW, Liu M, Petronis A. Quasi-linkage: a confounding factor in linkage analysis of complex diseases? *Hum Genet*. 2004; 114:588–593. [PubMed: 15045582]
4. Graber JH, Churchill GA, Dipetrillo KJ, King BL, Petkov PM, Paigen K. Patterns and mechanisms of genome organization in the mouse. *J Exp Zool A Comp Exp Biol*. 2006; 305:683–688. [PubMed: 16902958]
5. Petkov PM, Graber JH, Churchill GA, DiPetrillo K, King BL, Paigen K. Evidence of a large-scale functional organization of mammalian chromosomes. *PLoS Genet*. 2005; 1:33.
6. Petkov PM, Graber JH, Churchill GA, DiPetrillo K, King BL, Paigen K. Evidence of a large-scale functional organization of Mammalian chromosomes. *PLoS Biol*. 2007; 5:127.
7. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. Interchromosomal interactions and olfactory receptor choice. *Cell*. 2006; 126:403–413. [PubMed: 16873069]
8. Schierding W, Cutfield WS, O'Sullivan JM. The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. *Front Genet*. 2014; 5:39. [PubMed: 24600475]
9. Dai Z, Xiong Y, Dai X. Neighboring genes show interchromosomal colocalization after their separation. *Mol Biol Evol*. 2014; 31:1166–1172. [PubMed: 24505120]
10. Fritz AJ, Stojkovic B, Ding H, Xu J, Bhattacharya S, Berezney R. Cell type specific alterations in interchromosomal networks across the cell cycle. *PLoS Comput Biol*. 2014; 10:1003857.
11. Johnston RJ Jr, Desplan C. Interchromosomal communication coordinates intrinsically stochastic expression between alleles. *Science*. 2014; 343:661–665. [PubMed: 24503853]

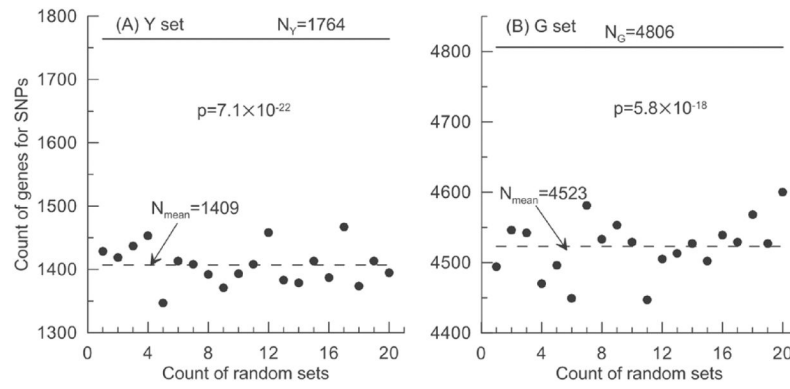
12. Williams A, Spilianakis CG, Flavell RA. Interchromosomal association and gene regulation in trans. *Trends Genet.* 2010; 26:188–197. [PubMed: 20236724]
13. Spilianakis CG, Flavell RA. Molecular biology. Managing associations between different chromosomes. *Science.* 2006; 312:207–208. [PubMed: 16614205]
14. Clegg MT, Allard RW, Kahler AL. Is the gene the unit of selection? Evidence from two experimental plant populations. *Proc Natl Acad Sci U S A.* 1972; 69:2474–2478. [PubMed: 4506768]
15. Miké V. Theories of quasi-linkage and “affinity”: some implications for population structure. *Proc Natl Acad Sci U S A.* 1977; 74:3513–3517. [PubMed: 269409]
16. Sapre AB, Deshpande DS. Spontaneous Emergence of Parents from the F1 Interspecific Hybrids of Coix-L. *J Hered.* 1987; 78:357–360.
17. Payne F. The chromosomes of *Grylotalpa borealis* Burm. *Archiv für Zellforschung.* 1912; 9:141–148.
18. Metz CW. Genetic Evidence of a Selective Segregation of Chromosomes in *Sciara* (Diptera). *Proc Natl Acad Sci U S A.* 1926; 12:690–692. [PubMed: 16587135]
19. Schrader F. A study of the chromosomes in three species of *Pseudococcus*. *Arch Zellforsch.* 1923; 17:45–62.
20. Kubai DF, Wise D. Nonrandom chromosome segregation in *Neocurtilla* (*Grylotalpa*) *hexadactyla*: an ultrastructural study. *J Cell Biol.* 1981; 88:281–293. [PubMed: 7204494]
21. Wise D, Sillers PJ, Forer A. Non-random chromosome segregation in *Neocurtilla hexadactyla* is controlled by chromosomal spindle fibres: an ultraviolet microbeam analysis. *J Cell Sci.* 1984; 69:1–17. [PubMed: 6386834]
22. Schwander T, Beukeboom LW. Non-random autosome segregation: a stepping stone for the evolution of sex chromosome complexes? Sex-biased transmission of autosomes could facilitate the spread of antagonistic alleles, and generate sex-chromosome systems with multiple X or Y chromosomes. *Bioessays.* 2011; 33:111–114. [PubMed: 21154781]
23. Forer A. Do Chromosomes Segregate Randomly during Meiosis?: Key Articles by Fernandus Payne Were Ignored, and Perhaps “Suppressed”. *Proc Am Philos Soc.* 1996; 140:148–163.
24. de Magalhães JP, Church GM. Genomes optimize reproduction: aging as a consequence of the developmental program. *Physiology (Bethesda).* 2005; 20:252–259. [PubMed: 16024513]
25. Barker DJ. The developmental origins of well-being. *Philos Trans R Soc Lond B Biol Sci.* 2004; 359:1359–1366. [PubMed: 15347527]
26. Kulminski AM. Complex phenotypes and phenomenon of genome-wide inter-chromosomal linkage disequilibrium in the human genome. *Exp Gerontol.* 2011; 46:979–986. [PubMed: 21907271]
27. Kulminski AM, Culminskaya I. Genomics of human health and aging. *Age (Dordr).* 2013; 35:455–469. [PubMed: 22174011]
28. Kulminski AM, Culminskaya I, Yashin AI. Inter-chromosomal level of genome organization and longevity-related phenotypes in humans. *Age (Dordr).* 2013; 35:501–518. [PubMed: 22282054]
29. Dawber, TR. *The Framingham study: the epidemiology of atherosclerotic disease.* Cambridge, Mass: Harvard University Press; 1980.
30. Govindaraju DR, Cupples LA, Kannel WB, O’Donnell CJ, Atwood LD, D’Agostino RB Sr, et al. Genetics of the Framingham Heart Study population. *Adv Genet.* 2008; 62:33–65. [PubMed: 19010253]
31. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ. The Third Generation Cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.* 2007; 165:1328–1335. [PubMed: 17372189]
32. Cupples LA, Heard-Costa N, Lee M, Atwood LD. Framingham Heart Study Investigators. Genetics Analysis Workshop 16 Problem 2: the Framingham Heart Study data. *BMC Proc.* 2009;3.
33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]

