

# Metabolic resilience is encoded in genome plasticity

**Leandro Agudelo** (✉ [agudelo@mit.edu](mailto:agudelo@mit.edu))

Massachusetts Institute of Technology

**Remy Tuyeras**

Massachusetts Institute of Technology

**Claudia Llinares**

Institute of Neuroscience, Universidad Miguel Hernandez-CSIC

**Alvaro Morcuende**

Institute of Neuroscience, Universidad Miguel Hernandez-CSIC

**Yongjin Park**

Massachusetts Institute of Technology

**Na Sun**

Massachusetts Institute of Technology

**Suvi Linna-Kuosmanen**

Massachusetts Institute of Technology <https://orcid.org/0000-0002-3278-2727>

**Naeimeh Atabaki-Pasdar**

Lund University

**Li - Lun Ho**

Computer Science and Artificial Intelligence Laboratory , Massachusetts Institute of Technology ,  
Cambridge, MA, USA

**Kyriakitsa Galani**

Massachusetts Institute of Technology

**Paul Franks**

Lund University <https://orcid.org/0000-0002-0520-7604>

**Burak Kutlu**

Novo Nordisk (United States) <https://orcid.org/0000-0001-6854-327X>

**Kevin Grove**

Novo Nordisk

**Teresa Femenia**

Institute of Neuroscience, Universidad Miguel Hernandez-CSIC

**Manolis Kellis**

Massachusetts Institute of Technology <https://orcid.org/0000-0001-7113-9630>

---

## Biological Sciences - Article

**Keywords:** metabolism, molecular evolution, genome plasticity

**Posted Date:** July 22nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-716965/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Metabolic resilience is encoded in genome plasticity

Leandro Z. Agudelo<sup>1,2, †, \*</sup>, Remy Tuyeras<sup>1,2, †, \*</sup>, Claudia Llinares<sup>4</sup>, Alvaro Morcuende<sup>4</sup>,  
Yongjin Park<sup>1,2</sup>, Na Sun<sup>1,2</sup>, Suvi Linna-Kuosmanen<sup>1,2</sup>, Naeimeh Atabaki-Pasdar<sup>1,6</sup>,  
Li-Lun Ho<sup>1,2</sup>, Kyriakitsa Galani<sup>1,2</sup>, Paul W. Franks<sup>6,7</sup>, Burak Kutlu<sup>5</sup>,  
Kevin Grove<sup>5</sup>, Teresa Femenia<sup>3,4, \*</sup>, and Manolis Kellis<sup>1,2, \*</sup>

<sup>1</sup> Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA, USA

<sup>2</sup> Broad Institute of MIT and Harvard, Cambridge MA, USA

<sup>3</sup> Department of Neuroscience, Karolinska Institutet, Biomedicum, Stockholm, Sweden

<sup>4</sup> Instituto de Neurociencias, Universidad Miguel Hernández-CSIC, Alicante, Spain

<sup>5</sup> Novo Nordisk Research Center Seattle, Washington, USA

<sup>6</sup> Department of Clinical Sciences, Lund University, Lund, Sweden

<sup>7</sup> Department of Nutrition, Harvard School of Public Health, Boston MA, USA

† These authors contributed equally to this work

\* Corresponding authors: [agudelo@mit.edu](mailto:agudelo@mit.edu), [rtuyeras@mit.edu](mailto:rtuyeras@mit.edu), [tfemenia@umh.es](mailto:tfemenia@umh.es), [manoli@mit.edu](mailto:manoli@mit.edu)

## Abstract

Metabolism plays a central role in evolution, as resource conservation is a selective pressure for fitness and survival. Resource-driven adaptations offer a good model to study evolutionary innovation more broadly. It remains unknown how resource-driven optimization of genome function integrates chromatin architecture with transcriptional phase transitions. Here we show that tuning of genome architecture and heterotypic transcriptional condensates mediate resilience to nutrient limitation. Network genomic integration of phenotypic, structural, and functional relationships reveals that fat tissue promotes organismal adaptations through metabolic acceleration chromatin domains and heterotypic PGC1A condensates. We find evolutionary adaptations in several dimensions; low conservation of amino acid residues within protein disorder regions, nonrandom chromatin location of metabolic acceleration domains, condensate-chromatin stability through cis-regulatory anchoring and encoding of genome plasticity in radial chromatin organization. We show that environmental tuning of these adaptations leads to fasting endurance, through efficient nuclear compartmentalization of lipid metabolic regions, and, locally, human-specific burst kinetics of lipid cycling genes. This process reduces oxidative stress, and fatty-acid mediated cellular acidification, enabling endurance of condensate chromatin conformations. Comparative genomics of genetic and diet perturbations reveal mammalian convergence of phenotype and structural relationships, along with loss of transcriptional control by diet-induced obesity. Further, we find that radial transcriptional organization is encoded in functional divergence of metabolic disease variant-hubs, heterotypic condensate composition, and protein residues sensing metabolic variation. During fuel restriction, these features license the formation of large heterotypic condensates that buffer proton excess, and shift viscoelasticity for condensate endurance. This mechanism maintains physiological pH, reduces pH-resilient inflammatory gene programs, and enables genome plasticity through transcriptionally driven cell-specific chromatin contacts. In vivo manipulation of this circuit promotes fasting-like adaptations with heterotypic nuclear compartments, metabolic and cell-specific homeostasis. In sum, we uncover here a general principle by which transcription uses environmental fluctuations for genome function, and demonstrate how resource conservation optimizes transcriptional self-organization through robust feedback integrators, highlighting obesity as an inhibitor of genome plasticity relevant for many diseases.

## Introduction

Dynamic 3D genome organization plays important roles in guiding gene regulation, and cellular response to environmental stimuli. It also confers robustness to perturbations, and stochastic control to programmed differences between cells and between genotypes (1). Biophysical phase separation, on the other hand, enables both rapid formation of chromatin domains and transcriptional assemblies by phase transitions (2–5), in which transcriptional co-activators mediate multi-enhancer condensate-chromatin assemblies (3, 6). This multi-molecular cooperation has helped define hub-models of transcription, where genome activity drives cell-specific usage of chromatin architecture (6). The flexibility of genome organization is guided by biophysical principles including polymer-polymer interactions, diffusional motion, phase separation, and architectural constraints (5). Indeed, composition of transcriptional regulators along with architectural principles grant cell-specific function in response to environmental fluctuations (3). The ubiquitous conservation of genome structural features and the stochasticity of gene expression are hallmarks of self-organized systems. In this regard, cellular states can provide temporal stability of the system, while state transitions, influenced by environment oscillation, offer tunable feedback integration (5). Therefore, composition and regulation of heterotypic nuclear assemblies are important for feedback tuning of self-organization (4). This is seen during cellular adaptations,

where compartmentalization of proteins by phase separation reduces noise from biological stochasticity (7). This mechanism, for example, ensures cell survival during phenotype transitions by reinforcing self-organization tuning-strategies, on top of an evolved and flexible framework (5). Then, transcriptional organization allows the cells to transition from different states without losing their functional identity (1) by coupling 3D dynamic transcriptional processes with environmental signals and with modulatory features of the chromatin-hardware (2–5).

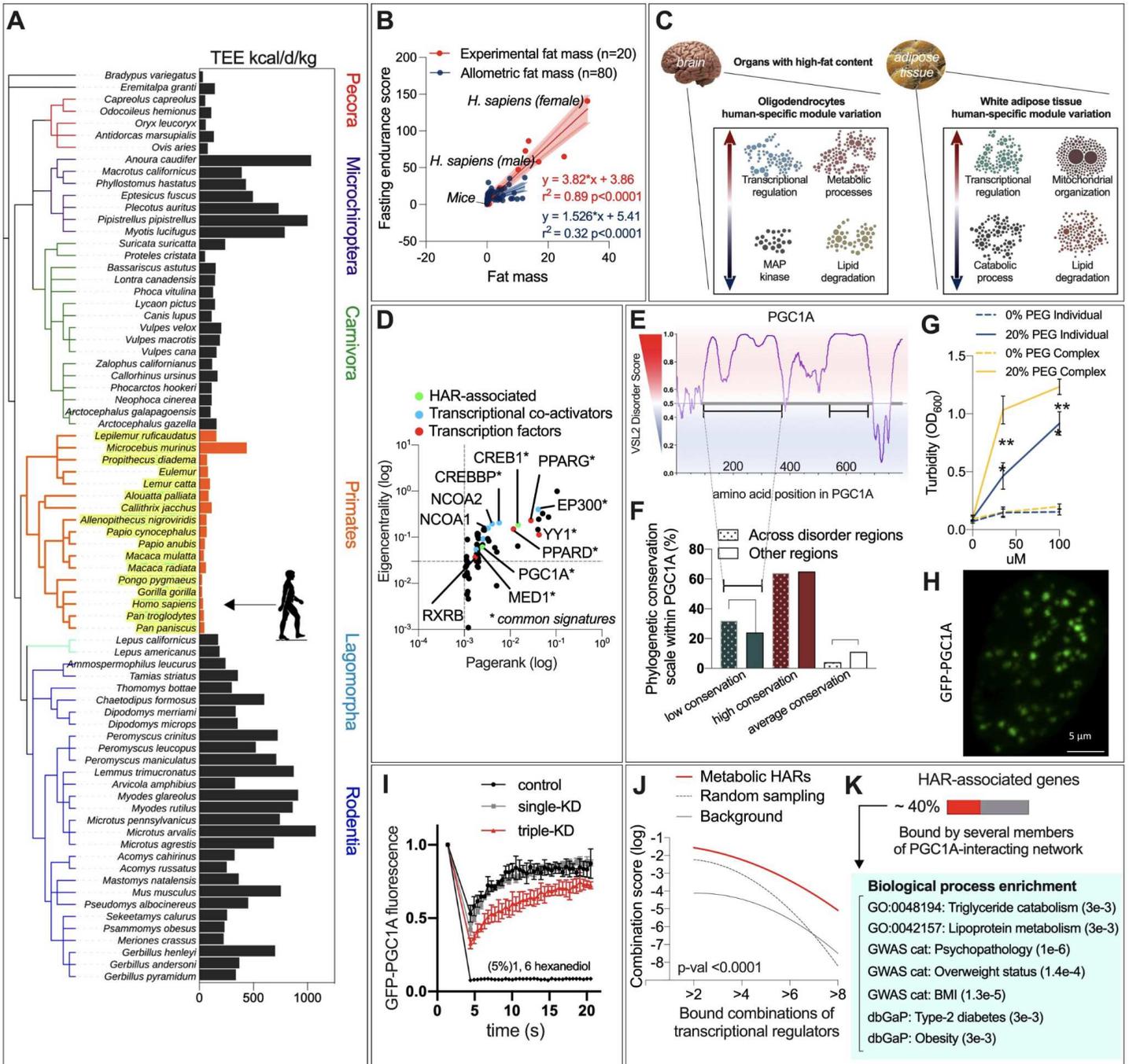
Formation of self-organized assemblies synchronizes with metabolic fluctuations to better use thermodynamic states favouring phase transitions (8, 9). Coupling the degree of self-organization with favourable states such as cold and fasting has implications on their tuning, and more broadly on cellular and evolutionary adaptations (8, 9). Cold and fasting, through reduced cytosolic volume, promote cell crowding and less molecular diffusion, which increases molecular interactions and assembly of condensates (9–13). Since their formation is driven by entropic-forces and less by ATP (9, 14, 15), elucidating their range of parametrization is crucial to understand cellular adaptations (8). Indeed, cellular states condition the material properties of condensates, which impact the collective behavior of interactions and their viscoelasticity (16). For example, during energy depletion, RNA-demixing of P-granules promotes liquid-like to gel-like transitions for fitness (8, 17). A process optimized by evolution through the selection of protein residues favouring multi-molecular assemblies (8, 18). This is in line with resource-driven selection, where preservation of energy during nutritional limitation acts as a selective pressure for evolutionary adaptations (19). This selection is also encoded in genomic-rearrangements or cis-regulatory substitutions (20) (also known as human accelerated regions, HARs), and 3D optimization of chromatin interactions (21). It remains unknown, at small scales of cellular adaptation, how energetic states integrate tuning of transcriptional condensate transitions with genome architecture, and at larger scales, how it drives evolutionary innovation.

The nuclear architecture provides a structural framework that guides the concerted action of multi-dimensional players for genome function (5). Here, we use the genome architecture to supervise the integration of multi-layered information into transcriptional hub-models rendering transcriptional condensate processes. By using the probabilistic and combinatorial nature of genome activity, along with its architecture as topological coordinates, we show that integration of multi-omic information into genomic territorial hubs identifies functional dependencies between evolutionary, phenotypic, structural and transcriptional data. Our integrative topological approach reveals metabolic human acceleration domains with nonrandom higher-order location, transcriptional regulatory networks, and functional chromatin interactions mediating resilience to nutritional stress. This is regulated by heterotypic peroxisome-proliferator-activated receptor- $\gamma$  coactivator-1 $\alpha$  (PGC1A) condensates, and through the incorporation of environmental signals into genome plasticity. This mechanism optimizes transcriptional bursts of lipid-cycling genes, buffers proton accumulation and promotes endurance of cell-specific condensate-chromatin conformations. These findings uncover a general mechanism of environmental control of genome adaptations with far-reaching implications.

## Phenotypic and molecular adaptations reveal metabolic human acceleration domains

To discover human-specific adaptations in nuclear compartmentalization, including phenotypic, genetic, structural and transcriptional adaptations we used an integrative topological approach (fig. S1 and S2). Comparison of metabolic parameters among placenta-derived mammals has shown that primates optimized basal energy expenditure (22) (Fig. 1A and S3A). Great apes, including homo sapiens, displayed the lowest basal calorie consumption per day among mammals (Fig. 1A and Table S1), while humans showed the highest body fat percentage among great apes (23) suggesting a lineage-specific variation (fig. S3B-D). Specific energetic allocation and surplus of energy resources has played a crucial role in human evolution (23, 24). To understand the impact of fat mass on energy budget, we compared its effect on resilience to energy deficit as described before (25). Remarkably, compared to allometric fat mass, experimental fat mass showed species-specific acceleration to energy-deficit endurance (Fig. 1A and Table S1). These results indicate metabolic phenotypic differences among mammals with human lineage-specific acceleration of fat-associated endurance. To assess lineage-specific molecular signatures between humans and primates, we compared transcriptomic and epigenomic data from organs with high fat content such as brain and adipose tissue (26, 27), followed by functional annotation in tissue-specific networks (28). Consistent with phenotypic data, homo sapiens displayed enhanced transcriptional regulatory mechanisms and decreased lipid degradation (Fig. 1C and Table S2). To study the evolutionary and functional impact of non-coding regulatory elements, we interrogated molecular signatures from genes associated with regions with human-specific substitutions (20, 29, 30) (fig. S3E and Table S3). We found overlapping tissue-specific representations of transcriptional regulation and lipid metabolism for genes within 1Mb of HARs (fig. S3F-H). This suggests that phenotypic optimization of metabolism in human lineage has been parallel with molecular evolutionary adaptations in organs with high-fat content. Given the enrichment of transcriptional mechanisms, we investigated whether transcriptional regulators of HAR-genes display physical interactions, share biological processes or are HAR-associated (fig. S4A). ChIP-seq- and transcription factor binding sites (TFBS) enrichment analysis revealed promoters of HAR-genes are bound by physically interacting transcriptional regulators displaying common biological signatures (fig. S4B-D and Table S4). Targeted (1+ interactors) and untargeted (full) network reconstitution and prioritization of bound interactors showed overrepresentation of co-activators and nuclear receptors (Fig 1D, fig. S4D-F and Table S5). Remarkably, only CREB1 and PGC1A displayed HAR association (Fig. 1D), whereas annotation of these cooperative regulators revealed they control adipocyte differentiation and lipid metabolism (fig. S4G).

Combinatorial regulation of transcription includes condensates as “integrators” of heterogeneous protein organizations and genome architecture (3) (**fig. S5A**). Protein phase separation can be studied with the identification of low complexity intrinsically disordered regions (IDRs) (31). Given that PGC1A is a promiscuous interacting co-activator and a master regulator of metabolism (32, 33), we derived computationally, individual and cumulative phase separation scores for PGC1A-network. PGC1A is a highly disordered protein (>70% of aa sequence; **Fig. 1E, fig. S5B and Table S6**), that interacts with highly disordered co-regulators (**fig. S5C-D**). We estimated ongoing evolutionary adaptations within intrinsically disordered regions, by examining the phylogenetic conservation of PGC1A residues (**fig. S6A-B and Table S7**) (34). Remarkably, the proportion of residues with low conservation was higher within IDRs compared to other protein domains (**Fig 1F**). We used tissue-specific networks to assess PGC1A-network and its condensate compositional dependencies to adipose tissue and brain (28). More than 80% of its core members were predicted or have been validated to form phase separation (**fig. S6C**) (35). Further, physically interacting transcriptional regulators displayed more cumulative phase separation scores compared to random interacting permutations (Chi-square <0.001; **fig. S6D**). Turbidity assays of recombinant proteins confirmed phase separation of PGC1A and one of its interacting members RARA, individually and in complex (**Fig 1G and fig. S7A**). PGC1A Fluorescence recovery after photobleaching (FRAP) experiments in living cells activated with forskolin and perturbed with PGC1A-interactors revealed that heterotypic TF-co-activator complexes display increased stability and different dynamics than homogeneous assemblies (**Fig 1H-I and fig. S7A-B**). These results indicate evolutionary pressure within the IDRs of promiscuous co-activators, and that heterogeneous TF-co-activator complexes are more stable compositions. This prompted us to assess combinatorial transcriptional binding of top PGC1A interacting regulators. To this end, we used ChIP-seq data from the ChIP-atlas repository (36). Combinatorial analysis showed that PGC1A interacting regulators are co-binding the promoters (10Mb to TSS) of several HAR gene clusters (**fig. S7C-D and Table S8**), with a large number of co-regulated genes showing co-activator dependencies (**fig. S7D-E**). This indicates physically interacting regulators with common biological signatures display specific binding density distribution to co-regulated genes (**Fig 1J**). This refinement step revealed that PGC1A-network co-regulates around 40% of genes contiguous to HARs, whose biological annotation is related to lipid metabolism (from now on metabolic-HAR domains; **Fig 1K**). Collectively, these results indicate there are evolutionary constraints affecting chromatin functionality as HARs are related to genes with similar function, and that transcriptional condensate hubs such as co-activators display low conservation of residues within IDRs.



**Fig. 1 | Evolutionary entanglement of metabolic phenotype and molecular adaptations.** (A) Total energy expenditure across mammals (22, 23). (B) Comparison of energy deprivation score between allometrically scaled fat mass and experimentally measured fat mass across mammals. (C) Network-derived discovery of biological processes (28) in human enriched gene sets from tissues with high-fat content. Data derived from differential expression by RNAseq and ATACseq comparing human and primate tissues (26, 27). (D) Network-based node prioritization of recomposed interacting transcriptional regulators binding to HAR-associated genes. In green, regulators in proximity to HAR elements. (E) VSL2 disorder score for the transcriptional coactivator PGC1A amino acid sequence. (F) Proportion of amino acid conservation by phylogenetic analysis, within disorder regions and other protein domains in PGC1A. (G) Validation of phase separation by turbidity assays at different concentrations of PGC1A and RARA recombinant proteins with 0 or 20% polyethylene glycol, individually and in complex (n = 3 replicates). (H) Confocal microscopy image for GFP-PGC1A in living cells. (I) Fluorescence recovery and photobleaching analysis (FRAP) of GFP-PGC1A condensates in living cells treated with forskolin for transcriptional induction. Comparison of knockdown (KD) perturbations of PGC1A-interacting regulators, PPARG alone, and in combination PPARG, RXRA RARA, along with 1,6 hexanediol. (J) ChIP-seq integration, combinatorial regulation of HAR-associated genes by PGC1A-interacting network. (K) Biological enrichment of HAR-associated genes bound by a combination of several members of the PGC1A-interacting network. Data show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and

ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value < 0.05.

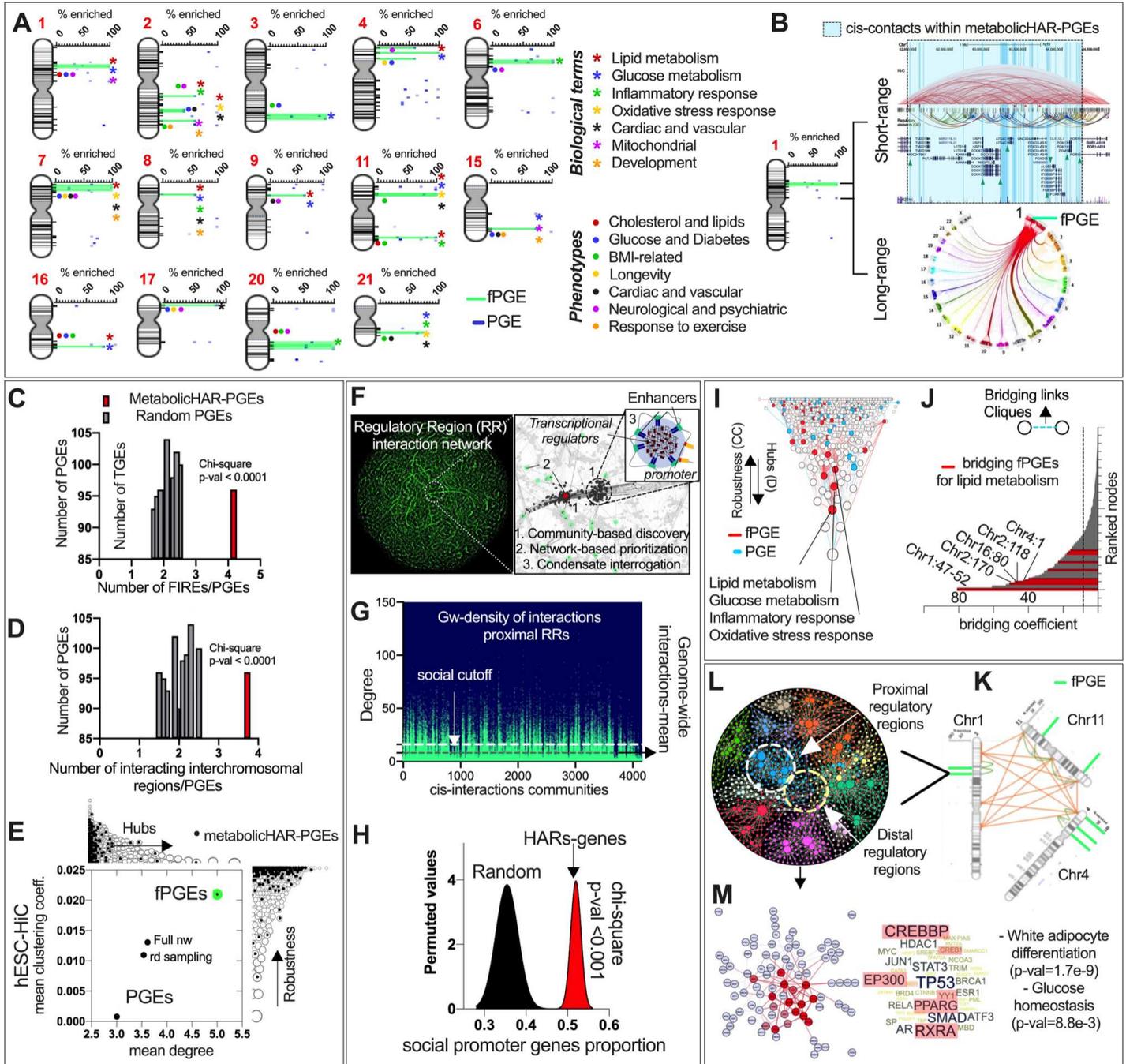
## Metabolic-HARs exploit functional features of spatial genome architecture

Metabolism couples environmental variation to cytosolic signals controlling cellular adaptations (37). These functional adaptations entail the integration of genome architecture and nuclear condensates into transcriptional hubs (5, 38). To dissect evolutionary constraints on functional chromatin regions, we determined genome-wide structural relationships of metabolic HAR domains (**fig. S8A**). We first mapped metabolic HAR genes into topological domains using positional gene enrichments (PGEs) (39). Remarkably, biologically derived gene sets always recapitulated highly enriched genomic clusters as an indication of active transcriptional hubs (**fig. S8B-D and Table S9**). Functional annotation of PGEs revealed chromatin regions of similar function and related phenotype association (fPGEs) (**Fig 2A and Table S9**) To uncover genome-wide structural relationships for functional topological clusters, we interrogated their relational behavior in integrative networks for long- and short-range chromatin interactions (**Fig 2B**). Using a compendia of chromatin contacts (40), we found that metabolic-HAR domains are preferentially located in intrachromosomal and interchromosomal interactive regions as well as in regions with conserved topological associated domains (TADs) (**Fig 2C-D, fig. S8E-F and Table S10**). Network integration of human embryonic stem cells HiC data (41, 42) revealed functional metabolic-HAR domains (fPGEs) are located within hubs and compartment (cliques) forming regions. (**Fig 2E, fig. S8G and Table S11**). The location within hubs and domains improving resilience of the structural-network indicates that, throughout evolution, they have exploited topological features of genome architecture. We next investigated short-range constraints, through network integration of interactions among cis-regulatory regions (43) (**Fig 2F and Table S12-13**). We used module-based hierarchical models of short-range interaction frequencies (**Fig 2F and fig. S8H-I**), to assess the density of interactions, and categorized regulatory regions as social or isolated (**Fig 2G and fig. S8J-K**). Interestingly, metabolic HAR enhancers were preferentially linked to genes with social promoters (**Fig 2H**), while active enhancers displayed more elements with one interaction than genome-wide averages (**fig. S8L**). These results indicate evolutionary constraints on local genome architecture as enhancers with human-specific substitutions are especially linked to social promoters, suggesting fine-tuning of highly regulated hubs.

Chromatin interactions cluster multiple genes with related functions into transcriptional hubs (44). We obtained functional nuclear compartments by assessing interactions between metabolic HAR regions of similar biological function (**Fig. S9A-B and Table S11**). Indeed, regions exhibiting high node degree and high clustering coefficients harbour HAR associated genes with lipid, glucose, inflammation and oxidative stress function (**Fig 2I**). In graph theory and network evolution, modulation of network information flow is better done through nodes bridging densely connected modules (45–47), rather than highly-connected nodes, as hub modulation can be costly for network integrity (47–49). Therefore, we investigated critical bridging HAR regions with similar biological function to hub structures. This showed several regions around chr1:48.5mb housing metabolic HAR genes with lipid and glucose function (**Fig 2J, fig. S9B and Table S11**). In agreement, chromosome interactions around other top functional connectors were pervasive in other HiC datasets (**fig. S9C-D**). To isolate critical regions that exploit information flow, and influence global genome architecture, we recomposed functional nuclear compartments enriched in top bridging nodes, guided by 3D homotypic hESC-HiC reconstruction (**Fig 2K and fig. S10**). Functional ranking showed 2 subnetworks built around top bridging nodes with fPGEs for lipid metabolism (**fig. S11A**). Collectively, these indicate that evolution optimizes metabolic functional hubs by intervening chromatin domains with potential systemic effects.

We found that, among all metabolic-HAR regions, approximately 84% harbour experimentally validated non-coding elements (50) (**fig. S11B**). Within the reconstructed nuclear compartment, every metabolic-HAR region showed active elements locally (**fig. S11C**). We next assessed topological hierarchies and local chromatin states, by network modular analysis of regulatory interactions (**Fig 2L, fig. S12A-B and Table S14-15**). Node prioritization showed that interaction frequencies directly correlate with chromatin states, as isolated enhancers are enriched in repressive epigenetic tags such as H3K9me and H3K27me (**Fig 2M and fig. S12C-D**). Consequently, we used ChIP-seq enrichment of transcriptional regulators bound to both social enhancers and promoters to assess regulators within the topological domains (**fig. S12E-F**). We then filtered binding regulators that physically interact, followed by network functional enrichment (**Fig 2M and fig. S12G**). This revealed that local topological territories can be classified by function as an indication of regional transcriptional activity (**Fig 2M and Table S14**). Subregions with metabolic HARs were bound by interacting regulators controlling lipid and glucose metabolism, with several members of the PGC1A-network (**Fig 2M and Table S14**). Interestingly, every submodule showed local interaction decay with distance constraints for submodular communication (**fig. S12H**). To assess if modular-territories in the regulatory network are representations of structure-based domains such as TADs and chromatin contact domains (CCDs), we investigated global and local CTCF-mediated 3D reconstruction (**fig. S13A-B**) (51, 52). Locally, chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and CTCF-mediated reconstruction from several cell lines (51) revealed a clear overlap between CTCF-cluster graphs and network submodule territories (**fig. S13C**), along with intra and inter-domain (e.g. TAD) 3D effects of structural variants on the metabolic-HAR region (**fig. S13D-E**) The overlap between structural regions and regulatory frequency networks suggests the existence of intrinsic properties of genome architecture.

Functional enrichment of chromatin territories and interacting regions produced complex functional hierarchies (**Fig 2K**). This suggests that structural dependencies might confer genome robustness as novel substitutions can be tolerated while being tested. Likewise, substitutions that lead to DNA misplacement affect evolutionary adaptations and disease (53, 54). We thus assessed territorial functional and phenotype association, including structural variants (55) and genomic-range GWAS (56) (**fig. S14A-D and Table S16**). This revealed interacting hierarchical maps of trait enrichments including lipid, glycemic and BMI in metabolic-HAR regions of selected nuclear compartments (**fig. S14D-F**). To obtain genomic-range trait enrichments with deeper resolution, including tissue-specific enhancer modules with trait enrichments and dependencies, we used data from EpiMap (epigenomic integration across multiple annotation projects) (57). Hierarchical network analysis yielded the polyfactorial penetrance of metabolic and lipid-related traits, and locally with traits in our reconstructed nuclear compartments (**fig. S15A-B and Table S17**). For those traits within the same modules, we built hub models of topologically related traits and tissues (**fig. S15A-C**). This showed tissues with lipid functions are drivers of disease and multifactorial lipid related traits are bridging across other traits with hierarchies following structural functionality (**fig. S15B-D**).



**Fig. 2 | Metabolic-HAR domains exploit functional features of spatial genome architecture. (A)** Metabolic-HAR genes positional gene enrichments (PGEs) and functional domain annotation (fPGEs). **(B)** fPGEs and PGEs domains in chromosome 1 with short- (chr1:62-65mb) and long-range structural relationships. **(C)** Overlapping frequently-interacting

regions (FIREs) within metabolic-HAR regions. **(D)** Overlapping trans-interchromosomal regions from hESC-HiC data within metabolic-HAR regions. **(E)** Mean clustering coefficient (robustness) and node-degree (hubs) across nodes of trans-interacting loci-loci network from hESC-HiC interactome. Labelled, node-loci with colocalized fPGEs, PGEs, full network and random sampling. **(F)** Network-mediated interrogation of *cis*-genome-wide relationships among regulatory regions (43). **(G)** Node-degree of proximal regulatory regions across diverse network communities from F. **(H)** Genes associated with metabolic-HARs display highly social promoters (node-degree >16) compared to random gene sets. **(I)** Clustering coefficient (robustness) and degree (hubs) hierarchies from hESC-HiC network. Genes within top fPGE regions display metabolic function. **(J)** Bridging coefficient in hESC-HiC network. In red, top bridging chromatin regions harbouring metabolic-HAR genes with lipid metabolic function. **(K)** Hub-bridging mediated chromosome reconstruction of nuclear compartments for lipid metabolism displaying selected fPGEs. **(M)** Cis-regulatory network within selected fPGE-modules. **(N)** Module-mediated ChIP-seq binding enrichment within fPGE followed by PPI-network of regulators, and functional biological annotation.

## Metabolic-HAR activity and heterotypic nuclear compartments control fasting endurance

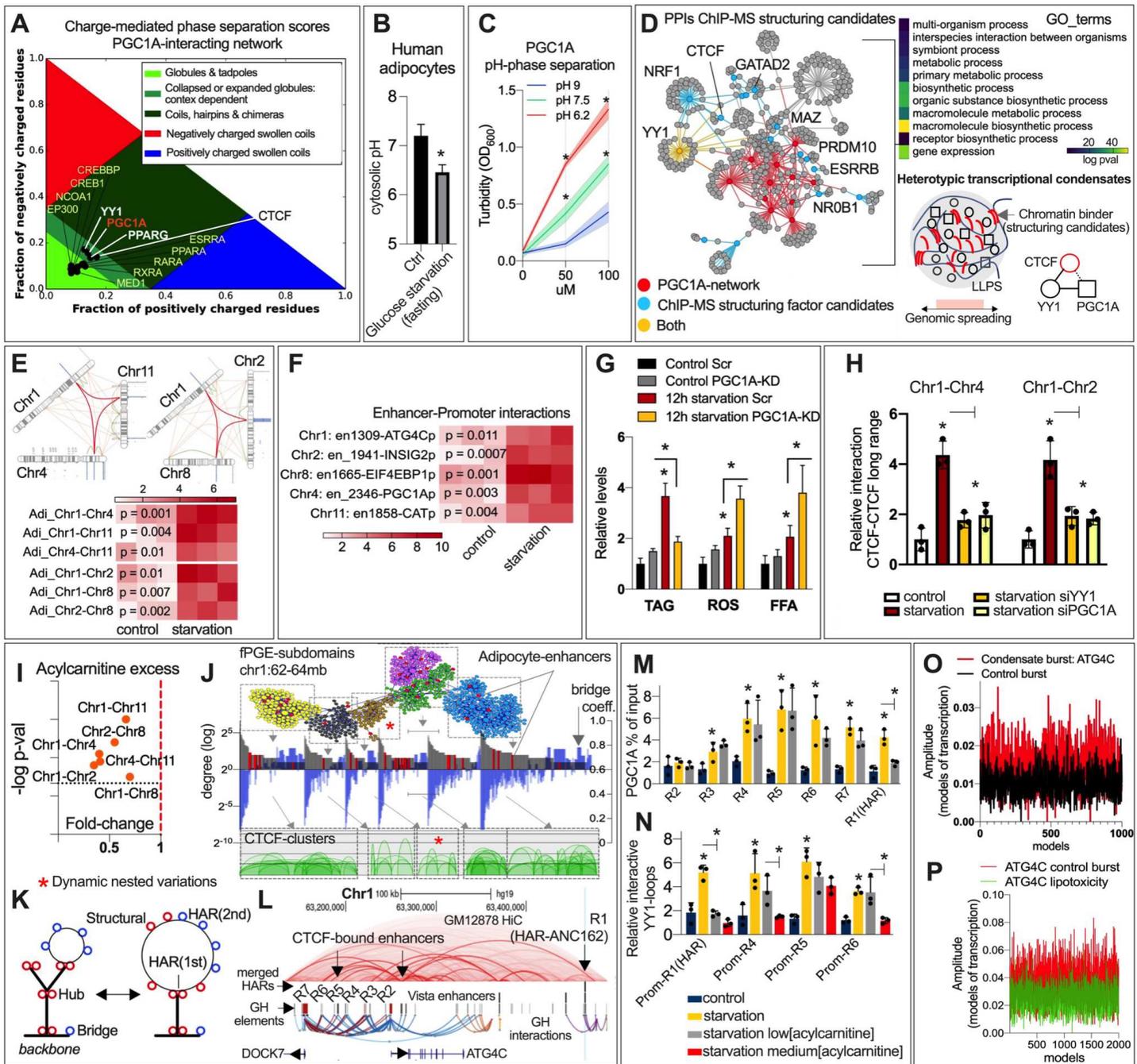
During cellular adaptations, phase transitions entrain the integration of environmentally mediated signals into self-organized conformations (5, 58). Environmental signals such as solvation and salt concentration, for example, affect self-organization through intrachain electrostatic repulsion-attraction (59). We thus used CIDER to dissect context-dependent condensate conformations from amino acid sequences (59). Among members of the PGC1A-network, YY1, PPARG and PGC1A displayed context-dependent globule conformations (**Fig 3A**) with well distributed charged residues and increased proportion of basic residues in PGC1A IDRs (**fig. S16A-B**). Given that PGC1A coactivators and interacting regulators control energy homeostasis (**fig. S5C**) and that species-specific variations of metabolic parameters are related to energy endurance and fat tissue (**Fig 1**), we interrogated nuclear compartmentalization during fasting in human adipocytes. First, we observed that glucose fasting promotes cytosolic pH acidification (**Fig 3B**). This is in line with protonation-mediated phase separation of PGC1A, individually and in complex (**Fig 3C and fig. S16D**), along with increased expression of PGC1A and its interacting members in fasting (**fig. S16E**). Transcriptional condensates control chromatin interactions by limiting stochastic motion and increasing persistent interactions (60). For this reason, we used structuring proteins identified by ChIP-MS bound to both enhancers and promoters (61, 62). In PPI networks, we found that PGC1A-network associates with structuring candidates CTCF, and YY1, both mediating enhancer-promoter loops (62) (**Fig 3D and Table S18**). Given that genomic clusters formed by binders within heterogeneous condensates might be larger than isolated chromatin binders (6, 63–65) (**Fig 3D and fig. S16F**), we tested PGC1A influence on hub models of nuclear compartmentalization. This revealed that fasting promotes CTCF expression and CTCF-mediated interactions around metabolic-HAR of selected nuclear compartments (**Fig 3E and fig. S16G**). In the same conditions, PGC1A bound to nCHARs within chromatin regions of interest (**fig. S16H**), while HAR-associated genes showed increased regional transcriptional activity for top mediators such as ATG4C, PGC1A, INSIG2, EIF4EBP1 and CAT (**fig. S16H-I**). Accordingly, fasting increased YY1-mediated nCHAR enhancer-promoter loops (**Fig 3F**).

ATG4C promotes fast generation of fatty acids through lipophagy (66). DGAT1, on the other hand, mediates fatty acid re-esterification as a protective mechanism to lipotoxicity (67, 68). Interestingly, DGAT1 loci is located at the tail of chromosome 8, a highly interacting region within the nuclear compartment for lipid metabolism (**fig. S16K**). Accordingly, fasting increased DGAT1 expression, PGC1A-binding to DGAT promoter and CTCF-mediated interactions between metabolic-HAR fPGE and DGAT region (**fig. S16K**). Reflecting DGAT1 enzyme activity, functional assays showed elevated esterification of fatty acids (**Fig 3G**). In agreement, knockdown of PGC1A blunted this effect while leading to excessive levels of oxidative stress (**Fig 3G**). To assess if other metabolic-HAR genes within the compartments control lipotoxicity, and whether this response is dependent on mitochondrial CPT1A, we performed functional assays after gene perturbations and after CPT1A-inhibition. Perturbation of genes within the compartments led to similar phenotypes (fatty acid accumulation and oxidative stress), which were exacerbated by CPT1A-inhibition (**fig. S16L-M**). To determine if compartment formation is dependent on heterogeneous condensates enriched in PGC1A, we investigated the time after fasting when compartment formation precedes phenotypic malfunction. We found that, although PGC1A gene perturbation has no effect on functional assays after 6h of fasting (**fig. S16N**), it blunts both long- and short-range chromatin interactions (**Fig 3H and fig. S16O**). Nuclear compartment formation is dependent on liquid-like condensates, as 1,6 hexanediol blocked chromatin interactions (**fig. S16P**). Similar results were obtained by blocking glycolysis, which led to increased ROS, FFA, and more acidic pH (**fig. S17A-B**).

Incomplete oxidation of fatty acids leads to accumulation of acylcarnitines (67), which are associated with obesity and insulin resistance (69). Indeed, acylcarnitine excess inhibited fasting-mediated long- and short-range chromatin interactions (**Fig 3I and fig. S17C**). This effect was dependent on uncontrolled cytosolic acidification as activators of proton pump extruders improved chromatin interactions (**fig. S17D-E**). To investigate the topological penetrance of acylcarnitine and pH variations on heterogeneous looping of regulatory regions, we assessed whether structural hierarchies within TADs or subdomains contribute to genomic and condensate stability. To this end, we focused on the ATG4C-fPGE domain (**Fig 3J**). Besides the convergence between CTCF-cluster graphs and network submodules (**Fig 3J**), our integration revealed some pervasive architectural principles such as interaction frequency specialization of regulatory regions (**Fig 3J**). Decay on interaction probabilities was paralleled by an increment on bridging score for nodes

within submodule boundaries (**Fig 3J and Table S14**). Since this is in line with polymer interaction models where nested-tree structures represent intra TAD variability (70) (**Fig 3K**), we evaluated if regulatory regions bound by structural proteins condition resilience to metabolic stressors. Interestingly, a large proportion of genes controlled by CTCF are promiscuous or social promoters, supporting a role for structuring candidates as anchors of transcriptional hub stability (**fig. S17F**). We first evaluated, during starvation, cell-type specific enhancers (57, 71) (**Fig 3J and Table 19-20**). Adipocyte-specific enhancers within the fPGE displayed higher levels of H3K27Ac epigenetic modification, despite not being bound by PGC1A during fasting (**fig. S17G-H**). This indicates that transcriptional activity is fine-tuned at different levels despite domain stochasticity and cell-type variation. For example, ATG4C structural enhancers bound by PGC1A exhibited different sensitivity to acylcarnitine excess (**Fig 3L-M**). HAR-element linked to ATG4Cp was highly sensitive to low levels of acylcarnitine, which might indicate an exposed 3D conformation during starvation (**Fig 3M**). On the other hand, loops formed on structural enhancers were resilient to elevated levels of acylcarnitine (**Fig 3M**), and, similar to local cell-specific enhancers, were bound by histone acetyltransferase EP300 and CTCF (**fig. S17I-J**). Perturbation of EP300 and CTCF during starvation led to loss of PGC1A-binding to HAR enhancer, and increased sensitivity to fatty acids as structural loops lost their resilience along with target gene expression (**fig. S17K-M**). Similar to previous reports (72), these results indicate structural hubs limit stochastic chromatin motion while following hierarchical principles of organization during genome function.

