UMIACS-TR-2020-02 **February 2020**

Nervous system maps on the C. elegans genome

Christopher Cherniak[1*], Zekeria Mokhtarzada[2], Raul Rodriguez-Esteban[3]

[1] University of Maryland; cherniak@umd.edu
[2] zekitm@gmail.com
[2] raul.rodriguez.esteban@gmail.com
* Correspondence: cherniak@umd.edu; Tel: 1 202 667-9224
Fax: 1 301 405-5690

1. Introduction

Abstract: This project begins from a synoptic point of view, focusing upon the large-scale (global) landscape of the genome. This is along the lines of combinatorial network optimization in computational complexity theory [1]. Our research program here in turn originated along parallel lines in computational neuroanatomy [2,3,4,5].

Rather than mapping body structure onto the genome, the present report focuses upon statistically significant mappings of the Caenorhabditis elegans nervous system onto its genome. Via published datasets, evidence is derived for a "wormunculus", on the model of a homunculus representation, but on the C. elegans genome. The main method of testing somatic-genomic position-correlations here is via public genome databases, with r^2 analyses and p evaluations.

These findings appear to yield some of the basic structural and functional organization of invertebrate nucleus and chromosome architecture. The design rationale for somatic maps on the genome in turn may be efficient interconnections. A next question this study raises: How do these various somatic maps mesh (interrelate, interact) with each other?

Key Terms: Somatic map of H. sapiens body on genome, Somatic map of H. sapiens body on chromosomes, Cell map on genome, Cell map on chromosomes, Homunculus, "Cellunculus," "Wormunculus," Connection optimization.

2. Review: Somatic mappings on the genome

Computation theory concepts can be used for understanding the structure and function of organism DNA. In particular, the genome itself can be treated like a "nano-brain" or pico-computer to see whether similar connection minimization strategies also appear in gene networks. For decades we have reported wiring optimization in the brain that begins to approach some of the most precisely confirmed predictions in neuroscience.

Two meta-models of the genome compete today: One is an atomistic "genome as hairball" idea, effectively possessing minimal structure. (E.g., our genome is a mess. It is "in an alarming state of disarray" [6].) The

alternative picture, examined here, is structuralist — that the genome itself has largescale global patterns.

A brief review of the context and format for body maps on the genome may be useful. We started by focusing on body organs mapped to tissue-specific genes on the genome (see Figure 1), and on cell organelles mapped to the genome (see Figure 2).

We have been exploring a connection-minimization model for the genome. Information transmission does not appear to be cost-free even within a cell, nucleus, or genome. For instance, genes strongly expressed in particular tissues are not just randomly distributed in the genome. Rather, the arrangement of such tissue-specific gene positions in the complete chromosome set mirrors the antero-posterior, and dorso-ventral, configuration of tissue-locations in the body. A statistically significant supra-chromosomal "genome homunculus" — a global, multi-dimensional, somatotopic mapping of the human body — appears to extend across chromosome territories in the entire sperm cell nucleus [7]. See Figure 1. Such a mapping is a strategy for connection cost-minimization (e.g., cf. body maps reported in sensory and motor cortex since the 19th century). Also, corresponding finer-scale somatotopic mappings seem to occur on each individual autosomal chromosome [8].

Furthermore, organelle sub-structure of the typical individual eukaryotic animal cell also turns out to map similarly as a "cellunculus" onto the total genome, via organelle-specific genes that express more strongly in particular organelle types [9]. See Figure 2. So, genome as palimpsest: multiple maps, at different scales, seem superimposed upon the genome.