34. Xu J, Kelly R, Fang H, Tong W. ArrayTrack: a free FDA bioinformatics tool to support emerging biomedical research--an update. *Hum Genomics*. 2010; 4:428–434. [PubMed: 20846933]
35. Alterovitz G, Xiang M, Hill DP, Lomax J, Liu J, Cherkassky M, et al. Ontology engineering. *Nat Biotechnol*. 2010; 28:128–130. [PubMed: 20139945]
36. Rhee SY, Wood V, Dolinski K, Draghici S. Use and misuse of the gene ontology annotations. *Nat Rev Genet*. 2008; 9:509–515. [PubMed: 18475267]
37. Sun H, Fang H, Chen T, Perkins R, Tong W. GOFFA: gene ontology for functional analysis--a FDA gene ontology tool for analysis of genomic and proteomic data. *BMC Bioinformatics*. 2006;7. [PubMed: 16401345]
38. Zeeberg BR, Qin H, Narasimhan S, Sunshine M, Cao H, Kane DW, et al. High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics*. 2005; 6:168. [PubMed: 15998470]
39. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]
40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B*. 1995; 57:289–300.
41. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 29:1165–1188. 2001.
42. Skunca N, Altenhoff A, Dessimoz C. Quality of computationally inferred gene ontology annotations. *PLoS Comput Biol*. 2012; 8:e1002533. [PubMed: 22693439]
43. Nadeau JH, Dudley AM. Genetics. Systems genetics. *Science*. 2011; 331:1015–1016. [PubMed: 21350153]
44. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet*. 2010; 11:843–854. [PubMed: 21085203]
45. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol*. 2003; 4:R28. [PubMed: 12702209]
46. Solnica-Krezel L, Sepich DS. Gastrulation: making and shaping germ layers. *Annu Rev Cell Dev Biol*. 2012; 28:687–717. [PubMed: 22804578]
47. Castellucci M, Kosanke G, Verdenelli F, Huppertz B, Kaufmann P. Villous sprouting: fundamental mechanisms of human placental development. *Hum Reprod Update*. 2000; 6:485–494. [PubMed: 11045879]
48. Niehrs C. On growth and form: a Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. *Development*. 2010; 137:845–857. [PubMed: 20179091]
49. Noden DM. Origins and patterning of craniofacial mesenchymal tissues. *J Craniofac Genet Dev Biol*. 1986; 2:15–31.
50. Monsoro-Burq AH, Fletcher RB, Harland RM. Neural crest induction by paraxial mesoderm in *Xenopus* embryos requires FGF signals. *Development*. 2003; 130:3111–3124. [PubMed: 12783784]
51. Noden DM, Trainor PA. Relations and interactions between cranial mesoderm and neural crest populations. *J Anat*. 2005; 207:575–601. [PubMed: 16313393]
52. Graham A. Development of the pharyngeal arches. *Am J Med Genet A*. 2003; 119:251–256.
53. Lumsden A, Krumlauf R. Patterning the vertebrate neuraxis. *Science*. 1996; 274:1109–1115. [PubMed: 8895453]
54. Muhr J, Jessell TM, Edlund T. Assignment of early caudal identity to neural plate cells by a signal from caudal paraxial mesoderm. *Neuron*. 1997; 19:487–502. [PubMed: 9331343]
55. Levine AJ, Brivanlou AH. Proposal of a model of mammalian neural induction. *Dev Biol*. 2007; 308:247–256. [PubMed: 17585896]
56. Bigler ED, Mortensen S, Neeley ES, Ozonoff S, Krasny L, Johnson M, et al. Superior temporal gyrus, language function, and autism. *Dev Neuropsychol*. 2007; 31:217–238. [PubMed: 17488217]
57. Zald, DH., Rauch, SL. *The orbitofrontal cortex*. Oxford; New York: Oxford University Press; 2006. p. 1