Since multi-enhancer hubs regulate the kinetics of transcriptional bursts (73, 74), we used a probabilistic approach to model burst profiles from dynamic gene expression. This showed that dynamic fluctuations of gene expression can be captured by nonlinear sine-wave models and Monte Carlo simulations. With this, we discriminated, in similar conditions, transcriptional burst kinetics for genes within the same TAD (**fig. S175N**). Compared to genes with isolated promoters such as ANGPTL3, dynamic fluctuations of HAR-associated genes with social promoters such as ATG4C, display higher amplitude and more frequent burst kinetics (**Fig 3O and fig S175O-Q**). Accordingly, these were susceptible and blunted by metabolic overload through acylcarnitine excess (**Fig. 3P**). Finally, we observed similar dynamic nuclear conformation during fasting in human neurons (**fig. S18A-G**). Transcriptional profiles from human adipocyte precursor cells (APC) (75) and snRNAseq from human cortex (76), showed that high lipid turnover APCs and oligodendrocytes display elevated territorial expression of HAR-fPGEs genes (**fig. S18H-I and Table S21**). Collectively, these results underscore an ongoing evolutionary pressure for burst stability, as nHARs are linked to social promoters with specific burst kinetics. At the molecular level, these topological constraints allow the identification of functional feedback transcriptional hubs controlling metabolic endurance. (**fig S17Q**) (61, 62).



**Fig. 3 | Metabolic-HAR nuclear compartmentalization controls fasting endurance.** (A) Context-dependent phase separation score by CIDER (59) of PGC1A-interacting regulators, and CTCF chromatin binder. (B) Cytosolic pH variation after 12h of glucose starvation in human adipocytes. (C) Turbidity assays of recombinant PGC1A phase separation mediated by pH variations. (D) Transcriptional regulator protein-protein interactions for PGC1A-network and structuring transcriptional candidates selected by ChIP-MS (61), their functional annotation and illustration of cooperative heterotypic regulation of condensates. (E) CTCF-mediated ChIP-loop of fPGE-fPGE interaction within selected nuclear compartments as in B. (F) YY1-mediated enhancer-promoter loops between HAR-elements and starvation-regulated genes within fPGEs as in B. (G) Triglycerides, reactive oxygen species, and free fatty acid functional assays in human adipocytes after starvation and with siRNA for PGC1A. (H) CTCF-mediated long-range chromatin interactions by ChIP-loop after 6 hours of starvation and with targeted siRNA. (I) CTCF-mediated long-range ChIP-loop interactions in human adipocytes after starvation with acylcarnitine (AC) excess (20uM). (J) Network-based model of regulatory region interactions (top panel) and submodules overlap with CTCF-mediated chromatin contact domains (CCDs) clusters (top down panel). Chr1-fPGE submodules with adipocyte-specific enhancers in red and ATG4C submodule with an asterisk. Hub degree and bridge divergent specialization of regulatory regions (middle panel). (K) Tree nested model representation showing dynamic chromatin variations with hierarchies of structural (CTCF-bound) and non-structural regulatory regions. (L) Human genome track showing nCHARs, regulatory elements and interactions, GM12878-Hi-C data, promoter locations, and structural and non-structural enhancers (R1 to R7). (M) ChIP-qPCR binding of PGC1A to regulatory elements in human

adipocytes during starvation  $\pm$  low concentration of acylcarnitines (1 $\mu$ M). **(N)** ChIP-loop enhancer-promoter interactions in the same conditions as in **M** (medium AC = 10 $\mu$ M). **(O)** Amplitude of Monte Carlo models of transcription burst built from dynamic expression data for ATG4C and control gene in the same conditions as in **M**. **(P)** Amplitude of transcriptional burst models as in **O** but with high concentration of acylcarnitine (20 $\mu$ M). Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value < 0.05.

### Metabolic-HAR domains shows species-specific convergence and tuning

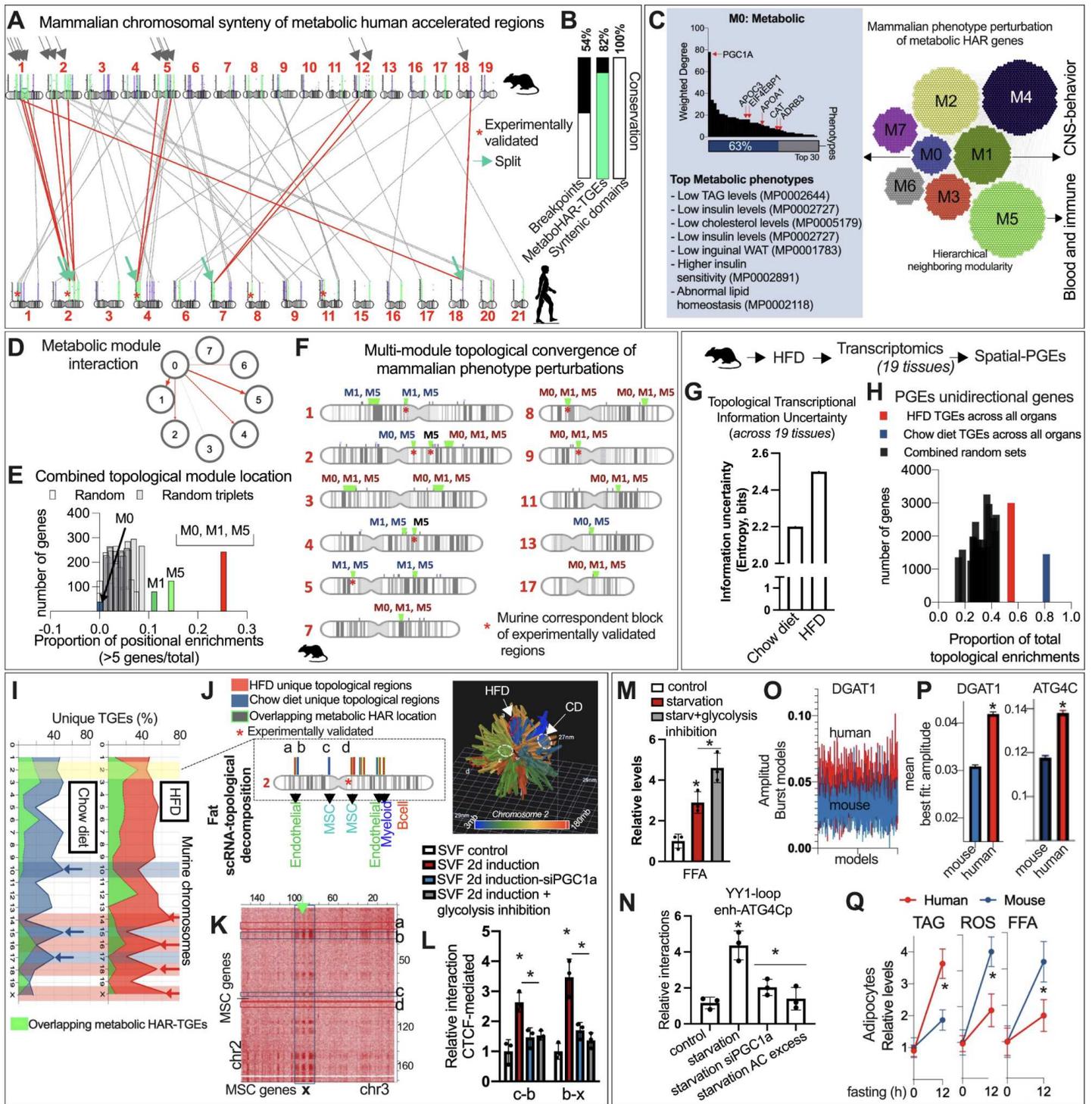
To assess principles of transcriptional hub activity across mammals, we first determined syntenic conservation of metabolic HAR domains in mice. This revealed a high degree of chromosomal synteny (**fig S19A and Table S22**), with some domains being split in murine chromosomes (**Fig. 4A-B**). Interestingly, metabolic HAR regions displayed higher conservation than chromosomal TAD boundaries (breakpoints; around 54% of human boundaries shared with mice) (41), and less than syntenic blocks (**Fig.4B**). Regarding the transcriptional networks mediating fasting endurance in humans, regions in chromosome 2 and 4 were split in murine chromosomes. These results indicate a high degree of preservation of metabolic-HAR blocks but also convergent rearrangements in humans for distant HAR-regions in mice (hChr2, 4, 7 and 18). Since compartmentalized hubs coordinate territorial transcriptional activity (3) (**Fig 3**), we interrogated whether phenotypes from mammalian perturbation data represent territorial activity with congruent phenotypes. Using mice knockout data displaying phenotype and gene relationships followed by network integration (**fig S19B and Table S22**), we found that metabolic HAR genes form hierarchical modules of overlapping phenotypes (**Fig 4B and S19B-C**). Accordingly, metabolic-HAR genes mediating metabolic endurance are located within metabolic, immune and CNS modules (**Fig 4B and fig. S19D**). Unbiased node prioritization confirmed several metabolic phenotypes among the most influential (**fig. S19E-F**). On the other hand, top drivers of metabolic, immune and CNS phenotype clusters were PGC1A, NFE2L2, and EN1 respectively (**Fig 4C and fig. S19E**), whereas module interaction analysis showed their topological dependency (**Fig 4D**). We thus assessed whether this phenotype relationship translates into specific genomic clusters. Remarkably, compared to random combinations or individual enrichments, combined positional enrichment of gene drivers for these related phenotypes showed increased topological coalescence with an increased number of regions harbouring > 5 genes (**Fig 4E and fig. S19G**). These highlight the heterotypic nature of transcriptional hub models where stochasticity, chromatin interactions, and condensate composition translate into territorial genome function.

To understand the influence of diet on localized transcriptional gene-programs, we performed transcriptomic profiles in 19 tissues after diet-induced obesity. Differentially expressed genes per tissue were then interrogated for spatial positional enrichments (**fig. S19H and Table S23**). Comparison of tissue-specific and global PGEs demonstrated a territorial transcriptional coalescence as the proportion of chromatin regions with > 5 genes increased drastically in combined enrichments for HFD and chow diet (**fig. S19I and Table S24**). These analyses also revealed that HFD exhibits higher topological enrichments for regions with 2-genes and isolated gene positions (**fig. S19H, and Table S24**). To compare the HFD impact on transcriptional stochasticity and uncertainty, we derived Shannon entropy using tissue-specific ratios between PGEs harbouring > 5 genes and the total number of regulated genes. HFD promoted a more pronounced transcriptional uncertainty observed in random positional enrichments (**Fig 4G**). By filtering gene-programs specific to HFD or chow diet (**fig. S19J**), and combined positional enrichments, HFD-specific genes were more randomly located, while chow diet genes formed transcriptional hubs (**Fig 4H**). A chromosomal map with unique highly enriched topological regions per phenotype revealed diet constraints on overlapping metabolic-HAR domains and a larger number of unique positions in small chromosomes (**Fig 4I and fig. S19K-L**). We found more divergence of unique regions in large chromosomes (**fig. S19K**), along with experimentally validated metabolic HARs mediating endurance (**fig. S19K**). To evaluate tissue-specific discrepancies in transcriptional hubs, we focused on gene-programs in adipose tissue subtypes. We found overlap of top positional enrichments across adipose tissues and tissue-specificity for less enriched domains (**fig. S20A and Table S25**). Furthermore, HFD constrained adipose transcription as there was a global shift on hub regions, with an increased proportion of inefficient PGEs harbouring only 2 genes (**fig. S20A**).

### Species- and cell-specific tuning of fasting endurance

Despite the heterogeneity of local genome architecture during transcription in neighboring single cells (1, 77), single cell transcriptional maps have revealed a degree of homogeneity. Given the relationship between global genome architecture and genome function, we interrogated if transcriptional regions can be decomposed into cell-specific hubs using scRNAseq data (78) (**fig. S20B and Table S26**). Compared to random permutations, chow-diet hubs overlap with cell-specific positional hubs derived from fat scRNAseq data (**fig. S20C**). In agreement with regional hub models, this approach confirmed cell-specific and shared clusters of topological activity: around 70% of chow-diet domains and 51% of HFD regions intersected cell-specific clusters (random overlap is 20-35%; **fig. S20D and Table S26**). Interestingly, mesenchymal stem cells and endothelial cells displayed the highest topological convergence with HFD-mediated hubs (**fig. S20D**). This result confirmed the pervasive nature of diet-excess on random transcription overload, at the tissue- and cell-specific level (**Fig 4G**). To validate diet loss of cell-specific hubs harbouring syntenic metabolic-HARs, we focused on the topological behavior of chromosome 2 in fat tissue (**Fig 4I**). Indeed, diet-dependent topological regions showed 3D discrepancies despite being neighboring, and cell-enriched (**Fig 4J and fig. S21A**). For example, closely located cell-specific domains for endothelial and mesenchymal stem cells displayed diet-mediated shifts (**Fig 4J and fig. S21A**).

Using HiC data, we found that these neighboring regions form homotypic interactions with distant domains harbouring homogeneous cell-specific hubs and metabolic-HARs (**fig. S21B**). Remarkably, HFD promoted a cell-specific shift in transcriptional activity from heterotypic interacting regions to neighboring isolated domains (**fig. S21C**). For example, interacting interchromosomal regions (chr3-chr2) with cell-specific hubs (MSCs) showed transcriptional activity in chow-diet, which was lost after HFD (**Fig 4K and fig. S21C**). Accordingly, heterotypic and homotypic interactions increased in adipose stromal vascular fraction after induction of differentiation (**Fig 4L**). This effect was blunted by perturbation of PGC1A and inhibition of glycolysis (**Fig 4L**). Underscoring its fatty acid dependency, PGC1A knockdown caused further accumulation of fatty acids, which was exacerbated after glycolysis inhibition (**fig. S21D**). Similar functional results were observed in differentiated murine adipocytes (**Fig 4M**). Interestingly, formation of syntenic nuclear compartments mediating metabolic endurance in human cells after fasting was incomplete in split domains in mice (**fig. S21E**). For regulated loop interactions, condensate hubs and chromatin interactions were abolished by PGC1A knockdown and acylcarnitine excess (**Fig 4N**). To elucidate species-specific burst kinetics, we focused on local top regulatory regions harboring ATG4C, DGAT1, and PGC1A, which showed PGC1A-binding to linked HARs was absent in murine orthologue regions after fasting (**fig S21F**). Burst kinetics showed that frequency and amplitude are elevated in humans compared to murine adipocytes (**Fig 4O-P and fig. S21G-H**). This human-specific tuning translated into more fasting-endurance with increased fatty acid esterification, less oxidative stress and less acidic cytosolic pH (**Fig 4Q and fig. S21I**). Overall, these results imply evolutionary optimization of burst kinetics during nutrient limitation. Indeed, tuning of species-specific regulatory enhancers in the form of anchors seems to increase the stability of self-organized mechanisms that improve the functional control of cellular endurance (**fig. S21J**).



**Fig. 4 | Species-specific convergence and tuning of nuclear compartmentalization.** (A) Chromosomal synteny of human metabolic-HAR domains in mice. In red lines, split chromosomal regions. (B) Rate of mice-human conservation across different chromosome domains. (C) Network integration of mammalian genetic perturbations for metabolic-HAR genes. Hierarchical modularity (right panel) and node degree of gene knockouts within modules (left panel). (D) Metabolic inter-module interactions. (E) Proportion of PGEs (> 5 genes in genomic cluster) for genes in individual modules and metabolic interacting multi-module PGEs. (F) Multi-module topological convergence of metabolic interacting modules in murine chromosome regions. (G) Mammalian HFD perturbation followed by RNAseq in 19 tissues and spatial-PGEs analysis. Shannon entropy for number of PGEs (>5 genes) and total number of genes with genomic-cluster enrichment (PGE) across all organs. (H) Proportion of total PGEs for genes regulated by either HFD or chow-diet. (I) Murine chromosomal proportion of unique PGE hubs by HFD or chow-diet and overlapping syntenic metabolic-HARs. (J) Chromosome 2 location of HFD and chow-diet unique hubs (left panel) with 3D reconstruction and positional location of discrepant hotspots (right panel). scRNA-topological decomposition in fat using tabula muris dataset (78). (K) mESC-HiC chr2-chr3 contact frequencies showing hubs for mesenchymal stem cells from fat (fat msc) scRNA-decomposition. Light

green arrow shows the location of metabolic-HAR regions. (L) CTCF-mediated compartments for fat-MSCs PGEs shown in K from induced stromal vascular cell fraction from subcutaneous adipose tissue. (M) Free Fatty Acid levels in differentiated murine adipocytes. (N) ChIP-loop YY1-mediated enhancer-promoter interactions for ATG4C gene in murine adipocytes. (O) Amplitude variation of transcriptional burst Monte Carlo models from dynamic gene expression data for murine and human DGAT1 expression. (P) Mean amplitude of burst models for DGAT1 and ATG4C. (Q) Murine and human divergent response to starvation by functional assays in adipocytes. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value < 0.05.

### **Metabolic-HAR domains share metabolic disease-variants hubs**

Obesity and type 2 diabetes are multifactorial diseases caused by the cumulative effect of lifestyle and small genetic variants dispersed throughout the genome. To understand structure-function relationships of polygenic metabolic traits, we interrogated the topological hub-contribution of causal genes mediating the effects of genetic variations in non-coding regulatory regions (fig. S22A). To this end, we used our approach (fig. S2B) on data derived from novel causal inference methods that integrate GWAS traits with tissue-specific cis-eQTLs to build trait-, tissue- and gene module-specific maps for metabolic disease polygenic dependencies (79) (fig. S22A-B and Table S27). We focused on commonalities and specificities between BMI and T2D polygenic and mediation maps among diverse tissues (fig. S22B-C). This showed multi-tissue contribution to polygenic scores, with 44% of gene modules coming from tissues with metabolic function (fig. S22C-D). Among those, we found trait- and tissue-specific contributions in visceral fat, skeletal muscle, stomach, cerebellum and cortex (fig. S22D-E). Given our observations on territorial transcriptional activity in diet and genetic perturbations (Fig 4), as well as genomic-range genetics (fig. S14-S15), we determined global and tissue-specific territorial hubs for metabolic disease (fig. S22F) In agreement, we found clusters of topological coalescence as tissue-specific genes further enriched optimal PGEs with > 5 genes (fig. S22G). Interestingly, we observed trait-specific discrepancies as global BMI-hubs from combined multi-tissue genes converge more into high-risk topological hubs (70% for T2D and > 95% for BMI; fig. S22G-H and Table S27). Given the global territorial difference between BMI and T2D polygenicity and their risk-disease causal association, we interrogated local chromosomal divergence of merged-PGEs (fig. S23A). This revealed unique and shared positional hubs as well as colocalization dependencies with metabolic HAR PGEs (Fig 5A and fig. S23A-B). BMI-PGEs and metabolic-HAR regions mediating stress-resilience were enriched in larger chromosomes while T2D-PGEs and shared-PGEs in small chromosomes (Fig 5A and fig. S23B). Chromosome size and trait-specific territorial hubs suggest structure-function principles of radial chromatin organization as BMI-T2D divergence could potentially explain risk factor-to-disease functional topological transitions.

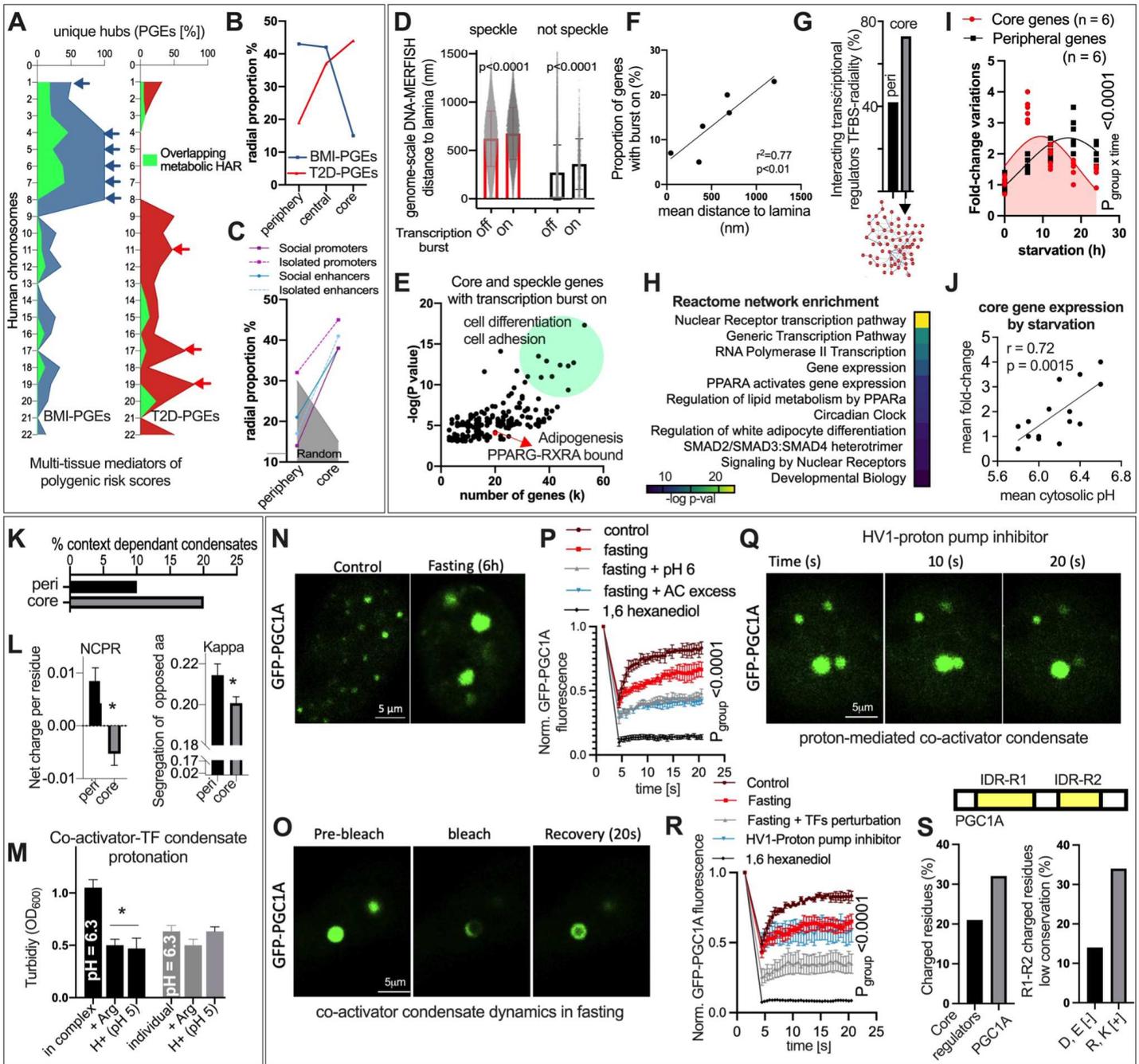
### **Radial genome organization pairs structure and function in metabolic adaptations**

To evaluate the impact of radial nuclear divergence on trait-specificity, we used high-resolution data of chromatin radial positions (80) (fig. S23C and Table S28). Remarkably, BMI-PGEs were preferentially located in the nuclear periphery while T2D-PGEs were enriched in the nuclear core (Fig 5B). Furthermore, HARs displayed both peripheral and nuclear specific regions, while validated metabolic-HARs localized in central to peripheral regions (fig. S23D). Indicating a higher degree of tuning during transcriptional activity, clusters of social regulatory regions were enriched at the nuclear core, with some exceptions for peripheral isolated promoters (Fig 5C). To integrate functional radial divergence during transcriptional activity, we used multiplexed FISH imaging data derived from single cells active loci (77). This revealed direct relationships between nuclear organelle association, distance to nuclear lamina and transcriptionally active loci (Fig 5D-E, fig. S23E-F and Table S28). Interestingly, functional annotation of active nuclear core loci associated with nuclear speckles showed transcription of genes regulating cell differentiation, commitment, and cell adhesion (Fig 5E). Similar to functional enrichments for metabolic-HARs, we observed PPARG-mediated adipogenesis in core active genes (Fig 5E). We next determined functional network enrichment of interacting transcriptional regulators sharing TFBS-radial correlations (80) (fig. S23G). Regulators with core TFBS-associations were highly social proteins as they display a larger proportion of physical interactions (Fig 5G and Table S29) Network reconstitution and functional enrichment revealed that core transcriptional regulators are related to nuclear receptors, PPARs, lipid metabolism, circadian clock, cell commitment and adipocyte differentiation (Fig 5H and Table S29). Notably, interacting peripheral regulators were enriched in anti-inflammatory pathways mediated by several cytokines such as IL21, IL7, IL15, IL4 and IL13 (fig. S23H). Interrogation of radial dynamic expression during metabolic challenge revealed transcriptional peak differences between core and peripheral genes, with acute core activity and prolonged transcriptional behavior for peripheral genes (Fig 5I and fig. S23I-J). Similar to nuclear-cytosolic pH discrepancies (81) and supporting a nuclear core susceptibility to excessive protonation, their dynamic behavior was concomitant to elevated acidification of cytosolic pH and a threshold-dependent correlation of core activity with pH variations (Fig 5J). Collectively, these results indicate that structure-function dependencies translate environmental variations into topological specialized adaptations. This is supported by observations on radial divergence of risk-to-disease polygenicity (peripheral-to-core), location of metabolic-HAR endurers and spatial dynamic bursts influenced by pH-gradient.

### **Transition from liquid-like to gel-like condensates modulates fasting endurance**

We next determined radial condensate composition by assessing context-dependent conformations from regulators with TFBS-radial enrichment. Core-enriched regulators showed a higher number of sequences whose condensate globule

behavior depends on local environment, such as pH variation and heterotypic interactions (**Fig 5K and Table S30**). In agreement with evolutionary adaptations to protonation (8, 16), correlation of IDR-ensemble parameters showed that core regulators display less segregation of charged residues and more negatively charged amino acids (**Fig 5L and fig. S23JK-L**). These enrichments highlight radial adaptations to environmental variations without precluding their activity in both compartments as observed in some metabolic-HAR regions and in DNA-MERFISH data. Among the environmental parameters, we confirmed that protonation regulates heterotypic co-activators and transcription factors condensates, which were perturbed by excessive protonation (**Fig 5M**). These results suggest that nutrient deprivation promotes adaptations at different levels to preserve cellular function (**Fig 3**). For example, fasting-mediated cytosolic acidification prompts heterotypic condensate formation guided by genome architecture to regulate the expression of genes dampening excessive cellular acidification (**Fig 3**). With both aspects showing evolutionary optimization (**Fig 1, 2 and 4**). This, however, does not fully explain if additional nuclear adaptations give mammals their ability to withstand chronic periods of nutrient deprivation. To elucidate this physiological process, we assessed condensate formation and dynamic properties during nutritional stress in living cells. Remarkably, FRAP experiments revealed that 6h of fasting promotes the formation of large and still dynamic PGC1A condensates (**Fig 5N-O**). Formation of condensates with multivalent interactions and different material properties have been previously reported during diverse stress conditions (8, 82, 83). For example, gel-like phase transitions by p-granules promote cellular fitness (8). Strikingly, fasting shifted PGC1A condensates dynamic behavior and recovery (**Fig 5P**), suggesting a fasting transition from liquid-like to gel-like conformations (8, 16). To confirm that these liquid-to-gel conformations are regulated by environmentally tuned electrostatic interactions, we found that their recovery decreased with addition of 1, 6 hexanediol, excessive acidification and acylcarnitine surplus (**Fig 5P**). In addition, we observed that, during fasting, co-activator condensates are much larger in size, and are regulated by pH-mediated ionic strength and salt concentration (**fig. S24A**). Similar to our observations, high-to-low entropic shifts suggest that gel-like formation ensures less diffusion of the transcriptional machinery required for physiological endurance (8, 16). To test if protonation alone can modulate these phase transitions, we blocked the cell-surface proton pump extruder HV1, previously shown to display metabolic functions (84). Similar to fasting, HV1 inhibition provoked cytosolic pH acidification (**fig. S24B**), dynamic gel-like condensate formation (**Fig 5Q**), and activity of selected metabo-HAR genes (**fig. S24C**). Next, we tested the heterotypic nature of these assemblies after perturbation of interacting transcriptional regulators. This revealed that, similar to fasting, HV1 inhibition promoted PGC1A gel-like organizations with longer times of recovery, which further decreased when interacting structuring regulators were perturbed (**Fig 5R and fig. S24D**). Suggesting a threshold-dependent effect of protonation on co-activator assemblies, condensates in prolonged fasting alone (18 hours) were still dynamic, while together with proton pump inhibition blunted their recovery (**fig. S24D**). By inhibiting condensate formation with hydrotropes such as ATP, we further showed that they buffer cytosolic pH acidification during resource conservation (**fig. S24E-F**). This threshold susceptibility indicates that optimization of heterotypic organizations buffering environmental variations might be favoured by evolution. Indeed, our results on PGC1A IDRs low conservation (**Fig 1F**), along with core-specific transcriptional adaptations (**Fig 5J**), prompted us to assess the fraction of charged residues and their degree of conservation (**fig. S24G-I**). Remarkably and contrary to core-enriched regulators (**Fig 5L**), PGC1A-IDRs displayed a higher proportion of positively charged residues (**Fig 5S and fig. S24H-I**), which, compared to negative residues, were 3 times as likely to exhibit low conservation (**Fig 5S**). This suggests evolutionary innovation on charge optimization for hub co-activators helps withstand acidic environments while preserving dynamic cooperative behavior.



**Fig. 5 | Chromatin radial divergence and gel-like condensate transitions control fasting endurance.** (A) Chromosomal proportion of unique disease hubs from multi-tissue mediator genes of polygenic risk for BMI and T2D disease variants. (B) Radial proportion of disease variants hubs for BMI and T2D using GPseq-radial loci data (80). (C) Radial proportion enrichment for network-based categories of regulatory regions interaction density. (D) Distance to lamina of speckle associated genes with transcriptional burst on/off from genome-scale multiplexed FISH (77). (E) Functional annotation of core and speckle associated genes with transcriptional burst on from multiplexed FISH data (77). (F) Correlation between the proportion of genes (not associated with nuclear lamina) with transcriptional burst on and their distance to nuclear lamina from multiplexed DNA-FISH (77). (G) Proportion of protein-protein interactions among transcriptional regulators with TFBS radial enrichment (80). (H) Network reactome enrichment for core interacting transcriptional regulators. (I) Fold-change variation for genes with specific radial enrichment during starvation at different time scales in human adipocytes. (J) Correlation between mean fold-change variation and mean cytosolic pH during starvation in human adipocytes. (K) Proportion of transcriptional regulators (with TFBS radial enrichment) with amino acid sequences showing context dependant phase separation scores using CIDER (59). (L) Net charge per residue (left panel) and segregation of oppositely charged residues (right panel) in amino acid sequences from core and peripheral regulators using CIDER (59). (M) Phase separation by turbidity assay of PGC1A and RARA recombinant proteins in complex, individually, and with arginine and acidic-pH excess. (N) GFP-PGC1A confocal microscopy in living cells. (O) Confocal microscopy of FRAP-experiments in the same conditions as in N. (P) FRAP time of recovery and PGC1A condensate

dynamics in living cells in control and starvation states. **(Q)** GFP-PGC1A confocal microscopy in living cells after HV1-proton pump inhibition. **(R)** FRAP time of recovery and PGC1A condensate dynamics in living cells in control, starvation, target perturbations and HV1-inhibitor. **(S)** Top panel shows PGC1A protein and consensus disorder regions. Proportion of charged residues and degree of conservation for positively and negatively charged residues in PGC1A. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value < 0.05.

## Metabolic resilience by robust genome plasticity

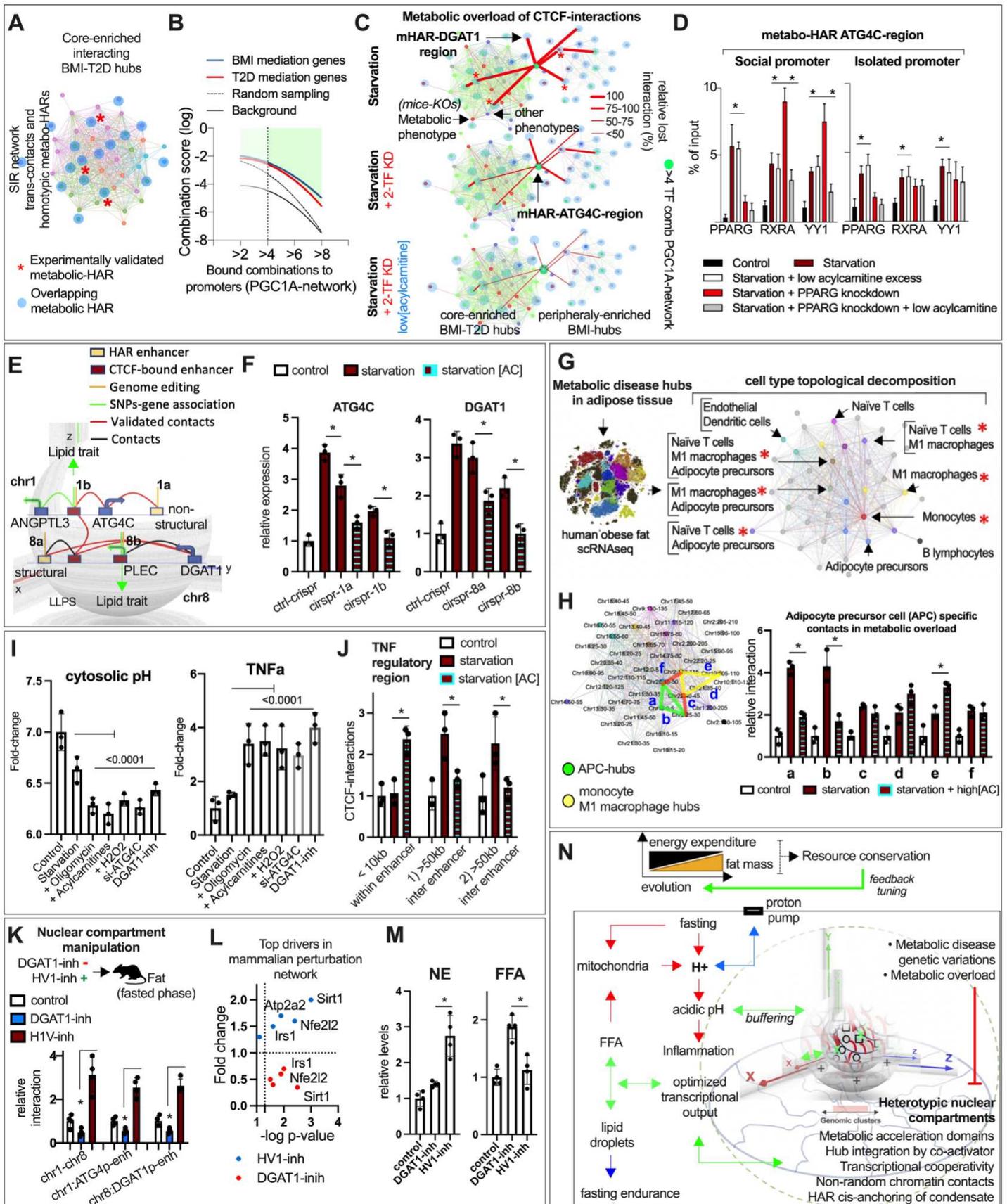
Genome architecture acts as a framework to guide the activity of transcriptional assemblies in response to environmental stimulus (**Fig 3-5**). Association of transcriptional factories with trans-interacting chromatin domains increases cell-specific contacts, ensuring the preservation of cellular function through the regulation of distant gene programs (85) (**Fig 2, 3 and 5**). To address the dependencies between disease hubs and dynamic architectural regions, we interrogated hub-regions of polygenic risk in relation to trans-interacting chromatin regions (**fig. S25A**). First, we defined social interchromosomal regions (SIRs), by using Hi-C data (42, 86) and data from crosslinked-based genome-wide detection of interaction hubs (87) (see structural prioritization of metabolic disease hubs). We complemented this by curating trans-interacting domains across different cell types, and obtaining a likelihood-based score for these social chromatin regions (**fig. S25B and Table S31**). Similar to metabolic-HAR regions, disease hubs from polygenic metabolic traits are preferentially located in SIRs (**fig. S25C**). Reconstruction of SIR locations into a chromatin-chromatin loci network of interacting disease hubs and metabolic-HAR regions (**fig. S25D-E**), revealed hierarchical maps of codependent disease hubs with potential systemic effects (**Fig 6A and fig. S25F**). Interestingly, metabolic associated FTO loci was found to interact with top trans-interacting nodes and top driver regions of metabolic-related phenotypes (**fig. S25D-G and fig. S15D**). BMI and T2D GWAS loci, located distantly, were topologically related between them and with validated metabolic-HAR domains (**fig. S25H**). This underscores the dependencies of polygenic metabolic hubs, their inter-trait relationships and their hierarchies. In addition, we found an overall enrichment for binding combinations of the PGC1A-interacting network to metabolic diseased genes within interacting hubs (**Fig 6B and Table S32**), and with overlapping metabolic-HAR regions (**fig. S26A**). In agreement with our observations (**Fig 2 and 5**), ATG4C-associated domain was found to integrate peripherally-enriched BMI hubs and shared BMI-T2D hubs with nuclear core locations (**fig. S26A**). For PGC1A-regulated genes during fasting, we found that the transcriptional activity within interacting disease hubs is higher compared to active genes within neighboring non-interacting regions (**fig. S26B-C**). Evaluation of disease variant hubs with mammalian phenotype-gene perturbation networks in syntenic domains revealed metabolic clusters, with several gene drivers (**fig. S26D-H and Table S31**). Module interaction analysis, syntenic conservation, and combined positional gene enrichments revealed territorial coalescence for metabolic phenotypes in specific regions (**fig. S26I-K**). Interestingly, this showed species-specific domains as only 15% of human disease hubs display syntenic regions with metabolic phenotypes in mice (**fig. S26L-N**).

To assess the effects of metabolic overload and combinatorial perturbation of regulators on interacting disease hubs, we evaluated long-range contacts after fasting. Remarkably, nuclear compartment formation helped withstand both metabolic overload by acylcarnitine and gene perturbations of cooperative regulators (**Fig 6C and fig. S27A-B**). Targeted validations on ATG4C-domain showed that, in social promoters, single perturbation of PGC1A-interacting transcription factor increased the binding of other cooperative regulators (**Fig 6D**). Notably, this resilient response was lost by added metabolic overload with acylcarnitine (**Fig 6D**). Consistent with our results (**Fig 2-3**), isolated promoters do not display this genome plasticity (**Fig 6D**). Strikingly, genes from mediation analysis of polygenic risk scores and those regulated by PGC1A-network were preferentially social promoters (**fig. S27C-D**). Transcriptional burst profiles with Monte Carlo models showed that genes with social promoters are more susceptible to acylcarnitine excess during fasting (**fig. S27E**). Genes with social promoters and located at the nuclear core exhibited resilience to moderate acylcarnitine and a steep loss of burst amplitude by increased metabolic overload (**fig. S27F**). On the other hand, genes with social promoters located at the periphery displayed resilience to high acylcarnitine levels during fasting (**fig. S27F**). Moreover, perturbation of ATG4C and DGAT1 genes sensitized adipocytes to metabolic overload, impairing chromatin contacts and functional response to starvation (**fig. S27G-K**). Further dissection of this circuit by genome editing of HARs, as well as interacting CTCF-bound enhancers with related GWAS associations (**Fig 6E and figs. S28-S29**), showed that HARs enable the cells to withstand metabolic overload by keeping key structural enhancers in place for gene expression (**Fig 6E and fig. S30A-C**). Mutation of key structural enhancers impaired functional response to starvation, low expression of feedback regulators, and promoted a transition from transcriptional hub regulation to expression of genes with isolated promoters (**fig. S30I-J**). Remarkably and underscoring the functional plasticity of this regulatory switch, these genes are pro-inflammatory, inhibitors of lipolysis, and are causal mediators of lipid accumulation disorders (88, 89) (**fig. S30J-K**). Thus, perturbation of social genome plasticity revealed a switch to transcription within remote areas, seen locally in ANGPTL3, and in non-interacting neighboring segments (chr1:55mb) with PCSK9 expression (**fig. S30K**). This shows nutritional stress integrates structure-function adaptations, and how perturbation of social genome plasticity leads to expression of disease associated genes.