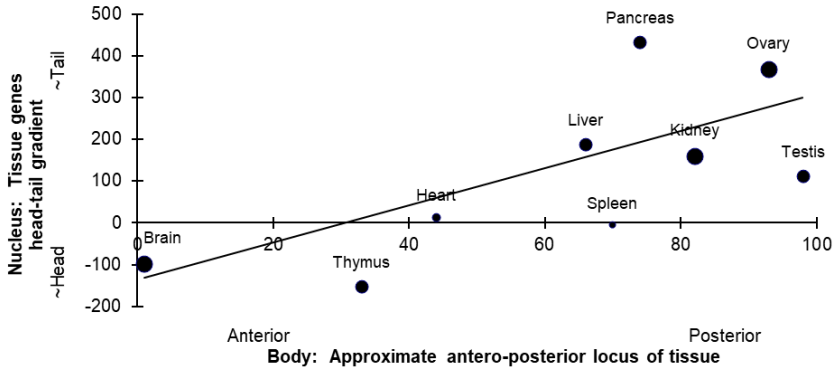

Figure 1. Antero-posterior "gradient of gradients" in nucleus. Tissue location in human body correlates significantly with pattern of tissue genes' positions in cell nucleus. (For 9 datapoints each weighted by their own significance, $r^2 = 0.62$; $p < 0.01$, 2 tail.) That is, tissue location-in-body relates to its genes' distribution-gradient in the complete genome. The more forward-placed a tissue in the body, the more forward-placed its genes on chromosomes in nucleus. — The head of the genome homunculus is at the head of the spermcell nucleus. A corresponding body-genome mapping also holds for the dorso-ventral body axis [7] . In addition, the body similarly maps onto individual chromosomes [8].

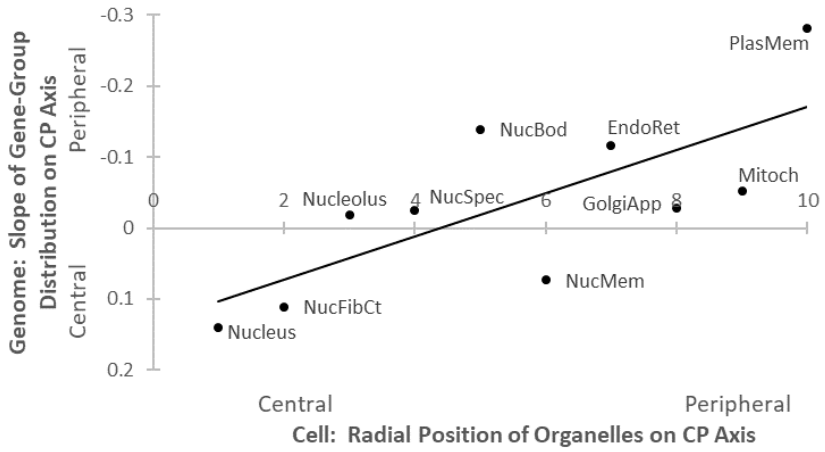

Figure 2. Isomorphism of cell microanatomy and largescale human genome structure: Components positioned more centrally in a cell tend to have their genes correspondingly concentrated on chromosomes sited more toward the center of genome. In a plot of 10 organelles, this cell-genome correlation is significant ($r^2 = 0.540$; $p < 0.015$, 2 tail). Each of the datapoints is labelled with its organelle-name [9].

Our prior goal had been to uncover a "body → genome" somatic map for the entire organism. (See, e.g., Figure 1 above.) The project of the present report is then to explore some qualitatively finer-grained somatic mappings, zooming in on a complete single organ system (e.g., the 11-ganglion invertebrate nervous system of the nematode (see Figure 3)).

3. "Graphical Abstract" (master diagram).

See **Figure 3** below: C. elegans "nervous system → genome" mapping.