58. Aguilar EJ, García-Martí G, Martí-Bonmatí L, Lull JJ, Moratal D, Escartí MJ, et al. Left orbitofrontal and superior temporal gyrus structural changes associated to suicidal behavior in patients with schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry*. 2008; 32:1673–1676. [PubMed: 18657587]
59. Rolls ET. The orbitofrontal cortex. *Philos Trans R Soc Lond B Biol Sci*. 1996; 351:1433–1443. [PubMed: 8941955]
60. Kringelbach ML. The human orbitofrontal cortex: linking reward to hedonic experience. *Nat Rev Neurosci*. 2005; 6:691–702. [PubMed: 16136173]
61. Trainor PA, Krumlauf R. Hox genes, neural crest cells and branchial arch patterning. *Curr Opin Cell Biol*. 2001; 13:698–705. [PubMed: 11698185]
62. Guthrie S. Patterning and axon guidance of cranial motor neurons. *Nat Rev Neurosci*. 2007; 8:859–871. [PubMed: 17948031]
63. Bostan AC, Dum RP, Strick PL. Cerebellar networks with the cerebral cortex and basal ganglia. *Trends Cogn Sci*. 2013; 17:241–254. [PubMed: 23579055]
64. Robinson SR, Smotherman WP. Fundamental motor patterns of the mammalian fetus. *J Neurobiol*. 1992; 23:1574–1600. [PubMed: 1487750]
65. Salihagic-Kadic A, Kurjak A, Medi M, Andonotopo W, Azumendi G. New data about embryonic and fetal neurodevelopment and behavior obtained by 3D and 4D sonography. *J Perinat Med*. 2005; 33:478–490. [PubMed: 16318610]
66. Goldberg JL. How does an axon grow? *Genes Dev*. 2003; 17:941–958. [PubMed: 12704078]
67. Noden DM, Francis-West P. The differentiation and morphogenesis of craniofacial muscles. *Dev Dyn*. 2006; 235:1194–1218. [PubMed: 16502415]
68. Mousavi K, Jasmin BJ. BDNF is expressed in skeletal muscle satellite cells and inhibits myogenic differentiation. *J Neurosci*. 2006; 26:5739–5749. [PubMed: 16723531]
69. Arber S, Burden SJ, Harris AJ. Patterning of skeletal muscle. *Curr Opin Neurobiol*. 2002; 12:100–103. [PubMed: 11861171]
70. Christensen RK, Petersen AV, Perrier JF. How do glial cells contribute to motor control? *Curr Pharm Des*. 2013; 19:4385–4399. [PubMed: 23360271]
71. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Szelinger S, Joshipura K, et al. Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N Engl J Med*. 2007; 357:775–788. [PubMed: 17671248]
72. Chiò A, Schymick JC, Restagno G, Scholz SW, Lombardo F, Lai SL, et al. A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum Mol Genet*. 2009; 18:1524–1532. [PubMed: 19193627]
73. Haidet-Phillips AM, Hester ME, Miranda CJ, Meyer K, Braun L, Frakes A, et al. Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. *Nat Biotechnol*. 2011; 29:824–828. [PubMed: 21832997]
74. Prell T, Grosskreutz J. The involvement of the cerebellum in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener*. 2013; 14:507–515. [PubMed: 23889583]
75. Nishijima H, Tomiyama M, Suzuki C, Kon T, Funamizu Y, Ueno T, et al. Amyotrophic lateral sclerosis with demyelinating neuropathy. *Intern Med*. 2012; 51:1917–1921. [PubMed: 22821113]
76. Federoff, S., Hertz, L. *Advances in cellular neurobiology*. Academic Press; New York: 1982. p. 3
77. Powers HJ. Riboflavin (vitamin B-2) and health. *Am J Clin Nutr*. 2003; 77:1352–1360. [PubMed: 12791609]
78. Lucius R, Mentlein R, Sievers J. Riboflavin-mediated axonal degeneration of postnatal retinal ganglion cells in vitro is related to the formation of free radicals. *Free Radic Biol Med*. 1998; 24:798–808. [PubMed: 9586810]
79. Kim S, Chiba A. Dendritic guidance. *Trends Neurosci*. 2004; 27:194–202. [PubMed: 15046878]
80. Roberts A, Conte D, Hull M, Merrison-Hort R, al Azad AK, Buhl E, et al. Can simple rules control development of a pioneer vertebrate neuronal network generating behavior? *J Neurosci*. 2014; 34:608–621. [PubMed: 24403159]
81. Goodman CS, Bastiani MJ, Doe CQ, du Lac S, Helfand SL, Kuwada JY, et al. Cell recognition during neuronal development. *Science*. 1984; 225:1271–1279. [PubMed: 6474176]