## Metabolic overload inhibits tissue- and cell-specific genome plasticity

To evaluate the cell-specificity of interacting polygenic hubs, we decomposed BMI-T2D fat hotspots with scRNAseq data from obese individuals (90) (**fig. S31A-B and Table S33**). Single cell hubs overlapping with the disease-SIR network revealed cell-specific, heterogeneous and promiscuous transcriptionally active trans-domains (**Fig 6G, fig. S31C-D and Table S33**). Given the overrepresentation of immune inflammatory hotspots, shared with APCs, we tested ectopic transitions during metabolic overload (**Fig 6G and fig. S31D**). Remarkably, during fasting, cell-specific transcriptional activity increased the interactions in contiguous trans-domains, which were lost and switched to immune hubs by metabolic overload (**Fig 6H**). This suggests cell-specific transcriptional regulators condition the contact probability of underlying architecture where ectopic expression of nonspecific genes might lead to tissue pathology (91). Given that ectopic expression of inflammatory genes in stromal or precursor cells affects tissue homeostasis (92), we determined APC function in additional harsh environments. Interestingly, metabolic overload, mitochondrial inhibition, ROS, ATG4C perturbation and DGAT1 inhibition, all led to excessive cytosolic pH acidity and elevated inflammatory gene expression (**Fig 6I**). Interestingly, TNFA, an inflammatory cytokine expressed in adipose tissue and related to insulin resistance (93) is located within the only nuclear core genomic-region of Chr6 (**fig. S12D**) Consistent with environmental regulation of inflammatory gene expression, burst kinetics of TNFA were similar to isolated genes, despite being located on a social regulatory region (**fig. S31E**). This indicates inflammatory genes exhibit transcriptional resilience to harsh environments likely due to homogeneous transcriptional regulation and specific genome plasticity. Remarkably, fasting metabolic overload by acylcarnitine excess showed a shift from long-range interactions to short-range enhancer-promoter entanglement in TNFA social regulatory region, suggesting again a social to nonsocial structural plasticity in metabolic overload (**Fig 6J and fig. S31F**).

To test heterotypic nuclear compartment formation in vivo, specifically in adipose tissue and APCs, mice undergoing fasting were treated intraperitoneally with an acute dose of HV1 proton pump and DGAT1 inhibitors. Fasting promoted metabolic-HAR ATG4C-DGAT1 domain interactions, and short-range contacts (**fig. S31G**). These were blunted by co-treatment with proton pump and DGAT1 inhibition, suggesting a gradient protonation threshold (**fig. S31G**). Chronic manipulation of compartment formation with 7 days daily-dose of either inhibitor revealed proton pump inhibitor increases selected metabolic-HAR compartments in adipose tissue (**Fig 6K**). Given that phenotype-gene perturbation networks of syntenic diseased hubs revealed top mediators genes including SIRT1, NFE2L2, and IRS1 (**fig. S26F**), we found that, in agreement with their function on metabolic homeostasis (94–96), HV1 inhibition increased their expression in adipose tissue (**Fig 6L**). We also observed opposing results with chronic HV1 inhibitor improving sympathetic tone and metabolic parameters, along with anti-inflammatory, mitochondrial and metabolic-HAR gene expression profiles (**Fig 6L-M and fig. S31H-L**). Chronic DGAT1-inhibition promoted a robust shift to pro-inflammatory and ER-stress gene expression (**fig. S31L**). These results confirm that re-esterification of lipids during metabolic overload is a mammalian response to increase resilience and fitness through protection of nuclear compartment tuning (**Fig 6N**).



**Fig. 6 | Resilience through robust genome plasticity.** (A) Cell-naïve chromatin loci-loci network of trans-domains interactions with colocated metabolic disease hubs, and related metabolic-HAR regions (in blue). With an asterisk, experimentally validated HAR-domains. (B) Combinatorial regulation of mediator genes of polygenic risk for BMI and T2D by PGC1A-interacting regulators. (C) Human adipocytes CTCF-mediated long-range contacts in starvation with siRNA

knockdowns (PPARG and RXRA) and low acylcarnitine overload (1uM). **(D)** ChIP-qPCR of social and isolated promoters within ATG4C-PGE region. Human adipocytes after starvation, with siRNA-knockdowns and metabolic overload with acylcarnitine (1uM). **(E)** Network reconstruction of structural dependencies between metabolic-HARs ATG4C and DGAT1 regions, location of related metabolic variants with associated genes, regulatory elements and coordinates (1a, 1b, 8a and 8b) for genome editing with CRISPR/Cas9 system. **(F)** Gene expression in human adipocytes displaying targeted mutations, in starvation and with acylcarnitine (10uM). **(G)** Single-cell topological decomposition of interacting (through trans-domains) metabolic disease hubs with scRNAseq from obese human subjects (90). With an asterisk, inflammatory cell hotspots. **(H)** Left panel, selection of cell-specific hubs subnetwork for adipose precursor cells (APCs in green) and topological neighboring subnetwork for inflammatory cells (in yellow). CTCF-mediated long-range interaction validation in human adipocytes in starvation and with acylcarnitine excess (20uM). **(I)** Cytosolic pH measurement in human adipocytes after starvation and starvation with oligomycin, acylcarnitine, H<sub>2</sub>O<sub>2</sub>, DGAT1-inhibitor and ATG4C-knockdown. Left panel, TNF $\alpha$  gene expression in human adipocytes in the same conditions. **(J)** CTCF-mediated interactions in TNF regulatory region. **(K)** Top panel shows illustration for chronic in-vivo intervention in mice by intraperitoneal injections with either DGAT1-inhibitor or HV1-inhibitor (2mg/kg). In the lower panel, CTCF-mediated (HAR regions ATG4C and DGAT1, and local enhancer-DGAT1\_promoter) and YY1-mediated (enhancer-ATG4C\_promoter) interactions in adipose tissue. Gene expression profile **(L)**, norepinephrine and free-fatty acids **(M)** in adipose tissue in the same conditions as in K. **(N)** Summary illustration. Genome plasticity encodes metabolic resilience. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value < 0.05.

## Discussion

In conclusion, we provide evidence of evolutionary coupling between phenotypic and molecular adaptations. Human acceleration of energy storage in fat tissue demonstrates the impact resource conservation has had on organismal adaptations seen in foraging behavior, cognition, physical endurance and reproduction (97). Coupling metabolic endurance with molecular optimization across transcriptional condensates, 3D architecture and cis-regulatory regions provides a framework for understanding evolutionary innovation, for example, in tuning nuclear self-organization. By studying these transcriptional dimensions, in relation to domains harbouring human specific substitutions, we show that environmental variation integrates components of metabolism into genome function. In addition, we show this has direct effects on genome plasticity and robustness, which could have many implications on a wide range of disease processes. We also demonstrate that fasting can recapitulate at small scales molecular adaptations taking shape throughout evolution. The pervasive nature of genome architecture coupled to cell-specific transcriptional machinery offer a way to manipulate specific subnetworks with systemic effects. It also provided us with a structural framework for multi-modal data integration. By using the probabilistic and combinatorial behavior of genome function, including protein-protein, protein-chromatin, chromatin-chromatin interactions, we use graph networks and genome proximity for integration of multi-genomic data. These nested structural representations allow us to model and funnel computational insights into experimentally testable hypotheses. With these, we demonstrate how genome function follows intrinsic architectural principles and drives cell-specific usage of the interacting architecture. For example, by integrating metabolic-driven pH variations into large heterotypic assemblies, cells optimize the transcriptional output by increasing genomic spreading through cooperative behavior with structural binders CTCF and YY1. This optimization is coupled with the usage of trans-interacting chromosomal regions harbouring functional related gene programs. One of these programs shown here, lipid-cycling genes, also harbouring human specific substitutions, is relevant for resilience to nutrient limitation. Indeed, fasting-mediated fatty acid mobilization leads to fatty acid oxidation, which cannot be sustained unhindered. This promotes the accumulation of fatty acids, inhibition of mitochondrial activity, and further decrease of cytosolic pH. Integrating proton excess to chromatin-condensate stability for the regulation of gene programs controlling cytosolic acidification provides evidence of a robust feedback integrator. The evolutionary tuning of chromatin structure and heterotypic condensates for optimal bursts of lipid cycling genes promotes metabolic and oxidative stress endurance. It also demonstrates the encoding of resource conservation in genome function, and more broadly evolutionary innovation in parametrization of self-organization.

Mammalian comparison of syntenic HAR-domains show some degree of convergence in long genomic-ranges, and acceleration of tuning in short-range scales. It also shows that territorial transcriptional activity is preserved in mice and is congruent with genomic-range phenotype associations, similar to metabolic disease variant hubs. Our topological hub integration reveals that high-fat diet promotes loss of transcriptional control by metabolic overload through accumulation of fatty acids, and excess of acidification. Moreover, we provide evidence that radial chromatin organization pairs structure and function. Resource-driven adaptation to protonation is observed in the enrichment of negatively charged amino acids in nuclear core transcriptional regulators. Despite this radial plasticity, the co-activator PGC1A, acting as a condensate-integrator in acidic milieu during energy conservation, shows more evolutionary innovation for positively charged residues. This suggests tuning of condensate endurance as excessive protonation impairs its dynamic properties. These features license the formation of large heterotypic organizations with gel-like properties that buffer proton excess, withstand less molecular diffusion and remain reversible self-organizations. We show that their reversibility and dynamic properties depend on the presence of heterogeneous interactions as perturbation of interactors reduces their transition ability. Moreover, HAR substitutions have been shown to be enriched in AT-to-GC transitions near telomere

regions, which is in line with observations on nuclear core enrichments (high in GC-content). Interestingly, specialized metabolic-HAR regions regulating fasting endurance are located towards central-peripheral nuclear regions. Their radial 3D location implies their need to withstand acidic excess during activity to promote a resilient phenotype. Below a physiological threshold, acidity not only dissolves heterotypic assemblies and chromatin interactions, but also leaves unaffected pH-resilient inflammatory gene programs. The latter is relevant as transcriptional assemblies have been found in proximity to interchromosomal regions. We therefore modeled the topological dependencies of metabolic disease hubs from mediator genes of polygenic risk with a cell-naïve trans-interacting chromatin network. This shows the pervasive nature of structural architecture in classifying hierarchical disease hubs with systemic effects and their inter-trait relationships. We further decompose these hotspots into cell-specific transcriptional hubs using single-cell RNA seq data. At this resolution, we show that cell-specific transcriptional activity drives cell-specific usage probability of underlying chromatin interactions. This cooperative plasticity, in both long-range convergence and short-range tuning, was dependent on metabolic overload and cytosolic pH. We further demonstrate, *in vitro* and *in vivo*, that pH reduction within physiological ranges through HV1 pump inhibition can promote a metabolic resilient response with heterotypic condensates, and genome architecture, which promotes organismal, tissue and cell-specific homeostasis. On the other hand, inhibition of fatty-acid esterification, and perturbation of social genome plasticity through HAR-circuit genome editing, block selected nuclear compartments in adipocytes and adipose tissue, promoting inflammation and metabolic malfunction.

In summary, our work uncovered mechanisms of evolutionary innovation in tuning transcriptional self-organization. While these mechanisms are recorded through evolution into transcriptional hubs within chromatin architecture, dynamic mechanisms such as stress-dependent phase-separation allow the integration of these uni-dimensional regions into multi-dimensional platforms consisting of metabolic-HAR domains, their master regulators and local thermodynamic states. The present study therefore highlights the existence of structural and topological mechanisms unraveling hierarchically-organized functions stored within the genome. In particular, it revealed robust feedback integrators of resource-driven adaptations, and general approaches for manipulating genome function, for example, through the use of compartmentalized hub-models for precision medicine and machine learning. Such computational approaches could benefit from a better understanding of the principles governing functional compartmentalization, for example, of stress-dependent transcriptional hubs and how these hubs are used as a feedback to stress. In parallel work (98), we have investigated mathematically and computationally, principles through which cellular self-organization, and more generally compartments, can be manipulated in stress conditions at different entropic states, to optimize resilience to pressure inputs, while acquiring properties of self-supervised organization (98). In all, our work suggests that metabolic input overload offers a therapeutic opportunity for obesity, metabolic disorders and many pathological processes by promoting stress resilience through genome plasticity.

## References and Notes:

1. E. H. Finn, T. Misteli, Molecular basis and biological function of variability in spatial genome organization. *Science*. **365** (2019), doi:10.1126/science.aaw9498.
2. A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, G. H. Karpen, Phase separation drives heterochromatin domain formation. *Nature*. **547**, 241–245 (2017).
3. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A Phase Separation Model for Transcriptional Control. *Cell*. **169**, 13–23 (2017).
4. J. A. Riback, L. Zhu, M. C. Ferrolino, M. Tolbert, D. M. Mitrea, D. W. Sanders, M.-T. Wei, R. W. Kriwacki, C. P. Brangwynne, Composition-dependent thermodynamics of intracellular phase separation. *Nature*. **581**, 209–214 (2020).
5. T. Misteli, The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*. **183**, 28–45 (2020).
6. B. R. Sabari, A. Dall’Agnese, A. Bojja, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. **361** (2018), doi:10.1126/science.aar3958.
7. A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, C. Zechner, Phase separation provides a mechanism to reduce noise in cells. *Science*. **367**, 464–468 (2020).
8. J. A. Riback, C. D. Katanski, J. L. Kear-Scott, E. V. Pilipenko, A. E. Rojek, T. R. Sosnick, D. A. Drummond, Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell*. **168**, 1028–1040.e19 (2017).

9. A. W. Fritsch, A. F. Diaz-Delgadillo, O. Adame-Arana, C. Hoege, M. Mittasch, M. Kreysing, M. Leaver, A. A. Hyman, F. Jülicher, C. A. Weber, Local thermodynamics governs the formation and dissolution of protein condensates in living cells, *BioRxiv*, doi:10.1101/2021.02.11.430794.
10. A. Tantos, P. Friedrich, P. Tompa, Cold stability of intrinsically disordered proteins. *FEBS Lett.* **583**, 465–469 (2009).
11. R. P. Joyner, J. H. Tang, J. Helenius, E. Dultz, C. Brune, L. J. Holt, S. Huet, D. J. Müller, K. Weis, A glucose-starvation response regulates the diffusion of macromolecules. *Elife.* **5** (2016), doi:10.7554/eLife.09376.
12. M. A. Mourão, J. B. Hakim, S. Schnell, Connecting the dots: the effects of macromolecular crowding on cell physiology. *Biophys. J.* **107**, 2761–2766 (2014).
13. L. B. Persson, V. S. Ambati, O. Brandman, Cellular Control of Viscosity Counters Changes in Temperature and Energy Availability. *Cell.* **183**, 1572–1585.e16 (2020).
14. H. Mostafavi, S. Thiyagarajan, B. S. Stratton, E. Karatekin, J. M. Warner, J. E. Rothman, B. O’Shaughnessy, Entropic forces drive self-organization and membrane fusion by SNARE proteins. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5455–5460 (2017).
15. R. Hancock, K. W. Jeon, *New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals* (Academic Press, 2013).
16. L. Jawerth, E. Fischer-Friedrich, S. Saha, J. Wang, T. Franzmann, X. Zhang, J. Sachweh, M. Ruer, M. Ijavi, S. Saha, J. Mahamid, A. A. Hyman, F. Jülicher, Protein condensates as aging Maxwell fluids. *Science.* **370**, 1317–1323 (2020).
17. A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, J. P. Taylor, Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell.* **163**, 123–133 (2015).
18. G. Hultqvist, E. Åberg, C. Camilloni, G. N. Sundell, E. Andersson, J. Dogan, C. N. Chi, M. Vendruscolo, P. Jemth, Emergence and evolution of an interaction between intrinsically disordered proteins. *Elife.* **6** (2017), doi:10.7554/eLife.16059.
19. L. Shenhav, D. Zeevi, Resource conservation manifests in the genetic code. *Science.* **370**, 683–687 (2020).
20. J. A. Capra, G. D. Erwin, G. McKinsey, J. L. R. Rubenstein, K. S. Pollard, Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130025 (2013).
21. I. E. Eres, K. Luo, C. J. Hsiao, L. E. Blake, Y. Gilad, Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet.* **15**, e1008278 (2019).
22. H. Pontzer, D. A. Raichlen, A. D. Gordon, K. K. Schroepfer-Walker, B. Hare, M. C. O’Neill, K. M. Muldoon, H. M. Dunsworth, B. M. Wood, K. Isler, J. Burkart, M. Irwin, R. W. Shumaker, E. V. Lonsdorf, S. R. Ross, Primate energy expenditure and life history. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1433–1437 (2014).
23. H. Pontzer, M. H. Brown, D. A. Raichlen, H. Dunsworth, B. Hare, K. Walker, A. Luke, L. R. Dugas, R. Durazo-Arvizu, D. Schoeller, J. Plange-Rhule, P. Bovet, T. E. Forrester, E. V. Lambert, M. E. Thompson, R. W. Shumaker, S. R. Ross, Metabolic acceleration and the evolution of human brain size and life history. *Nature.* **533**, 390–392 (2016).
24. A. Navarrete, C. P. van Schaik, K. Isler, Energetics and the evolution of human brain size. *Nature.* **480**, 91–93 (2011).
25. S. L. Lindstedt, M. S. Boyce, Seasonality, Fasting Endurance, and Body Size in Mammals. *The American Naturalist.* **125** (1985), pp. 873–878.
26. D. Swain-Lenz, A. Berrio, A. Safi, G. E. Crawford, G. A. Wray, Comparative Analyses of Chromatin Landscape in White Adipose Tissue Suggest Humans May Have Less Beigeing Potential than Other Primates. *Genome Biol. Evol.* **11**, 1997–2008 (2019).
27. S. Berto, I. Mendizabal, N. Usui, K. Toriumi, P. Chatterjee, C. Douglas, C. A. Tamminga, T. M. Preuss, S. V. Yi, G. Konopka, Accelerated evolution of oligodendrocytes in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24334–24342 (2019).
28. C. S. Greene, A. Krishnan, A. K. Wong, E. Ricciotti, R. A. Zelaya, D. S. Himmelstein, R. Zhang, B. M. Hartmann, E. Zaslavsky, S. C. Sealfon, D. I. Chasman, G. A. FitzGerald, K. Dolinski, T. Grosser, O. G. Troyanskaya, Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47**, 569–576 (2015).

29. K. S. Pollard, S. R. Salama, N. Lambert, M.-A. Lambot, S. Coppens, J. S. Pedersen, S. Katzman, B. King, C. Onodera, A. Siepel, A. D. Kern, C. Dehay, H. Igel, M. Ares Jr, P. Vanderhaeghen, D. Haussler, An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. **443**, 167–172 (2006).
30. S. Uebbing, J. Gockley, S. K. Reilly, A. A. Kocher, E. Geller, N. Gandotra, C. Scharfe, J. Cotney, J. P. Noonan, Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2007049118.
31. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science*. **357** (2017), doi:10.1126/science.aaf4382.
32. V. Martínez-Redondo, A. T. Pettersson, J. L. Ruas, The hitchhiker's guide to PGC-1 $\alpha$  isoform structure and biological functions. *Diabetologia*. **58**, 1969–1977 (2015).
33. P. Puigserver, Z. Wu, C. W. Park, R. Graves, M. Wright, B. M. Spiegelman, A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell*. **92**, 829–839 (1998).
34. H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–50 (2016).
35. K. You, Q. Huang, C. Yu, B. Shen, C. Sevilla, M. Shi, H. Hermjakob, Y. Chen, T. Li, PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.* **48**, D354–D359 (2020).
36. S. Oki, T. Ohta, G. Shioi, H. Hatanaka, O. Ogasawara, Y. Okuda, H. Kawaji, R. Nakaki, J. Sese, C. Meno, ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19** (2018), doi:10.15252/embr.201846255.
37. O. Argüello-Miranda, Y. Liu, N. E. Wood, P. Kositangool, A. Doncic, Integration of Multiple Metabolic Signals Determines Cell Fate Prior to Commitment. *Mol. Cell*. **71**, 733–744.e11 (2018).
38. H. A. Shaban, A. Seeber, Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Res.* **48**, 3423–3434 (2020).
39. K. De Preter, R. Barriot, F. Speleman, J. Vandesompele, Y. Moreau, Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. *Nucleic Acids Res.* **36**, e43 (2008).
40. A. D. Schmitt, M. Hu, I. Jung, Z. Xu, Y. Qiu, C. L. Tan, Y. Li, S. Lin, Y. Lin, C. L. Barr, B. Ren, A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
41. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. **485**, 376–380 (2012).
42. S. Kaufmann, C. Fuchs, M. Gonik, E. E. Khrameeva, A. A. Mironov, D. Frishman, Inter-chromosomal contact networks provide insights into Mammalian chromatin organization. *PLoS One*. **10**, e0126125 (2015).
43. S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, D. Cohen, GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. **2017** (2017), doi:10.1093/database/bax028.
44. C. H. Eskiw, N. F. Cope, I. Clay, S. Schoenfelder, T. Nagano, P. Fraser, Transcription factories and nuclear organization of the genome. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 501–506 (2010).
45. D. J. Watts, S. H. Strogatz, Collective dynamics of “small-world” networks. *Nature*. **393**, 440–442 (1998).
46. Z. Zhang, J. Zhang, A big world inside small-world networks. *PLoS One*. **4**, e5686 (2009).
47. A.-L. Barabási, *Network Science* (Cambridge University Press, 2016).
48. G. Lawyer, Understanding the influence of all nodes in a network. *Sci. Rep.* **5**, 8665 (2015).
49. G.-S. Peng, S.-Y. Tan, J. Wu, P. Holme, Trade-offs between robustness and small-world effect in complex networks. *Sci. Rep.* **6**, 37317 (2016).
50. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser--a database of tissue-specific human

- enhancers. *Nucleic Acids Res.* **35**, D88–92 (2007).
51. Z. Tang, O. J. Luo, X. Li, M. Zheng, J. J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Rusczycki, P. Michalski, E. Piecuch, P. Wang, D. Wang, S. Z. Tian, M. Penrad-Mobayed, L. M. Sachs, X. Ruan, C.-L. Wei, E. T. Liu, G. M. Wilczynski, D. Plewczynski, G. Li, Y. Ruan, CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell.* **163**, 1611–1627 (2015).
  52. M. Wlasnowolski, M. Sadowski, T. Czarnota, K. Jodkowska, P. Szalaj, Z. Tang, Y. Ruan, D. Plewczynski, 3D-GNOME 2.0: a three-dimensional genome modeling engine for predicting structural variation-driven alterations of chromatin spatial structure in the human genome. *Nucleic Acids Res.* **48**, W170–W176 (2020).
  53. J. Weischenfeldt, O. Symmons, F. Spitz, J. O. Korb, Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* **14**, 125–138 (2013).
  54. G. H. Perry, F. Yang, T. Marques-Bonet, C. Murphy, T. Fitzgerald, A. S. Lee, C. Hyland, A. C. Stone, M. E. Hurles, C. Tyler-Smith, E. E. Eichler, N. P. Carter, C. Lee, R. Redon, Copy number variation and evolution in humans and chimpanzees. *Genome Res.* **18**, 1698–1710 (2008).
  55. R. L. Collins, H. Brand, K. J. Karczewski, X. Zhao, J. Alföldi, L. C. Francioli, A. V. Khera, C. Lowther, L. D. Gauthier, H. Wang, N. A. Watts, M. Solomonson, A. O'Donnell-Luria, A. Baumann, R. Munshi, M. Walker, C. W. Whelan, Y. Huang, T. Brookings, T. Sharpe, M. R. Stone, E. Valkanas, J. Fu, G. Tiao, K. M. Laricchia, V. Ruano-Rubio, C. Stevens, N. Gupta, C. Cusick, L. Margolin, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, K. D. Taylor, H. J. Lin, S. S. Rich, W. S. Post, Y.-D. I. Chen, J. I. Rotter, C. Nusbaum, A. Philippakis, E. Lander, S. Gabriel, B. M. Neale, S. Kathiresan, M. J. Daly, E. Banks, D. G. MacArthur, M. E. Talkowski, A structural variation reference for medical and population genetics. *Nature.* **581**, 444–451 (2020).
  56. C. J. Willer, Y. Li, G. R. Abecasis, METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* **26**, 2190–2191 (2010).
  57. C. A. Boix, B. T. James, Y. P. Park, W. Meuleman, M. Kellis, Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature.* **590**, 300–307 (2021).
  58. H. Yoo, C. Triandafillou, D. A. Drummond, Cellular sensing by phase separation: Using the process, not just the products. *J. Biol. Chem.* **294**, 7151–7159 (2019).
  59. A. S. Holehouse, J. Ahad, R. K. Das, R. V. Pappu, CIDER: Classification of Intrinsically Disordered Ensemble Regions. *Biophysical Journal.* **108** (2015), p. 228a.
  60. N. Khanna, Y. Zhang, J. S. Lucas, O. K. Dudko, C. Murre, Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions. *Nat. Commun.* **10**, 2771 (2019).
  61. X. Ji, D. B. Dadon, B. J. Abraham, T. I. Lee, R. Jaenisch, J. E. Bradner, R. A. Young, Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3841–3846 (2015).
  62. A. S. Weintraub, C. H. Li, A. V. Zamudio, A. A. Sigova, N. M. Hannett, D. S. Day, B. J. Abraham, M. A. Cohen, B. Nabet, D. L. Buckley, Y. E. Guo, D. Hnisz, R. Jaenisch, J. E. Bradner, N. S. Gray, R. A. Young, YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell.* **171**, 1573–1588.e28 (2017).
  63. B. R. Sabari, A. Dall'Agnese, R. A. Young, Biomolecular Condensates in the Nucleus. *Trends Biochem. Sci.* **45**, 961–977 (2020).
  64. F. Erdel, K. Rippe, Formation of Chromatin Subcompartments by Phase Separation. *Biophys. J.* **114**, 2262–2270 (2018).
  65. M. Conte, L. Fiorillo, S. Bianco, A. M. Chiariello, A. Esposito, M. Nicodemi, Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nat. Commun.* **11**, 3289 (2020).
  66. B. Levine, G. Kroemer, Biological Functions of Autophagy Genes: A Disease Perspective. *Cell.* **176**, 11–42 (2019).
  67. T. B. Nguyen, S. M. Louie, J. R. Daniele, Q. Tran, A. Dillin, R. Zoncu, D. K. Nomura, J. A. Olzmann, DGAT1-Dependent Lipid Droplet Biogenesis Protects Mitochondrial Function during Starvation-Induced Autophagy. *Dev. Cell.* **42**, 9–21.e5 (2017).

68. C. Chitraju, N. Mejhert, J. T. Haas, L. G. Diaz-Ramirez, C. A. Grueter, J. E. Imbriglio, S. Pinto, S. K. Koliwad, T. C. Walther, R. V. Farese Jr, Triglyceride Synthesis by DGAT1 Protects Adipocytes from Lipid-Induced ER Stress during Lipolysis. *Cell Metab.* **26**, 407–418.e3 (2017).
69. C. S. McCoin, T. A. Knotts, S. H. Adams, Acylcarnitines--old actors auditioning for new roles in metabolic physiology. *Nat. Rev. Endocrinol.* **11**, 617–625 (2015).
70. K. Huang, Y. Li, A. R. Shim, R. K. A. Virk, V. Agrawal, A. Eshein, R. J. Nap, L. M. Almassalha, V. Backman, I. Szleifer, Physical and data structure of 3D genome. *Sci Adv.* **6**, eaay4055 (2020).
71. T. Gao, J. Qian, EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.* **48**, D58–D64 (2020).
72. Y. Guo, Q. Xu, D. Canzio, J. Shou, J. Li, D. U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M. Q. Zhang, B. Ren, A. R. Krainer, T. Maniatis, Q. Wu, CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell.* **162**, 900–910 (2015).
73. A. J. M. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, Å. Segerstolpe, C. M. Rivera, B. Ren, R. Sandberg, Genomic encoding of transcriptional burst kinetics. *Nature.* **565**, 251–254 (2019).
74. T. Fukaya, B. Lim, M. Levine, Enhancer Control of Transcriptional Bursting. *Cell.* **166**, 358–368 (2016).
75. A. Raajendiran, G. Ooi, J. Bayliss, P. E. O'Brien, R. B. Schittenhelm, A. K. Clark, R. A. Taylor, M. S. Rodeheffer, P. R. Burton, M. J. Watt, Identification of Metabolically Distinct Adipocyte Progenitor Cells in Human Adipose Tissues. *Cell Rep.* **27**, 1528–1540.e7 (2019).
76. H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang, A. J. Martorell, R. M. Ransohoff, B. P. Hafner, D. A. Bennett, M. Kellis, L.-H. Tsai, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature.* **570**, 332–337 (2019).
77. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell.* **182**, 1641–1659.e26 (2020).
78. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature.* **562**, 367–372 (2018).
79. Y. Park, A. Sarkar, L. He, J. Davilla-Velderrain, P. L. De Jager, M. Kellis, Causal gene inference by multivariate mediation analysis in Alzheimer's disease. *bioRxiv* (2017), p. 219428.
80. G. Girelli, J. Custodio, T. Kallas, F. Agostini, E. Wernersson, B. Spanjaard, A. Mota, S. Kolbeinsdottir, E. Gelali, N. Crosetto, M. Bienko, GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* **38**, 1184–1193 (2020).
81. H. Hou, Y. Zhao, C. Li, M. Wang, X. Xu, Y. Jin, Single-cell pH imaging and detection for pH profiling and label-free rapid identification of cancer-cells. *Sci. Rep.* **7**, 1759 (2017).
82. P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q.-X. Jiang, B. T. Nixon, M. K. Rosen, Phase transitions in the assembly of multivalent signalling proteins. *Nature.* **483**, 336–340 (2012).
83. P. Chien, L. M. Gierasch, Challenges and dreams: physics of weak interactions essential to life. *Mol. Biol. Cell.* **25**, 3474–3477 (2014).
84. H. Pang, J. Li, H. Du, Y. Gao, J. Lv, Y. Liu, S. J. Li, Loss of voltage-gated proton channel Hv1 leads to diet-induced obesity in mice. *BMJ Open Diabetes Res Care.* **8** (2020), doi:10.1136/bmjdr-2019-000951.
85. J. Dekker, T. Misteli, Long-Range Chromatin Interactions. *Cold Spring Harb. Perspect. Biol.* **7**, a019356 (2015).
86. A. Belyaeva, S. Venkatachalapathy, M. Nagarajan, G. V. Shivashankar, C. Uhler, Network analysis identifies chromosome intermingling regions as regulatory hotspots for transcription. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 13714–13719 (2017).
87. S. A. Quinodoz, N. Ollikainen, B. Tabak, A. Palla, J. M. Schmidt, E. Detmar, M. M. Lai, A. A. Shishkin, P. Bhat, Y.

- Takei, V. Trinh, E. Aznauryan, P. Russell, C. Cheng, M. Jovanovic, A. Chow, L. Cai, P. McDonel, M. Garber, M. Guttman, Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*. **174**, 744–757.e24 (2018).
88. M. Abifadel, M. Varret, J.-P. Rabès, D. Allard, K. Ouguerram, M. Devillers, C. Cruaud, S. Benjannet, L. Wickham, D. Erlich, A. Derré, L. Villéger, M. Farnier, I. Beucler, E. Bruckert, J. Chambaz, B. Chanu, J.-M. Lecerf, G. Luc, P. Moulin, J. Weissenbach, A. Prat, M. Krempf, C. Junien, N. G. Seidah, C. Boileau, Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).
89. J. Wang, M. R. Ban, G. Y. Zou, H. Cao, T. Lin, B. A. Kennedy, S. Anand, S. Yusuf, M. W. Huff, R. L. Pollex, R. A. Hegele, Polygenic determinants of severe hypertriglyceridemia. *Hum. Mol. Genet.* **17**, 2894–2899 (2008).
90. J. Vijay, M.-F. Gauthier, R. L. Biswell, D. A. Louiselle, J. J. Johnston, W. A. Cheung, B. Belden, A. Pramatarova, L. Biertho, M. Gibson, M.-M. Simon, H. Djambazian, A. Staffa, G. Bourque, A. Laitinen, J. Nystedt, M.-C. Vohl, J. D. Fraser, T. Pastinen, A. Tchernof, E. Grundberg, Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. *Nat Metab.* **2**, 97–109 (2020).
91. D. Nicetto, G. Donahue, T. Jain, T. Peng, S. Sidoli, L. Sheng, T. Montavon, J. S. Becker, J. M. Grindheim, K. Blahnik, B. A. Garcia, K. Tan, R. Bonasio, T. Jenuwein, K. S. Zaret, H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science*. **363**, 294–297 (2019).
92. T. Krausgruber, N. Fortelny, V. Fife-Gernedl, M. Senekowitsch, L. C. Schuster, A. Lercher, A. Nemc, C. Schmidl, A. F. Rendeiro, A. Bergthaler, C. Bock, Structural cells are key regulators of organ-specific immune responses. *Nature*. **583**, 296–302 (2020).
93. G. S. Hotamisligil, N. S. Shargill, B. M. Spiegelman, Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science*. **259**, 87–91 (1993).
94. H.-C. Chang, L. Guarente, SIRT1 and other sirtuins in metabolism. *Trends Endocrinol. Metab.* **25**, 138–145 (2014).
95. F. He, X. Ru, T. Wen, NRF2, a Transcription Factor for Stress Response and Beyond. *Int. J. Mol. Sci.* **21** (2020), doi:10.3390/ijms21134777.
96. A. C. P. Thirone, C. Huang, A. Klip, Tissue-specific roles of IRS proteins in insulin signaling and glucose transport. *Trends Endocrinol. Metab.* **17**, 72–78 (2006).
97. H. Pontzer, M. H. Brown, D. A. Raichlen, H. Dunsworth, B. Hare, K. Walker, A. Luke, L. R. Dugas, R. Durazo-Arvizu, D. Schoeller, J. Plange-Rhule, P. Bovet, T. E. Forrester, E. V. Lambert, M. E. Thompson, R. W. Shumaker, S. R. Ross, Metabolic acceleration and the evolution of human brain size and life history. *Nature*. **533**, 390–392 (2016).
98. R. Tuyeras, L. Z. Agudelo, S. P. Ram, A. Loon, B. Kutlu, K. Grove, M. Kellis, Cellular intelligence: reaching dynamic specialization through non-equilibrium multi-scale compartmentalization, BioRxiv, doi.org/10.1101/2021.06.25.449951

**Acknowledgments:** We thank all the members of the computational biology group at MIT, the Femenia lab members at the institute for Neuroscience in Alicante, and the Novo Nordisk research unit at Seattle for discussions and help throughout the research activities described here.

**Funding:** LZA, RT and MK are supported by the Novo Nordisk foundation, Novo Nordisk research in Seattle. TF, CL, AM are supported by Ramon y Cajal grants by the Spanish state research agency, and the “Severo Ochoa” programme for Centres of Excellence in R&D (SEV-2017-0723).. SLK is supported by the Academy of Finland (grant no: 342074), Finnish Foundation for Cardiovascular Research, Maud Kuistila Memorial Foundation, Orion Research Foundation, Sigrid Jusélius Foundation, and Yrjö Jahnsson Foundation. BK and KV are supported by Novo Nordisk. NAP and PWF are supported by IMI DIRECT (under grant agreement n°115317) and an IRC award from the Swedish Foundation for Strategic Research and a European Research Council award ERC-2015-CoG 681742\_NASCENT. NAP is supported in part by Henning och Johan Throne-Holsts Foundation, Hans Werthe’n Foundation.

**Author contributions:** Project design and Conceptualization: LZA, RT. Computational Methodology: LZA with feedback from RT, MK, PWF, NAP, YP, TF, NS, BK, KG. Polygenic risk scores and mediation analysis: YP with feedback from LZA, MK. Experimental design and validation: LZA and TF with contributions from CL, AM, RT, SKL, LH, KG, PWF, KG, BK, MK. Supervision of the work: LZA, MK, RT, TF. Funding acquisition: LZA, MK, RT, TF. Writing of original draft: LZA. Writing & editing: LZA, TF, RT, MK. All authors reviewed the manuscript.

**Competing interests:** PWF has received consulting honoraria from Eli Lilly and Novo Nordisk A/S. He has also received research grants from multiple pharmaceutical companies and is a consultant and stock owner in Zoe Global Ltd. He is currently the Scientific Director in Patient Care at the Novo Nordisk Foundation. Other authors declare non competing interests.

**Data and materials availability:** Data is available in the supplementary materials. Accession numbers to any data relating to the paper will be deposited in a public repository

**Supplementary Materials:**

Materials and methods

Supplementary text

Figures S1-S31

Tables S1-S33

Captions for tables

## Supplementary materials for

### Metabolic resilience is encoded in genome plasticity

Leandro Z. Agudelo<sup>1,2, †, \*</sup>, Remy Tuyeras<sup>1,2, †, \*</sup>, Claudia Llinares<sup>4</sup>, Alvaro Morcuende<sup>4</sup>,  
Yongjin Park<sup>1,2</sup>, Na Sun<sup>1,2</sup>, Suvi Linna-Kuosmanen<sup>1,2</sup>, Naeimeh Atabaki-Pasdar<sup>1,6</sup>,  
Li-Lun Ho<sup>1,2</sup>, Kyriakitsa Galani<sup>1,2</sup>, Paul W. Franks<sup>6,7</sup>, Burak Kutlu<sup>5</sup>,  
Kevin Grove<sup>5</sup>, Teresa Femenia<sup>3,4, \*</sup>, and Manolis Kellis<sup>1,2, \*</sup>

<sup>1</sup> Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA, USA

<sup>2</sup> Broad Institute of MIT and Harvard, Cambridge MA, USA

<sup>3</sup> Department of Neuroscience, Karolinska Institutet, Biomedicum, Stockholm, Sweden

<sup>4</sup> Instituto de Neurociencias, Universidad Miguel Hernández-CSIC, Alicante, Spain

<sup>5</sup> Novo Nordisk Research Center Seattle, Washington, USA

<sup>6</sup> Department of Clinical Sciences, Lund University, Lund, Sweden

<sup>7</sup> Department of Nutrition, Harvard School of Public Health, Boston MA, USA

† These authors contributed equally to this work

\* Corresponding authors: [agudelo@mit.edu](mailto:agudelo@mit.edu), [rtuyeras@mit.edu](mailto:rtuyeras@mit.edu), [teresa.femenia@goumh.umh.es](mailto:teresa.femenia@goumh.umh.es), [manoli@mit.edu](mailto:manoli@mit.edu)

#### This pdf file includes

Materials and Methods

- Experimental methods
- Computational methods

Figures S1 to S31

Tables S1 to S33

Captions for Tables S1 to S33

References and notes

## Materials and Methods

### 1. Experimental material and methods

#### Animal experiments

All animal experiments were performed in accordance with internationally accepted guidelines and principles for the use of laboratory animals, and were approved by the Novo Nordisk Research Centre Seattle Institutional Animal Care and Use Committee and the Novo Nordisk Ethical Review Committee, Novo Nordisk Research Center in China, and by the regional ethics committee of Alicante, and Universidad Miguel Hernández-CSIC Spain.

Adult Wild-type male C57BL/6J mice aged 6-12 weeks were obtained from Jackson Laboratories (Stock # 000664). Mice were housed 4-8 per cage with ad libitum access to water and food (Type IV Scanbur Cages; 1800 cm<sup>2</sup> floor area). Animal holding rooms temperature ranged from 20-26 °C and humidity ranged from 30-70%. Lighting was on an automated light/dark cycle with lights on between 6:00 AM and 6:00 PM. Mice were provided with Alpha Dri bedding material to just cover the bottoms of cages cage, one Nestlet square, and two 8 g Bed-r'Nest "pucks" or similar. Environmental enrichment included a plastic hut, a paper hut, one wood gnawing block, and a tunnel suspended from the top of cages. Mice received a hydrogel cup upon arrival until familiarized with the automatic water system which supplied reverse osmosis purified water via the Scantainer rack system. Standard rodent chow diet [PicoLab Rodent Diet 20 Lab Diet #5R53 (24.5% Protein, 13.1% Fat, 62.4% Carbohydrate)] was provided to the mice upon arrival until diet was switched from chow to high fat (described below). All food and water were available ad libitum. Mice were acclimated to housing for one week and then marked with 2D barcode ear tags.

High fat diet experiments: At 4 weeks of age mice were placed on a 60% high fat diet (High-Fat Diet #D12492 (20.0% Protein, 60% Fat, 20% Carbohydrate) during 10 weeks and they reached dietary induced obese status (DIO). After this point, mice were moved from group housing to 2 mice per Type IV Scanbur cage, separated by an opaque cage divider. Bedding and environmental enrichment was as described above. Food and water were provided ad lib regardless of housing condition. Mice were sacrificed and tissues of interest were extracted and kept at -80 °C until RNA extraction protocol. Once RNA per tissue was obtained it was sent for RNA-sequencing at the Covance Genomics Lab, Redmond, WA. cDNA libraries were prepared with Illumina TruSeq stranded mRNA kit and sequenced on an Illumina HiSeq. 2000 using paired-end, 100 nucleotide reads with an average depth of 20 million reads per sample. Transcriptomic profile from 19 tissues, comparing control chow diet and HFD diet interventions were carried out. Differential expressed genes from each tissue were obtained following previously established protocols (98). This was followed by integrative topological genomic protocols described in this paper.

Wild-type mice of 12 weeks received intraperitoneally 1 daily dose during 1 week of either HV1 proton pump inhibitor (2.mg/kg 2GBI, Sigma Aldrich), DGAT1 inhibitor (2mg/kg Sigma Aldrich, CatNo:A922500) or control. On day 7, during their fasting phase mice were sacrificed. This was followed by tissue extraction and further processing for molecular experiments. In addition, wild-type mice of 12 weeks were fasted for 12 hours. In parallel, a subgroup of mice received intraperitoneally 1 single dose of either HV1 proton pump inhibitor (2.mg/kg 2GBI, Sigma Aldrich), DGAT1 inhibitor (2mg/kg Sigma Aldrich, CatNo:A922500) or control. At the end of the fasting period mice were sacrificed. This was followed by tissue extraction and further processing for molecular experiments.