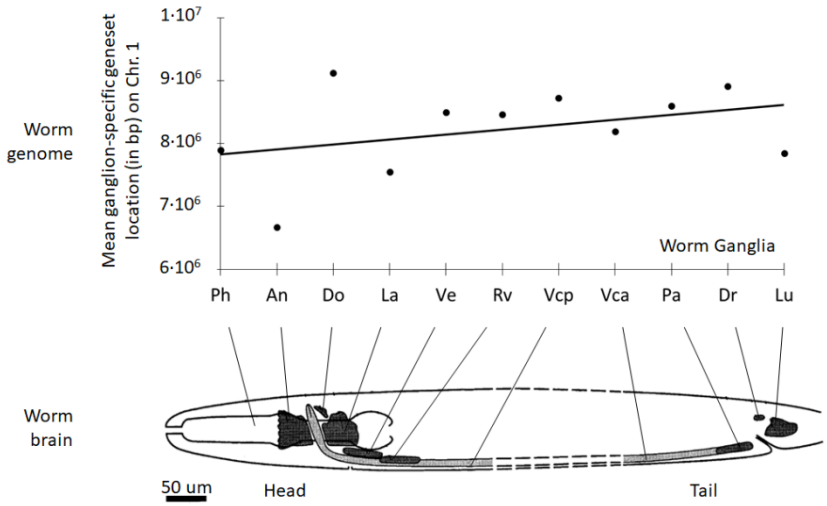

Figure 3. Diagram of ganglia of the ~1 mm-long Caenorhabditis elegans hermaphrodite: their body locations and schematized shapes. The 11 ganglia (Ph – Lu) constitute the C. elegans nervous system, with ~300 neurons. - It approximates a "1-dimensional (antero-posterior) nervous system". Master diagram (Figure 3): [= worm nervous

system (below) + genome plot (above)]. Best-fit line for the ganglion-genesets on Chr 1 above: $y = 78581x + 8E + 06$ [with log transforms of scores]. (How nervous system and genome and mesh together: Ring (between An & Do ganglia) constitutes the main crossbar matrix for interconnections, maximizing wiring configuration flexibility.))

The ganglia are positioned approximately end-to-end, with partial overlap of some contiguous ganglia. The anterior and lateral (largest) ganglia and the circumpharyngeal ring in turn surround portions of the pharynx muscles and neurons. The anterior, dorsal, and lateral ganglia directly abut the ring neuropil. For neuron counts of ganglia, see Table S1 below [2]. Total length of body: ~1300 um. Derived in part from [10 – 13]. Worm neuroanatomy based on [2]. See also Figure 2, [2] for a complete worm nervous system map at the individual neuron, and synapse, level.

Our original goal had been to uncover a "body ➞ genome" somatic map for the entire organism. (See, e.g., Figure 1.) The project of the present report is then to explore somatic mappings, zooming in on a complete single organ system (e.g., the 11-ganglion invertebrate nervous system of the nematode (see Figure 3)).

The nervous system of C. elegans [13] includes 11 ganglionic components, which have 11! (= ~40,000,000) alternative possible anteroposterior orderings. In fact, the actual layout happens to require the minimum total possible wirelength, a predictive success story [2].

Stages of connection-tracing a ganglion to its gene-sites on a chromosome:
1. Ganglion
➞ **2**. Its neuron set [see supplementary Table S1, below]
➞ **3**. All their genes [https://wormbase.org/species/all/anatomy_term#1-0-5]v.WS276][14].
[https://wormbase.org/about/wormbase_release_WS276]
➞ **4**. And, their chromosome loci (see also wormbase.org).

The nervous system of the nematode C. elegans [13] includes 11 ganglionic components, which have 11! (40,000.000) alternative possible orderings. In fact, the actual ganglion layout happens to require the minimum possible total wirelength. [2]. Furthermore, all 6 chromosomes [I, II, II, IV, V, X] each also show a log transformed positive slope – as for Chr I in Figure 3 above. ("Log Transform" is a conventional treatment to correct skewed distribution of data.)

The "brain ➞ genome" mapping-slopes are all positive [+]. Significance of a binomial test for the 6 out of 6 trials is $p < 0.0313$ (2 tail). (See Table 1.)

Table 1. All 6 C. elegans chromosomes:
Each has a positive slope ($p < 0.003$, 2 tail).

Chromosome	Raw Slope	Log Transformed
I	68321	78581
II	125226	125131
III	3130	67247
IV	-1796	244
V	62703	10027
X	37675	65319

It is not easy to compare chromosome slope-signs across species. A moment's study of supplementary Table S2 (below) indicates that the 22 H. sapiens chromosomes' slope signs [+ vs -] are evenly distributed, vs the 6 C. elegans chromosome slope signs, which are all positive – a striking vertebrate/ invertebrate difference. Perhaps the parallel map-slopes on the 6 C. elegans chromosomes could be interpreted as a structural strategy simply to assist in maintaining the separate chromosome maps meshed in "registration" (i.e., synchrony).