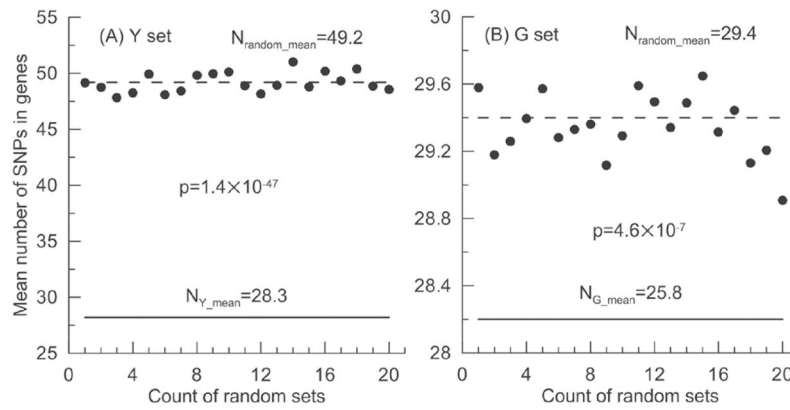
82. Stiles J, Jernigan TL. The basics of brain development. *Neuropsychol Rev.* 2010; 20:327–348. [PubMed: 21042938]
83. Batista-Brito R, Close J, Machold R, Fishell G. The distinct temporal origins of olfactory bulb interneuron subtypes. *J Neurosci.* 2008; 28:3966–3975. [PubMed: 18400896]
84. *Encyclopedia of the eye.* 1. Boston, MA: Elsevier; 2010.
85. Kelley, MW. *Springer handbook of auditory research.* New York: Springer; 2005. Development of the inner ear; p. 5
86. Schlosser G. Making senses development of vertebrate cranial placodes. *Int Rev Cell Mol Biol.* 2010; 283:129–234. [PubMed: 20801420]
87. Sanders, EJ., Lash, JW., Ordahl, CP. *The origin and fate of somites.* IOS Press; 2001. p. 329
88. Normén C, Tammela T, Petrova TV, Alitalo K. Biological basis of therapeutic lymphangiogenesis. *Circulation.* 2011; 123:1335–1351. [PubMed: 21444892]
89. Conway EM, Collen D, Carmeliet P. Molecular mechanisms of blood vessel growth. *Cardiovasc Res.* 2001; 49:507–521. [PubMed: 11166264]
90. Ribatti D, Crivellato E. “Sprouting angiogenesis”, a reappraisal. *Dev Biol.* 2012; 372:157–165. [PubMed: 23031691]
91. Marcucio RS, Cordero DR, Hu D, Helms JA. Molecular interactions coordinating the development of the forebrain and face. *Dev Biol.* 2005; 284:48–61. [PubMed: 15979605]
92. Kinder SJ, Tsang TE, Wakamiya M, Sasaki H, Behringer RR, Nagy A, et al. The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. *Development.* 2001; 128:3623–3634. [PubMed: 11566865]
93. Dykes IM. *Left Right Patterning, Evolution and Cardiac Development.* *J Cardiovasc Dev Dis.* 2014; 1:52–72.
94. Dressler GR. Advances in early kidney specification, development and patterning. *Development.* 2009; 136:3863–3874. [PubMed: 19906853]
95. Cunha GR, Hom YK. Role of mesenchymal-epithelial interactions in mammary gland development. *J Mammary Gland Biol Neoplasia.* 1996; 1:21–35. [PubMed: 10887478]
96. Wells KL, Gaete M, Matalova E, Deutsch D, Rice D, Tucker AS. Dynamic relationship of the epithelium and mesenchyme during salivary gland initiation: the role of Fgf10. *Biol Open.* 2013; 2:981–989. [PubMed: 24167707]
97. Radisky DC. Epithelial-mesenchymal transition. *J Cell Sci.* 2005; 118:4325–4326. [PubMed: 16179603]
98. Leese HJ. Metabolism of the preimplantation embryo: 40 years on. *Reproduction.* 2012; 143:417–427. [PubMed: 22408180]
99. Grigoryan T, Wend P, Klaus A, Birchmeier W. Deciphering the function of canonical Wnt signals in development and disease: conditional loss-and gain-of-function mutations of beta-catenin in mice. *Genes Dev.* 2008; 22:2308–2341. [PubMed: 18765787]
100. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol.* 2004; 20:781–810. [PubMed: 15473860]
101. Sonderegger S, Pollheimer J, Knöfler M. Wnt signalling in implantation, decidualisation and placental differentiation--review. *Placenta.* 2010; 31:839–847. [PubMed: 20716463]
102. Schmidt C, Patel K. Wnts and the neural crest. *Anat Embryol (Berl).* 2005; 209:349–355. [PubMed: 15891909]
103. Choe CP, Collazo A, Trinh le A, Pan L, Moens CB, Crump JG. Wnt-dependent epithelial transitions drive pharyngeal pouch formation. *Dev Cell.* 2013; 24:296–309. [PubMed: 23375584]
104. Aulehla A, Pourquie O. Signaling gradients during paraxial mesoderm development. *Cold Spring Harb Perspect Biol.* 2010; 2:a000869. [PubMed: 20182616]
105. Cisternas P, Henriquez JP, Brandan E, Inestrosa NC. Wnt signaling in skeletal muscle dynamics: myogenesis, neuromuscular synapse and fibrosis. *Mol Neurobiol.* 2014; 49:574–589. [PubMed: 24014138]
106. Zhou H, Mak W, Kalak R, Street J, Fong-Yee C, Zheng Y, et al. Glucocorticoid-dependent Wnt signaling by mature osteoblasts is a key regulator of cranial skeletal development in mice. *Development.* 2009; 136:427–436. [PubMed: 19141672]