#### Mammalian cell culture

Human embryonic kidney 293 cells (HEK293) were cultured on glass-bottomed MatTek dishes. Cells were grown in Dulbecco's modified Eagle's medium (Gibco, Thermo Fisher Scientific), 10% fetal bovine serum (Thermo Fisher Scientific) and 10U/mL penicillin-streptomycin (Thermo Fisher Scientific) and incubated at 37°C and 5% CO<sub>2</sub> in humidified incubator. Cells were transfected when they reached 70% confluence with plasmid DNA GFP-PGC1 (Addgene, Massachusetts, USA) using Lipofectamine 3000 (Invitrogen, Thermo Fisher Scientific) and removed 24h post-transfection, according to standard manufacturer protocols. Transfected cells were imaged 24-48 h post-transfection. Human primary pre-adipocytes (ATCC) were differentiated into mature adipocytes and used for fasting and chemical treatment experiments. For mice primary adipocytes, SVF from inguinal fat depots of 6-12-week-old male mice were isolated and prepared as previously described (99). The differentiation cocktail was used during the first 2 days of culture, followed by rosiglitazone and Insulin for 6 to 8 days.

#### Cell and chemical treatments

Indicated cell cultures were kept in different condition media: Dulbecco's without glucose for fasting condition or Dulbecco's high glucose medium for control. Adipocytes under differentiation were transfected using lipofectamine 3000 (Invitrogen, Thermo fisher technologies) with siRNA for PPAR $\alpha$  (Applied biosystems), with siRNA for YY1 (Applied biosystems), with siRNA for PPARG (Applied biosystems), with siRNA for PGC1A (Applied biosystems), siRNA for RARA (Applied biosystems), siRNA for RXRA (Applied biosystems), with siRNA for CTCF (Applied biosystems), with siRNA for

EP300 (Applied biosystems), with siRNA for ATG4C (Applied biosystems), with siRNA for DGAT1 (Applied biosystems), with siRNA for INSIG2 (Applied biosystems), with siRNA for CAT (Applied biosystems), and scramble control. Once they were differentiated, they were processed for analysis of gene expression, or follow up experiments (as indicated). HEK293 cells were transfected as previously indicated with siRNAs or scramble control 24h previous to downstream experiments. Cells were treated when indicated with 2-Deoxy-Glucose at 0.25mM (Sigma Aldrich) to inhibit glycolysis. 2-guanidinobenzimidazole at 200uM (2GBI, Sigma Aldrich), was used to inhibit the proton pump extruder HV1 as indicated. DGAT1 specific inhibitors were used as indicated (Sigma Aldrich, CatNo:A922500). C:16-acylcarnitine was used as representative long-chain acylcarnitines for cellular experiments at different concentrations as indicated (Sigma Aldrich 61251 - Palmitoyl-L-carnitine). Etomoxir (Sigma Aldrich, 50uM) was used to block CPT1A-B mitochondrial fatty acid import channel as indicated. 1,6 Hexanediol at 2.5mM was used to dissolve condensates (240117, sigma aldrich)

### **Analysis of Gene Expression**

Total RNA was isolated from cells or tissues using Isol-RNA Lysis Reagent (5 PRIME) according to manufacturer's instructions. Amplification Grade DNase I (Life Technologies) was used to treat 1 µg of RNA, from which, 500 ng were used for cDNA preparation following Applied Biosystem Reverse Transcription Kit (Life Technologies). Quantitative Real-Time PCR was performed in a ViiA 7 Real-Time PCR system thermal cycler with SYBR Green PCR Master Mix (both Applied Biosystems). Analysis of gene expression was performed using the  $\Delta\Delta C_t$  method and relative gene expression was normalized to hypoxanthine phosphoribosyltransferase (HPRT) mRNA levels. Gene expression analyses were expressed as mRNA levels relative to controls. Primer sequences available upon reasonable request.

### **ChIP immunoprecipitation**

ChIP experiments were performed as previously described (100) with the following modifications: Cells were plated and (5-10million) pooled before homogenization. Cells were homogenized in PBS and crosslinked with 1% formaldehyde during 10min. For tissue, 50mg were homogenized in PBS and crosslinked with 1% formaldehyde for 10 min. Glycine (125 mM) was added, followed by centrifugation at 4000 rpm for 5 min. After aspirating the supernatant, the pellet was washed with 1 ml cold PBS followed by centrifugation at 4000 rpm for 5 min. This step was repeated with buffer 1 (0.25% Triton X-100, 10mM EDTA, 0.5mM EGTA, 10mM HEPES, pH 6.5) and buffer 2 (200mM NaCl, 1mM EDTA, 0.5mM EGTA, 10mM HEPES, pH 6.5). The pellet was then resuspended in 300 µl lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, 1X protease cocktail inhibitor, pH 8). Samples were sonicated on ice using a Bioruptor (Diagenode) 30 times for 30 seconds ON and OFF intervals. After the lysates were sonicated to shear the DNA to fragment lengths of 300bp, the complexes were immunoprecipitated with antibodies specific for CTCF (Abcam), YY1 (Abcam), PGC1A (Abcam), PPARG (Abcam), RXRA (Abcam), EP300 (Abcam), and H3K27Ac (Abcam). 10-20 ul of antibody was used for each immunoprecipitation. No antibody controls were also included for each ChIP assay and no precipitation was observed by quantitative Real-Time PCR (qPCR) analysis. Input samples were processed in parallel. The antibody/protein complexes were collected by either salmon sperm DNA/protein A agarose slurry or Protein A/G PLUS agarose beads (Santa Cruz sc-2003) and washed several times. The immuno complexes were eluted with 1% SDS and 0.1 M NaHCO<sub>3</sub> and samples were treated with proteinase K for 1 hour and DNA was purified by phenol/chloroform extraction, ethanol precipitation and resuspended in 20 ul H<sub>2</sub>O. Input DNA and immunoprecipitated DNA were analyzed by quantitative PCR. Primer sequences used to detect each DNA region are available upon reasonable request. Primers are available upon reasonable request.

### **ChIP-loop immunoprecipitation of chromatin interactions**

ChIP-loop-qPCR protocol is based on ChIP immunoprecipitation assay previously described with the following modifications: After sonication of the complexes, the lysate was digested HindIII (NEB) following manufacturer's recommendation. Then the digested complexes were immunoprecipitated with specific antibodies (see ChIP-immuno). The immunoprecipitated complexes were resuspended in 200ul of 1x T4 ligase buffer with 400 U of T4 ligase (NEB). Then the samples were collected by protein A/G Plus agarose beads, and DNA samples were purified with phenol/chloroform extraction and ethanol precipitation. Ligation efficiencies were determined using the qPCR assay compared with control templates. Specific primers for computationally selected loops were designed to validate chromatin-chromatin interactions. When comparing different biological samples, highly interactive chromatin regions were used as a control for normalization of sample effect. Positive interactions were matched by ChIP-binding of the regulator. Primers are available upon reasonable request.

### **Kit measurements**

Triglyceride (Abcam ab65336), Free Fatty Acid (ab65341), ROS Detection (ab139476), NE elisa kits (myBiosource), and cellular pH (thermo fisher) quantification from biological samples were done using commercially available kits, following manufacturer's instructions.

### **Protein sample preparation and turbidity assays**

Protein aliquots were thawed above the UCST to ensure the protein was soluble. Phase separation was induced by mixing (stored in 500 mM NaCl and 20 mM Tris, pH 7.5) with no-salt buffer (0 mM NaCl and 20 mM Tris, pH 7.5) to obtain a solution containing 150 mM NaCl and 20 mM Tris, pH 7.5. Protein concentrations were adjusted as necessary by diluting further with the same buffer conditions. Crowding was induced with different concentrations of PEG 8000 (Sigma Aldrich). Proteins were dissolved in potassium phosphate buffer (pH 7) to a concentration of 160  $\mu$ M/800  $\mu$ M as stock and diluted with potassium phosphate buffer (pH 7) in the assay. Then, PGC1A and RARA were mixed with 0, and 20% PEG 8000 solution with a volume ratio of 1:1 at room temperature to obtain various concentrations. Arginine (R) was dissolved in a potassium phosphate buffer (pH 7) to a concentration of 2, 1.5, and 2 M as stock, respectively. PGC1A-RARA combined or individual solutions were mixed with amino acid solution first to obtain various concentrations of R. Then, 20% PEG 8000 was added to the mixture to a volume ratio of 1:1. PGC1A-RARA combined or individual solutions were desalted into a potassium phosphate buffer of pH 5, pH 6.2, pH 7.5, or pH 9. Protein solutions were then concentrated into 320  $\mu$ M/800  $\mu$ M as stock and diluted with corresponding buffers in the assay. Then, protein solutions were mixed with 20% PEG 8000 solution in a volume ratio of 1:1 at room temperature to obtain various concentrations. Turbidity assays were performed on a UV-Vis spectrophotometer (Thermo Scientific, Evolution 350 UV-Vis Spectrophotometer). Samples (in quartz cuvettes with 1 cm path length; Thorlabs) were first equilibrated, and the instrument was blanked. Throughout the experiment, absorbance was measured at  $\lambda = 600$  nm. All samples were examined in triplicate ( $n = 3$ ).

### **Live cell imaging and Fluorescence Recovery After Photobleaching (FRAP)**

FRAP was performed using an inverted confocal super-resolution microscope Zeiss LSM 880-Airyscan Elyra PS.1 with a 63x oil immersion objective. Cells were maintained in the microscope chamber at 37°C and 5% CO<sub>2</sub>. A circular region of interest (ROI) 23  $\mu$ m was selected on GFP-PGC1 nuclear condensates and bleached with Argon laser at 488 nm wavelength by 10 consecutive bleaching iterations (1ms). Fluorescence intensity was recorded at baseline, during and after the bleaching and corrected by ROI corresponding to the background. Different media conditions were used as indicated; Dulbecco's without glucose (6-8h) for fasting condition or Dulbecco's high glucose medium. When indicated cells were transfected with targeted siRNA for knockdown experiments followed by the aforementioned protocol. 1,6 Hexanediol at 2.5mM was used to dissolve condensates (240117, sigma aldrich). ATP at 1mM was used as a hydrotrope to evaluate condensate dissolution when indicated. Image processing and analysis was performed with ImageJ.

### **Extracellular Flux Analysis (Seahorse) Assays**

Cells were differentiated and treated as described above. When indicated, mitochondrial oxidative phosphorylation from human adipocytes was analyzed using extracellular flux analysis (XF24; Seahorse Biosciences) in DMEM buffer (pH 7.4; Sigma Aldrich). Baseline oxygen consumption rates (OCR) were measured every 7 min for control and cells undergoing glucose fasting as previously described.

### **Genome editing of regulatory elements by CRISPR/Cas9**

Circuitry dissection: Our integrative genomic approach revealed chromatin segments with nonrandom higher-order locations. We have characterized their functional role in resource-conservation endurance across many experimental systems. We have focused on nuclear compartments with similar function, with functional interactions both at short- and long-range levels, along with their metabolic disease risk associations. The latter is observed in unbiased genomic range genetics, and in unbiased epigenomic integration of loci-trait-tissue relationships, where both underscored the importance of inter-trait-disease dependencies for metabolic-related traits, especially lipid associated modules. We next tested in polygenic risk scores and downstream causal genes, the topological relations between T2D and BMI variant-hubs, and how these colocalize with metabolic-HAR domains, interacting chromatin regions, cell-specific transcriptional hotspots, driver regions of metabolic disease, especially lipid disease, and within selected functional compartments. This approach showed functional nuclear compartments for lipid cycling genes, fine-tuned by cis-regulatory human accelerated regions, heterotypic condensates and local thermodynamic states, such as resource limitation. The same functional regions display strong topological relationships to T2D-BMI polygenic risk variant hubs, bridging segments of chromatin architecture and regions of inter-trait and inter-disease dependencies by lipid-related modules. Our experimental validation confirmed this, and helped find top functional dependencies during resource conservation, such as chr1-ATG4C region and chr8-DGAT1 region. During fuel restriction, we showed that HAR-enhancers are more sensitive to metabolic overload by acylcarnitines, while key structural enhancers (CTCF-bound) formed promoter loops that are resilient to acylcarnitine excess. Regarding the latter, even though they preserved basal transcriptional activity of target genes in starvation, fatty acid excess decreased their burst profile. For these reasons, we dissected the local circuitry of this heterotypic compartment using genome editing of key regulatory elements (HARs and structural) in human adipocytes, followed by functional, structural and transcriptional assessment of their genome perturbation in starvation states. Our initial goal is to test our previous findings and find causality as to how HARs offer stability to social genome plasticity, and how structural enhancers are key for intra-domain stability during resource conservation. Second, we want to find context-dependent causality for loss of genome plasticity transitions, and how these are relevant for phenotype outcomes of metabolic genetic variants within the circuitry. Thus, we focused on the local interacting circuitry between HAR elements and promoter regions of target genes, followed by isolation of key CTCF-bound enhancers (this is in addition to experimentally validated loops such as YY1-mediated), with genetic variations related to metabolic traits, especially lipid disorders (as previously described in genomic-range genetics, epimap integration and mediators of polygenic risk for

T2D-BMI). ATG4C domain circuitry: following our validations, we focused on the HAR ANC162 that forms interactions with the promoter region. We carried out C-to-T genome editing using CRISPR/Cas9 in the binding motif (see transcriptional pipeline for TFBS-assessment) of a PGC1A-interacting target, YY1. As shown in our results in fig. S28 on circuit dissection, there is a genetic variation rs17124210 in that position, related to waist-hip-ratio (mutation name CRISPR-1a). Next, we focused on our previously validated key enhancer linked to ATG4C promoter, R5 (fig. S28). In starvation, this enhancer forms YY1-loops, and it is bound by CTCF as well. Given our results on heterotypic cooperative behavior of interacting transcriptional regulators, and how CTCF as a chromatin binder participates in this plasticity, we focus on finding CTCF-binding sites with genetic variations or in proximity. Using the DNA sequence of the whole regulatory element span, we scanned it for CTCF-motifs using Ziebarth et al. approach (101). Considering the syntax effect and influence of neighboring motifs, we selected for genome editing the closest CTCF-motif to the strongest genetic variation to lipid disorders, rs12130333. This position is associated with the ANGPTL3 gene, which is located within the ATG4C submodule (Fig 2 and 3). In this motif, we also carried out C-to-T genome editing using CRISPR/Cas9 in the binding motif (mutation name CRISPR-1b). DGAT1 domain circuitry: we have shown in our experimental validations that DGAT1 location in the tail of chromosome 8, lies in a social interchromosomal segment, and belongs to our selected nuclear compartment. We found in starvation, long-range CTCF mediated interactions both homotypic to CAT-domain and heterotypic to metabolic-HAR ATG4C domain. By using the network reconstruction of short-range interactions (Fig 2), we obtained DGAT1 territorial module and submodules. We then focused on the HAR-HACNS71 element as a candidate for DGAT1-circuitry, which, due to distance constraints (1mbp), was validated in Fig 3 as CTCF-mediated interaction. In close proximity to this HAR element, lies an active validated element\_558 (VISTA), forming cis and trans interactions, as well as being bound by ChIP-seq experiments to CTCF and YY1, both members of PGC1A-heterotypic condensates. Therefore, we mapped top CTCF-motifs in this HAR element to perform C-to-T genome editing using CRISPR/Cas9 (fig. S29; mutation name CRISPR-8a). Next, we scanned regulatory elements within the DGAT1 submodule for genetic variations associated with phenotypes congruent with the trans-interacting ATG4C domain. We found in a regulatory cluster, a genetic variation rs55831924 associated with lipid disorders and related to PLEC gene. Unbiased mapping of CTCF-motifs in the regulatory element DNA sequence, revealed that the genetic variant colocalized with a top CTCF-motif. This motif was selected to perform C-to-T genome editing using CRISPR/Cas9 (mutation name CRISPR-8b). Full network reconstruction of circuitry dissection can be found in supplementary figures S28 and S29, and Fig 6E. Genome editing: we performed genome editing in human adipose cells as previously described (102). In brief, all mutations in selected regulatory elements (some with genetic variations; see previous section) were from CC to TT alleles. hCas9 and guide RNA (gRNA) vectors were used (Addgene). C-to-T change was done using site-directed mutagenesis Q5 kit (New England Biolabs). Guide RNAs were designed using the CRISPR design tool available at <http://crispr.mit.edu/>. GFP, Cas9 with sgRNAs, homology vector, and pMACs 4.1 plasmids were co-transfected in adipose progenitor cells using Amaxa-nucleofector (Lonza). Cell sorting was done using MACSelect™ (Miltenyi Biotec), followed by cell culture for 6 days, and propagation of selected clones for downstream experiments. These were functional assays (FFA, ROS, TG, pH, OCR), targeted chromatin interactions (cis and trans within circuitry), expression of target genes (ATG4C and DGAT1) and topological outcomes from loss of social genome plasticity in resource-conservation states. For the last one, we were guided by the interaction frequencies among regulatory elements within the domain (defined in Fig 2 as social or isolated regulatory regions). Network or topologically-based prioritization of interacting modules was used to extract isolated regulatory regions (fig. S30I), followed by functional annotation, and detection of strong genetic variants in contiguous segments such as PCSK9 in Chr1:55mb (rs11591147). We then tested in cells with mutations in key structural enhancers (1b and 8b), if loss of social genomic plasticity in nutrient limitation leads to transition of transcriptional activity to remote areas or isolated regulatory regions within domains of interest. We selected top isolated region genes and quantified their transcriptional activity in the aforementioned conditions. Primers are available upon reasonable request.

### **Intraperitoneal glucose tolerance test (IP-GTT)**

Intraperitoneal GTT was performed on overnight fasted mice. Blood glucose levels were measured at basal state (0 min) and then at 30, 60, and 120 min after i.p. injection of glucose (1.5 mg/kg body weight). Blood glucose concentrations were measured using the Accu-Chek Aviva monitoring system (Roche, Basel, Switzerland).

### **Mitochondrial mass**

Plated cells were pooled and homogenized with PBS and Mitochondria Isolation buffer (225 mM Mannitol, 75 mM Sucrose, 30 mM Tris-HCl pH 7.4, 0.1mM EGTA). For animal experiments, 10 g of tissue was used from each animal to measure mitochondrial mass. Tissues were homogenized with Mitochondria Isolation buffer. Tissues were slowly homogenized with strokes to break them up to 80%–90%. They were spun down at 500 xg for 10' to bring down unbroken cells and cell debris. The supernatant was then transferred to weighted Eppendorf tubes. This was followed by centrifugation at 7000xg for 10' at 4°C to bring down mitochondria. Finally, the Eppendorf was weighted again and the difference between the last weight and the previous one gave us the total mitochondria mass.

### **Mitochondria/Genomic DNA Ratio**

DNA was extracted using Isol-RNA Lysis Reagent (5 PRIME) according to the manufacturer's instructions. After DNA quantification, 3 nmol of DNA were used to quantitative real-time PCR using SYBR-green. We used primers for mitochondrial DNA and primers for hypoxanthine-guanine phosphoribosyltransferase (HPRT) as a gene specifically transcribed in the nucleus. We used the ratio between the mitochondrial DNA and genomic DNA as an indication of the mitochondrial DNA per sample. Primers are available upon reasonable request.

## 2. Computational methods

### Integrative topological genomics (ITG)

#### General description

Genome features including activity, structure and diversity are being measured at an exceptional pace. These multi-layered representations are modeled with single statistical approaches or with ensemble machine learning methods. Despite recent advances, biological systems, data generation, and analysis are imprinted with stochasticity, which coerces the generation of general models. On the other hand, multi-models of stacked single tasks are poised to better capture this diversity, as they funnel computational inferences into guided representations of genome proximities. To elucidate genome structure-function dependencies, and given that nuclear architecture provides a structural framework for genome function, we use here structural representations of genome distances to supervise the integration of multi-layered information (e.g. structural features, protein interactions, binding of regulatory regions, heterotypic condensates, tissue and cell-specific gene expression profiles, genome substitutions). These pipelines are categorized in several classes (evolutionary, transcriptional, structural, comparative genomics, disease hubs, and radial genome), each with its own focus, but complementing each other (**fig. S1 and S2**). Every pipeline has points of transition between them, and points complemented by experimental validation of computational inferences (**fig. S1 and S2**).

#### Integrative topological genomics I (ITG-I): evolutionary relationships

- **Multi-species metabolic phenotypes**

Data for multi-species comparisons of metabolic phenotype were obtained as follows: total energy expenditure (TEE) from, including hominidae fat and lean mass, and (22, 23), experimental fat mass measurement across mammals (24) (table S1).

- **Multi-species endurance score**

To calculate energy endurance score, we followed previous work by Lindstedt et al. (25), where they generated a fasting endurance score (survival days in nutritional deficit) as a function of body size in mammals. We used a similar approach but with allometric (10% of body weight) and experimentally measured fat mass. Briefly, endurance score is based on the relationship between fat mass energy levels (in kilojoules) and the amount of energy consumed per day from TEE data (table S2). It is then derived from resource availability, assuming energy intake is low. This gives a representation of the number of days a given species can last solely on the energy stored in fat tissue. After obtaining an endurance score per species where data was available, we plot a XY graph together with fat mass. Fat mass is derived from both experimentally measured and allometric fat mass. We then performed linear regression to compare both experimental and allometric models.

- **Primate-Human comparison**

We used previous published work to compare primate and human data: from brain cells (27) and adipose tissue (26). We re-analyzed the data following protocols described there (GEO: GSM3494237–GSM3494249), and used differential signatures for downstream analysis.

- **Network-based module discovery**

To assess functional gene-gene dependencies in tissue-specific networks, we performed a targeted approach of overlapping network representation using data from tissue-integrated genome-scale analysis (28) (archived in <http://giant.princeton.edu> and <https://hb.flatironinstitute.org/>). This was followed by functional clustering of dependencies.

- **Integration Human accelerated regions (HAR)**

Human accelerated regions and HAR associated genes were obtained from previous work (29) (table S3). Full HAR-associated genes were completed with subsequent studies (20). Enhancer coordinates and HAR-genes were used for downstream analysis. HAR-genes were used as input for network-based module discovery in specific tissues, followed

by interrogation of overlapping molecular signatures with results from primate-human high-throughput comparisons. Human gain enhancer positions were obtained from Uebbing et al. with TFBS enrichments and target regulated genes (30)

## **ITG-II: transcriptional relationships**

### **• Transcriptional regulators promoter-binding**

To assess transcriptional binding to promoters of HAR-associated genes, we used a combined enrichment approach including motifs and binding of regulators. For a given geneset we used conserved common regulatory motifs enrichment (archived in <http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). Enrichr in R was used to perform functional enrichment analysis based on the database: ChEA 2016 and ENCODE consensus. FDR <0.01 and p-value <0.05 were used as a threshold to select the significant enrichments (28, 103) (archived in <https://maayanlab.cloud/Enrichr/>). Enrichment results from each class were followed by venn-diagram filtering of overlapping regulators. These were used for downstream analysis. Prediction tool for TFBS from DNA sequences can be found at <http://algggen.lsi.upc.es/>. Tool for prediction of CTCF-binding sites can be found at <https://insulatordb.uthsc.edu/> (101).

### **• Functional interacting transcriptional regulators (f-ITR)**

Common regulators from our previous step were used as inputs to identify physical interactions in Protein-Protein interaction (PPI) data stored in several repositories. First, we obtained physical interactions from regulators using human protein interaction atlas (archived in <http://www.interactome-atlas.org/> (104)). This was followed by filtering of regulators displaying common molecular signatures. These were used for network recomposition

### **• Network recomposition and modelling of f-ITRs**

Next, we carried out a 1+ node reconstitution (partial) and full network reconstitution of all possible interactors from our filtered targets. To do so, we used both HuRI and Biogrid PPI-data (archived in <https://thebiogrid.org/> (105)). Partial and fully reconstituted networks were used for downstream analysis. We performed global and local network analysis with clustering and modularity followed by node-prioritization with eigenvector centrality and random-walks algorithms. Nodes with the highest influence on the network were used for functional annotation embedding using DAVID clustering (106) and HAR-association with our curated dataset. Top ranking nodes were filtered for downstream analysis.

### **• Individual phase separation scores**

We next identify computationally intrinsically disordered regions in the amino acid sequences of f-ITRs. Disorder scores from protein amino acid sequences were obtained using predictors of naturally disorder regions software (107, 108), inference of intrinsically unstructured proteins (IUPred) with Anchor (binding regions within IDRs) package (109)(110), multivalent pi-pi interactions in non-aromatic residues of folded proteins (111) and curated experimentally validated phase separation databases (stored at <http://www.pondr.com/>, <https://iupred3.elte.hu/>, <https://mobidb.bio.unipd.it/> and <http://db.phasep.pro/>) (35). From PONDR, we obtained VSL2- and hydrophathy-scores plotted against mean net charge of proteins of interest. The hydrophathy-score in PONDR uses a linear discriminate function to define boundaries between disordered and ordered proteins (112). We obtained genome-wide phase separation scores based on highly frequent pi-interactions in small amino acids with exposed peptide backbone within IDRs from Vernon et al. (111). These pi-contacts were found to be important for protein-protein interaction formation, hydrogen-bond stabilization and solvation, which is directly coupled to the metabolic environment of the cells. Based on these pi-pi contact frequencies in folded proteins, and their correlation with reversible self-aggregation, the authors developed a predictor of pi-mediated phase separation (111). We downloaded their proteome wide pi-scores and ranked the proteins with high likelihood of phase separation. As described by the authors, we cross-examined their prediction using other predictors of phase separation. For proteins with high pi-pi phase separation, we assessed whether they displayed more physical interactions and compared it with counterparts displaying the lowest phase separation score. This revealed, top pi-pi phase separated proteins formed more heterogeneous interactions by interrogating their physical partners using PPIs databases (see interacting transcriptional regulators). We next performed functional annotation of those top pi-pi proteins using Enrichr in R and found they displayed metabolic associated functions. Finally, we examined whether PGC1A-interacting network were among top pi-pi based phase separated proteins.

### **• Context-dependent phase separation scores**

Physical conformations between intrinsically disordered proteins vary depending on different principles and environments (59). Interestingly, solvent-mediated electrostatic repulsions and attractions have been shown to depend on linear sequence confirmation and charge segregation within intrinsically disordered regions (113). Then, the proportion of linear sequences with charge segregation in IDRs partitioned structural representations of proteins into categories such as

globules, context-dependent globules, coils and chimeras. These conformations form an ensemble of parameters for new phase separation classifiers that complement well established low-complexity regions predictors. We then used classification of intrinsically disordered ensemble regions (CIDER) to obtain parameters related to phase conformational transitions (archived in <http://pappulab.wustl.edu/CIDER/about/>). We implemented local-CIDER in python 3 (stored at <http://pappulab.github.io/localCIDER/>), to compute bulk- conformational variations in large data samples. We first focused on globules and context-dependent globule conformation classification (113). These were shown to be influenced by environmental variations such as salt, charge, ligand-binding, cis-interactions and solvation (59). Among the parameters obtained that are at the basis of their conformation classifier, charge-mediated transitions can be recapitulated by net charge and mixing of opposing residues within IDRs. A list of parameters evaluated by CIDER is as follows: K or Kappa represents segregation of opposing charge residues with 0 being well mixed or intertwined while 1 being separated stretches of positive or negative residues; Fraction of charge residues (FCR); net charge per residues (NCPR); f- negative fraction; f+ positive fraction; hydrophathy (114) and disorder promoting scores. Fasta files of protein sequences of interest were downloaded from the NCBI repository.

- **Cumulative phase separation scores of f-ITRs**

To quantify phase separation score in the network of interacting transcriptional regulators, we used adipose-tissue and brain tissue-specific network background to interrogate their collective associations. We then discriminated between individual nodes that have been experimentally validated in phase separation and nodes that are computationally predicted to display a high degree of IDRs. We next quantify network z-enrichment with nodes predicted or validated in phase separation (p-value <0.05). To complement this approach, we used data containing a global genome-wide disorder score based on planar pi-pi interactions (111). For the selected f-ITRs with observed association in tissue-specific networks we summed up their individual phase separation score to a cumulative phase separation score. We then carried out permutations with random proteins and quantified their cumulative phase separation score. This was followed by Chi-square tests comparison and p-value determination.

- **Experimental validation of heterotypic condensates**

See turbidity assay and live imaging FRAP.

- **Phylogenetic amino acid conservation:**

Phylogenetic analysis was performed on amino acid sequences of proteins of interest such as PGC1A. We used the ConSurf software package to obtain conservation scores and multi-species comparison (34). Briefly, it estimates the degree of conservation of amino acid sequences based on iterative comparison with homologous sequences. The conservation rate of a sequence depends on its importance for the structural function. The evolutionary relationship is dictated by the similarity between species comparison in the substitution matrix using previously established Bayesian or maximum likelihood (ML) approaches (115). Here below is a brief description of ConSurf stepwise analysis. First, given your amino acid sequence, consurf uses a heuristic algorithm to search default homologous sequences using BLAST across different databases and with known 3D structural features (116). A multiple sequence alignment is constructed with these followed by phylogenetic tree building using neighbor-joining algorithms (115). Conservation score or evolutionary rate in each residue is imputed using rate4site, and divided in different scales for the degree of preservation. Residues evolving slowly are referred to as conserved and those evolving rapidly are referred to as variable. The degree of preservation depends on different levels of purifying selection, which can be folding constraints, enzymatic activity, ligand-binding or protein interactions. For the rate of evolution, the Bayesian approach takes both into account the stochasticity of evolutionary change and the phylogenetic comparison (117). This probabilistic model of amino acid replacement estimates how likely a residue is to influence structure and function of protein domains. 3D structures of proteins are obtained from updated databases such as the protein data bank (118). It uses simulation based estimation of amino acid substitutions within the tree branch and compared to known structural dependencies of the protein (117). For each residue within the sequence, consurf gives a normalized score of evolutionary conservation from 1 with lowest conservation to 9 with the highest conservation. In addition, consurf maps the estimated conservation score within proteins, to known 3D models of protein structure (118), giving a homology approximation of 3D structural features for the residues. It uses HHPred, which exploits a hidden markov model to look for 3D templates from homologous sequences or similar proteins using PDB (119) and extrapolate them to the query sequence with structural features. The features for each residue are the degree of 3D exposure, hiddenness and functionality. The evolutionary phylogenetic tree is calculated using the WASABI platform (120) which is integrated with multi-species display of amino acid sequences and the degree of evolutionary conservation based on the aforementioned methods. We then evaluated the phylogenetic tree and the degree of conservation for PGC1A sequence. We obtained for each amino acid a conservation scale, with structural features from known 3D representations. We used this to compare the degree of preservation of intrinsically disordered regions within PGC1A. The IDRs were defined as consensus disordered domains within the protein. We then extracted structural features and the degree of evolutionary rate. To compare this domain with the rest of the protein we first defined following consurf estimations: conserved residues scoring as higher than 7 and low conservation rates as lower than 3. As previously described, from the consensus disordered predictions, we obtained 2 IDR domains in PGC1A.

One close to the N-terminal and the other in proximity to the RNA-binding domains. For both IDRs we quantified the number of residues with low conservation score, high conservation, and average conservation. We then obtained the proportion for each category relative to the total number of residues. We performed the same approach for the rest of the protein and compared the estimation of evolutionary rate between IDRs vs the rest of the amino acid sequences. A similar approach was carried out to assess evolutionary innovation in charged residues within the IDR domains. Final scoring was obtained as the proportion of charged residues (positively charged: arginine[R], lysine[K], and negatively charged: glutamate[E], aspartate[D]) within IDRs with low conservation score, in function of the total number of residues.

- **Combinatorial binding density distribution**

To assess binding combinations of f-ITRs to promoters of HAR-associated genes, we used an integrative ChIP-seq database containing >100000 experiments (archived in <https://chip-atlas.org/>) (36). First, we selected top ranked nodes from the interacting transcriptional regulator network. Next, we parsed and integrated all ChIP-seq experiments for each regulator, binding  $\pm 10$ kb from transcription start site (TSS) of target genes. Given that the peak calls in each experiment are based on MACS2 score as previously described (36), we used the average MACS2 score from the available experiments. To filter out all possible combinations of co-regulated target genes, we used a cutoff with an average score >50. Next, we built overlapping maps using venn-diagrams in R from filtered target genes. With this, we quantified the number of co-regulated genes by pairwise combinations of f-ITRs. To assess the frequency and dependencies of pairwise combinations we hot-encoded them and used manifold unsupervised learning to cluster their relationships. We complemented this approach with the use of sieve diagrams or contingency tables as previously described (121). With the sieve diagrams, we obtained the dependencies of transcriptional regulators combinations ( $n=881$  and dependency discrimination  $p$ -value < 0.001). We next selected top combinations for downstream analysis by overlapping manifold clusters and top-dependencies from sieve diagrams. We then used these frequent pairwise-combinations for f-ITRs to select transcriptional regulators with most dependencies and derive a combinatorial-score for a given gene set. As a background set are all possible combinations with all possible target genes for the selected transcriptional regulators. We then used venn-diagrams in R to quantify the number of genes for each combination (classes are >2, >4, >6, and >8 combinations). The combinatorial score for the background set is then the log fraction between the number of combinations in each combination class and the number of overlapping target genes for selected f-ITRs (comb score = Class # of combinations / Total number of overlapping genes with >2 combinations). We then evaluated the combinatorial score from the binding of the selected f-ITRs to HAR-associated genes as follows: HAR-genes comb-score is the log fraction between the number of combinations in each binding combination class and the overlapping HAR-associated genes (number of HAR-genes with >2 combinations). We compared our target gene set with random sampling gene sets with the same number of genes to our query set. Finally, we performed hypergeometric enrichment to compare our query to the background and random sampling, and defined statistical significance at  $p$ -value <0.01. HAR-associated genes bound by several combinations of the selected f-ITRs were functionally annotated by using EnrichR in R followed by selection of common biological terms. We also calculated the PGC1A-network combinatorial score for mediators of polygenic risk for BMI and T2D (see combinatorial disease hubs pipeline).

- **ChIP-MS structuring candidates**

To evaluate regulators that bind active enhancers and promoters and their relationships with PGC1A-network, we used previously published datasets (61), and a similar implementation (62). Briefly, structuring candidates were identified with chromatin immunoprecipitation followed by mass spectrometry. Antibodies targeting epigenetic tags of active enhancers (H3K27ac) and promoters (H3K4me3) (122), were used to precipitate binding regulators. Mass spectrometry was used in those complexes to identify putative candidates. As previously described by Weintraub et al. binding candidates were filtered using the log<sub>2</sub> ratio of the IP over the IgG with a cutoff of log<sub>2</sub> > 1.

- **PPIs structuring candidates and PGC1A-network**

For the resulting structuring candidates binding both active enhancer and promoters, we performed network propagation with PGC1A-network candidates whose expression was increased after fasting in adipocytes. To this end, we used protein-protein-interaction reference networks (at HuRi) filtered for adipose specific enrichments. Network recomposition (node + 1 interactor) was performed for all direct interactors in both datasets, and network analysis parameters and functional enrichment were obtained as previously described (see ITG for transcriptional relationships).

### **ITG-III: functional relationships within genome architecture**

- **Functional embedding of positional gene enrichments (PGE)**

To identify spatial chromatin regions with genes displaying topological dependencies, we used previously established methods of positional gene enrichments (39, 119). Briefly, PGEs exploit topological features of gene locations to

extrapolate dosage sensitive active chromatin regions. This approach is based on calculating the hypergeometric distribution along genomic distances for gene sets inputs. For a given region, it corresponds to the probability of having observed genes in that region. To test statistical significance for that region, cumulative p-value distributions are used on random simulations or false discovery rate on large gene-sets. This works as the probability of achieving p-value enrichments that are better than chance estimates and use this to define topological enrichments. The additional constraints to positively categorize a genomic region displaying positional enrichments for a given set of genes are as follows: having at least 2 genes; no smaller region was found by random permutations and cumulative p value distributions; no bigger regions with more target genes were found. This algorithm defines genomic regions by distance of query genes, followed by estimation of p-value distribution to compare chance expectation. This implies that query lists containing genes located in the same chromosome but distantly, might give false positive enrichments. For this reason, we defined target genomic ranges for positional enrichment following structural information related to the approximate size of homotypic highly interacting TADs (123). The range for topological regions was then set to 15-mb, where PGEs with larger domains were filtered out. These topological dependencies were defined as enriched regions within the aforementioned span. Given that local homotypic chromatin interactions are more enriched within cytogenetic bands (123), we filtered out PGEs without cytogenetic band enrichment, using a curated list of cytogenetic band-gene dependencies (archived in <http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). To obtain a global algorithmic performance determined by the number of genes being queried, we additionally performed permutations with random gene sets (same number of inputs), followed by statistical comparisons using Chi-square test and p-value discrimination. For every topological gene cluster in the form of PGEs, we derived functional enrichments using Enrichr in R. FDR <0.01 and p-value < 0.05 was used as a threshold to select the significant enrichments for biological terms and phenotypes associated with a given region. PGEs with common biological enrichment across domain-catalogues stored in Enrichr, were categorized as class fPGEs (functional PGEs) for downstream analysis. These hierarchical topological chromatin domains were used as seeds for prioritization of structural chromatin data when indicated, and as inputs for determining long- and short-range structural relationships. **Combinatorial-topology PGEs:** To identify colocalized enrichments or topological improvements of enrichment between single gene sets and combined gene sets we devised a combinatorial pipeline variation. For example, for cross-tissue hubs of transcription within each biological group, we combined gene sets composing PGEs per group and evaluated a combined-PGE hub over-enrichment. To this end, we used optimal-PGEs, which were defined as PGEs with more than 5 genes (see transcriptional information uncertainty with Shannon entropy and mice phenotype networks). We then compared the proportion of PGEs with >5 / total regions, in combined by group, individual by group and combined random permutations of PGEs with similar parameters for either comparison. To discriminate between the average of expected proportions for individual and random combinations, with the proportions from biological combinations we used chi-square test and p-value estimation.

- **Hi-C Data Analysis**

Representative Hi-C data from diverse cells was procured from (4DN hESC-REF, human lymphoblastoid cells, colon cancer, hs2-hi-c) (40, 124) and uploaded using straw (125). To determine A/B compartments in chromosomes, we performed previously established protocols (126). To identify TADs, we used previously established protocols (40, 127). Briefly, boundaries were defined at 40-kb, 1-mb genomic regions, and 200-kb window for delta vector calculation. Identification of significant Hi-C contacts at 40-kb resolution was done with a previously established protocol called Fit-Hi-C (128). Within a 2-Mb genomic distance p- and q-value were calculated for each bin pair with FDR<1e-5 threshold for peak-calling.

- **Conserved TADs and PGEs embedding**

We used a comprehensive published compendium of 21 human tissues and primary cells Hi-C data downloaded from the GEO database with accession number GSE87112 (129), along with previously established protocols described therein (129). Briefly, data was normalized using HiCNorm (130), vanilla coverage (124) or ICE (131). For comparative analyses, quantile normalization was used to normalize differences. First line of analysis including compartment A/B and TAD calling were done as previously described (see Hi-C data analysis). TAD boundaries were previously described using established protocols (129). Briefly, TAD boundary regions were defined at 40-400-kb, as previously described (41). TAD overlaps were defined between samples if 80-kb TAD boundaries were shared, followed by chi-square test to discriminate significance among TAD overlaps. Once conserved segments were obtained, these conserved TADs were plotted as chromosome positions in each chromosome. We next quantified conserved TADs within chromatin domains defined by PGEs as previously described. These chromatin regions were in this case defined by the chromatin regions derived from metabolic-HAR genes or metabo-HAR PGEs. We next calculated the mean of conserved TADs within PGEs and compared it to the mean number of conserved TADs within PGEs from random gene sets. We defined each random permutation with the same number of genes composing the metabo-HAR PGEs. This was followed by chi-square test to evaluate statistical significance and area under the curve of ROC to evaluate the discrimination threshold between comparisons.

- **Frequently interacting regions (FIREs) and PGEs embedding**

FIREs regions have been previously described by Schmitt et al (129), and were obtained from the same integrative compendia of human Hi-C data (GEO: GSE87112). Briefly, to identify FIRE bins a poisson regression model (HiCNormCis) was fitted as previously described (130), where intrinsic data biases are taken into account as defined by Yaffe et al. (132). With this approach the total normalized cis-interactions were obtained, and a gaussian distribution approach was used to estimate the local cis contacts, followed conversion of values to  $-\ln(p\text{-value})$  denoted as FIRE score and used in our pipelines for downstream analysis. Once conserved segments were obtained, these FIREs were plotted as chromosome positions in each chromosome. We next quantified FIREs within chromatin domains defined by PGEs as previously described. These chromatin regions were in this case defined by the chromatin regions derived from metabolic-HAR genes or metabo-HAR PGEs. We calculated the mean of FIREs within PGEs and compared it to the mean number of FIREs within PGEs from random gene sets. We define each random permutation with the same number of genes defining the metabo-HAR PGEs. This was followed by chi-square test to evaluate statistical significance and area under the curve of ROC to evaluate the discrimination threshold between comparisons.

- **Trans-interacting regions in hESC Hi-C data**

To identify the relationship between metabo-HAR domains and trans-interacting regions in human chromosomes we used data from hESC GEO: GSE35156 (41). To assess the trans-interacting interchromosomal regions, we used previously established methods with a few modifications (42). Briefly, Hi-C data was normalized with hicpipe, and a probabilistic model was followed to calculate chromatin segment contact maps (132). Contact maps are based on quantifying systematically intrinsic biases within the Hi-C data. For contact call significance a binomial distribution p-value estimation was used as previously described (133). Bins were defined at 500-kb and contact probabilities were normalized against chromosome length. Trans-interacting networks were then formed with segments as nodes and edges as presence of interaction. An undirected graph was built and used to interpolate metabo-HAR domains. Where a metabo-HAR was present in a given chromatin segment, we labeled that node as either PGE and fPGEs if they had a functional embedding. In addition, we quantified the number of trans-interacting bins within chromatin domains defined by PGEs as previously described. These chromatin regions were defined by the chromatin regions derived from metabolic-HAR genes or metabo-HAR PGEs. We next calculated the mean of the number of trans-interacting bins within PGEs and compared it to the mean number of bins within PGEs from random gene sets. We define each random permutation with the same number of genes defining the metabo-HAR PGEs. This was followed by chi-square test to evaluate statistical significance and area under the curve of ROC to evaluate the discrimination threshold between comparisons.