4. Conclusion. "Genome without structure"?

Somatic mappings onto the genome relate to their functioning. However, much theorizing rejects such order in the genome: For instance, in general, "The genome is a junkyard." [15]; similarly, the human genome "seems to be in an alarming state of disarray" [6]. On the one hand, such structurelessness has costs — access to particular genes then requires direct brute force search. In this way, the very vehicle of innateness itself is denied largescale structure; chaos rules. ("And then a miracle occurs." [16]) On the other hand, if realworld genetic systems had unlimited capacities (e.g., to squeeze code rapidly through genomic bottlenecks), such mappings would be unneeded. Highly idealized models are widespread in biology, and simplify theorizing.

Another caveat: Some decades ago, our laboratory reported that the economy of C. elegans nervous system wiring was effectively highly optimized [17,2]. "Save wire" turns out to yield correct predictions of brain and genome structure, sometimes down to one-in-a-billion precision [18]. Correspondingly, the present report argues that the worm's 11-ganglion nervous system significantly maps to the worm's genome: an image of this wormbrain appears on its genome [Figure 3]. However, some similarly well-optimized vertebrate cerebral cortexes are not yet shown to map well onto vertebrate genomes [3]. Similarly, comparatively weaker significant body-genome maps can be identified on mammalian genomes [Figure 1]. Thus far, body-genome maps appear relatively independent of brain optimization.

*** * ***

Christopher Cherniak

Zekeria Mokhtarzada

Raul Rodriguez-Esteban

Committee for Philosophy and the Sciences

Department of Philosophy

University of Maryland College Park, MD 20742 USA

http://terpconnect.umd.edu/~cherniak/

Table S1a. Supplementary table for neuron-content in ganglia of the first (anterior) half [PH - VE] of the C.elegans nervous system.

PHARYNX (20 neurons) PH					
Ila	Ilb	I2a	I2b	I3	I4
I5	I6	M1	M2a	M2b	M3a
M3b	M4	M5	Mca	MCb	MI
NSMa	NSMb				
ANTERIOR (36 neurons) AN					
BAGL	BAGR	CEPVL	CEPVR	IL1DL	IL1R
IL1DL	IL1DR	IL1VL	IL1VR	IL2L	IL2R
IL2DL	IL2DR	IL2VL	IL2VR	OLLL	OLLR
OLQDL	OLQDR	OLQVL	OLQVR	RIPL	RIPR
RMEL	RMER	URADL	URADR	URAVL	URAVR
URBL	URBR	URYDL	URYDR	URYVL	URYVR
[sh, so]					
DORSAL (6 neurons) DO					
ALA	CEPDL	CEPDR	RID	URXL	URXR
LATERAL (64 neurons) LA					
ADFL	ADFR	ADLL	ADLR	AFDL	AFDR
AIBL	AIBR	AINL	AINR	AIZL	AIZR
ASEL	ASER	ASGL	ASGR	ASHL	ASHR
ASIL	ASIR	ASJL	ASJR	ASKL	ASKR
AUAL	AUAR	AVAL	AVAR	AVBL	AVBR
AVDL	AVDR	AVEL	AVER	AVHL	AVHR
AVJL	AVJR	AWAL	AWAR	AWBL	AWBR
AWCL	AWCR	RIAL	RIAR	RIBL	RIBR
RICL	RICR	RIML	RIMR	RIVL	RIVR
RMDL	RMDR	RMDVL	RMDVR	SAAVL	SAAVR
SIBDL	SIBDR	SMDVL	SMDVR		
VENTRAL (32 neurons) VE					
AIAL	AIAR	AIML	AIMR	AIYL	AIYR
AVKL	AVKR	AVL	RIH	RIR	RIS
RMDDL	RMDDR	RMFL	RMFR	RMHL	RMHR
SAADL	SAADR	SIADL	SIADR	SIAVL	SIAVR
SIBVL	SIBVR	SMBDL	SMBDR	SMBVL	SMBVR
SMDDL	SMDDR				

(Compiled by C. Cherniak, 1990 - 2016.)

Table S1p. Supplementary table for neurons in ganglia of the last (posterior) half [RV - LU] of the nervous system.