107. Petersen CP, Reddien PW. Wnt signaling and the polarity of the primary body axis. *Cell*. 2009; 139:1056–1068. [PubMed: 20005801]
108. Martindale MQ, Hejnal A. A developmental perspective: changes in the position of the blastopore during bilaterian evolution. *Dev Cell*. 2009; 17:162–174. [PubMed: 19686678]
109. Zou Y, Lyuksyutova AI. Morphogens as conserved axon guidance cues. *Curr Opin Neurobiol*. 2007; 17:22–28. [PubMed: 17267201]
110. Maro GS, Klassen MP, Shen K. A beta-catenin-dependent Wnt pathway mediates anteroposterior axon guidance in *C. elegans* motor neurons. *PLoS One*. 2009; 4:e4690. [PubMed: 19259273]
111. Fenstermaker AG, Prasad AA, Bechara A, Adolfs Y, Tissir F, Goffinet A, et al. Wnt/planar cell polarity signaling controls the anterior-posterior organization of monoaminergic axons in the brainstem. *J Neurosci*. 2010; 30:16053–16064. [PubMed: 21106844]
112. Pei Y, Brun SN, Markant SL, Lento W, Gibson P, Taketo MM, et al. WNT signaling increases proliferation and impairs differentiation of stem cells in the developing cerebellum. *Development*. 2012; 139:1724–1733. [PubMed: 22461560]
113. Salinas PC, Zou Y. Wnt signaling in neural circuit assembly. *Annu Rev Neurosci*. 2008; 31:339–358. [PubMed: 18558859]
114. Cao F, Yin A, Wen G, Sheikh AM, Tauqeer Z, Malik M, et al. Alteration of astrocytes and Wnt/ $\beta$ -catenin signaling in the frontal cortex of autistic subjects. *J Neuroinflammation*. 2012; 9:223. [PubMed: 22999633]
115. Ille F, Atanasoski S, Falk S, Ittner LM, Märki D, Büchmann-Møller S, et al. Wnt/BMP signal integration regulates the balance between proliferation and differentiation of neuroepithelial cells in the dorsal spinal cord. *Dev Biol*. 2007; 304:394–408. [PubMed: 17292876]
116. Pinto C, Cárdenas P, Osses N, Henríquez JP. Characterization of Wnt/ $\beta$ -catenin and BMP/Smad signaling pathways in an in vitro model of amyotrophic lateral sclerosis. *Front Cell Neurosci*. 2013; 7:239. [PubMed: 24348333]
117. Sánchez-Camacho C, Bovolenta P. Emerging mechanisms in morphogen-mediated axon guidance. *Bioessays*. 2009; 31:1013–1025. [PubMed: 19705365]
118. Charron F, Tessier-Lavigne M. The Hedgehog, TGF-beta/BMP and Wnt families of morphogens in axon guidance. *Adv Exp Med Biol*. 2007; 621:116–133. [PubMed: 18269215]
119. Hashiguchi M, Mullins MC. Anteroposterior and dorsoventral patterning are coordinated by an identical patterning clock. *Development*. 2013; 140:1970–1980. [PubMed: 23536566]
120. Robinson SR, Kleven GA, Brumley MR. Prenatal Development of Interlimb Motor Learning in the Rat Fetus. *Infancy*. 2008; 13:204–228. [PubMed: 20198121]
121. Walhovd KB, Fjell AM, Brown TT, Kuperman JM, Chung Y, Hagler DJ Jr, et al. Long-term influence of normal variation in neonatal characteristics on human brain development. *Proc Natl Acad Sci U S A*. 2012; 109:20089–20094. [PubMed: 23169628]