- **PGE-domains in trans-interacting-network**

Trans-interacting networks were formed with segments as nodes and edges as presence of interaction as previously described (42). An undirected graph was built and used to interpolate metabo-HAR domains. Where a metabo-HAR was present or colocalized in a given chromatin segment, we labeled that node as either PGE and fPGEs if they had a functional embedding. We used the NetworkX library in python to perform subsequent topological analysis of the embedded network. We quantified average node degree, mean clustering coefficient, centrality measurements such as betweenness and closeness, information flow parameters such as bridging centrality, and node-prioritization with pagerank and eigenvector centrality. Graph layout such as force atlas, openord, circular, were used for representations when indicated (in figure legends). We performed randomization with similar identified parameters and assessed global structural properties with hESC trans-interacting network. To compare the topological location properties between PGEs and fPGEs, we averaged the clustering coefficients and node degrees for the interpolated chromatin segments or nodes. Again, we compared it to full network averages and random sampling with a similar number of nodes to either PGEs or fPGEs. With the embedded functional annotation obtained before, we identified top hubs and robust nodes with a specific functional cluster. The functional annotation of top clusters were obtained with the functional embedding of PGEs (see above). To filter nodes that both display a specific functional annotation and highly influence information flow through the network, we used bridging coefficient and overlapping functional annotation to hubs and robust segments. Top bridging chromatin segments harbouring fPGEs of interest were selected for downstream nuclear compartment reconstruction. Complementing the network-based prioritization, we functionally annotated genes within top trans-interacting chromatin segments using DAVID functional clustering analysis (<https://david.ncifcrf.gov/>) (106).

- **Integration of short-range chromatin relationships**

To study genomic regulatory element relationships at high resolution, we used previously established methods (43). Briefly, GeneHancer map is a genome-wide regulatory elements relationships, comprising around 284000 integrated elements from different databases: ENCODE, the Ensembl regulatory build, the functional annotation of the mammalian genome (FANTOM) project, the VISTA Enhancer Browser, dbsuper super-enhancers, EPDnew species-specific databases of experimentally validated promoters, and UCNEbase ultra-conserved noncoding elements. The approach by the Cohen Lab links regulatory elements to genes, using a combination of relationships: correlation between genes and enhancer RNAs (134), enhancer-targeted transcription factor genes, expression quantitative trait loci (eQTLs) (135), promoter-capture Hi-C (136), and distance-based associations based on immediate adjacency (43)(50). GeneHancer map of regulatory elements uses a likelihood-based score for each enhancer–gene association. This generates a hierarchical

pairing that is defined by the number of sources supporting it (43). Given our interest in studying both global and local topological representations and the fact that human-accelerated regions are elements that might not have known links, we included all linking data for downstream analysis.

- **Relationships in short-range linking network**

Cis-interacting short-range chromatin networks were formed as an undirected graph with regulatory regions defined as follows: enhancer elements as distal regulatory regions and genes or promoters as proximal regulatory regions. Regulatory regions were defined as the nodes while the edges were the presence or linking association between them. We then interpolated metabo-HAR domain information as follows: HAR-associated genes described before (see pipeline ITG-I) were labeled in the proximal regulatory regions, while non-coding HAR element coordinates were overlapped in the distal regulatory elements. We next used the NetworkX library in python to layout the graph and perform subsequent topological analysis of the network. We quantified average node degree, mean clustering coefficient, centrality measurements such as betweenness and closeness, information flow parameters such as bridging centrality, and node-prioritization with pagerank and eigenvector centrality. We also performed a modularity based clustering approach to find communities of regulatory-region dependencies. When indicated, the graph layout was hierarchical neighboring modularity. In addition, we performed randomization with similar identified parameters and assessed global structural properties. Using hub-based and influence-based node prioritization, we ranked regulatory regions by their density of interactions. We derived a genome-wide density of interactions for proximal regulatory regions as the global mean degree. We categorized this threshold as a social cutoff, and defined nodes as social if they had 2 times the genome-wide mean degree ( $>16$ ). Likewise, isolated nodes were defined as nodes displaying  $< 8$  degree interaction. To compare the topological location properties for HAR-associated genes, we calculated a hypergeometric enrichment from the metabo-HAR-genes on the list of social promoters derived from our genome-wide social network ranking approach. We compared this with permutations of PGEs built from random gene sets of equal parameters. This was followed by chi-square tests to evaluate statistical significance. To evaluate distal regulatory dependencies of non-coding active elements, we first classified all regulatory elements by their density of interactions in function to the proportion of enhancers in each group. We then compared the genome-wide behavior of enhancer linking with the linking density of non-coding active elements (see selection from vista enhancer database). We next fitted gaussian curves in the data followed by sum-of-squares F-test to assess variability between the 2 models.

- **Nuclear compartment reconstruction and functional embedding**

Once we evaluated global relationships for the trans chromatin segments of interest harbouring gene clusters from PGE decomposition, we took advantage of the network-based analysis approach to select with high confidence structural representations valuable for the network global dependencies, yet sharing functional annotation. This protocol can be used to study any PGE constructed from any given gene set and it is based on selecting trans-interacting chromatin segments that both display topological qualities and harbour functional-PGEs (fPGE) of similar function (see PGEs in trans-interacting-network). Briefly, top bridging trans-interacting segments harbouring fPGEs that shared similar functional annotation to top hub trans-interacting segments were chosen for nuclear compartment reconstruction and further experimental validation. Given that our combinatorial transcriptional based selection (see ITG-I) and our topological decomposition (see PGE functional embedding) both showed lipid metabolism as a top functional candidate, highly trans-interacting segments (with top bridging scores) harbouring fPGEs for lipid metabolism were chosen for reconstruction. Following any possible interaction in our trans-interacting chromatin network from hESC Hi-C data, we iteratively sought for other segments with high bridging score and with a functional embedding for lipid metabolism. To evaluate if these interactions were present in other Hi-C datasets, we tested their presence in other cell types (see Hi-C data analysis). Because we were looking for likelihood of interactions for downstream experimental validations in cell- and tissue-specific context, we only highlighted their behavior as binary, either present or absent without any other integrative method. They were displayed as shared contacts across studies. To fully reconstruct highly ranked and functionally embedded trans structural relationships, we used the top bridging segment, harbouring metabolic-HAR genes, Chr1:48.5mb to derive other local homotypic interacting segments, and other interacting heterotypic segments with high bridging coefficient and related function, such as Chr1:62mb, Chr2:170mb, Chr2:118mb, Chr4:42mb, Chr4:1mb and Chr11:34mb. Interestingly, we found a high density of interacting homotypic regions around chr1:48, including chr1:43, chr39, displaying top bridging coefficient scores in the network. This highlights this intertwined region of the genome as a modulator of information flow within chromatin architecture. As shown later, this region contains genes related to lipid metabolism (fig. S10), and displays high-density genetic variations related to lipid metabolism as well (fig. S14 and S15). We displayed our selection as Chr1-based sub-compartment triplets with Hi-C derived homotypic and heterotypic interactions, as well as the location of PGEs and fPGEs with similar functional annotation within the composing chromosomes. This subnetwork of compartments was used for downstream genomic-range topological genetics and experimental validation in biological systems of interest.

- **PGE-domains at high resolution in short-range linking subnetworks**

Once we evaluated global relationships for the regulatory elements of interest within the short-range network, we took

advantage of a sub-module clustering approach to study at high resolution structural representations, yet keeping its global dependencies. This method can be used to study any PGE constructed from a given gene set. First, we selected functional-PGEs (fPGE) following our topological filtering method (see PGEs in trans-interacting-network and nuclear compartment reconstruction). Briefly, top bridging trans-interacting segments harbouring fPGEs that shared similar functional annotation to top hub trans-interacting segments were chosen for nuclear compartment reconstruction and further experimental validation. For the selected trans-compartments, we labeled the modules where local fPGEs reside, using the short-range linking network, and used these modules for downstream analysis. To embed both network and biological qualitative information in the modules of interests, we used a social network based approach supported by epigenetics and transcriptional binding from experimental data. First, every module was further classified by sub-modules using the module cluster algorithm. This generated territorial clusters of associations for HAR-domains (or any PGE) with intra-submodular and inter-submodular interactions, as well as homotypic inter-modular interactions. We then interrogated iteratively binding enrichments of epigenetic chromatin tags given territorial composing regulatory elements (for each module and submodule). To do this, we used the integrative ChIP-atlas data (see combinatorial binding density distribution), to calculate binding enrichments for enhancer coordinates that form topological clusters of associations. When necessary we used liftover function to match coordinates between different assemblies (USCS genome browser at <https://genome.ucsc.edu/cgi-bin/hgLiftOver>). To calculate background ChIP-enrichments per module, we used histone modifications (n=16641 experiments) across all cell types (n=61679 experiments), with a threshold of significance of 50 and random permutations for comparison. Data was subsequently presented as fold-enrichment to p-value followed by word-cloud generator showing most frequent histone modifications. We next embedded topological qualitative information from our network-based analysis to classify social and isolated enhancers (see relationships in short-range linking networks). We next differentiated chromatin states between social and isolated enhancers. First, we quantified from the peaks with significant binding enrichment the number of peak differences for all histone marks between both enhancer classes. Given that histone repressive tags govern cell-specific transcriptional identity by blocking unspecific gene programs (137), we subsequently compared its binding enrichment for social and isolated enhancers. We quantified the total number of significantly enriched repressive peaks (H3K9me and H3K27me) for social and isolated enhancers in a given module, followed by comparison of the total number of enhancers with the repressive tags. This step allowed us to prioritize social enhancers to further dissect active transcriptional elements and transcriptional regulators for a given module. We then selected coordinates for social enhancers to study binding enrichment analysis using the ChIP-atlas dataset. For this we used TFs and regulators (n=15217 experiments) and similar parameters as indicated before. Binding enrichment was defined as fold-change > 2 and p-value <0.05. Once we established the top regulatory binders to territorial social enhancers (modules or submodules), we assessed the binding enrichment for proximal regulatory regions ( $\pm 10$ kb from TSS) for genes within the module or submodule of interest. For a given submodule or module, we then selected top ranked transcriptional regulators binding both social enhancers and promoter regions, and generated a hierarchical list of likely active promoter-enhancers links with their intertwined inter-submodular dependencies. This unbiased approach exploited both integrative and topological associations, which allowed us to generate qualitative information for downstream analysis. In this case, we used for each selected module our previously described pipeline for interrogation of interacting transcriptional regulators (see ITG-II for transcriptional dependencies). Finally, for interacting transcriptional regulators territorially enriched, we highlighted those that recapitulated locally our global transcriptional inferences. This was followed by word-cloud representation (based on number of interactions) and network functional enrichment. We complemented this approach by interrogating local topological behavior of metabo-HAR domains as follows: with a targeted manually curated strategy, we used the ucsc genome browser (archived in <https://genome.ucsc.edu/>) to find the metabo-HAR domain window containing genes and enhancers of interest. We highlighted active elements and assessed by ChIP-regulatory-binding and TFBS motifs, whether members of the PGC1A-interacting network were present among the regulators.

- **Selection of active non-coding elements within PGEs**

To select PGE or fPGE regions that contained experimentally tested activity of regulatory elements, we used the VISTA enhancer resource containing in-vivo validated enhancer information (archived in [https://enhancer.lbl.gov/frnt\\_page\\_n.shtml](https://enhancer.lbl.gov/frnt_page_n.shtml)) (50). For any given PGEs, in this case, PGEs derived from HAR-associated genes (table S3), we evaluated the number of active non-coding elements within the chromatin region. For example, around 84% of metabo-HAR PGE regions displayed active elements. We next assessed the number of active elements for the fPGEs that reside within chromosomes composing nuclear compartments of interest (see nuclear compartment reconstruction). We then labeled these elements for experimental validation in biological systems of interest. Using the short-range linking network, we were able to prioritize genes (promoters or proximal regulatory regions) that were associated with active elements. We next overlapped nCHARs (archived in <http://docpollard.org/research/>) (20), with top ranked active regulatory elements within the network module and submodule of interest. We used these coordinates for downstream experimental validation.

- **Chromatin 3D-reconstruction**

To assess global and local 3D conformations from topological data of chromosomes composing the reconstructed nuclear compartments and within chromatin domains derived from PGEs, we used previously established protocols (138)(139)(52),

139) (archived in as follows: GSDB at <http://sysbio.mnet.missouri.edu/3dgenome/GSDB/>; Spacewalk at <http://3dg.io/spacewalk/>). Briefly, for global chromosome reconstruction GSDB and spacewalk modelling was used to explore the architecture and homotypic interacting clusters where PGEs reside. GSDB is a comprehensive repository of 3D modeling algorithms and Hi-C data. Spacewalk is an application interface with interacting animation for the navigation of genomic-ranges. Using the GSDB application, we first compare the performance between 2 distance-based algorithms for 3D Hi-C reconstruction. We obtained structure evaluation for representative Hi-C at 40-kb resolution of 3D reconstruction using 3D maximum likelihood algorithm (138), and LorDG which uses a nonlinear Lorentzian objective function to optimize realistic interactions (140). In addition to the same Hi-C data previously used for isolating trans-interacting segments (hESC HiC; GSE35156), we used colorectal adenocarcinoma cells (GEO: GSE105318), human brain pericytes (GSE105513), astrocytes from the spinal cord (GSE105957), human fibroblast IMR90 (GSE35156), brain microvascular endothelial cells (GSE105544), and human hepatic endothelial sinusoid cells (GSE105988). To compare the structure reconstruction between these distance-based algorithms, we used the average spearman correlation of Expected Distance vs. Reconstructed Distance (139) on the chromosomes that composed our selected nuclear compartments and the small chromosome 19 as a control. Next, we represented the data in a XY plot as performance score (expected vs reconstructed distances) vs chromosome size (log). We then evaluated the performance across different cell types by performing a linear regression followed by F-test statistics. This algorithmic selection showed that 3DMax maximum likelihood algorithms represent better in 3D Hi-C data including long chromosomes. Therefore, we next evaluated in spacewalk 3DMax pdb-files to assess the exposure of metabo-HAR PGEs, homotypic interactions, homo-heterotypic relationships, as well as the genomic range of their cluster association.

- **CTCF- and RNAPII-mediated chromatin contact domains**

To complement the global reconstruction of 3C chromatin interactions, we evaluated targeted chromatin contact domains with previously published approach (51). Briefly, CTCF-mediated modeling is done from chromatin interaction analysis and paired-end tag sequencing (ChIA-PET) from several cell lines (GEO:GSE72816). Tang et al. employed a hierarchical method to integrate specific structure candidates such as CTCF ChIA-PET data with all chromatin contacts from Hi-C data into higher resolution chromatin folding structures. First, PET inter-ligation products with a genomic span of 8-kb were used for inter-region interactions. These were expanded 500-kb along the genome to calculate the interaction frequencies and to define PET clusters for sequencing as previously described (51). These inter-ligation PET cluster products were used to call singletons as enriched interactions between discrete chromatin loci (141) (142). To identify binding peaks MACS pipeline was used (143). For intra-chromosomal normalization data was partitioned into 1-mb bins, where targeted quartile normalization was applied as described by Tang et al. (51). CTCF ChIP-seq of ENCODE data and STORM CTCF motif interrogation was also used to find overlapping foci for filtering of CTCF loops. To identify chromatin contact domains (CCDs), data was clustered into loops and gaps, followed by aggregate calculation of CTCF loop coverage along all chromosomes and filtering out of regions with a span <10-kb. CTCF anchors were defined using CTCF binding peaks and motifs as previously described (51). To map transcription functionality into CCDs defined by CTCF, RNAPII (RNA-pol II) ChIA-PET was analyzed similarly to the CTCF approach as previously described. RNAPII loops were filtered following GENCODE promoter (TSS), transcription end site (TES) and enhancers overlapped data. Briefly, RNAPII loops were then overlapped with CTCF-CCDs and found an 85% approximate inclusion of RNAPII loop clusters within CTCF domains. Single- and multi-gene loop complexes were defined and identified as previously described (144) (145). This depends on the amount of promoters involved; 1 for single and more for multiple complexes. In addition, this finding allowed us to define with confidence the rule for node notation in our short-range linking network, where genes were identified as proximal regulatory regions and enhancers as distal regulatory regions. Multi-RNAPII complexes showed promoter regions in heterotypic clusters can behave as distant regulators of transcription within CTCF-domains (51). The topological domains are displayed as CTCF-graph clusters with loops as arcs within the genomic range. Because depending on the genomic position CTCF-cluster might vary, statistics on the genomic range of interest are displayed as loci distribution and cluster length distribution. We used these PET-based domains and the superimposed RNAPII-complexes to compare the distribution and boundaries of these CCDs with our short-range linking network (see above). More specifically, we compared the submodule boundaries of the integrative short-range network with CCDs-based boundaries, and found a high degree of overlapped boundaries within our PGE domains. Altogether, the integrative topological comparison by Tang et al. between a targeted CTCF-mediated and untargeted methods for topological domain reconstruction (CCDs and TADs respectively), showed a high degree of superposition between both methods. This underscores the relevance of CTCF cis- and trans chromatin binder regulation while regulating local complexes for transcriptional tuning through multi-RNAPII loops. Given that functional domain of transcription are CCDs and TADs, and that intra-domain stochasticity have been represented in loop-extrusion and polymer-inspired models (70), we used submodular hierarchical node ranking from our high-resolution short-range network to interrogate loop anchors with CTCF enhancer binding discrimination (see hierarchical enhancers prioritization).

- **CTCF-mediated 3D-reconstruction**

To complement the global reconstruction using 3C chromatin interaction frequency, we evaluated multi-scale local 3D modeling of PEGs and segments of interests with previously established protocols (51, 52, 146) (archived in <https://3dgenome.cent.uw.edu.pl/>). Briefly, the authors used a hierarchical tree based model of ChIA-PET derived CCDs

(ChIA-PET data stored at GEO: GSE72816). The chromosomes are then modeled following a tree-based representation where parent-child node relationships are defined by the resolution. They used a beads-on-a-string polymer model where beads are genomic coordinates of 1-2 mb genomic range with PET-clusters within. Anchors and subanchors were defined by CCDs and gaps between them respectively. The hierarchical clustering algorithm aggregate then neighboring interacting segments following genomic size constraints. This approach allowed it to go from different levels of resolution; for example from low with singleton units or CCDs to high resolution with intra-CCDs ChIA-PET defined loops. For structure reconstruction a monte carlo simulated annealing function is used to minimize the energy in each level. At low level the singleton data is used to minimize simulations while at higher resolution PET-loops are used to optimize local interactions. To translate loop arc data to physical genomic distances a inverse power law approach was used, where the number of interactions in a given PET genomic range reflects physical distances. Finally, 3D models are tuned depending on loop directions and singletons size. With this approach we then obtained CTCF-mediated 3D models of local PEG domains and their relative structural genomic range.

- **CTCF-bound regulatory regions**

To identify gene promoters bound by the structuring protein CTCF we used the ChIP-atlas dataset as previously described (see combinatorial binding density distribution). For the genes identified as CTCF-bound, we assessed whether they were classified as social proximal regulatory regions by obtaining their association degree from our integrative regulatory network (see short-range linking network). This was followed by hypergeometric enrichment of CTCF-bound promoters within promoters defined as social. We used the Chi-square test to evaluate statistical significance. In parallel, we interrogated ChIPseq-enrichment of CTCF in enhancers defined as isolated or social (see short-range linking network), using again the ChIP-atlas dataset. In addition, we used other database compendia of enhancers for interrogating CTCF-bound elements in local chromatin regions (see structural enhancer prioritization).

- **Structural enhancers prioritization**

We next sought to identify structural enhancers within genomic ranges defined by fPGEs (in the present study as metabolic-HAR regions). We used enhancer coordinates within submodule territories of the integrative short-range linking network. To prune our global network approach with enhancer datasets including structuring-based experiments (ChIA-PET and Hi-C) and identify TF binding candidates across several cell lines, we used the HACER (human active enhancer to interpret regulatory variants) database (147) (archived in <http://bioinfo.vanderbilt.edu/AE/HACER/index.html>). In brief, HACER is a comprehensive database that incorporates interaction frequency experiments to the linking association derived from GROseq, PROseq and CAGE profiles. Enhancers in HACER are cross-referenced with VISTA enhancers (50), ENCODE elements, the ensembl regulatory build (148), and chromHMM (149). Enhancer interactions in HACER are defined using similar parameters to the GeneHancer regulatory network used here to perform network propagation of short-range association. These are genome distance, eQTLs, and FANTOM. Besides these, the authors used structural-mediated linking to improve their associations with experimental interactions from 4D genome consortium (150) and chromatin contacts derived datasets (51, 136, 145). Moreover, HACER has integrated ENCODE ChIP-seq data from several cell lines and tissues. This includes > 700 000 TF-enhancer binding and 156 TFs. In order to interrogate TF-binding of structuring candidates as well as experimental structural validations of enhancer-gene associations, we used as inputs the coordinates of regulatory regions within fPGEs. We then used HACER assemblies as test datasets for our global networks, where enhancers bound by CTCF in HACER were classified as social or isolated with the network parameters of the short-range linking networks (see short-range linking network). With this approach we were able to estimate the average number of links that structuring enhancers display within selected chromatin regions. In sum, this was a qualitative step to filter short-range network representations within regions of interest, in submodules for downstream experimental validation. For the filtered enhancers within fPGEs, we used HACER GWAS-integration of regulatory regions displaying SNPs for experimental validations (151).

- **Cell-specific regulatory regions**

To identify cell-specific regulatory elements in our global short-range association networks, we used 2 different approaches relying on previously established protocols and datasets. *Enhancer atlas*: First, we used the enhancer atlas (archived in <http://www.enhanceratlas.org/index.php>) (71). Briefly, this atlas contains regulatory elements for > 500 tissue and cell types across several species. It incorporates high-throughput experiments including chromatin states, DNase-seq, ATAC-seq, ChIA-PET, GRO-seq, STARR-seq and MPRA among others. To combine different tracks from each dataset and generate a consensus assembly the authors used an unsupervised approach that weights each track for the global track (152). Peak filtering is done at 2500bp and merging with the average summit in the size of the average peak width (ASW) (153). To focus on the degree of overlap between regulatory elements, Tianshun et al. used the Jaccard index with intersection over union (71). Using this resource we downloaded cell-specific enhancer coordinates and mapped them with our short-range regulatory network, highlighting their location and network statistics. We then used this qualitative filtering step for experimental validation of cell-specific activity during fasting. To this end, we performed ChIP-qPCR with epigenetic tags of enhancer activation (H3K27Ac), and transcriptional regulatory binding. Other tissue-specific modules with motif enrichments were obtained from Epimap (57).

- **Structural variants density within PGEs**

Structural variants (SVs) promote large genomic rearrangements that have profound impact on disease and evolution (54)(55). These rearrangements can affect dosage of protein-coding genes and cis-regulatory function (55). Until recently, single-nucleotide variants (SNVs) have been the reference in genome-sequencing population genetics. However, new efforts to generate maps of SVs from diverse populations have highlighted their potential impact on functional genomics. With this in mind, we have integrated into our pipelines approximate 15630 SVs data from the reference resource compendia developed by Collins et al. (55) (archived in <https://gnomad.broadinstitute.org/>). As previously described and for our purposes, we used common SVs (allele frequency >1%) that displayed strong linkage disequilibrium (LD) ( $R^2 \geq 0.8$ ) with common short variants with a reported association to GWAS catalogue and UK-biobank (55). These selected SVs in LD with GWAS variants were found to be enriched in genic and intergenic regulatory regions. To evaluate the density of SVs within PGEs located in the chromosomes composing nuclear compartments of interest, we partitioned SVs location by cytogenetic band genomic range (resolution 850 bphs). We then displayed the number of SVs per cytogenetic band within the selected nuclear compartment. We next located the band with fPGEs harbouring HARs, as well as other neighboring and interacting bands composing both topological domains and 3D interacting homotypic regions.

- **Genomic region enrichment annotation tool (GREAT)**

To assess how pervasive the functional annotation of HAR genes, and specifically the metabo-HAR regions obtained by our ITG-pipeline, we performed GREAT analysis (154), to find ontologies of the genes within the same genomic regions. Briefly, HAR coordinates (table S3) were used as an input into the GREAT software (available at <http://great.stanford.edu/public/html/index.php>), which evaluates biological process enrichment for neighboring genes compared to genomic background, followed by bonferroni-corrected test to estimate p-values.

- **Genomic-range genetics**

To find metabolic disease associations for a given genomic-range, we used the curated common metabolic disease portal (archived in <https://hugeamp.org/>) This integrated database allowed the analysis of human genetics linked to common metabolic diseases. To evaluate genetic association enrichments for fPGEs or genomic regions of interest, we used previously established methods (currently used at hugeamp) (56). Briefly, this protocol is based on meta-analysis of associations that provide an estimate of variant effect by integrating information from diverse studies. By taking into account the sample overlap among datasets, this approach weights in the study contribution for the final estimate. The meta-analyses are implemented in METAL which is a tool for genome-wide scans that was described by Willer et al. (56). Their approach is based on p-values and subsequent z-statistics for the impact of the association. Given that all studies used are aligned to the same allele, z and p-value are obtained from the weighted sum of individual statistics. The weights are then the square-root of the number of individuals in each sample. Following their streamline implementation of METAL and genome-wide scans we interrogated the cumulative association of genomic-regions to metabolic disease traits. For this, we used fPGEs in chromatin segments composing heterotypic compartments including those in Chr1, Chr2, Chr4, Chr8 and Chr11. Next, we assessed genetic associations to homotopically interacting chromatin segments and displayed p-value of association for a given region. We obtained from this, trait groups and top traits within groups. We displayed the integrative results for all genomic-range traits from metabo-HAR regions as a wordcloud representing the number of appearances of groups and trait descriptions.

- **Epimap integration**

To integrate epigenomic data with high-resolution into our pipelines, we used previously established methods developed recently (archived in <http://compbio.mit.edu/epimap/>) (57). In brief, Epimap includes > 800 biosamples with > 10000 epigenomic maps from Encode and Roadmap epigenome consortium (155, 156), all processed uniformly and integrated following a machine learning approach to impute chromatin states, tissue-specific enhancer modules, upstream and downstream regulators (57). This comprehensive compendium of functional enhancer modules with motifs, and GWAS trait enrichments revealed patterns of disease dependencies and comorbidity. We downloaded results from this resource including enhancer modules with tissue-specific enrichments, enhancer-gene links, GWAS enrichments, tissue-trait, and trait-enhancer relationships for downstream network analysis. We next built undirected relational graphs and performed hierarchical neighboring analysis to identify module dependencies. Each module was classified depending on its phenotype composition, and ranked by node-degree. We next used multidimensional scaling and node prioritization to uncover module drivers and information flow bridging nodes (bridging coefficient). We focused on metabolic- and lipid related traits to carry out network propagation and find out the degree of trait permeation within modules and with interacting modules. For top associated enhancers enriched in driver tissues, we extrapolated enhancer-gene information. In addition, we displayed inter-related trait dependencies in function of bridging coefficient score (mean per module) within obtained modules, which showed their lipid-trait hierarchical dependencies for general network-flow bridging score coefficients. For related genes with modular, or bridging dependencies we assessed topological enrichments to identify spatial positional hubs. For each inter-related module, we quantified positional hubs of driver genes, individually (per module), and combinatorially (for top interacting modules; see functional embedding of positional gene enrichments). We

then assessed if individual module hubs, and topological top driver hubs, enriched in relational traits obtained here, were colocalized with other interacting metabolic-HAR hubs in general, and those within reconstructed nuclear compartments.

- **Aggregated topological expression from scRNAseq and bulkRNAseq**

For aggregate topological expression we used previously published data from human brain snRNA (76) and human APC-bulk expression data (75) (see human brain single nuclei RNA seq analysis and transcriptomic profile of human adipose precursor cells). We then partitioned the datasets into fPGEs of interest, in this case our selected metabo-HAR domains within the reconstructed nuclear compartments. For the genes composing a given fPGE or PGE, counts of expression were aggregated per biosample or cell subtype defined by the clustering representation. We next represented local transcriptional activity within the PGE by the aggregate expression number for each gene (y axis) vs an approximated location of corresponding genes (x axis) following the human genome assembly Hg19. Gene location number was defined depending on the appearance following a 5' to 3' direction. To compare topological transcriptional activity between groups we used Area under the curve (AUC). To give an estimate of topological regional expression, we simulated the aggregated counts within each biological group.

### ITG-IIIa: transcriptional burst models

- **Dynamic gene expression data**

To quantify dynamic gene expression changes in cellular systems of interest, we used intrinsic fluctuations captured by targeted gene expression (see analysis of gene expression) between challenged (i.e. fasting) and control states (RNA extraction and cDNA library preparation protocol was performed as previously described). The intrinsic fluctuations of dynamic gene expression were used to build statistical models using a Monte Carlo approach (see Monte Carlo modeling). Modifications in cell culture for broad capture of gene expression were as follows: cells were plated in 96 well plates in equal cell numbers ( $1 \times 10^4$  cells/well). Induction and differentiation were carried out as previously described (see mammalian cell culture). Once differentiated, fasting (low glucose medium) was induced in half of the wells followed by dynamic time tracking, where every plate was destined to track 1 different time point. After 3 hours of fasting, cells were collected (1 plate per time point) and RNA extraction protocol was performed every 10 minutes for at least 10 different times ( $T_0 + 10$  additional measurements; where  $T_0 = 3$  hours after fasting induction). We collected at least 6 biological replicates per time point followed by qPCR gene expression analysis as previously described. To compare the expression of genes within the same TADs, CCDs or submodules as previously defined (see ITG for structural chromatin relationships and short-range link networks). For the selected TADs within PGEs of interest, we quantified fold-change distribution across time variations in genes with social or isolated promoters. These were defined by network integration and propagation analysis of short-range link networks. In brief, we obtained fold-change variations (and distributions) by comparing the expression between control and fasting in every time point evaluated. These dynamic fold-change fluctuations between fasting-induced genes were used to build downstream statistical models. We were interested in studying, for differentially expressed genes within the same TADs during fasting, how much their gene expression fluctuates over time comparing it with control bursting fluctuations. We observed that, for genes regulated during fasting, there were time variations in fold-change over control, indicative of specific bursting patterns (**fig. S17N**). Indeed, genes that were defined as isolated by our network approach did not display dramatic fluctuations as those defined as having social promoters (**fig. S17N**). Their fold-change variation over time was stable and constant during fasting (**fig. S17N**). On the other hand, active genes with social promoters recurrently exhibited bursting fluctuations in several experimental setups indicative of different kinetics of expression. To control for primer effect we tested several sets and obtained consistent results. In addition, we observed that perturbing genomic plasticity during fasting, blunted this fluctuation in social genes, while preserving their constant and stable pushed expression.

- **Non-linear sinusoidal models on dynamic gene expression**

In order to capture these variations and model statistical differences we investigated non-linear models. Experimental observations showed an intricate oscillating pattern of expression for some genes within fPGE domains. Their behavior was reminiscent of sinusoidal spectral patterns where fluctuations seem to capture energy discrepancies at different frequencies. This oscillatory dynamic behavior has been observed in complex biological systems where synchronicity of population behavior has been exploited to extrapolate principles for non-stationary variability (157). Moreover, gene expression burst kinetics have been shown to differ between enhancer- and promoter-encoding burstings as well as models for condensate-control of transcription (73)(74)(74, 158). With this in mind, we used non-linear sinusoidal regression (sine-wave with nonzero baseline model in graph pad) to model our experimental observations. We fitted models for each target gene fluctuations to assess their best-fit parameter. Non-linear least squares model was used to fit the regression without weighting the data points (159). To discriminate the best-fit parameters between models we used the Akaike's information criteria (AICc), which assumes non-nested inferences. Goodness-of-fit was evaluated by  $R^2$  and standard error of the estimate. Parameters obtained from the sine-wave models are as follows: amplitude defines the

height of the waves from the baseline; frequency is the number of cycles per time unit; baseline is the Y value where the curve oscillates. We next compared the best-fit parameters from biological experiments with randomly generated values that preserved amplitude distributions. To this end, we generated random values per time point followed by sinusoidal non linear regression, and used AICc to evaluate goodness-of-fit between them. In addition, we also modeled basal expression detection of both house-keeping and unstimulated genes. This showed discrepancies between the models, with amplitude and frequencies displaying the largest variation. Burst profile models of social genes exhibited higher frequencies, amplitude and baseline parameters. We therefore used these quantified parameters to simulate in large scale sinewave behavior constrained by these model variations. We used an in-built simulation option in graph pad where XY table values were selected. Time-series variations were on the X values with similar intervals as our experimental data set (>1000 interval simulations). Y values were generated using sinewave non zero baseline model equations with amplitude, frequency, phaseshift and baseline constants from experimentally derived models. Gaussian random error was used to generate random scatter values. Once the simulations were obtained we performed nonlinear sinewave regression on simulated values to get a better estimation of confidence intervals and estimate best-of-fit and oscillatory parameters. We used these statistical approximations for Monte-Carlo simulation followed by outcome discrimination of the models recreated.

- **Monte Carlo modeling of dynamic gene expression**

To estimate the distribution and range of variation among oscillatory parameters in models from different experimental biological observations, we used a Monte-Carlo approach on large-scale simulated data from our previously described step (see non-linear sinusoidal models) (160) (161). From the simulated sinusoidal models, we obtained best-fit estimates of models constrained by experimental observations. Using these best-fit estimates, we modeled >1000 paralleled models using Monte-Carlo. We defined sinewave equation parameters for each group followed by nonlinear sinewave regression evaluation. Outcome discrimination on each models was carried out with AICc and standard error of the estimate from experimental groups and random generated models. We next tested distribution discrepancies for oscillatory parameters (amplitude, frequency, and baseline) from Monte-Carlo models. Mean distributions between biological groups from which the models were devised, were used to make statistical comparisons between model fluctuations and estimate p-value for model assumptions.

#### **ITG-IV: Topological comparative genomics: human-mice convergence and tuning of nuclear compartmentalization**

- **Mice-human chromosomal synteny**

To evaluate genomic conservation regions between humans and mice, positional gene enrichments of mice orthologues metabo-HAR genes were obtained following the same criteria as indicated before (see functional embedding of positional gene enrichments). We quantified the number of optimal-PGEs (> 5 genes) between human and mice topological enrichments and compared them to the number of PGE derived from random permutations. Subsequently, mice PGEs regions of < 15-Mb were selected for downstream analysis. We next evaluated whether metabolic-HAR regions (mPGEs and hPGEs) were located in syntenic domains, using the synteny portal (archived in [http://bioinfo.konkuk.ac.kr/synteny\\_portal/](http://bioinfo.konkuk.ac.kr/synteny_portal/)) (162). This revealed that 82% of PGEs were in syntenic regions and 18% were located in split regions in mice. Of these, some belonged to our network-based nuclear compartment for lipid metabolism and were experimentally validated. We also displayed the level of genomic conservation in breakpoint regions and in fully syntenic blocks (41).

- **Mice phenotype networks**

For a given gene set of interest (in this case, metabolic-HAR genes), we identify gene and phenotype relationships from data stored in Jackson laboratory containing mammalian gene perturbation experiments. This dataset has been parsed and is stored in ENRICH as mammalian perturbation catalogue (103). Network reconstruction was done with all phenotypes associated with gene knockout information followed by network integration and analysis. To identify phenotypic modules we used modularity and to identify gene drivers per module we used node degree, pagerank and eigencentality. To identify phenotype drivers we used node prioritization followed by quantification of phenotype representations by proportion frequencies (top 30 phenotypes) Intermodule communication was performed by quantifying interactions between modules. Positional gene enrichment coalescence of gene drivers within top interacting modules was performed using positional gene enrichments for individual genes within modules and combined genes from top interacting modules. Proportion of positional enrichments by coalescence was defined by comparing positional enrichments with > 5 genes per region (see functional embedding of positional gene enrichments).

- **Transcriptional hubs (PGEs) in chow-diet and High-fat-diet across 19 organs**

We performed positional gene enrichments of differentially expressed genes derived from RNAseq in mice models of obesity (see animal experiment; HFD intervention). In brief, mice either on chow-diet or HFD were sacrificed and 19 tissues were collected for RNAseq profiling. Multi-tissue differentially expressed genes were classified as HFD regulated or Chow-diet regulated. These gene sets were used as inputs for hub PGE-analysis. We displayed PGE hubs as the proportion of total PGE by dividing the number of regulated genes in topological enrichment with the total number of regulated genes. We carried out this approach in every tissue per biological group, and compared them with hub-PGEs formed from random sets (same distribution parameters as the number of genes composing PGEs of interest). To identify overlapping or cross-tissue hubs of transcription within each group, we combined gene sets per group and evaluated spatial hubs. We assessed optimal-PGEs, which were defined as PGEs with more than 5 genes in genomic cluster (see transcriptional information uncertainty with Shannon entropy and mice phenotype networks). We next compared the proportion of PGEs with >5 genes / total regions, in combinatorial hubs by group, individual and control random permutations (with similar parameters for comparison). To discriminate between the average of expected proportions for individual and random combinations, with the proportions from biological combinations we used chi-square test and p-value estimation. We also compared optimal-PGE hubs from genes that were only regulated by HFD or chow-diet (see transcriptional information uncertainty by Shannon entropy). To this end, we used proportions of total topological enrichment to get an estimate of full transcriptional burden by treatment. Proportion of total topological enrichment was defined as the number of genes in PGEs / total number of regulated genes. We plotted representative optimal-PGEs in each chromosome by group (genomic range defined by start and end of chromosomal PGE region) and defined colocalized regions of activity. Metabolic-HAR domains and experimentally validated HAR-regions in human cellular models were highlighted. In addition, we calculated the proportion of unique PGEs in each chromosome, by quantifying the total number of PGEs irrespective of treatment and the ratio contribution per state. To show in full-display this topological coalescence across similar tissues and by physiological state, we plotted all PGEs derived from adipose tissues (subcutaneous, visceral, mesenteric, retroperitoneal and brown fat) from chow-diet and HFD. This revealed both territorial cross tissue transcriptional activity and tissue-specific domains. For those chromatin regions showing convergence, we hierarchically ranked them by p-value of enrichment. Top hubs were indeed cross-fat-tissue convergent regions where transcriptional activity was present. Inter-physiological state comparison demonstrated a HFD-mediated topological shift as a large portion of PGEs were only present in the obesity model. The shift of regions between treatments was defined by a genomic-range transition of at least 500kb for those optimal-PGEs. This yielded evidence about both a global topological shift on spatial regions of transcriptional activity, and information about the spatial transcriptional burden derived from suboptimal PGEs in HFD (see transcriptional information uncertainty by Shannon entropy).

- **Topological PGE-mediated single-cell decomposition**

To decompose topologically bulk-mediated hubs into single-cell PGEs in tissues of interest, we first downloaded and re-analysed single cell RNAseq data from the tabula muris consortium (see adipose tissue scRNAseq from mice). After obtaining differentially expressed genes in adipose tissue, we performed PGEs per cell type (see functional embedding of positional gene enrichments). We observed cell-specific transcriptionally active regions as scPGEs from the scRNAseq data with some large PGEs with shared cell pair enrichments (e.g. mesenchymal stem cells and endothelial; B-cells and endothelial). Given that this dataset was generated in wild-type non-treated mice, we assessed whether scPGEs colocalize with hubs derived from chow-diet transcriptomic data. The overlap genomic-range was defined as > 200 kb between PGEs and we quantified the proportion of domains shared as transcriptionally active regions. To assess if the degree of overlap was higher than random expectation we performed permutations with PGEs formed with random gene sets. We showed that random overlap lingered between 20 to 35% while scPGEs displayed > 50% overlap for most cell-specific PGEs in bulk-derived chromatin domains. NK overlap was similar to random expectation while T cell overlap was around 40%. A similar approach was performed with PGEs from HFD in adipose tissue. We overlapped them with fat scPGEs and quantified the proportion of overlapping for each cell type with the HFD-mediated chromatin domains. This revealed that mesenchymal stem cells and endothelial cells scPGEs shared the highest proportion of HFD-regulated hotspots. We then assessed chow-diet specific hubs in fat that are lost after HFD and are located contiguous to syntenic metabolic-HAR domains. These genomic-regions of interest were interrogated for homotypic and heterotypic interaction using contact frequency Hi-C data (see Hi-C data analysis). We then mapped homotypic and heterotypic interactions from PGE-PGE chow or HFD locations that displayed cell-specific transcriptional activity by scPGEs. Metabolic-HAR regions were also mapped in contact frequency maps of interest. Those cell-specific topological variations induced by diet were selected for downstream experimental validation by ChIP-loop, expression and chromatin immunoprecipitation.