RETRO-V (20 neurons) RV					
AS1	AVFL	AVFR	AVG	DA1	DB1
DB2	DD1	RIFL	RIFR	RIGL	RIGR
SABD	SABVL	SABVR	VA	VB1	VB2
VD1	VD2				
VENTRAL Cord (58 neurons) VC					
AS2	AS3	AS4	AS5	AS6	AS7
AS8	AS9	AS10	DA2	DA3	DA4
DA5	DA6	DA7	DB3	DB4	DB5
DB6	DB7	DD2	DD3	DD4	DD5
VA2	VA3	VA4	VA5	VA6	VA7
VA8	VA9	VA10	VA11	VB3	VB4
VB5	VB6	VB7	VB8	VB9	VB10
VB11	VC1	VC2	VC3	VC4	VC5
VC6	VD3	VD4	VD5	VD6	VD7
VD8	VD9	VD10	VD11		
PRE-Anal (12 neurons) PA					
AS11	DA8	DA9	DD6	PDA	PDB
PVPl	PVPR	PVT	VA12	VD12	VD13
DORSO-Rectal (3 neurons) DR					
DVA	DVB	DVC			
VENTRAL (32 neurons) VE					
AIAL	AIAR	AIML	AIMR	AIYL	AIYR
AVKL	AVKR	AVL	RIH	RIR	RIS
RMDDL	RMDDR	RMFL	RMFR	RMHL	RMHR
SAADL	SAADR	SIADL	SIADR	SIAVL	SIAVR
SIBVL	SIBVR	SMBDL	SMBDR	SMBVL	SMBVR
SMDDL	SMDDR				
LUMBAR (24 neurons) LU					
ALNL	ALNR	LUAL	LUAR	PHAL	PHAR
PHBL	PHBR	PHCL	PHCR	PLML	PLMR
PLNL	PLNR	PQR	PVCL	PVCR	PVNL
PVNR	PVQL	PVQR	PVR	PVWL	PVWR
[sh, so]					
NON-GANGLIONIC (27 neurons) NG					
<u>Head:</u>					
ADAL	ADAR	ADEL	ADER	AQR	FLPL
FLPR	RMED	RMEV	RMGL	RMGR	
<u>Body:</u>					
AIML	AIMR	AVM	BDUL	BDUR	CANL
CANR	HSNL	HSNR	PDEL	PDER	PVDL
PVDR	PVM	SDQL	SDQR		

Table S2. Division of labor: Individual autosomal chromosomes show a statistically significant preference for a bodymapping that is either anteroposterior or dorsoventral. See **boldface** blocks below. (Chrs 5 and 3 are marginally AP.) For mean slope data of best-fit regression lines.

Chromo	Mean AP Slope	$p <$	Mean DV Slope	$p <$	$p <$
AnteroPosterior Axis Map					
16	**1,604,409**	**5.33E-05**	191,576	0.012381	1.41E-06
7	**-1,472,407**	**1.26E-05**	628,621	0.066309	1.44E-05
10	**-1,470,431**	**0.000119**	659,221	0.010798	2.3E-06
4	**1,431,861**	**0.003314**	462,782	0.013728	0.036203
13	**-1,394,665**	**0.012927**	-302,921	0.573604	0.014447
20	**1,016,725**	**2.36E-06**	71,342	0.526125	6.02E-06
3	**984,114**	**0.058667**	-63,595	0.883411	0.111664
5	**-586,941**	**0.074487**	-69,349	0.757337	0.178083
11	**567,410**	**0.000912**	26,274	0.865145	0.012288
19	**354,616**	**0.000632**	76,037	0.193754	0.007838
21	**-185,325**	**0.001703**	-119,651	0.241717	0.549434
22	**-166,935**	**0.042127**	-121,929	0.365256	0.765776
Means	56,869	0.01625	119,867	0.37588	0.13965
DorsoVentral Axis Map					
8	-44,453	0.933389	**3,414,430**	**0.000139**	0.000287
1	-857,225	0.003236	**-2,472,568**	**6.41E-07**	0.000115
18	508,226	0.007738	**-2,223,069**	**0.000287**	3.24E-05
2	127,471	0.602037	**-1,849,409**	**2.59E-05**	1.78E-05
15	420,705	0.012465	**-1,744,266**	**0.000244**	2.24E-05
9	-212,524	0.344678	**986,262**	**0.002161**	0.00149
12	7,051	0.979252	**703,600**	**0.00353**	0.045874
17	477,013	0.004346	**-553,692**	**5.83E-06**	2.63E-06
14	140,904	0.230478	**-524,061**	**0.026944**	0.01125
6	-224,168	0.321311	**472,382**	**0.003019**	0.012026
Means	34,300	0.34389	-379,039	0.00364	0.00711
X	310,901	0.40721	812,730	0.26145	0.52744