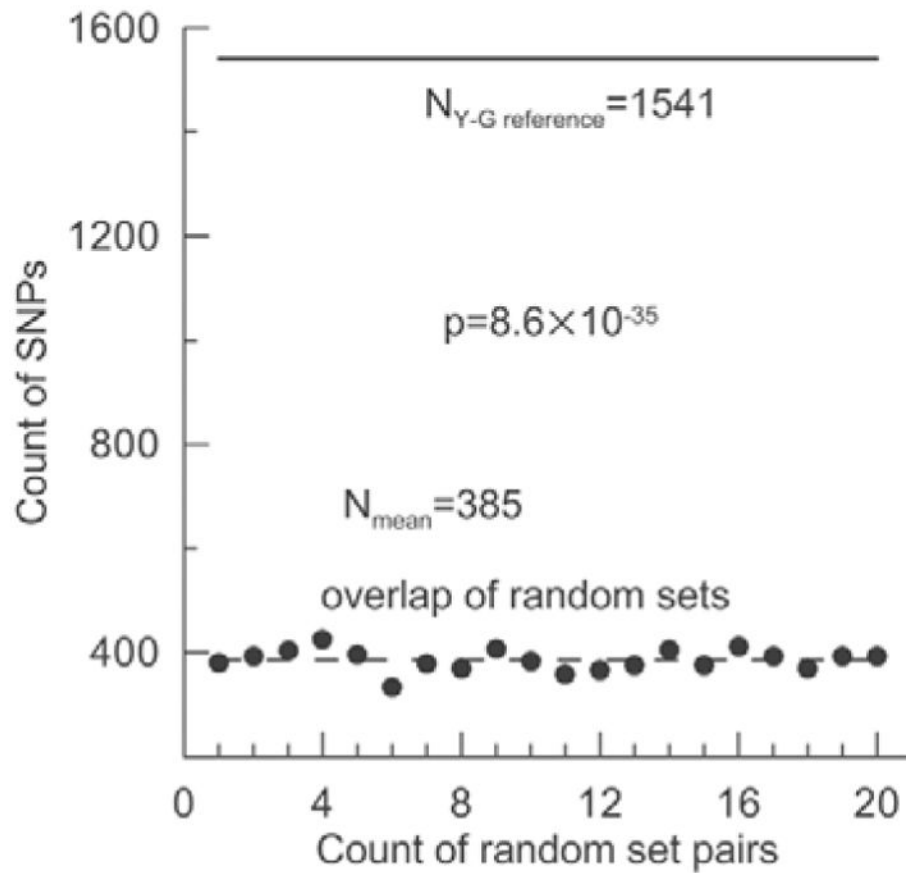
**Figure 1. Clustering of SNPs in genes**

Solid line denotes the number of genes for SNPs from the (A) Y and (B) G reference SNP sets. Filled dots show the number of genes for SNPs from each of the random SNP sets matching the (A) Y and (B) G reference SNP sets. The differences between mean numbers of genes in random sets (dashed line) and the number of genes in the reference sets (solid line) are highly significant for the Y and G sets.

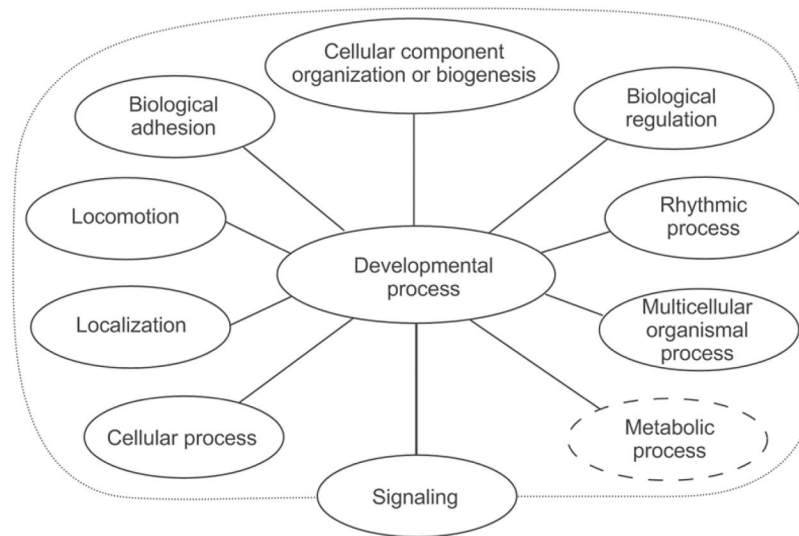


**Figure 2. Averaged length of genes in the reference and random sets**

Solid line denotes mean number of SNPs in genes (that is a proxy of gene length) for the (A) Y and (B) G reference sets. Filled dots show mean number of SNPs in genes from each of the random SNP sets matching the (A) Y and (B) G reference sets. The differences between mean number of SNPs in genes in random sets (dashed line) and the mean number of SNPs in genes in the reference sets (solid line) are highly significant for the Y and G sets.