- **Transcriptional information uncertainty by Shannon entropy**

In information theory, entropy is defined as the uncertainty of a random variable (163, 164). In other words, it quantifies the expected values of information in a given message, or how much information or surprise levels are related to an outcome. If we take event probabilities, information theory assumes that the lower the probability of an event the higher the uncertainty in the observable outcome. Now, if we have a variable taking many numbers (or events) with a finite quantifiable limit, we can then consider the distribution of its probabilities. Any time we observe an event of probability we gain information, which we quantified as bits. Thus, for whole probability distributions we can assess the average level of

uncertainty for given probability of events. To extrapolate information from spatial hubs of transcriptional activity, our assumption based on our observations and previously described observations on transcriptional assemblies, was that localized genomic domains of transcription in the form of positional gene enrichments or hubs (derived from high-throughput transcriptional profiles) can be used to estimate the level of randomness or uncertainty in transcription. Moreover, biophysical condensate formation decreases cellular entropy by localizing hubs of cellular activity (7). As shown before (see functional embedding of positional gene enrichments), compared to random gene sets, gene sets derived from high-throughput biological observations always displayed an elevated proportion of topological enrichments with > 5 genes within genomic clusters (defined before as optimal hubs). This approach can therefore be considered as an indirect measure of hub-models of transcriptional activity, where the number of observations (number of PGEs with > 5 genes / total number of genes in positional enrichments) for a given gene set defines here “event probabilities” of the variable transcription. We considered the total number of genes in positional enrichments as the full transcriptional output for a given state, which we used to obtain probabilities of localized transcriptional organization to subsequently derive Shannon entropy. The reason to use the total number of genes in positional enrichments rather than the total number of PGEs for a given state lies in our previous observations in Fig 4. These showed that, for a given state across many organs (chow-diet or high-fat diet), around >60% of genes are within PGE hubs. Of these, around 15% are located within top hub PGEs of >5 genes. This yielded evidence that the largest discrepancy in transcriptional burden lies on the number of genes being transcribed and located within suboptimal <5-genes PGEs. In that regard, qualitative hub-PGEs of >5 genes were defined by comparing the average p-values between optimal and suboptimal PGEs using all transcriptomic data (independent of state) vs random permutations. In addition, we plotted fold-enrichment vs p-value followed by non-linear regression and f-test statistics to compare random and biological models. We then perturbed optimal PGEs to show that random and biological models of suboptimal PGEs are similar. Therefore, for every organ or biological dataset we derived the “event probabilities” as the number of optimal PGEs with > 5 genes / total number of genes in positional enrichments. We used these proportions as observed event probabilities to calculate individual (per sample or organ) or global (event probabilities from organs in a given state) Shannon entropy (**Fig 4G**). This was performed using python library for scientific computing scipy. Finally, to show we can capture random transcription by single genes using PGEs in a given state, we first partitioned unidirectional genes by state. This was done by taking all differentially expressed genes across all 19 tissues, followed by venn-diagram intersections to define genes that were ‘only’ regulated by HFD or Chow-diet. These unidirectional gene sets were used to quantify PGEs per state and the proportion of total topological enrichments (**Fig 4H**). This showed HFD increased the transcriptional burden, similar to random-derived positional enrichments (**Fig 4H**).

## **ITG-V: Combinatorial disease hubs: polygenic risk scores (PRS) and causal mediators**

### **• Introduction**

Metabolic-related traits and other complex-traits are polygenic with >90% of variations located in non-coding regulatory regions. They also display pleiotropy as the same variant associations are shared between traits (165). Organismal and cellular metabolic states influence the genetic burden of other complex-traits, which highlights the ubiquitous and pervasive nature of metabolism on disease risk. Indeed, obesity and type-2 diabetes increase the risk of developing a plethora of pathological processes determined by individual genetics (165). It then becomes paramount to understand the intertwined relationships of shared variations among complex-traits with specific focus on metabolic disease maps. The systematic dissection of cis-regulatory-gene associations is a key step to decompose hierarchically metabolic disease and subsequently its multi-trait penetrance. To this end, we used results from previously published approaches where causal inference methods were developed to systematically assess the dependencies between genetic variations, genes, tissues and traits (79). Park et al. implemented an inference framework based on causal multivariate mediation analysis with extended linkage disequilibrium (CaMMEL) (79), to identify causal cis-regulatory genes mediating the effects of complex traits. This approach uses eQTLs in 49 tissues from Genotype-Tissue Expression (GTEx project) and jointly analysed them using summary statistics to define tissue-specific causal genes. Source code for CaMMEL is publicly available as a part of the R package for summary-based QTL / GWAS analysis (available at <https://github.com/YPARK/zqtl>; <https://github.com/YPARK/cammel-gwas>).

### **• Data preparation, PRS and TWAS statistics**

Here below, we give a general description of the method for multi-tissue PRS causal mediation developed by Park et al. with specific implementations in metabolic disease (BMI and T2D) by integrative topological genomics pipelines used in this study. For **data preparation**, Genomic locations of GTEx eQTL summary statistics and genotype matrix were used, linkage disequilibrium blocks and dosage matrix of GTEx data were done using published methods (166, 167). Additional variations on LD estimation can be found at Park et al. (79). For **Polygenic risk scores (PRS) and TWAS statistics**, a summary-based method (168) was used to predict PRS from GTEx individuals. Specific information and full results can be found at <https://github.com/YPARK/PRS>. After having the PRS of gene expression and GWAS, a linear regression analysis of the GWAS PRS on the expression PRS identifies slope parameters and standard error. These are then used

to estimate the shift from zero by building the Wald test statistic. This corresponds to summary-based TWAS statistics as previously described (169, 170).

- **Causal Mediation analysis**

To estimate causal loci across different complex traits by their indirect association through inferred intermediate molecular traits of gene expression and enhancer activity, CaMMEL was developed (79). It follows a Bayesian generalization of Mendelian Randomization (171, 172) and TWAS (169, 170) approach that uses eQTLs to infer the genetic component of intermediate phenotypes and correlate them with different phenotypes. To define causal mediation problem the authors assume that only cis-regulatory mechanisms drive the effects of mediation between genetic variation and phenotypes. Then the relationships between genotype, mediator and phenotype variables are generated by stepwise linear combination of genotype and mediator information with coefficients. In brief, CaMMEL focuses on finding non-causal effects or confounders, to underscore residual variation as causal genes. To correct for non-causal correlations, the algorithm uses inherent and invariant relationships related to LD blocks, and contrast the confounders with true cis-regulatory effects. Similar control-based approaches have been described before (173, 174). The projection of eQTL and GWAS z-score matrices onto independent LD blocks constitutes the control matrix. Confounding factors on mediators and phenotypes are removed by matrix factorization, followed by removing their effects on GWAS traits and estimation of residual effects. This is followed by summary-based multivariate regression of corrected GWAS z-scores on the cis-eQTLs z-scores for all genes and tissues (79). Full method details of CaMMEL implementation and benchmarking against TWAS methods with the systematic analysis of GWAS traits and cis-eQTLs from 49 tissues (GTEx datasets), can be found as follows:

- Source code can be found at [https://github.com/YPARK/GTEx\\_mediation](https://github.com/YPARK/GTEx_mediation)
- Results at [https://github.com/YPARK/GTEx\\_mediation/tree/master/docs/share/20190701](https://github.com/YPARK/GTEx_mediation/tree/master/docs/share/20190701)

- **Multi-tissue PRS and mediators of complex metabolic traits**

Systems genetics implementation of CaMMEL by Park et al. revealed a wide range of dependency combinations that exploit the conjoint dissection of multi-resolution data through one statistical framework. This stepwise prioritization revealed genotype--eQTL-gene--phenotype associations between GWAS traits, 49 GTEx and 17608 protein coding genes. CaMMEL identification of causal relationships reveals the pleiotropic and the pervasive nature of dependencies in complex traits. To study multi-trait, multi-tissue and multi-gene dependencies, and their global and local associations with complex metabolic traits, we performed 3 complementary approaches: network integration, hierarchical prioritization of topological based enrichments (cumulative-PGEs), and hierarchical prioritization of topological enrichments with integration of chromatin structural data.

- **Hierarchical prioritization of BMI and T2D PRS-mediation topological enrichments**

To investigate topological hubs representing territorial models of transcriptional activity shared across PRS-mediation tissues from metabolic traits, we focused on GIANT-BMI and T2D-BMI-adjusted. We downloaded GTEx mediation data for these traits across multiple tissues and filtered genes with a log odds ratio cutoff of 2 per trait-tissue and a TWAS adjusted z-score > 1. Of all the mediators in each trait within the afore-mentioned constraints, we quantified the proportion of unique genes in tissues defined as metabolic and the rest of the tissues. We compared the tissue contribution per trait (BMI and T2D) for the same metabolic tissues. We then performed positional gene enrichment of genesets derived from PRS-mediation. Multi-tissue PRS-mediation genes were classified as BMI or T2D causal mediators. These gene sets were used as inputs for PGE-analysis. We then displayed PGE hubs as the proportion of total PGE by dividing the number of regulated genes in topological enrichment by the total number of regulated genes. We carried out this approach in every tissue per trait, and compared them with PGEs formed from random sets (same distribution of number of genes building PGEs of interest). To identify overlapping or cross-tissue hubs within each trait, we combined gene sets per group and evaluated combinatorial PGE-hubs (see functional embedding of positional gene enrichments). We then assessed optimal-PGEs, which were defined as PGEs with more than 5 genes (see transcriptional information uncertainty with Shannon entropy and mice phenotype networks). We compared the proportion of PGEs with >5 genes / total regions, in combined sets by trait, individual by tissue and combined random permutations. To discriminate between the average of expected proportions for individual and random combinations, with the proportions from biological combinations we used chi-square test and p-value estimation. We also estimated the dependencies of the number of genes in optimal-PGEs in multi-tissue combinations versus random permutations, by assessing the linear regression with p-value enrichments between traits and random expectation. We plotted representative optimal-PGEs in each chromosome by group (genomic cluster range defined by start and end of chromosomal PGE hub region) and defined colocalized regions of activity. In addition, we calculated the proportion of unique PGEs in each chromosome, by quantifying the total number of PGEs irrespective of trait and the ratio contribution per PRS-mediation. Metabolic-HAR domains and experimentally validated HAR-regions in human cellular models were highlighted, and used for downstream analysis.

- **Structural prioritization of topological enrichments within trans-interacting chromatin regions**

Transcriptional factories of co-regulated genes are usually located in or in close proximity to homotypic and heterotypic interacting chromosome regions (44, 85). These models suggest that interacting chromatin regions serve as anchors that increase the likelihood of trans- co-regulation (85). This has been further supported by observations where non-random spatial proximity of adjacent chromosomes determine their DNA translocation (175). In addition, contiguous regions or regions within interacting chromatin hubs were found to be regulated by transcriptional factories or hotspots (85)(86). These interactions can, in a tissue- or cell-specific way, be more active from a wide range of conserved global interacting locations to define cell-specific factories (86). These events indicate a high-degree of plasticity, likely depending on cell-specific heterotypic transcriptional regulators (85). This is supported by our observations on spatially defined positional gene enrichments from high-throughput profiling (Fig 4 and Fig 6), and by the degree of higher-order conservation of structural data including TAD boundaries, frequently interacting chromatin regions (FIREs) (129) and trans-interacting regions (86, 126). These observations also suggest a high degree of conservation of trans-interacting domains across diverse tissues and cells. The dosage and usage of the conserved trans-architecture likely depend on cell-specific transcriptionally active regions conditioning the hierarchical combination of interactions, and perhaps the nonrandom location of nuclear speckles within transcriptional factories (176). Supporting this, transcriptional factories have been found to be located in proximity to trans-domains, perhaps as they allow less space constraints compared to intrinsic TAD-mediated forces of homotypic chromosomal folding (85) (44). Indeed, genome-wide crosslinked methods simultaneously mapping higher-order interacting regions have revealed that physical limitations within the nucleus constrained these domains within rather large interacting hub regions (87)(5). In order to build a map of all possible combinations, we used a naive approach to investigate trans-interacting chromatin regions. This global trans-network was exploited to generate hierarchical maps of tissue and cell-specific territorial activity suited for experimental validation. Then, to investigate the structural dependencies of transcriptional hub models (using for example the PRS-mediation cross-tissue cumulative location), we defined higher-order trans-interacting regions using Hi-C data from representative cell types, hESC (see Hi-C data analysis and trans-interacting regions in hESC), and data from SPRITE (split-pool recognition of interactions by tag extension) a method based on tagging crosslinked protein-DNA-RNA complexes followed by genome-wide detection of higher-order interaction hubs in the nucleus (87). We then used a two tier approach in which diversity of both data and algorithmic analysis is exploited to define large trans-domains. This is followed by manual curation across representative cells with a likelihood-based integration within curated genomic windows, as previously shown (177). We then tuned and cross-validated these recovered interactions using discovered hubs by SPRITE. First, to define Hi-C based domains we used previously described procedures with hESC (see trans-interacting regions in hESC). We also implemented the approach developed by Belyaeva et al. (86) using a large average submatrix algorithm (LAS) with a binary classifier to identify trans-domains in human lung fibroblasts (IMR90) (available at [https://github.com/anastasiyabel/functional\\_chromosome\\_interactions](https://github.com/anastasiyabel/functional_chromosome_interactions)). For those interacting domain peaks, we calculated in each chromosome the genomic-range average from the genomic ranges identified by both algorithmic implementations. Following these results, we then partitioned chromosome regions by 5-Mb genomic-windows and proceeded to manually curate the presence of trans-domains using HiC data from other representative cell types (see HiC data analysis). For each chromosome, we labelled genomic windows as interacting if they display trans-domain bins from the HiC algorithmic comparison analysis. We then proceeded to the manual curation of heterotypic interaction on analysed HiC data as follows. For each biological cell type, we used a naive binary approach in which positive genomic-windows with trans-domains were labeled as 1 in each chromosome if heterotypic interaction was found with every other chromosome (absence was labeled as 0). We then calculated the likelihood of trans-domain presence in each genomic-window by dividing the sum of interactions with the total number of possible interactions. Taking into account the results from each representative cell type, we obtained a global likelihood score of trans-domain presence by averaging the likelihood scores from the other biological cell types. This final score was used to cross-validate genomic-windows with analysed interacting hubs from SPRITE datasets and with the first step of trans-domain HiC interrogation. This cross-validation was used to tune and filter the nodes that overlapped with trans-domain regions discovered by SPRITE-hubs and algorithmic trans-domains. In brief, using the genomic-range windows as nodes and the presence of interaction as edges we built an integrative trans-domain network. As previously described, we performed network-based prioritization of influencer and critical connector nodes. We used this network as all possible combinations of global genome trans-architecture for downstream modeling of overlapping disease hotspots, metabolic-HAR regions, PGC1A-network regulatory regions, and cell-specific transcriptional activity. We focused on trans-interacting domains with overlapping disease hubs from the PRS-mediation approach for metabolic traits. We then filtered a more stringent trans-interacting network with only metabolic-disease hubs. These hubs were composed of shared BMI-T2D PRS-mediation PGEs locations. This graph representation of trans-architectural metabolic disease was used for downstream analysis. To assess the degree of overlapping from other biologically derived gene sets versus random expectation (randomly generated gene sets), we used a previously described approach. In brief, overlapping between PGEs locations (from different datasets) onto our trans-interacting network is quantified and compared with PGEs location generated from random gene sets. Overlapping was defined as PGEs within trans-regions or with at least 50% overlap. The mean value of the number of regions overlapped by random PGEs composes the random expectation number. This might differ in between experiments as every biologically derived gene set is composed of different gene numbers. For this reason, each comparison stands alone with their own random permutations from gene sets forming a similar number of PGEs. For each comparison made,

we calculated chi-square tests followed by p-value assessment. To assess the interpolation of loci positions from GWAS studies with our disease-hub trans-domain network, we used GWAS data for metabolic traits stored and analysed in <https://t2d.hugeamp.org/>. There we obtained GWAS association for BMI from GIANT UK biobank study (178) and for T2D BMI-adjusted from DIAMANTE study (179). We then assessed the overlap between GWAS locus with our metabolic disease hubs trans-domain network. The shortest distance path between superimposed loci intra-trait and inter-trait was estimated in our network, and topological distances were compared for dependencies.

- **Topological single-cell decomposition of structural metabolic disease hubs**

To decomposed topologically metabolic disease hubs (PGEs) overlapping trans-interacting domain network (see previous section) into fat-specific disease hubs and fat-specific single-cell PGEs from human obese subjects, we first downloaded and re-analysed published single cell RNAseq data from obese subjects (90) (see adipose tissue scRNAseq from obese human subjects). We then obtained differential expressed genes in cell-specific clusters, followed by PGEs assessments per cell type (see functional embedding of positional gene enrichments). In parallel, we obtained PRS-mediation results specific for adipose tissue (see PRS-mediation approach) and compared their spatial location with cumulative multi-tissue mediation hubs. From this, we assessed the overlap enrichment of fat-specific disease hubs within the trans-interacting domain network (see previous section). Our previous observations showed metabolic-HAR domains are enriched across chromatin interacting regions, PGE-hubs act as territorial domains of transcriptional activity, and heterotypic condensates are tuned dynamically to assemble nuclear compartments that exploit genome architecture and function. This is in line with observations on transcriptional factories assembled in proximity of interacting hubs and tuned cell-specifically with the dynamic activity of heterotypic transcriptional regulators (44). In the previous section, we argue that global trans-architecture could be used to infer the proximity of transcriptional factories and the cell-specific location of interacting hubs with transcriptionally active territories (see structural prioritization of topological enrichments within trans-interacting chromatin regions). Considering this, we obtained scPGEs from differentially expressed genes (see single cell analysis) and evaluated cell-specific domains in the trans-interacting disease hub network. We defined PGE overlap within each domain as previously described and highlighted cell-specific trans-interacting regions with transcriptional activity. This revealed cell-specific domains, domains with double cell-specificity, and promiscuous domains. Hierarchies of active trans-interacting domains per cell and cell-related were obtained by extracting overlapping scPGE and trans-domains. We used this decomposed network to filter regions for experimental validation of cell-specific chromatin contacts by ChiP-loop and transcriptional activity by gene expression. We evaluated APC-specific contacts, transcriptional activity and their interaction with metabolic-HAR domains mediating fasting endurance (ATG4C- and DGAT-domains). We compared the specificity and regulation of these transcriptional active interactions with neighboring and interacting regions during fasting and with metabolic overload. These revealed APC-ectopic inflammation-driven regulation of trans-regions (enriched for macrophage type 1 and monocytes) during fasting with metabolic overload.

## **ITG-VI: Functional impact of radial chromatin organization**

- **Integration of GPSeq data for radial chromatin organization**

Genome architecture determines proper function of genome plasticity. Chromosomes are indeed non-randomly positioned in the nucleus with specific territorial domains with respect to the nuclear lamina (180)(80). Several studies have revealed structural associations of radial chromatin organization with gene density (180), while genome-wide approaches have provided high resolution maps (100kb) of genome locations along the radial axis (80). However, functional interpretation of radiality with ensemble multi-resolution models aimed at understanding complex disease is still missing. Therefore, we have integrated previously published data derived from genome loci positioning by sequencing (GPSeq) method, which gives relative distances between genomic-loci and the nuclear lamina (80) (GEO:GSE135882; GPSeq protocol available at <https://doi.org/10.21203/rs.3.pex-570/v1>; data available at <https://github.com/ggirelli/GPSeq-source-data>). Briefly, radially directed enzymatic restriction of fragmented genomic DNA is paired with sequencing. To follow enzymatic radial diffusion with microscopy, Girelli et al. used a modified in situ hybridization in which adapters ligated to the restricted fragments are detected by fluorescently labeled oligos. The adapters guided next-generation sequencing, library generation and radial estimates were inferred by comparing radial locations with 3D DNA FISH methods (181). For genome-wide GPSeq score calculation the authors employed custom pipelines. In brief, after quality control of the sequencing data, the genome was binned in 1-Mb overlapping windows in steps of 100-kb, followed by calculation of digestion probability considering all restriction sites cut. GPSeq score from different experiments was generated by averaging the score in each window followed by log2 scaling (available at <https://github.com/ggirelli/gpseqc>). We then obtained genomic-loci radial scores at 100-kb resolution, which was used to superimpose previously generated PGEs. First, we divided each genomic position in ranges of radial probabilities by peripheral, central and core genome-bins. This was carried out by estimating the average GPSeq score of core-located chromosome 19, and peripherally located chromosome 4. *GPSeq-score in genomic-positions used in this study:* We plotted these 3 categories and assessed the overlap with PGEs derived from either BMI or T2D polygenic-risk score mediation hubs (see PRS-mediation PGEs for metabolic traits). To quantify the PGEs classification, we considered the average of the GPSeq-score for the given

genomic-range. To assess the radial proportion, we took into account all PGEs within each group and estimated the proportion of radial location following the range of radial probabilities. We also approximate random expectation likelihood of radial enrichments by assessing the radial proportions for PGEs derived from random gene sets (30% peripheral and 15% core). Similarly, we estimated the radial proportion location of active enhancers by VISTA, metabolic-HAR domains, frequently interacting regions (FIREs), and parameters derived from our network-based analysis of interactions including social and isolated regulatory regions.

- **TFBS-radial distribution and transcriptional PPI networks**

We used GPSeq-mediated transcription factor binding sites (TFBS) radial enrichments as previously described (80). In brief, predicted TFBS were obtained as previously described (182) and their appearances quantified within the genomic windows. GC-content in each TFBS was quantified. Correlation coefficient between the number of TFBS and GPSeq-score for each window computed against the GC TFBS content. For downstream analysis, we used transcriptional regulators with TFBS-radial peripheral or core correlation. We next assessed in each group their propensity to form heterogeneous physical interactions. To this end, we used protein-protein interaction data from the HuRi repository (see ITGs for transcriptional dependencies) (104), followed by network analysis of interaction parameters. We quantified the proportion of interacting transcriptional regulators in each group by dividing the number of interacting factors by the total of regulators per group. Given the differences in interaction probabilities, and our previous observations on co-activators being hub-assemblers of transcription, we quantified the number of co-activators and their interacting transcription factors within each radially enriched group. This showed co-activator-related radial enrichment in core factors. For each of the networks formed, either core-TR or peripheral-TR network, we assessed their functional embedding with reactome and gene ontology network enrichments (p-value < 0.01). This revealed interacting heterogeneity of radially enriched regulators and divergence of functional annotation.

- **Context-dependent phase separation score for radially enriched regulators**

We downloaded FASTA amino acid sequences from transcriptional regulators with radial enrichments. We then used them as input to localCIDER and obtained the ensemble parameters of phase separation and its conformational dependencies. We used the globule conformation classifiers to quantify the proportion of context-dependent phase separation among core or peripheral transcriptional regulators. We next compared both groups using the average value in all parameters obtained, including segregation of charge residues ( $\kappa$ ) and charge-related variations ( $f^-$ ,  $f^+$ , NCPR, FCR, hydrophathy). We built a correlation matrix to study their associations within each group.

- **Radiality of transcriptionally active loci in single cells by DNA-MERFISH**

To evaluate radial positions of transcriptional active loci in single cells and their structural relationships, we used datasets from established genome-scale approaches that use multiplexed FISH imaging technology (77) (available at <https://zenodo.org/record/3928890#.YDgeiZKiuV>). Briefly, this approach is based on systematic integration of sequential rounds of genomic-loci imaging using in situ hybridization (183) at single cell resolution (77). This is coupled to a combinatorial strategy previously described as multiplexed error-robust FISH (MERFISH) where imaging combination rounds determine chromatin loci identities (184). They coined DNA-MERFISH technique to specific modifications on combinatorial rounds following observations on chromosome territorial location and polymer models. It assigns to each genomic loci a 100-bit barcode, which determines the signal across imaging rounds. To detect the barcodes imprinted in genomic loci the authors used sequential hybridization of labeled probes. With this method, a comprehensive resource of nuclear architecture was generated. This displays single cell resolution of more than 1000 genomic loci information containing both nuclear structural relationships and nascent RNA transcriptional burst activity (77). This systematic method images each 30-kb genomic window size for each chromosome, in IMR-90 cells. Multi-resolution Information is then available from around 1000 genomic loci in > 5000 cells, including nascent transcripts from genes located within these loci and their distances to nuclear structures (e.g. nucleoli, nuclear speckles and nuclear lamina). Data processing and script functions can be found at [https://github.com/ZhuangLab/Chromatin\\_Analysis\\_2020\\_cell](https://github.com/ZhuangLab/Chromatin_Analysis_2020_cell). We used genome-scale 3D position of the measured genomic loci to assess distance to lamina of specific metabolic-HAR PGEs, and other domains composing reconstructed nuclear compartment for lipid metabolism. In addition, we used genome-scale transcription of genes contiguous to the loci imaged by DNA-MERFISH. This extended multimodal approach allows the quantification of transcriptional burst activity and their structural relationships. We used this dataset to obtain transcriptional burst profiles of genes with radial enrichments. We used the structural parameter “distance to nuclear lamina” as an estimate of radiality, and partitioned genomic loci by transcriptional burst on or off. For those genomic-loci with different structural associations (nucleoli, nuclear speckles and lamina), we quantified their mean distance to lamina when they are transcriptionally active on or silent off. We compared these mean distances and estimated their p-value. We next evaluated those genomic-loci with genes with bursting activity on, the correlation between their mean distance to nuclear lamina and the proportion of genes actively transcribed. This revealed that the larger the distance from the nuclear lamina the larger the amount of genes with active transcription. Finally, we performed functional annotation of genes with burst on and located at the core of the nucleus. This revealed cell-fate, differentiation, cell-adhesion genes are transcribed actively at the single cell level despite heterogeneity in local structural measurements (77).

## Human brain Single nuclei RNA sequencing data analysis

We downloaded snRNAseq data (gene expression and metadata) from 24 human brain control samples (76). We re-analyzed the data using SEURAT (185) and performed cluster-based separation of different subtypes, followed by identification of cell types based on the canonical marker genes and the annotation provided in the paper. The differentially expressed genes for each cell type were identified by FinderMarker function in Seurat package. These genes were used for further analysis in this study.

## Transcriptomic profile of human adipose precursor cells

To compare different human adipocyte progenitor cells (APC) we used previously defined subtypes: APC beige-like, APC low lipid turnover and APC high lipid turnover (75). These categories represent different phenotypic behavior such as being ability and lipid turnover activity (75), and were defined using FACS analysis, transcriptomic, proteomic and metabolic profiling. We therefore downloaded datasets from this study from GEO: GSE129042, which corresponds to expression profiling by high throughput sequencing of FACS-sorted APCs from adipose stromal vascular fraction of human subjects. Counts were processed and differentially expressed genes determined using the DEGSeq package in R as previously described (186) For differentially expressed genes in each APC subtype we performed functional annotation and aggregate topological expression within fPGEs composing our reconstructed nuclear compartments (see aggregate topological expression).

## Adipose tissue scRNAseq from obese human subjects

We downloaded the snRNAseq data (gene expression and metadata) from > 20 samples from adipose tissue biopsies of obese human subjects (90). We re-analyzed the data using Seurat (185) and separated cell types based on the canonical marker genes provided in the paper. The differentially expressed genes for each cell type were then identified by FinderMarker function in Seurat package. These genes were used for further analysis in this study.

## Adipose tissue scRNAseq from mice

We downloaded the scRNAseq data (gene expression and metadata) from different tissues of the tabula muris consortium (78). We re-analyzed data from tissues of interest using seurat (185) and separated cell types based on the canonical marker genes provided in the paper. The differentially expressed genes for each cell type were then identified by FinderMarker function in Seurat package. These genes were used for further analysis in this study.

## Quantification and Statistical Analysis

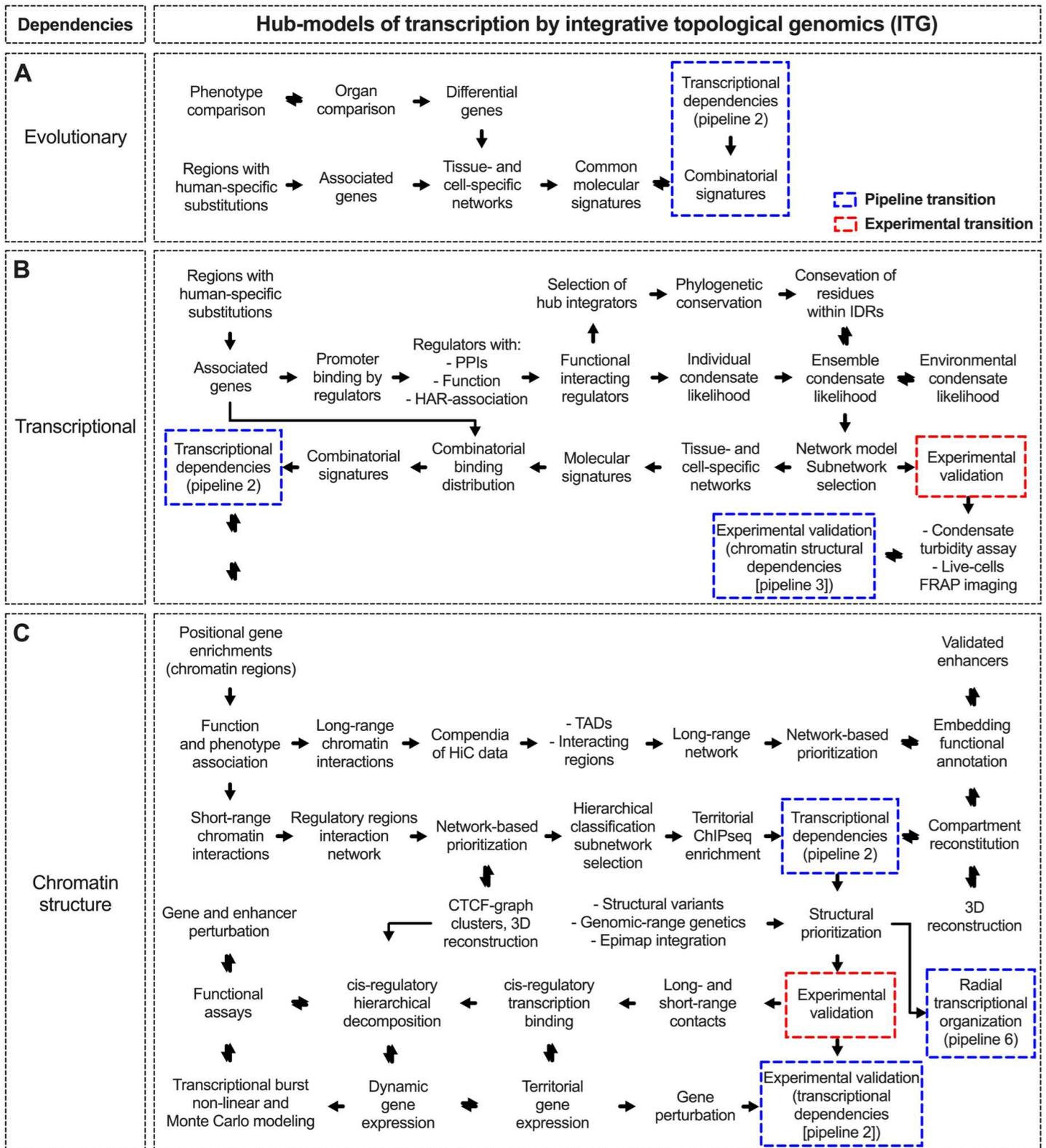
Statistical analyses related to experimental procedures were performed using GraphPad Prism Software (San Diego, CA) and all parameters are indicated in the corresponding figure legend. Quantitative data are presented as the mean  $\pm$  SEM and n is indicated for each experiment. All experiments were carried out with 3 biological replicates and 3 independent times. Animal experiments were performed in more than 3 animals per group. Unpaired Student's t test was used to determine statistical significance when two groups were compared. One-way ANOVA followed by Fisher's least significant difference (LSD) test for post hoc comparisons was used to determine statistical significance when multiple groups were compared. Two-way ANOVA was used to estimate significance between groups constrained by time-based measurements, followed by Tukey test for post-hoc multiple comparisons. Correlations were calculated by Spearman rank and Pearson correlation with a two-tailed test. Statistical significance was defined as p value < 0.05 by either test and is denoted with an asterisk as follows: \* p-value 0.05-0.01, \*\* p-value 0.01-0.001 and \*\*\* < 0.001 when indicated.

## Data and Code Availability

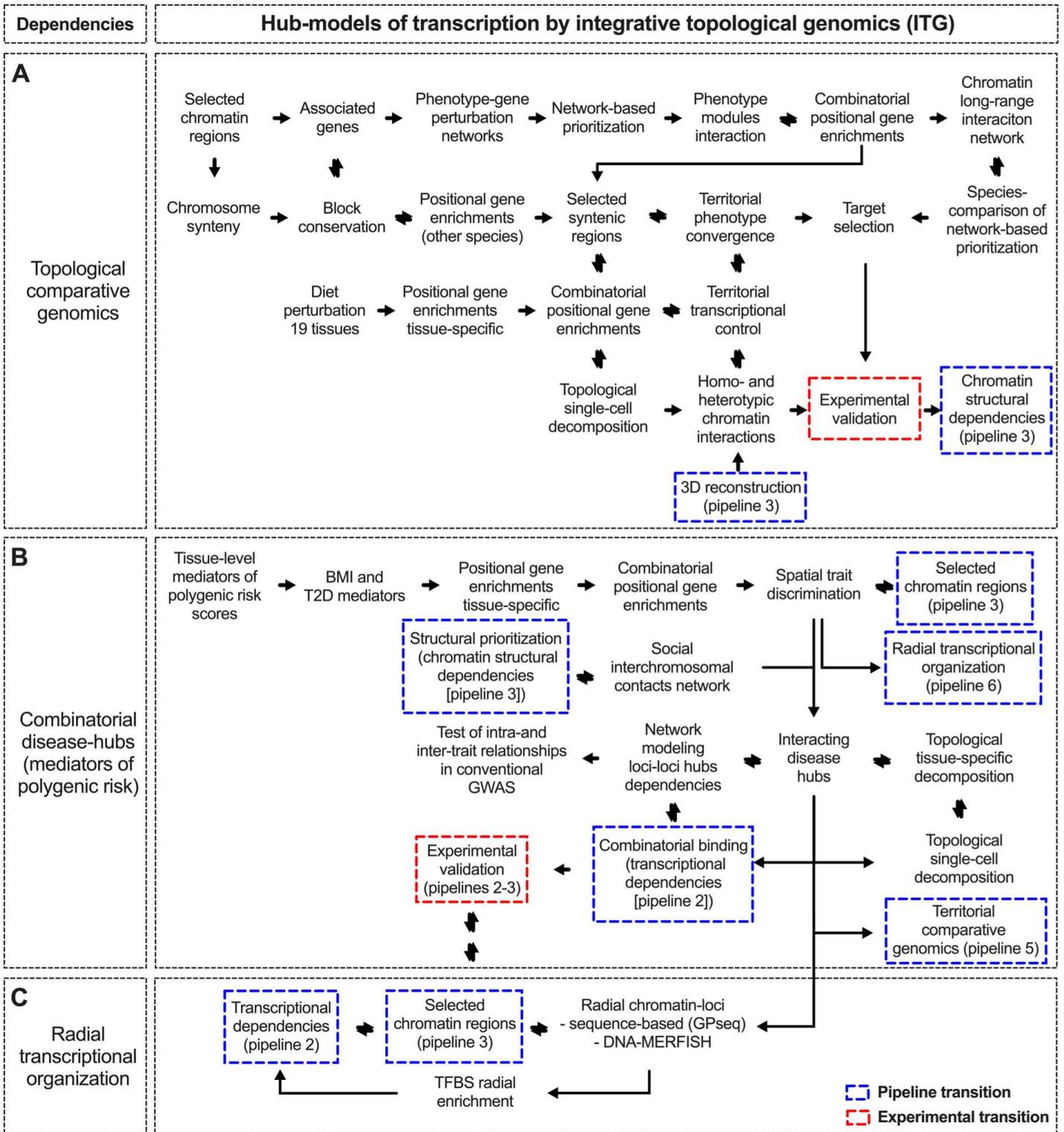
The analysis was performed in R (3.5 and 3.6) and python (2.7, 3.5 and 3.7). Libraries for data analysis include numpy, pandas, scikit-learn, scanpy and scipy. The datasets reported in this paper are available in supplementary materials. Analysis software will be released upon publication at <https://github.com/leandroagudelo189>. When indicated, network analysis and visualization were done using the python library NetworkX, and the software GEPHI. Functions to evaluate networks have been created and validated by the Ideker Lab (archived in [https://github.com/idekerlab/Network\\_Evaluation\\_Tools](https://github.com/idekerlab/Network_Evaluation_Tools)) and are stored on the network data exchange (NDEx) (187). Other functions to evaluate network topology can be found as inbuilt analysis tools in GEPHI. Tissue-specific networks for functional annotation of topological dependencies can be found at <http://giant.princeton.edu> and <https://hb.flatironinstitute.org/>. Positional gene enrichment analysis tool can be found at <http://silico.biotoul.fr/pge/>. Functional annotation tool from a large library of biological repositories can be found at <https://maayanlab.cloud/Enrichr/>. DAVID functional clustering analysis tools can be found at <https://david.ncifcrf.gov/>. To filter out PGE regions without cytogenetic band enrichment, we used a curated list of cytogenetic band-gene dependencies archived in <http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>. Prediction of TFBS from DNA sequences can be found at <http://alggen.lsi.upc.es/>. Prediction of conserved TFBS in promoter regions of genes of interest can be found at <http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>. Tissue-naïve and tissue-specific protein-protein interactions used for n+1 network reconstruction can be found at <http://www.interactome-atlas.org/>. Tissue-naïve protein-protein interactions

used for full network reconstruction can be found at <https://thebiogrid.org/>. In silico prediction of intrinsically disordered regions in amino acid sequences can be found at <http://www.pondr.com/>, <https://iupred3.elte.hu/>, <https://mobidb.bio.unipd.it/>. Curated experimentally validated phase separation database stored at <http://db.phasep.pro/>. Context-dependent phase separation score from ensemble parameters of conformation changes present in disordered proteins. They offered discrepancies of conformations that are related to the local environment such as salt, solvation, multi-molecular interactions and ligand presence. These can be found at <http://pappulab.wustl.edu/CIDER/about/> and can be implemented locally following instructions stored at <http://pappulab.github.io/localCIDER/>. Phylogenetic analysis tool information for protein residues conservation score and 3D modelling of residue exposure can be found at <https://consurf.tau.ac.il/>. Evolutionary domain conservation of protein domains with known structure can be found at <https://consurfdb.tau.ac.il/>. CHIP-atlas methodology, documentation and functions for analysis can be found at <https://github.com/inutano/chip-atlas>. Compendium of human chromatin contact maps has been created by the Ren Lab and is stored at GEO: GSE87112 (schmitt, bing ren 2016). GREAT software, available at <http://great.stanford.edu/public/html/index.php>, was used to evaluate biological process enrichment for neighboring genes compared to genomic background. Cis-regulatory elements relationship were obtained from GeneHancer, and can be found at <http://www.genecards.org/>. VISTA enhancer elements can be found at [https://enhancer.lbl.gov/frnt\\_page\\_n.shtml](https://enhancer.lbl.gov/frnt_page_n.shtml). HACER, human active enhancer to interpret regulatory variants database is archived in <http://bioinfo.vanderbilt.edu/AE/HACER/index.html>. A curated database of cell-specific active regulatory regions can be found at <http://www.enhanceratlas.org/index.php>. Global chromatin 3D reconstruction tool and code can be found as follows: GSDB at <http://sysbio.met.missouri.edu/3dgenome/GSDB/>; and Spacewalk at <http://3dg.io/spacewalk/>. 3D-multiscale Monte Carlo modeling code for global and local CTCF-mediated ChIA-PET sequencing can be found at [https://bitbucket.org/3dome/3dome\\_mmc/src/master/](https://bitbucket.org/3dome/3dome_mmc/src/master/). Code for integration of structural variants (SVs) to spatial chromatin interaction patterns can be found at [https://bitbucket.org/4dnucleome/spatial\\_chromatin\\_architecture](https://bitbucket.org/4dnucleome/spatial_chromatin_architecture). Data by Sadowski et al. on SVs spatial chromatin effects can be found at <https://zenodo.org/record/2837248#.YCQBvZNKhPM>. CTCF-binding site prediction tool can be found at <https://insulatordb.uthsc.edu/>. Structural variants data from the reference resource compendia developed by Collins et al. (55) is archived in <https://gnomad.broadinstitute.org/>. To isolate metabolic disease associations for a given genomic-range, we used the curated common metabolic disease portal, archived in <https://hugeamp.org/>. Maps of epigenomic integration of Encode and Roadmap consortium data, EpiMap can be found at <http://compbio.mit.edu/epimap/>. For topological comparative genomics and to evaluate chromosomal synteny, tools can be found at the synteny portal, archived in [http://bioinfo.konkuk.ac.kr/syntenyn\\_portal/](http://bioinfo.konkuk.ac.kr/syntenyn_portal/). CAMMEL mediation analysis and polygenic risk scores can be found at [https://github.com/YPARK/GTEEx\\_mediation](https://github.com/YPARK/GTEEx_mediation) and <https://github.com/YPARK/PRS>. We integrated in our pipelines previously published data derived from radial genome loci positioning by sequencing (GPSeq) method, which gives relative distances between genomic-loci and the nuclear lamina (80) (GEO:GSE135882; GPSeq protocol available at <https://doi.org/10.21203/rs.3.pex-570/v1>; data available at <https://github.com/ggirelli/GPSeq-source-data>). To evaluate radial positions of transcriptional active loci in single cells and their structural relationships, we used datasets from established genome-scale approaches that use multiplexed FISH imaging technology (77) (available at <https://zenodo.org/record/3928890#.YDgeiZNKiuV> and script functions can be found at [https://github.com/ZhuangLab/Chromatin\\_Analysis\\_2020\\_cell](https://github.com/ZhuangLab/Chromatin_Analysis_2020_cell))