From main Table [S1] data, for 13 SPM settings over range 0.3 - 0.9, by 0.05 increments [8]. Ordered by absolute magnitude of (significant) mean slope values.

p-values are from *t* tests (2 tailed). For comparison, sex chromosome X has neither significant AP nor DV bodymaps.

References

1. Cherniak, C. *Minimal Rationality*, Cambridge MA: MIT Press, 1986.
2. Cherniak, C. Component placement optimization in the brain, *J. Neurosci*. 1994b, 14, 2418-2427.
3. Cherniak, C.; Mokhtarzada, Z.; Rodriguez-Esteban, R.; and Changizi, B. Global optimization of cerebral cortex layout, *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 1081-1086.
4. Lewis, S.; Christova, P.; Jerde, T.; and Georgopoulos, A. A compact and realistic cerebral cortical layout derived from prewhitened resting-state fMRI time series: Cherniak's adjacency rule, size law, and metamodule grouping upheld, *Front. Neuroanat*. 2012, 36.
5. Cherniak, C. Bounded-Resource Models in Biology, Viale, R. in press in *Routledge Handbook of Bounded Rationality*, Viale, R., Ed.; Taylor & Francis, Abingdon, UK, 2021. Ch. 24.
6. Alberts, B; Johnson, A; Lewis, J.; Raff, M.; Roberts, K; Walter, P. *Molecular Biology of the Cell*, 6th ed. Garland Science, New York, U.S.A., 2015, Ch. 4.
7. Cherniak, C.; Rodriguez-Esteban, R. Body maps on the human genome, *Mol. Cytogenet.* 2013, *6*, (1) 61.
8. Cherniak, C.; Rodriguez-Esteban, R. Body maps on human chromosomes, College Park, MD: University of Maryland Institute for Advanced Computer Studies Tech Report, 2015, UMIACS-TR-2015-04.
9. Cherniak, C.; and Rodriguez-Esteban, R. Cell maps on the human genome, College Park, MD: University of Maryland Institute for Advanced Computer Studies Tech Report, 2018, UMIACS-TR-2018-01. And, Mol. Cytogenet. 2019, 12: 14.
10. Albertson, D.; Thompson, J. The Pharynx of Caenorhabditis elegans, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 1976, 275: 299-325.
11. White J.; et al., The Structure of the Ventral Nerve Cord of Caenorhabditis elegans, *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 1976, 275: 327-348.
12. White J., et al., The structure of the nervous system of Caenorhabditis elegans, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 1986, 314: 1-340.
13. Wood, W., ed. *The Nematode Caenorhabditis elegans*, New York, Plainview: New York, U.S.A., 1988 Cold Spring Harbor Laboratory.
14. Harris, T. WormBase (https://wormbase.org/) *Nucleic Acids Research*, 48, Issue D1, 08 Jan 2020, pp. D762–D767, https://doi.org/10.1093/nar/gkz920 (accessed 1 Jul, 2020).
15. Brenner,S. *Chronicles of Evolution*. Wildtype Media: Singapore, 2018.
16. Harris, S.; Then a Miracle Occurs, *American Scientist*, Sep-Oct 1986, 74, no. 5, 542-545.
17. Aaronson, S. Why philosophers should care about computational complexity. In Copeland,J.; Posy,C.; Shagrir, O. *Computability: Turing, Gödel, Church, & Beyond*, MIT Press Cambridge MA, U.S.A., 2013; pp. 261-327. (sec. 10.1).
18. Dean,S. Computational complexity theory. In Zalta, E. *Stanford Encyclopedia of Philosophy*, In Zalta, E.; 2016, pp. 261-327. (sec. 4.7).