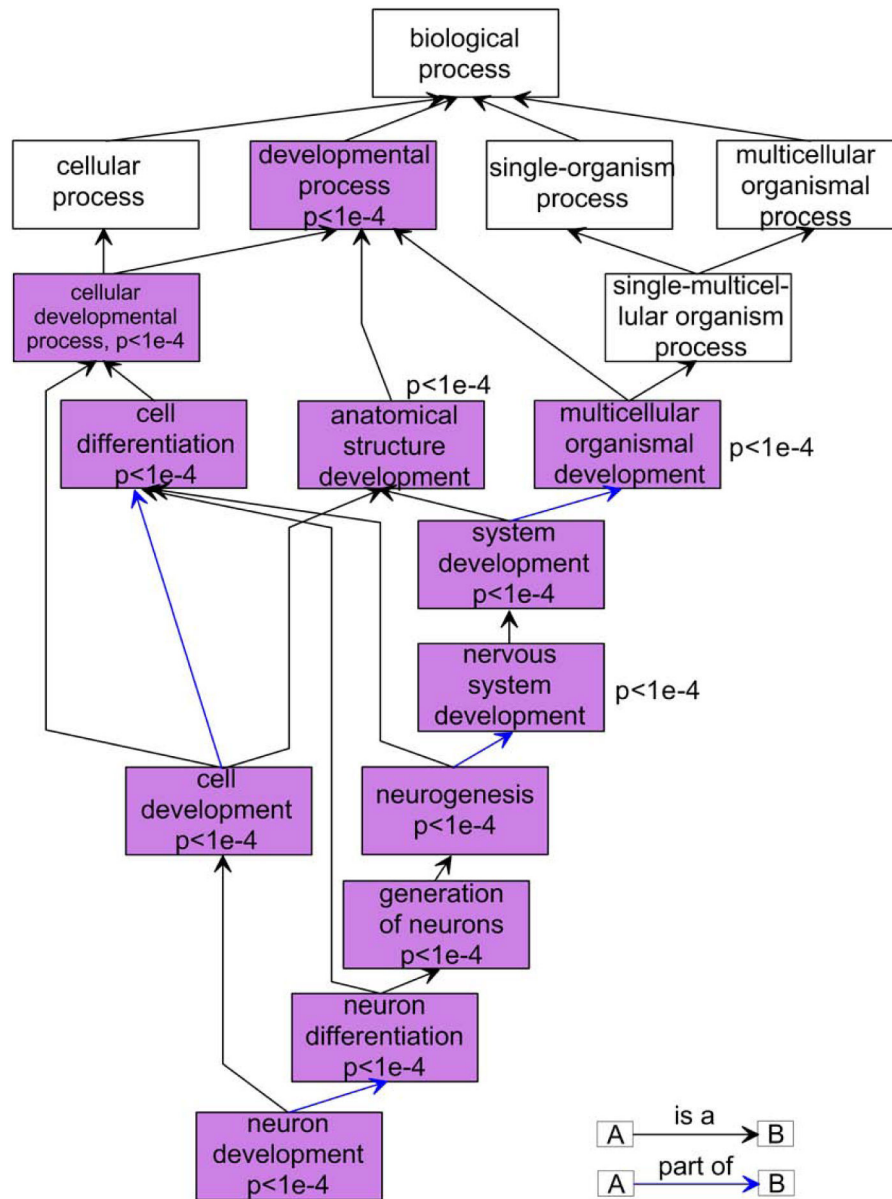


**Figure 3. Overlap of SNPs between the reference sets and between the random sets**  
 Solid line denotes the number of SNPs overlapping between the Y and G reference sets. Filled dots show the number of SNPs overlapping between pairs of random sets matching the Y and G reference sets. The difference between the overlap of the reference sets (solid line) and the mean overlap of pairs of random sets (dashed line) are highly significant.



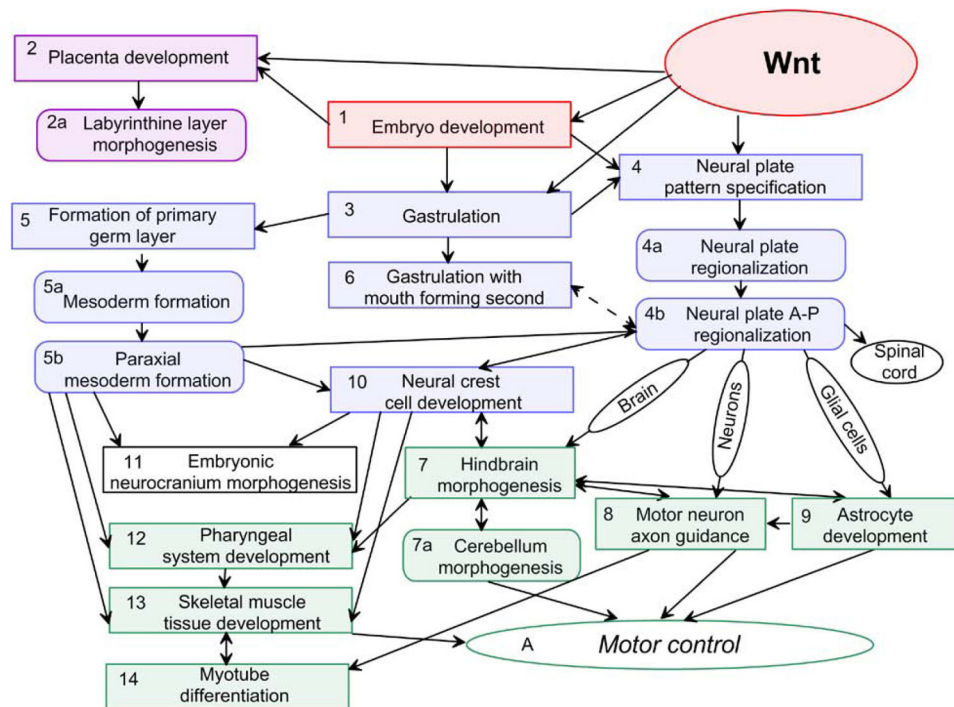
**Figure 4. General biological process GO terms significantly enriched by genes from the reference Y and G sets**

Embryo development encompasses the differentiation, morphogenesis and growth of cells, tissues, organs, and organ systems. The developmental processes characteristic for embryogenesis tightly cooperate with other complex, biologically important processes that drive many aspects of development. Multiple signaling events integrate the information, coordinate, trigger and ensure regulation of biological processes to correct and complete formation of fetus under various environments. Solid ovals show highly-significantly enriched processes. Although GO term metabolic process was not significantly enriched, its major sub-processes were highly significant.



**Figure 5. GO terms tree for neuron development**

Related higher level developmental BPs are significantly enriched (with p-values shown in the figure) by genes for SNPs from the Y set.