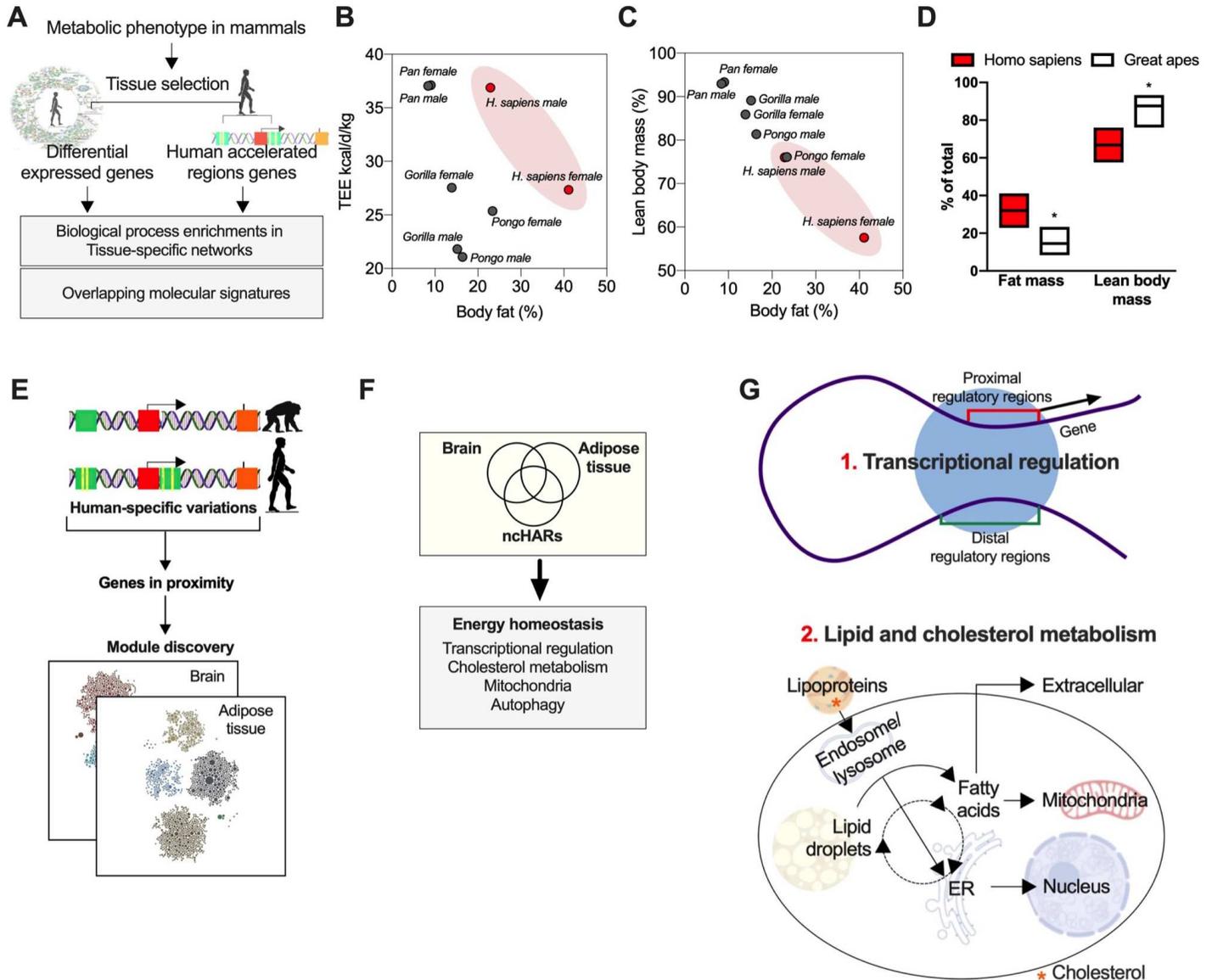
## Supplementary Figures



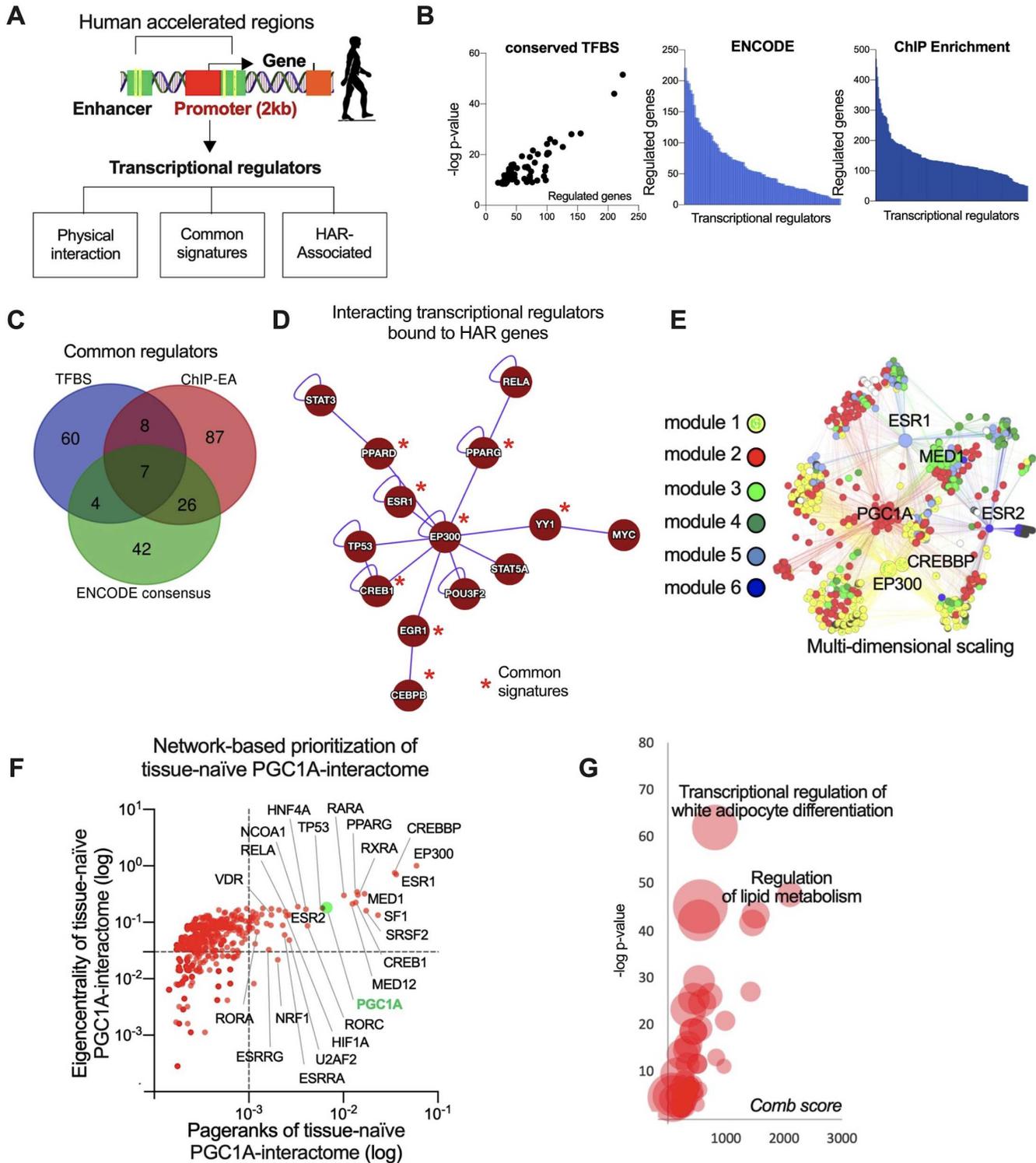
**Fig. S1. Integrative topological genomics overview: Part 1.** (A) Representation of pipeline steps for evolutionary relationships. (B) Representation of pipeline steps for transcriptional relationships. (C) Representation of pipeline steps for chromatin structural relationships



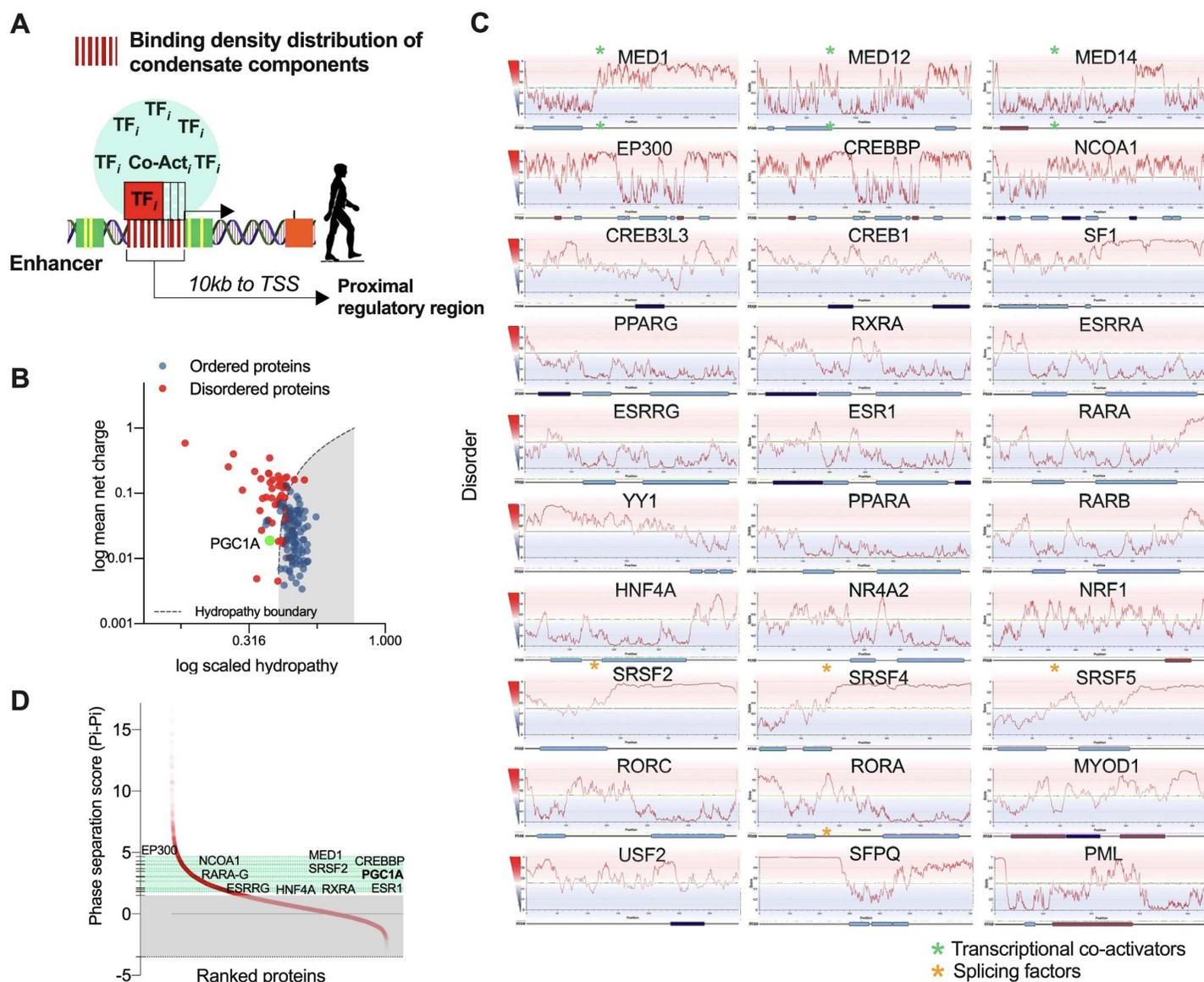
**Fig. S2. Integrative topological genomics overview: Part 2.** (A) Representation of pipeline steps for topological comparative genomics relationships. (B) Representation of pipeline steps for disease mediators of polygenic risk relationships. (C) Representation of pipeline steps for radial chromatin relationships



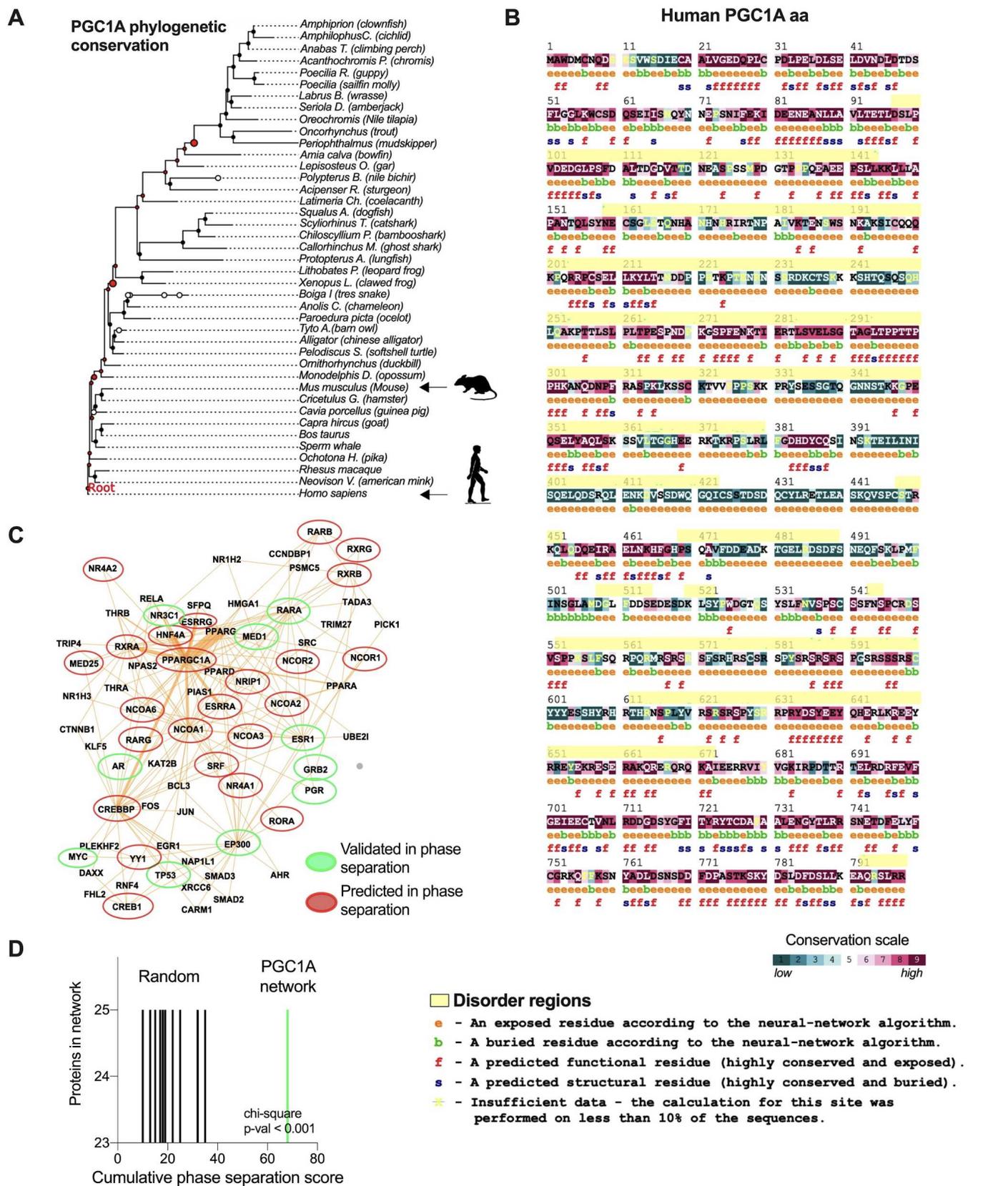
**Fig. S3. Phenotypic and genetic human accelerated metabolism** (related to Fig. 1). **(A)** Schematic for network-based integration of phenotype and genetic human enriched biological enrichments (part of evolutionary relationships pipeline). **(B)** Body fat composition in higher primates and relationship to total-energy expenditure and lean body mass (23). **(C)** Comparison of fat mass and lean body mass between homo sapiens and great apes (23). **(D)** Body fat composition in higher primates and relationship to total-energy expenditure and lean body mass. **(E)** Illustration showing non-coding human accelerated regions and genes in proximity followed by network based functional enrichments (part of evolutionary relationships pipeline). **(F)** Overlap of network based functional enrichments between molecular signatures increased in humans compared to primates and genes in proximity to HARs. **(G)** Schematic illustration of molecular signatures observed in **(F)** and in **Fig. 1C**. Bars show mean values. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.



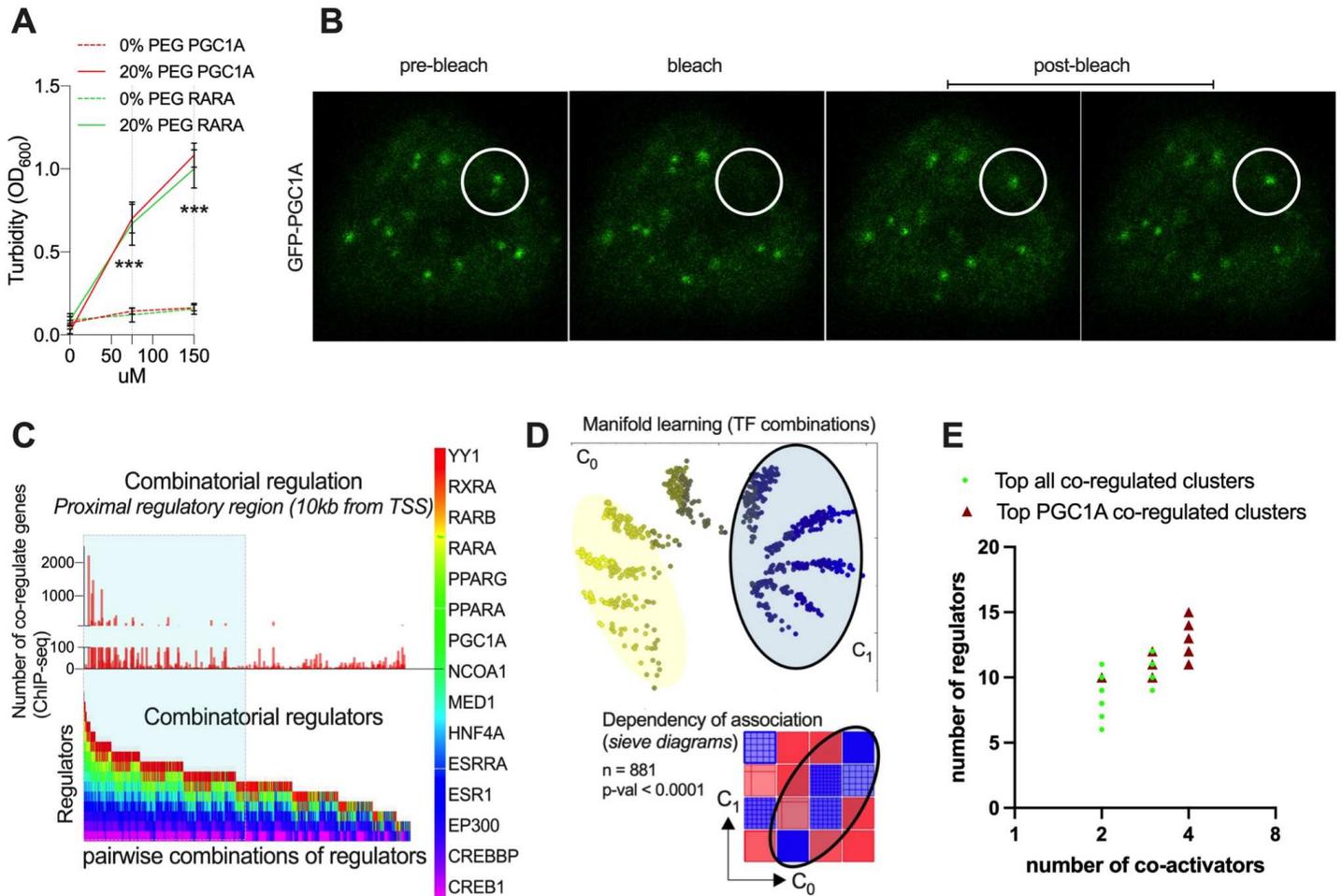
**Fig. S4. Transcriptional dependencies of human accelerated regions genes (related to Fig. 1).** (A) Illustration showing genes in proximity of HARs and interrogation of transcriptional dependencies on their promoter regions (part of transcriptional relationships pipeline). Bound regulators with physical interactions, similar function and HAR-association. (B) Binding dependencies of transcriptional regulators of nCHAR-genes by TFBS-conservation, ENCODE ChIP-seq binding, and ChIP-seq atlas binding enrichment. (C) Venn diagram of common regulators from (D). Network reconstitution of common regulators of HAR genes displaying physical interactions and similar function. (E) Full reconstitution of interactors of common regulators followed by network analysis and representation using multi-dimensional scaling. (F) Network-based prioritization of interacting regulators by eigencentrality and pagerank. (G) Functional enrichment of top nodes from (F).



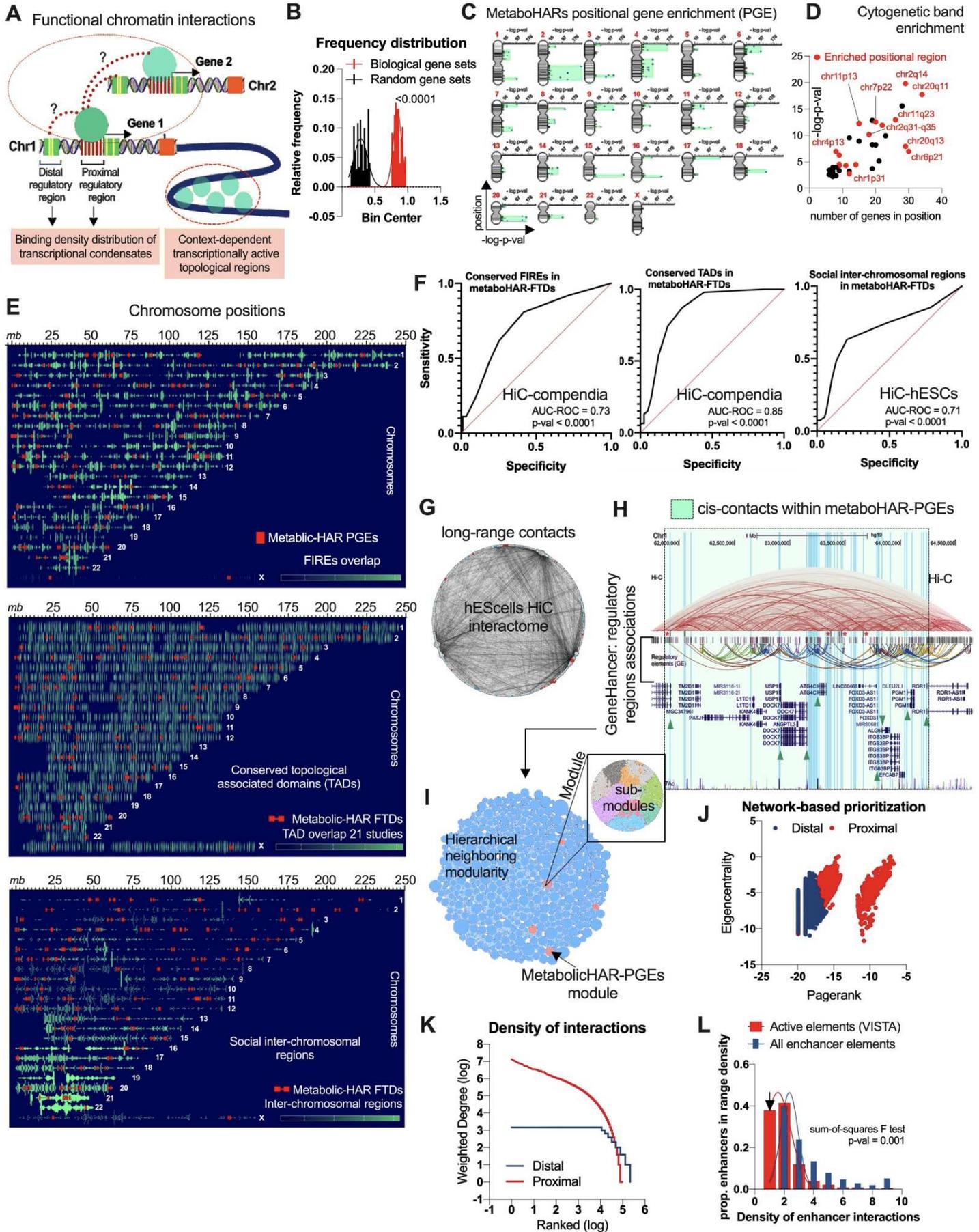
**Fig. S5. Disorder scores for interacting transcriptional regulators of HAR genes** (related to Fig. 1). **(A)** Illustration on regulatory binding density distribution for HAR-genes promoter regions (10kb) (part of transcriptional relationships pipeline). **(B)** net charge and scaled hydropathy of hub integrator the co-activator PGC1A, along with ordered and disordered regions from PONDR software. **(C)** Disorder scores for prioritized nodes from fully reconstituted interacting regulators network. **(D)** Genome-wide phase separation scores based on highly frequent pi-interactions in small amino acids with exposed peptide backbone (111) Highlighted in green range score for interacting transcriptional regulators of HAR genes.



**Fig. S6. Phylogenetic conservation of hub integrator PGC1A co-activator and cumulative phase separation scores of its network (related to Fig. 1).** (A) PGC1A amino acid residues phylogenetic conservation using CONSURF. (B) Conservation scores in PGC1A amino acids from phylogenetic analysis. In yellow consensus disordered regions. (C) Brain and adipose tissue backbone PPI network, displaying interacting PGC1A regulatory network, and proteins undergoing phase separation validated experimentally in green and predicted in silico in red. (D) Cumulative phase separation score for PGC1A-interacting network.

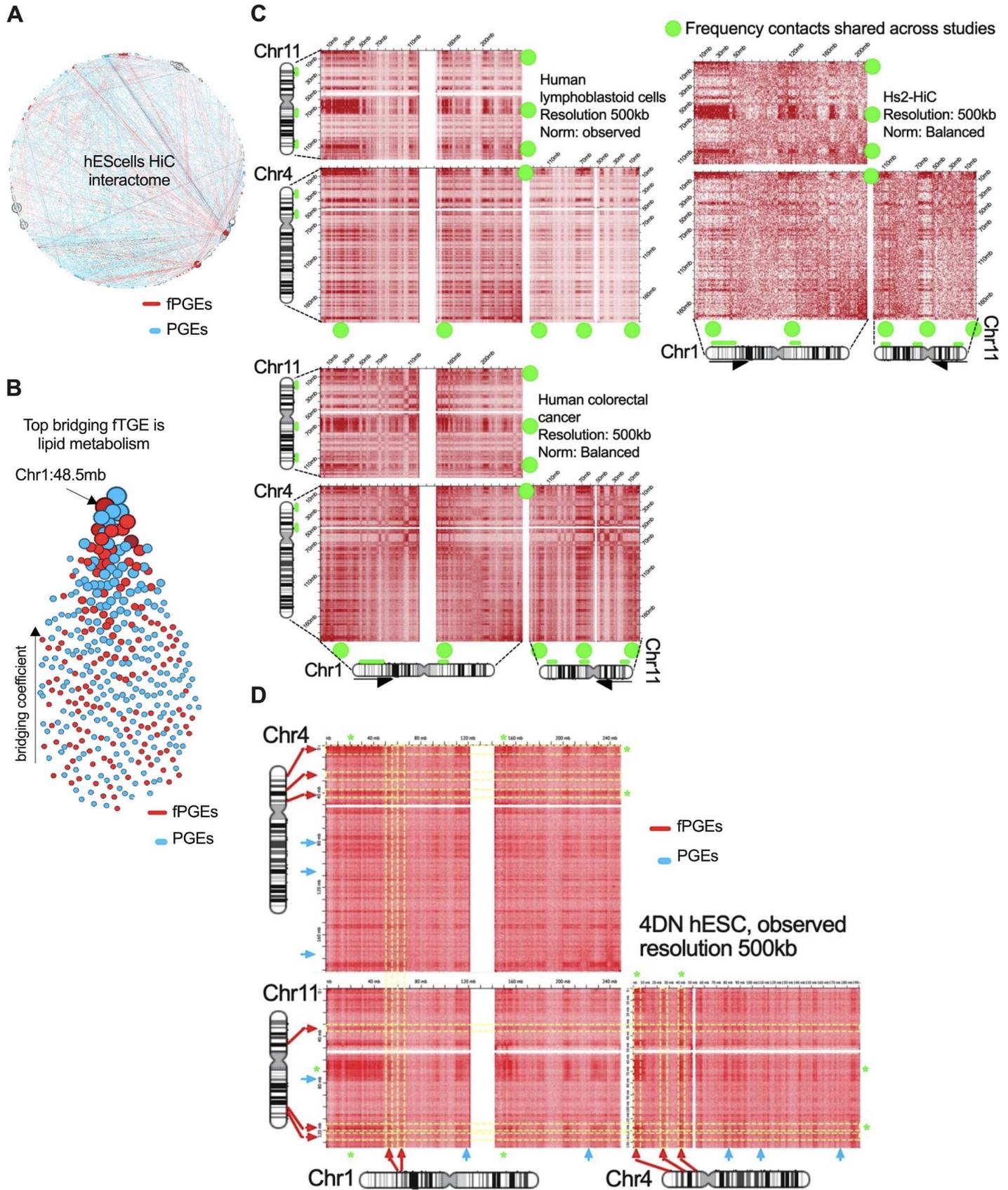


**Fig. S7. PGC1A phase separation and PGC1A-network combinatorial binding** (related to Fig. 1). **(A)** Turbidity assays of PGC1A and RARA protein at different concentrations and with and without polyethylene-glycol. **(B)** FRAP images of GFP-PGC1A condensates in HEK293 cells and treated with forskolin for 4h, and dynamic recovery of luminescence. **(C)** Combinatorial binding of PGC1A-interacting regulators showing the number of co-regulated genes from ChIP-seq experiments and pairwise combination of interactors. Candidates of regulators selected from network analysis. **(D)** Manifold representation of combinations of transcriptional regulators showing in blue top clusters with combinations co-regulating a higher number of genes. Below, sieve diagram representations assessing the dependence of association found in top clusters 1. **(E)** Top co-regulated clusters generally (all combinations) and targeted (combinations with PGC1A). Bars show mean values and standard error of the mean. Two-way anova was performed to compare the effect of group and concentration between groups, followed by two-tailed student's t-test by Tukey's multiple comparison test. P-value cutoff was set at <0.05. \*\*\* representing <0.001.



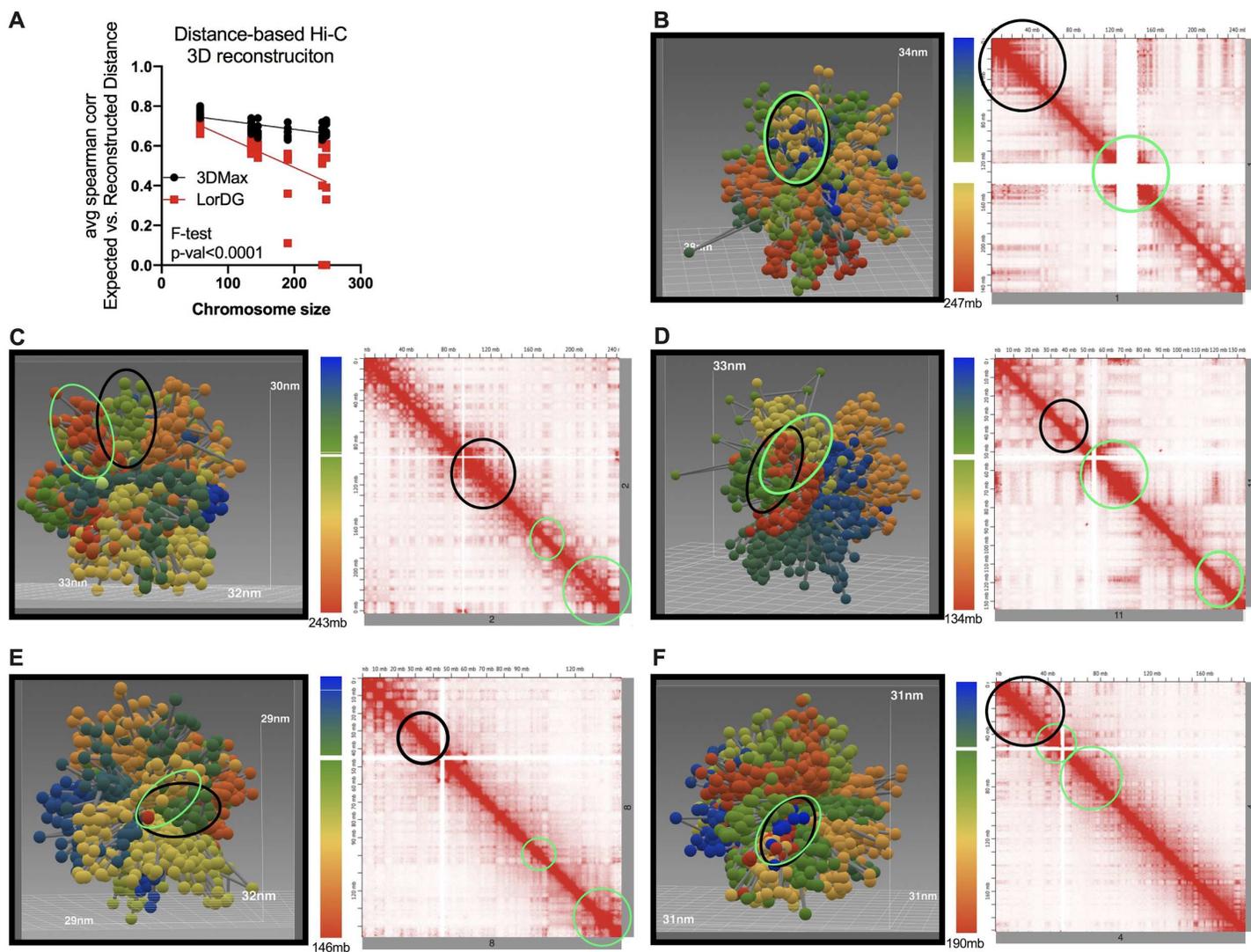
**Fig. S8. Interrogation of functional nuclear compartments through topological integration (related to Fig. 2).** (A) Illustration showing pipeline for chromatin structural relationships. (B) Frequency distribution for positional gene

enrichments (PGEs; > 5 genes per genomic cluster) comparing biologically derived data sets and random gene sets. **(C)** Metabolic HAR genes positional enrichments. **(D)** Cytogenetic band enrichments for metabolic-HAR genes. **(E)** Metabolic-HAR PGEs relationships with compendia of frequently interacting chromatin regions (FIREs), conserved TADs and social trans-interacting chromatin regions (see pipeline for chromatin structural relationships). **(F)** Area under the curve of the receiver operating characteristic (AUC-ROC) for analysis comparing the presence of regions from **(E)** in PGEs from metabolic-HAR genes set and random gene sets. **(G)** Trans-interacting chromatin region network from hESCs displayed in a circular layout. Node size represents pagerank analysis. **(H)** Metabolic HAR domain in chr1: 62-65mb, showing regulated genes by PGC1A-interacting network (green arrow), representative contact HiC frequency maps (GM12878), genehancer regulatory elements and interactions along with vista enhancers. **(I)** Genome-wide network of regulatory element interactions from Genehancer database, displayed in modules and with metabo-HAR regions in red. **(J)** Network-based prioritization of regulatory elements classified in proximal comprising promoter, and distal for enhancer regions. **(K)** Density of interactions by degree parameter for distal and proximal regulatory elements. **(L)** Overlapping consensus of transcriptional regulators. **(L)** Number of links among genome-wide distal regulatory regions and nCHAR elements.



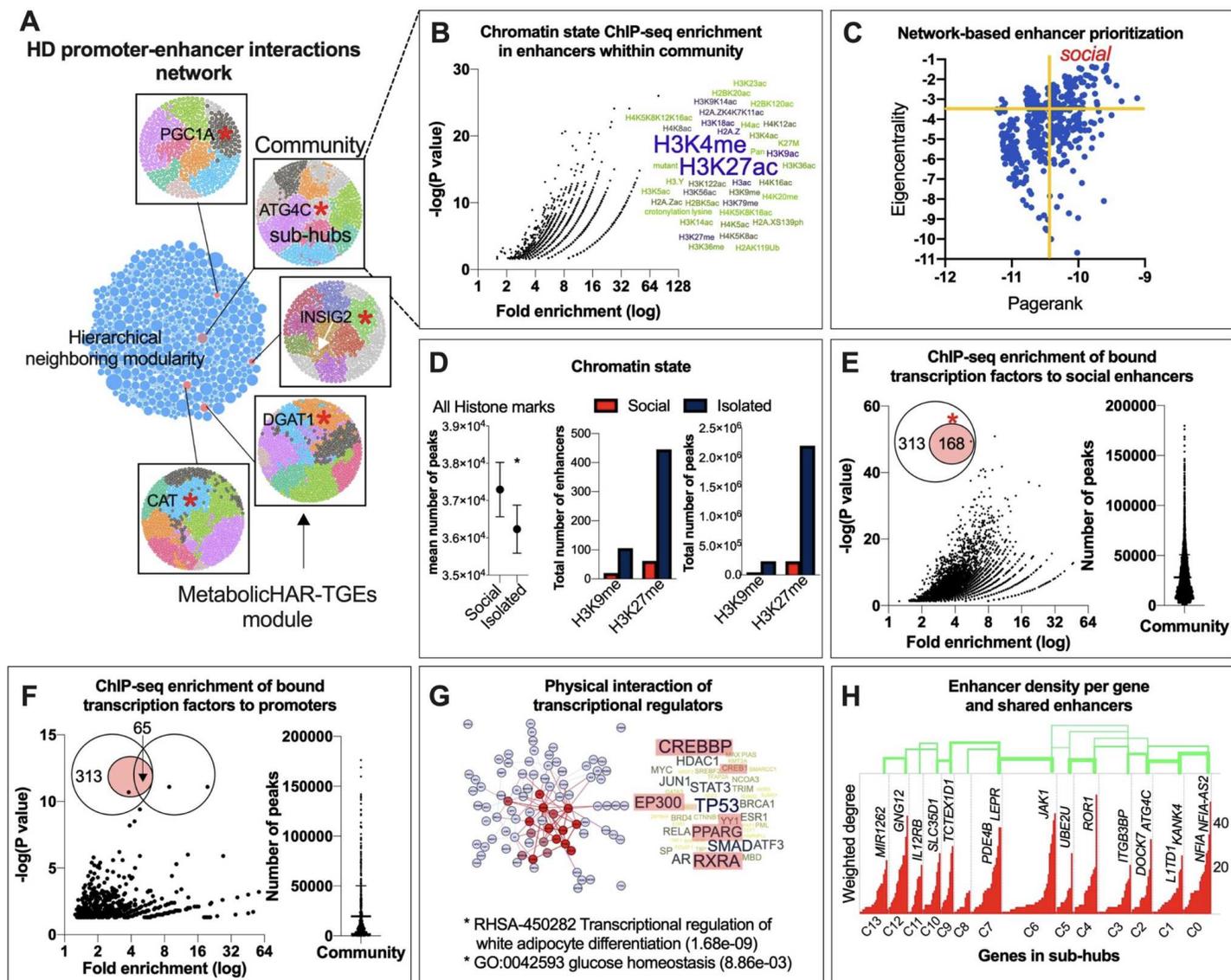
**Fig. S9. Top metabolic functional nuclear compartments** (related to Fig. 2). (A) hESCells HiC network of trans-interacting chromatin regions, displaying overlapping positional enrichments. In red metabolic-HAR PGEs with similar functional annotation and in blue for the rest. (B) Bridging coefficient for overlapping metabolic HAR PGEs in

trans-interacting chromatin regions. Layout upwards with top PGEs with highest score, and functional (red) or not (blue). (C) Pervasive of contact frequency maps across other studies for metabolic-HAR regions with similar function (top bridging coeff.) and in trans-interacting regions. (D) For selected chromosomes harboring metabolic-HAR regions with similar function, displaying the location of PGEs in relation to trans-interacting contacts.