**Figure 6. Scheme of a system of interlinked developmental processes controlled by the Wnt signaling pathway represented by the selected Y set-specific BPs**

Each of the numbered specific processes was neither enriched in the G set nor in majority (18+) of the Y set related 20 control random SNP sets. The entire system was not enriched in any random set. **Description. Wnt signaling** is essential for various aspects of embryo development (1) such as the early placental development (2), gastrulation (3), and pattern formation, including patterning of the neural plate (4). **The placenta** (2) develops from embryonic (1) and maternal tissues. The labyrinthine layer (2a) of the placenta provides a vital link between the embryonic and the maternal circulatory systems. **Gastrulation** (3) of all vertebrate embryos involves the development of three distinct primary germ layers (5), early embryonic patterning (e.g. regionalization of neural plate (4, 4a, 4b)) and formation of the primitive gut (e.g., gastrulation with mouth forming second (6)). **Mesoderm** (5a) is the middle of the three germ layers. The paraxial mesoderm (5b) is an early mesoderm derivative, that contributes extensively to many adult tissues. GO term gastrulation with mouth forming second (6) highlights the origin of a through gut of bilaterally symmetric organisms that initially regionalized along A-P axis. The neural plate, the CNS primordium, lengthens along the A-P axis. **Regionalization of neural plate** (4, 4a, 4b) is the first step in neural patterning. The neural plate (4b) can be patterned by signals from the paraxial mesoderm (5b). The neural plate transforms into the neural tube which forms the brain and spinal cord. The neural plate is also the source of the majority of neurons and glial cells. **The hindbrain** (7) is a key source of patterning information and the motor innervation in the developing head. The cerebellum (7a) arises from the dorsal surface of the hindbrain (7) and plays an important role in motor control (A). Hindbrain (7) neural progenitors give rise to cranial motor neurons and astrocyte as well. **Motor neurons** axon guidance (8) is the process by which motor neurons send out the nerve fibers or axons to reach the correct

targets. Astrocytes (9) are the most abundant glial cell type in the CNS. **Astrocytes** (9) contribute to axon guidance (8) and play a significant role in motor neuron survival and in cerebellar development (7, 7a). The neural plate (4b) contributes to neural crest (10) along with the underlying paraxial mesoderm (5b). The neural crest is first induced at neural plate border but then neural crest cells migrate elsewhere. The paraxial mesoderm (5b) and the neural crest (10) are both responsible for the development of the craniofacial skeleton, particularly **neurocranium** (11). Hindbrain-derived neural crest cells (10) and paraxial mesoderm (5b) that located superficially beside the hindbrain are the source of **the pharyngeal arches** (12). The pharyngeal arches develop into **skeletal muscle** (13) and cartilage. The cranial neural crest (10) plays an important role in patterning muscular organization (13). The unsegmented paraxial mesoderm (5b) of the head region gives rise to the head and neck skeletal muscles (13) as well. Muscle fibers are formed by the maturation of myotubes (14). Motor axons (8) contact myotubes (14) in skeletal muscles. The hindbrain motor neurons (8) innervate skeletal muscles (13) derived from the pharyngeal arches (12). Developmental relationships between these *characteristic for the Y set* specific processes controlled by Wnt signaling emphasizes organization in the head region and formation of basic apparatus for motor control (A). This is in line with crucial roles of Wnt signaling in synapse organization and neuromuscular junction development which occurs synchronously.

**Table 1**

Selected specific biological processes characteristic for the Y and G sets and enriched at  $p < 0.05$  and  $q < 0.10$ .

GO term	p-value	q-value	Gene Hit	Gene Total
<b>The Y set</b>				
neural plate anterior/posterior regionalization	2.7E-3	8.8E-2	3	4
pharyngeal system development	3.4E-3	1.0E-1	4	8
motor axon guidance	5.0E-4	1.8E-2	6	13
astrocyte activation	2.9E-3	8.6E-2	3	5
genetic imprinting	1.7E-3	5.4E-2	6	11
peptidyl-arginine modification	5.0E-4	2.1E-2	5	9
cellular carbohydrate metabolic process	4.0E-3	1.0E-2	58	444
O-glycan processing	2.7E-3	8.8E-2	3	4
<b>The G set</b>				
developmental cell growth	8.8E-3	9.2E-2	22	55
actin filament-based movement	2.3E-3	3.9E-2	23	54
cell cycle phase	7.3E-3	9.0E-2	205	728
prophase	8.5E-3	9.8E-2	8	14
meiotic prophase I	6.3E-3	8.8E-2	7	11
pachytene	3.5E-3	5.9E-2	4	4
stem cell division	8.5E-3	9.8E-2	8	14
DNA strand elongation	2.2E-3	3.6E-2	17	36
DNA strand elongation involved in DNA replication	2.1E-3	3.4E-2	16	33
cellular carbohydrate catabolic process	7.7E-3	9.0E-2	41	119
hexose catabolic process	9.4E-3	1.0E-1	31	86
L-fucose metabolic process	6.3E-3	8.8E-2	7	11
regulation of systemic arterial blood pressure mediated by a chemical signal	4.2E-3	7.4E-2	14	29

Specific biological processes characteristic for the Y (G) set were neither significant (at  $p < 0.05$  and  $q < 0.10$ ) in the G (Y) set nor in 18+ of 20 control random SNP sets.