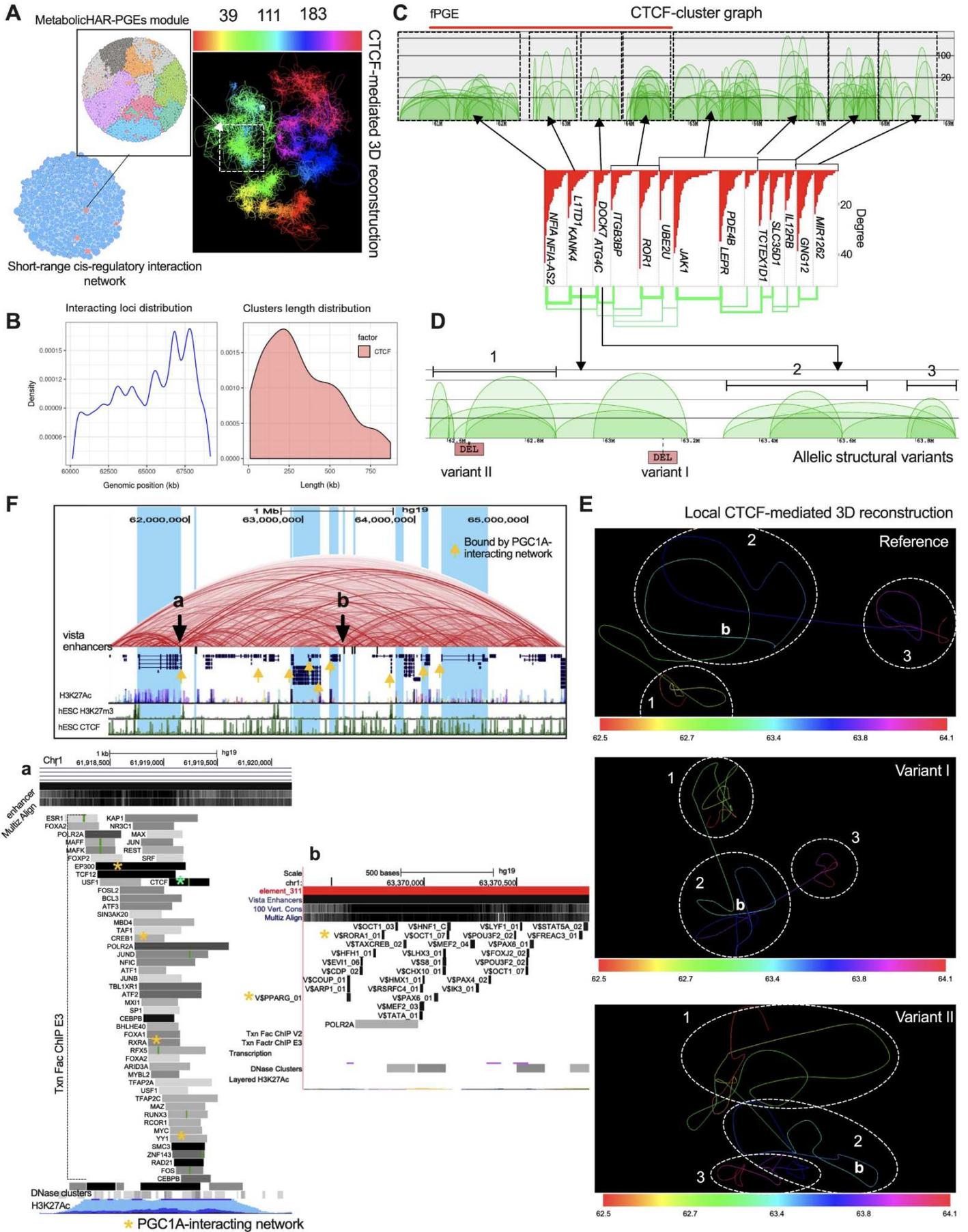


**Fig. S10. Global 3D reconstruction of chromosomes harboring metabo-HAR regions and their homotypic interactions** (related to Fig. 2). (A) Comparison of distance based 3D HiC reconstruction for chromosome size bias, between 3DMax and LorDG softwares. Displayed expected vs reconstructed distances and chromosome size. 3D picture reconstruction using 3DMax for chromosomes harboring selected metabolic-HAR regions (black circle) and top homotypic interaction (green circle) for Chr1 (B), Chr2 (C), Chr11 (D), Chr8 (E) and Chr4 (F).

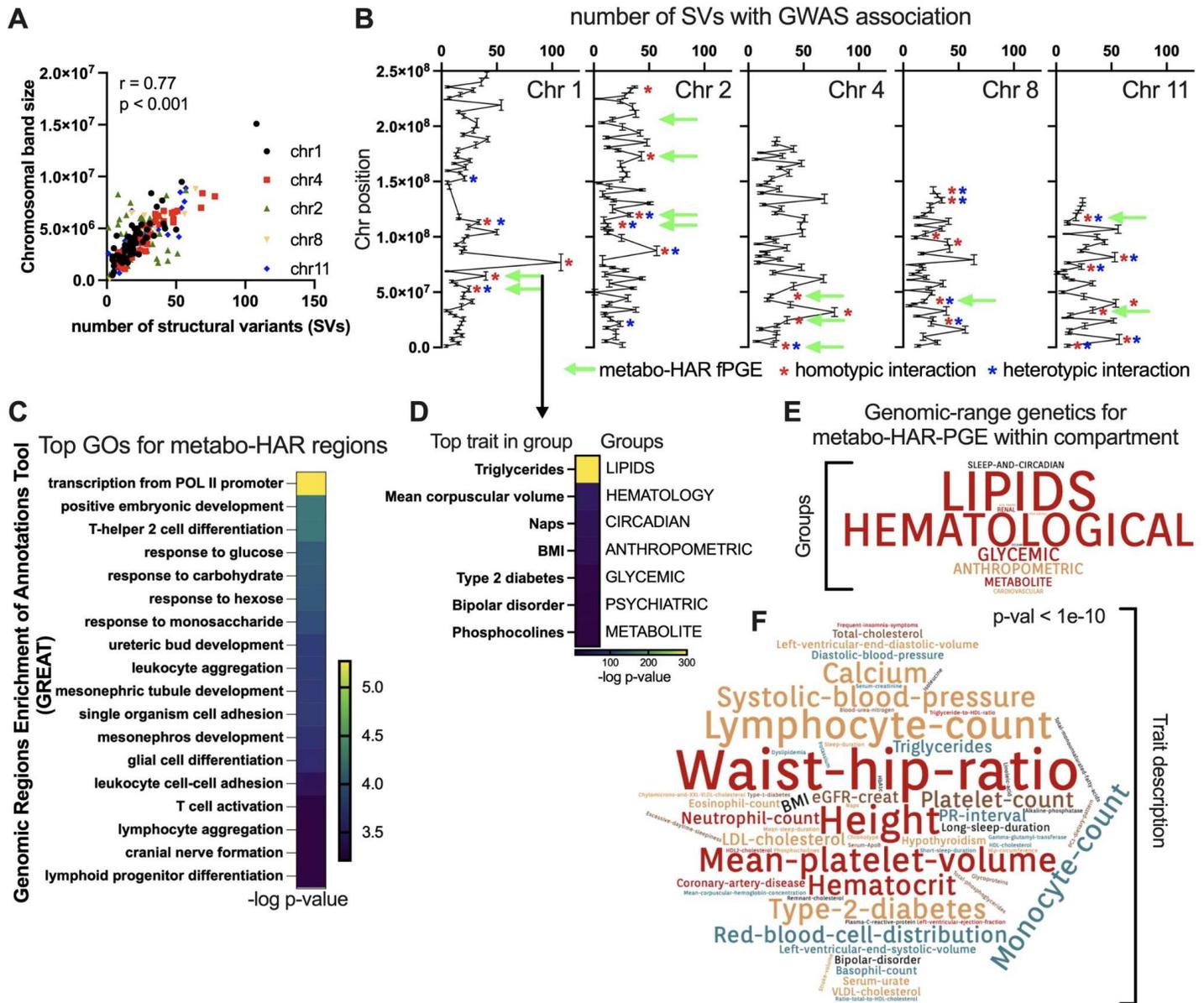




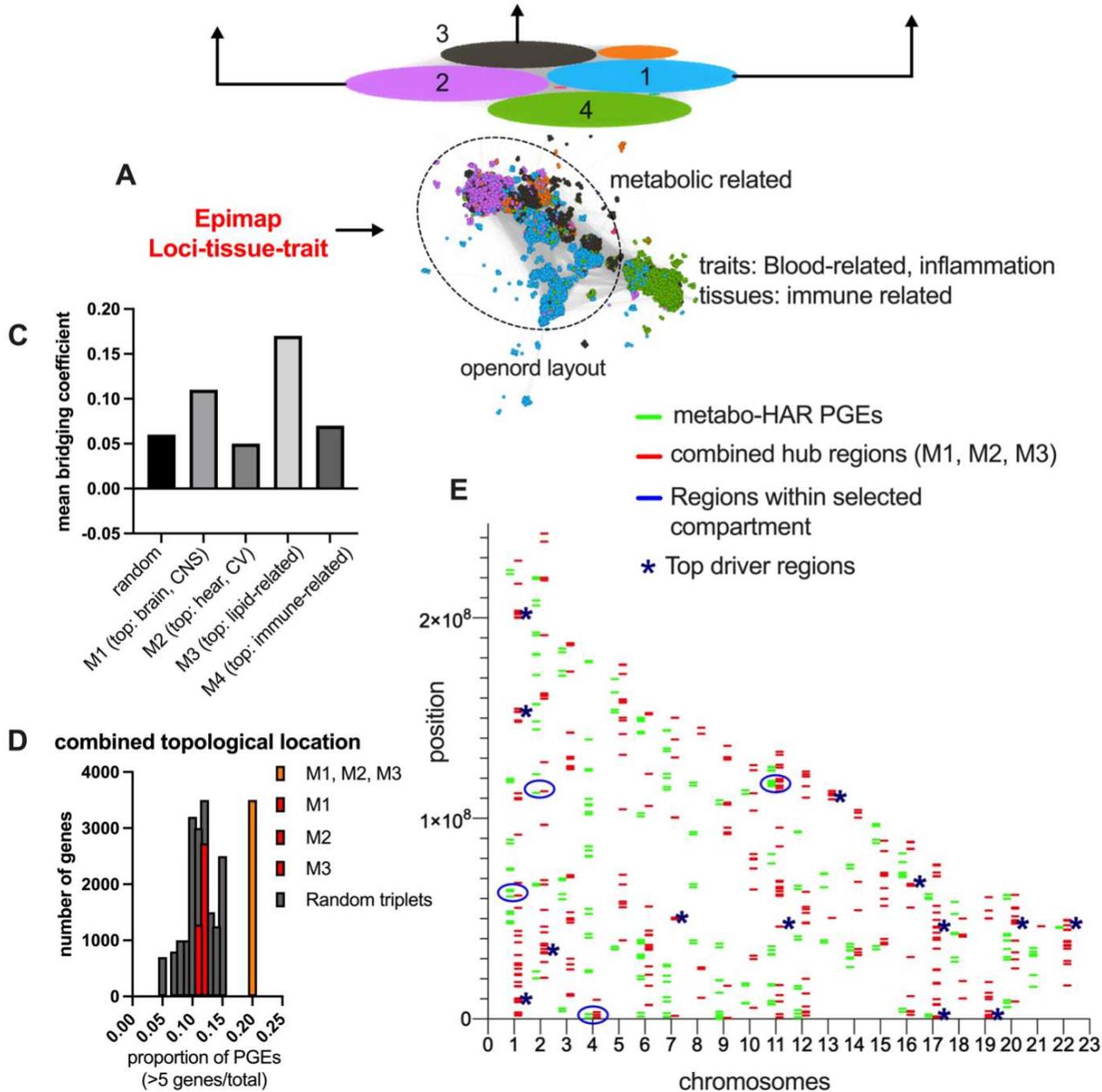
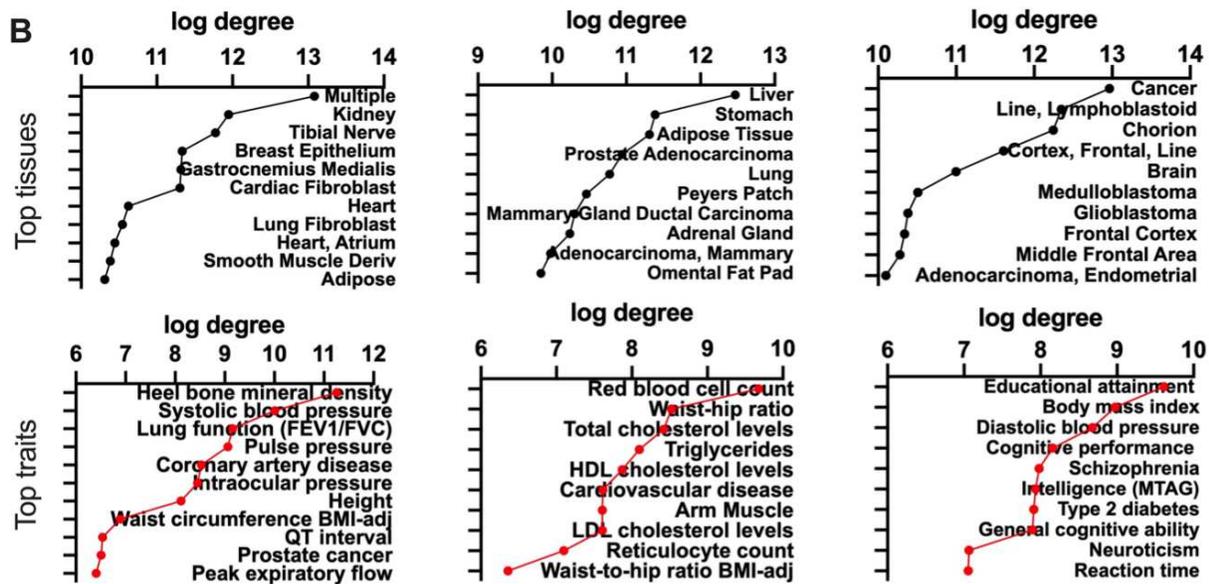
**Fig. S12. High-definition short-range territorial networks** (related to Fig. 2). **(A)** Short-range regulatory elements interaction network, displayed with a modularity based layout. Highlighted in red metabolic-HAR domains modules and submodule territories. Gene in proximity to HAR elements and submodule location (asterisk). **(B)** ChIP-seq enrichment analysis for elements composing metabolic-HAR domain and specific for chromatin states. In a word-cloud subpanel represented top chromatin modifiers within elements. **(C)** Network-based prioritization of metabolic-HAR regions module. **(D)** Chromatin state ChIPseq enrichment comparison for highly social and isolated enhancers. First right panel showing peaks for all histone marks, number of enhancers and number of peaks for repressive histone marks H3K9me and H3K27me. **(E)** Transcriptional regulator ChIP-seq enrichment analysis for social enhancers within metabolic-HAR module and number of peaks within module. Venn diagram showing number of regulators across all elements (313) and top regulators in social regulators (168) **(F)** Transcriptional regulator ChIP-seq enrichment analysis for promoters within metabolic-HAR module and number of peaks within module. Venn diagram showing number of regulators across all elements (313) and overlapping top regulators in promoter and social enhancers. **(G)** PPI Network for interacting transcriptional regulators selected in F. Word-cloud of top regulators (in red) and functional annotation. **(H)** Promoter gene degree distribution of submodules within metabolic-HAR region and density of shared enhancers between submodules.



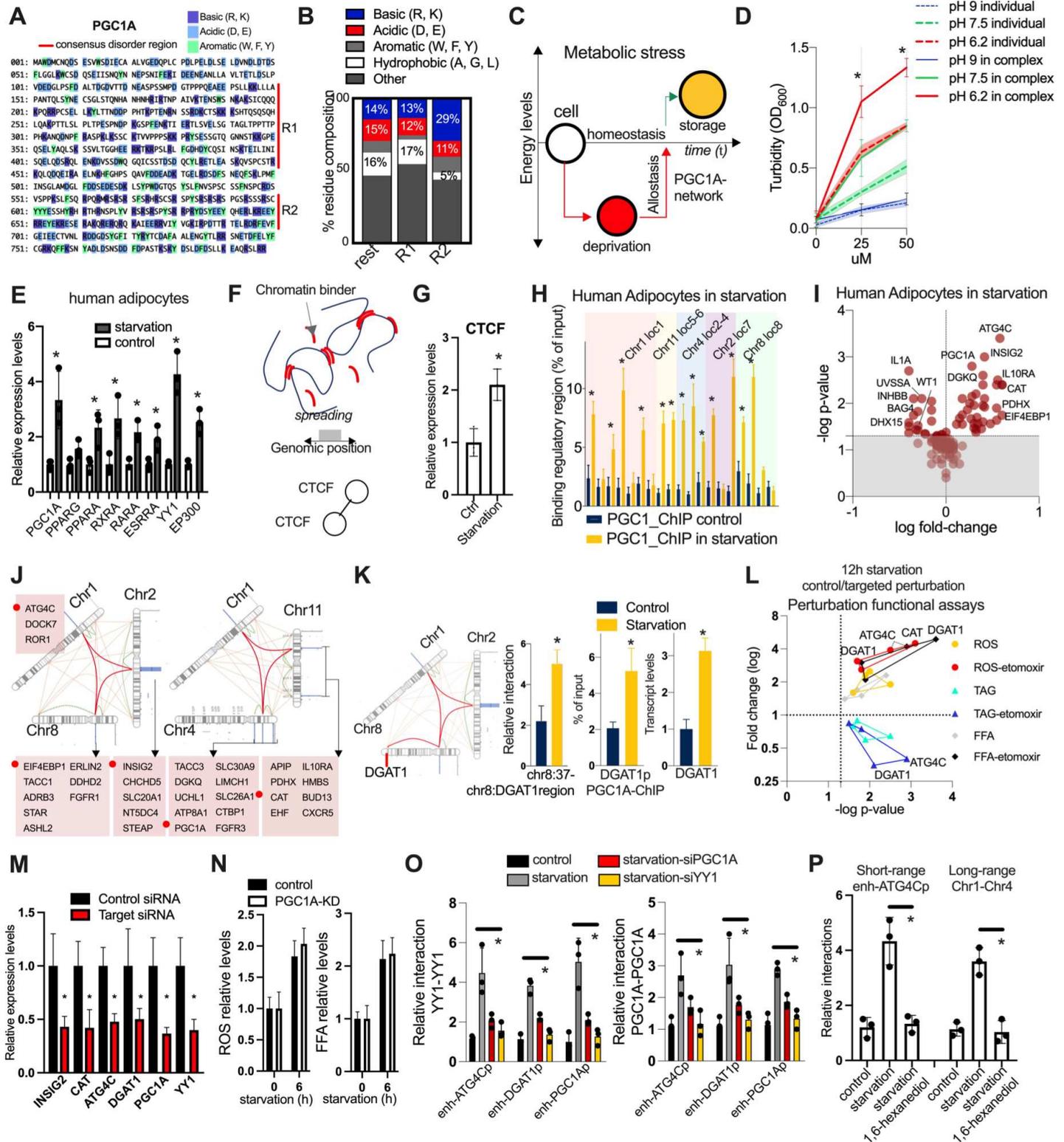
**Fig. S13. Local CTCF-mediated 3D reconstructions and territorial submodules as chromatin domains** (related to Fig. 2). (A) Short-range network and metabolic-HAR region submodules with CTCF-based 3D reconstruction and localization of HAR domain within chromosome 1. (B) Density of interacting loci in genomic range, and cluster length of CTCF interactions. (C) CTCF-based cluster graph density of interaction within metabolic-HAR domain (top panel and fPGE in red). Submodule territorial overlap between interacting frequency network and CTCF specific clusters. (D) CTCF-graph cluster in 2 neighboring submodules within metabolic-HAR domain, and allelic structural variant location. (E) Local CTCF-based 3D reconstruction for regions in D showing reference structure and both allelic structural variant effects on density of interactions. (F) Genome browser picture of metabolic-HAR region showing 2 selected enhancers (a, b) for downstream analysis with TFBS and ChIP-binding location of PGC1A-interacting network.



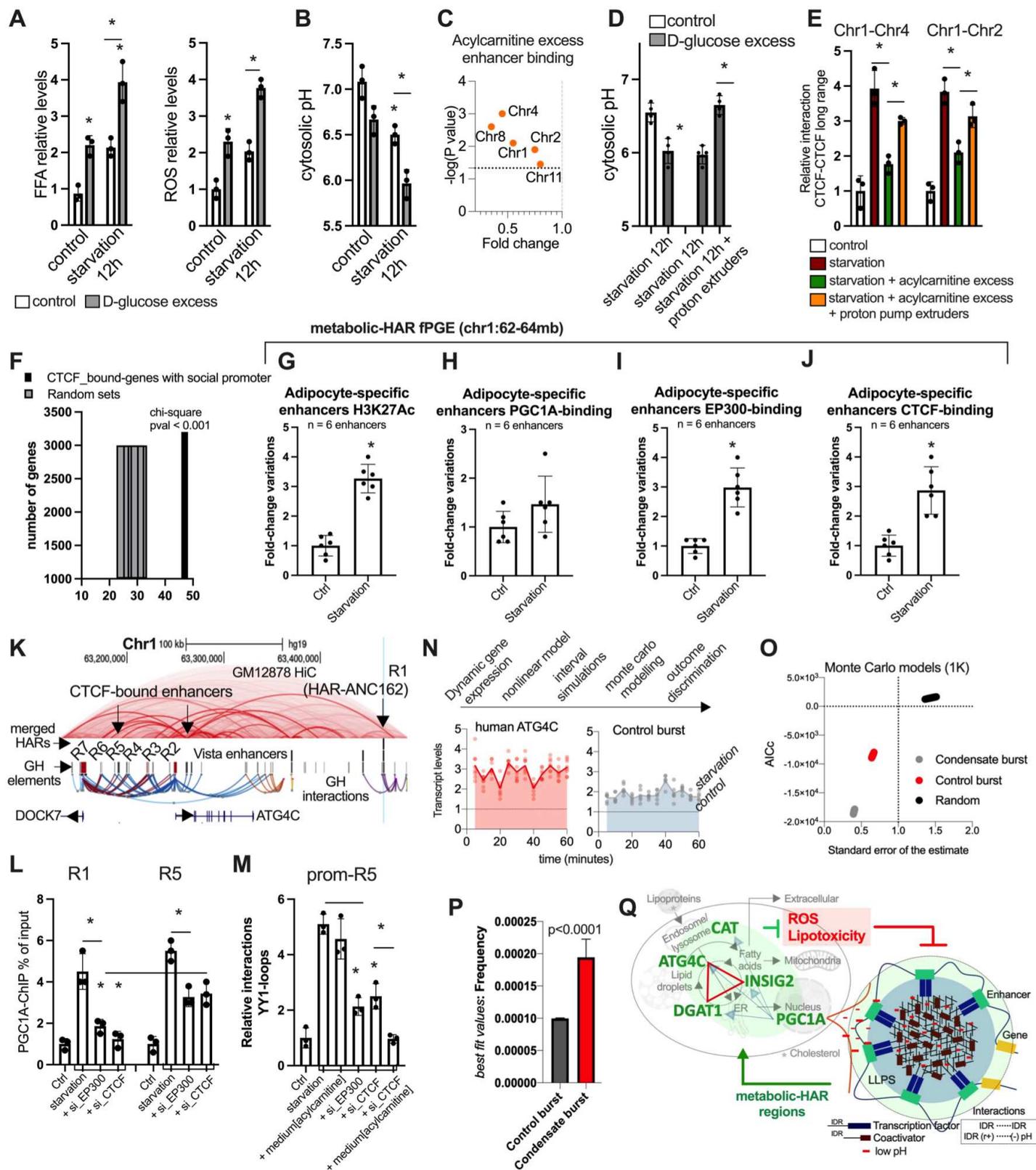
**Fig. S14. Structural variants density and genomic-range genetics in reconstructed nuclear compartments** (related to Fig. 2). (A) Correlation between number of structural variants with GWAS association and chromosome harboring selected metabolic-HAR domains. (B) Number of structural variants with GWAS association per cytogenetic band in chromosomes harboring metabolic-HAR. Green arrow shows metabolic-HAR location and red and blue asterisks for homo- and heterotypic interacting regions within compartments. (C) Genomic-range functional enrichment for metabolic-HAR domains using GREAT annotation tool. (D) Genomic-range genetics for selected metabolic-HAR regions with a representative example in Chr1:61-65mb range, using the HugeAmp tool. Displayed top trait within trait-group class. (E) Genomic range genetic results for selected metabolic-HAR regions within compartment displaying top trait-group repetition as a word-cloud. (F) Genomic range genetic results for selected metabolic-HAR regions within compartment. Displaying top traits description repetition as a word-cloud.



**Fig. S15. EpiMap network integration and characterization of lipid-traits relationships** (related to Fig. 2). **(A)** Epimap network reconstruction of loci with tissue and trait enrichments. Module classification in openord layout displaying metabolic trait relationships in interconnected modules (module 1-3) and immune-related traits. **(B)** Trait and tissue top degree scores in each module with strong metabolic relationships. **(C)** Mean bridging coefficient across modules and random network. **(D)** Combined topological coalescence of positional enrichments for top genes within each interconnected metabolic trait, followed by chromosomal landscape positioning and colocalization with metabolic-HAR regions, and top driver regions from combinatorial-hub assessment **(E)**.



**Fig. S16. Metabolic control of nuclear compartmentalization** (related to Fig. 3). **(A)** PGC1A protein sequence with consensus disorder regions, basic, acidic and aromatic residues. **(B)** Proportion comparison of residue composition features between disorder regions and the rest of PGC1A protein sequence. **(C)** Illustration of homeostatic and allostatic mechanisms regulated by PGC1A-network **(D)** Turbidity assays at different concentrations and pH variations for PGC1A alone or in complex with RARA. **(E)** Expression of transcriptional regulators in glucose starvation (12h). **(F)** Illustration showing genomic spreading between chromatin interactions mediated by structuring candidates such as CTCF. **(G)** CTCF expression as in (E). **(H)** ChIP-qPCR of PGC1A to HAR elements within selected metabolic-HAR compartments. **(I)** Expression of metabolic-HAR genes within selected compartments during glucose starvation. Top regulated genes in red and compartments in panel **(J)**. **(K)** Chromosomes 1-2-8 compartment and DGAT1 loci location, CTCF-mediated long range homotypic interaction in chr8, PGC1A binding to DGAT1 promoters and DGAT1 expression levels. All assays were done as in (E). **(L)** Functional assays in starvation with or without CPT1A-channel inhibitor etomoxir, in cells perturbed by siRNA for top regulated genes in metabolic-HAR regions. Etomoxir exacerbates targeted perturbation phenotypes. **(M)** Gene expression after siRNA knockdowns. **(N)** Functional assays in control and PGC1A knockdown cells after 6h of starvation induction. **(O)** Relative short-range interactions in starvation and in targeted siRNA downregulation for top regulated metabolic-HAR genes. **(P)** YY1-short- and CTCF-long-range mediated interactions in starvation and abolished by 1,6-hexanediol. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-va<0.05

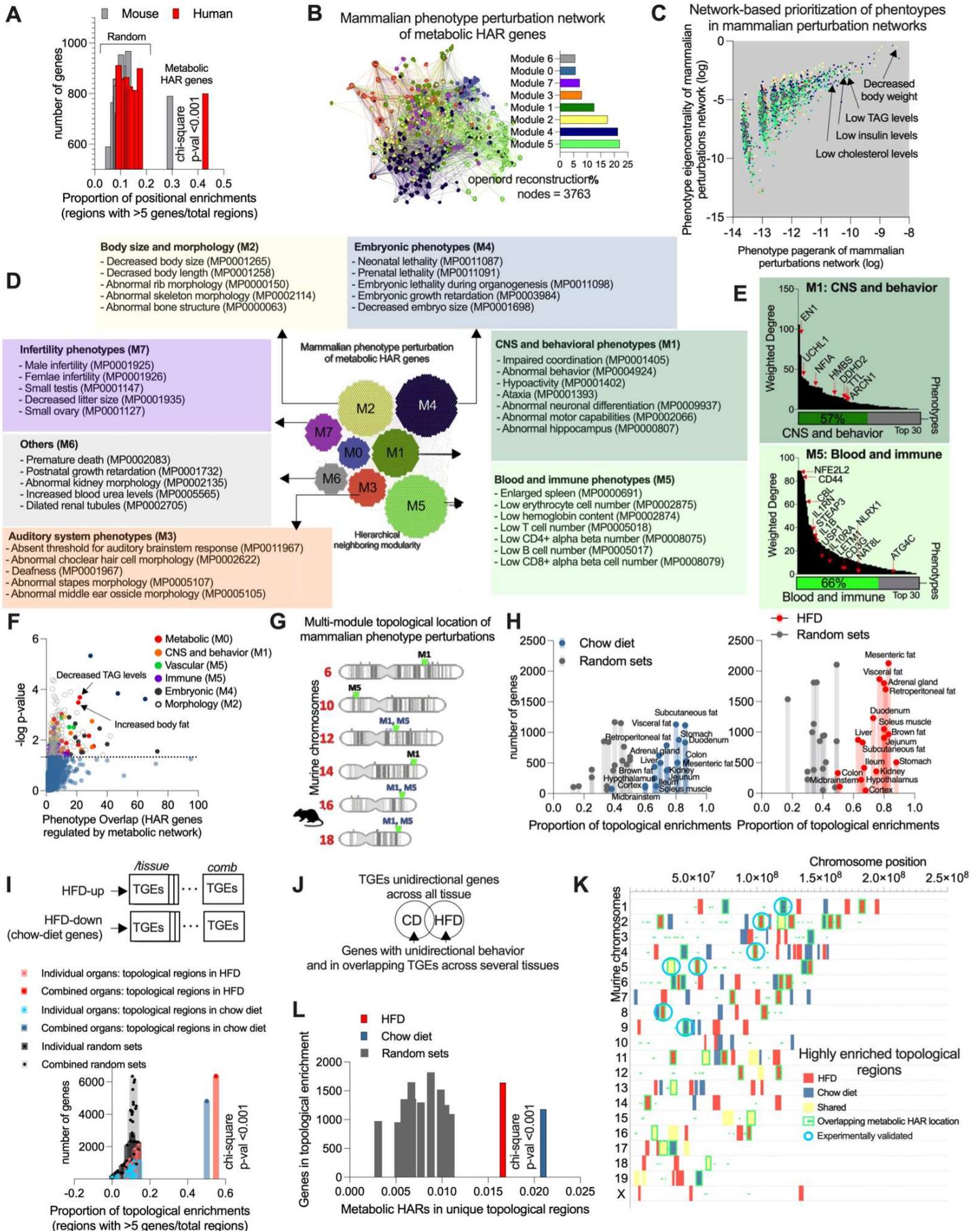


**Fig. S17. Metabolic control of nuclear compartmentalization** (related to Fig. 3). (A) FFA and ROS functional assays in human adipocytes in starvation and with glycolysis inhibition by D-glucose. (B) Cytosolic pH in the same conditions as in A. (C) PGC1A binding to selected HAR-elements during starvation and with acylcarnitine excess (20uM). (D) Cytosolic pH in starvation and after glycolysis inhibition and activators of proton pump extruders such as Zinc. (E) CTCF long range chromatin interactions in starvation, starvation with acylcarnitine excess as in C, and proton pump activators as in D. (F) Proportion of CTCF-bound genes displaying social promoters from our regulatory regions interaction network, compared to random permutations. (G) H3K27Ac ChIP-qPCR in adipocyte specific enhancers within selected metabolic HAR region

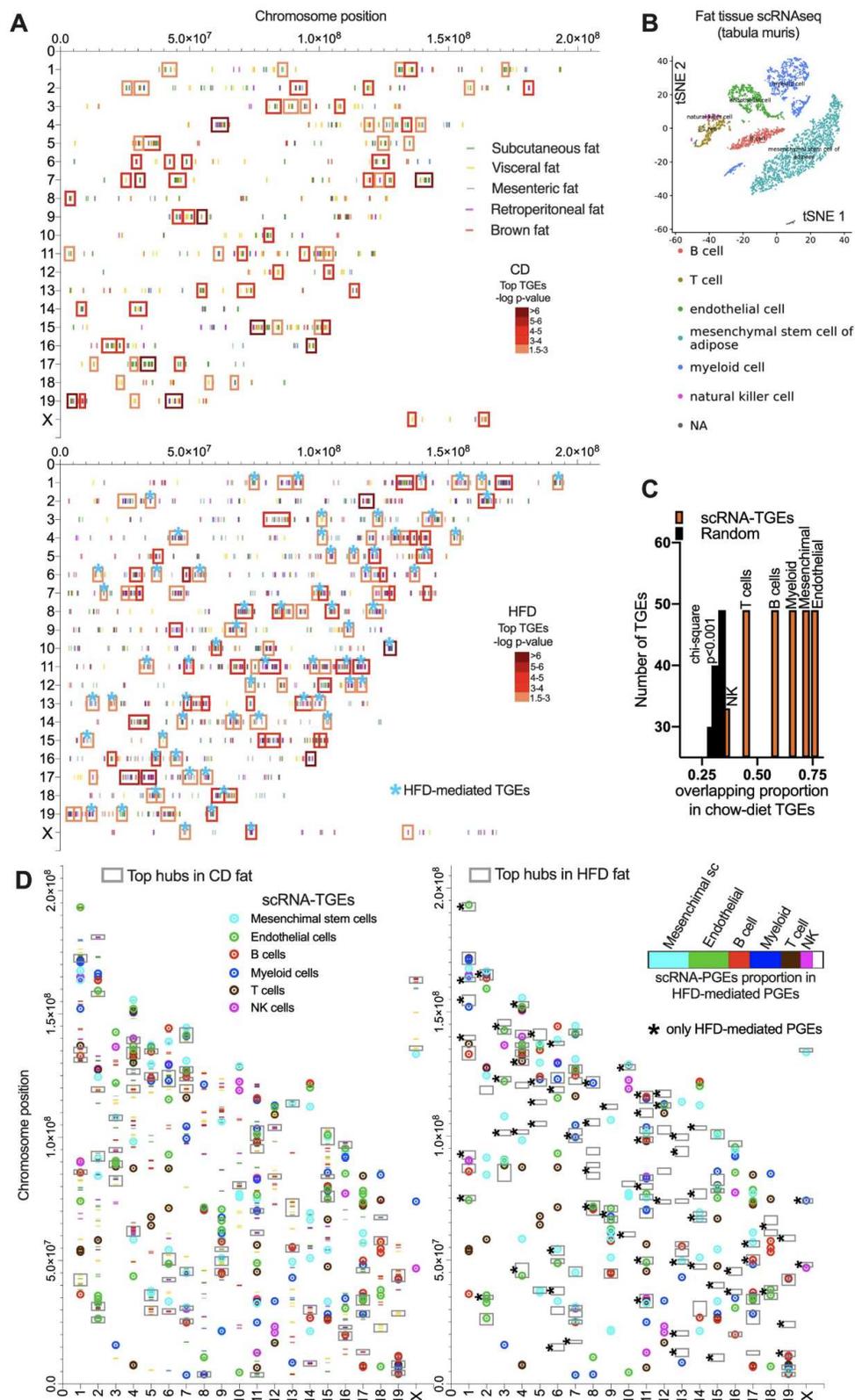
during starvation (12h). **(H)** PGC1A-binding by ChIP-qPCR in adipocyte specific enhancers as in G. **(I)** EP300 ChIP-qPCR as in G. **(J)** CTCF-binding by ChIP-qPCR as in G. **(K)** Human genome track showing zoom in top target within metabolic-HAR domain in chr1:62-64mb: ncHARs (ANC162), regulatory elements and interactions, GM12878-Hi-C data, CTCF-bound elements and promoter locations of ATG4C and DOCK7. **(L)** Exacerbated loss of PGC1A binding to HAR by EP300 and CTCF knockdown. PGC1A ChIP-qPCR during starvation and with EP300 and CTCF knockdown in R1(HAR) element and R5 (CTCF bound) element. **(M)** Loss of resilience and sensitization to metabolic overload in structural loops by EP300 and CTCF knockdown. YY1-mediated loop interactions between ATG4C-promoter and R5-CTCF bound element in starvation, in medium acylcarnitine excess (10uM), with EP300 and CTCF knockdowns and together with medium AC excess. **(N)** Pipeline steps description of Monte Carlo modeling of dynamic gene expression data in stimulated conditions (starvation). Panels below are dynamic transcript levels in starvation of regulated genes: ATG4C gene (social promoter by regulatory interaction networks) and ANGPTL3 (isolated promoter). **(O)** Akaike information criterion and standard error of the estimate for Monte Carlo burst models from dynamic gene expression comparing genes in N and random models. **(P)** Frequency best-fit-values parameters from Monte Carlo discrimination of burst models. **(Q)** Illustration of mechanistic insight for lipid cycling regulation of nuclear compartment formation. Top genes within interacting metabolic-HAR domains regulated by PGC1A-network during starvation and controlling lipid-derived oxidative stress and toxicity. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.



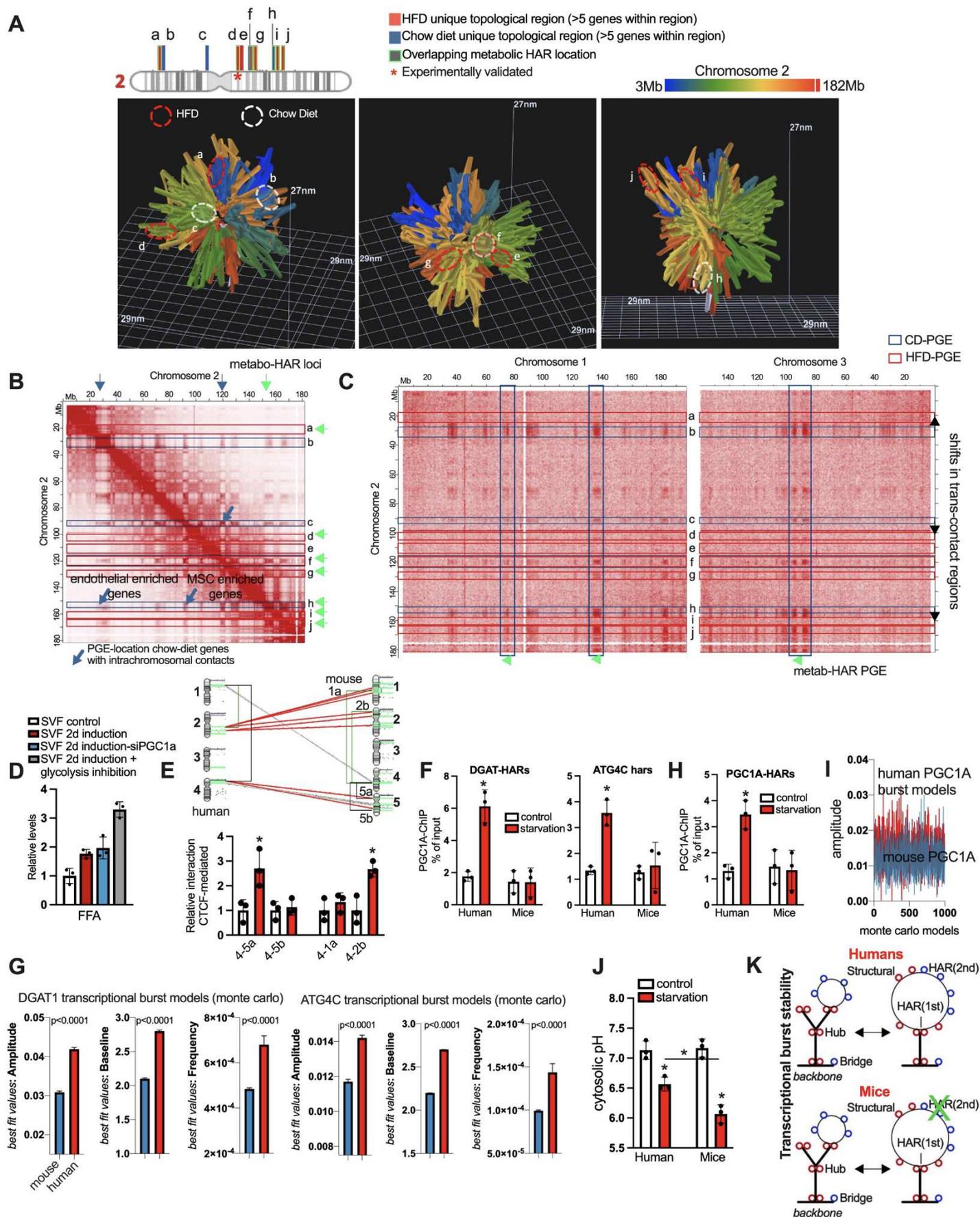
metabolicHAR domains. **(D)** ChIP-loop of metabolic-HAR interactions within selected nuclear compartments. **(E)** Expression of metabolic-HAR genes within selected compartments during glucose starvation. **(F)** Enhancer-promoter loops between HAR-elements and starvation-regulated genes within metabolic-HAR domain. **(G)** HAR-DGAT1 CTCF-mediated long-range interaction by ChIP-loop, DGAT1 expression levels, PGC1A-binding to DGAT1 promoter and CTCF-mediated short range HAR-promoter interactions. **(H)** Aggregate domain expression of genes composing metabolic-HAR domains in human-derived adipose precursor cells subtypes. **(I)** Aggregate domain expression of genes composing metabolic-HAR domains in human-derived single nuclei RNA seq brain cell subtypes. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.



**Fig. S19. Mammalian convergence and tuning of spatial genome architecture** (related to Fig. 4). **(A)** Proportion of positional gene enrichments (>5 genes per genomic cluster) for metabolic-HAR genes in humans, mice, and random gene sets. **(B)** Network integration of mammalian genetic perturbations for metabolic-HAR genes in openord layout and modularity (right panel). **(C)** Network-based prioritization of mammalian perturbation relationships for metabolic-HAR genes showing top phenotypes. **(D)** Top phenotypes in each module. **(E)** Node degree classification of gene drivers per phenotype in CNS and blood-immune modules. **(F)** Phenotype enrichment for metabolic-HAR genes showing preservation of network modules and metabolic phenotype as top enrichments. **(G)** Multi-module topological convergence of metabolic interacting modules in murine chromosome regions. **(H)** Mammalian HFD perturbations followed by RNAseq in 19 tissues and spatial-PGEs analysis. Differentially expressed genes for both chow diet and HFD were used for hub-positional enrichments compared to random sets. **(I)** Combinatorial hub-positional enrichments for diet-induced genes comparing organ-effect and cumulative hub positions across all organs assessed. **(J)** Venn diagram illustration of topological positional enrichments (TGE or PGE) for differential genes specifically regulated by either diet intervention. **(K)** Chromosomal landscape hub-positions for PGEs regulated by diet intervention, with overlapping syntenic metabolic-HAR regions and those that were experimentally validated. **(L)** Quantification of metabolic-HAR domain overlap with diet induced cumulative hubs.

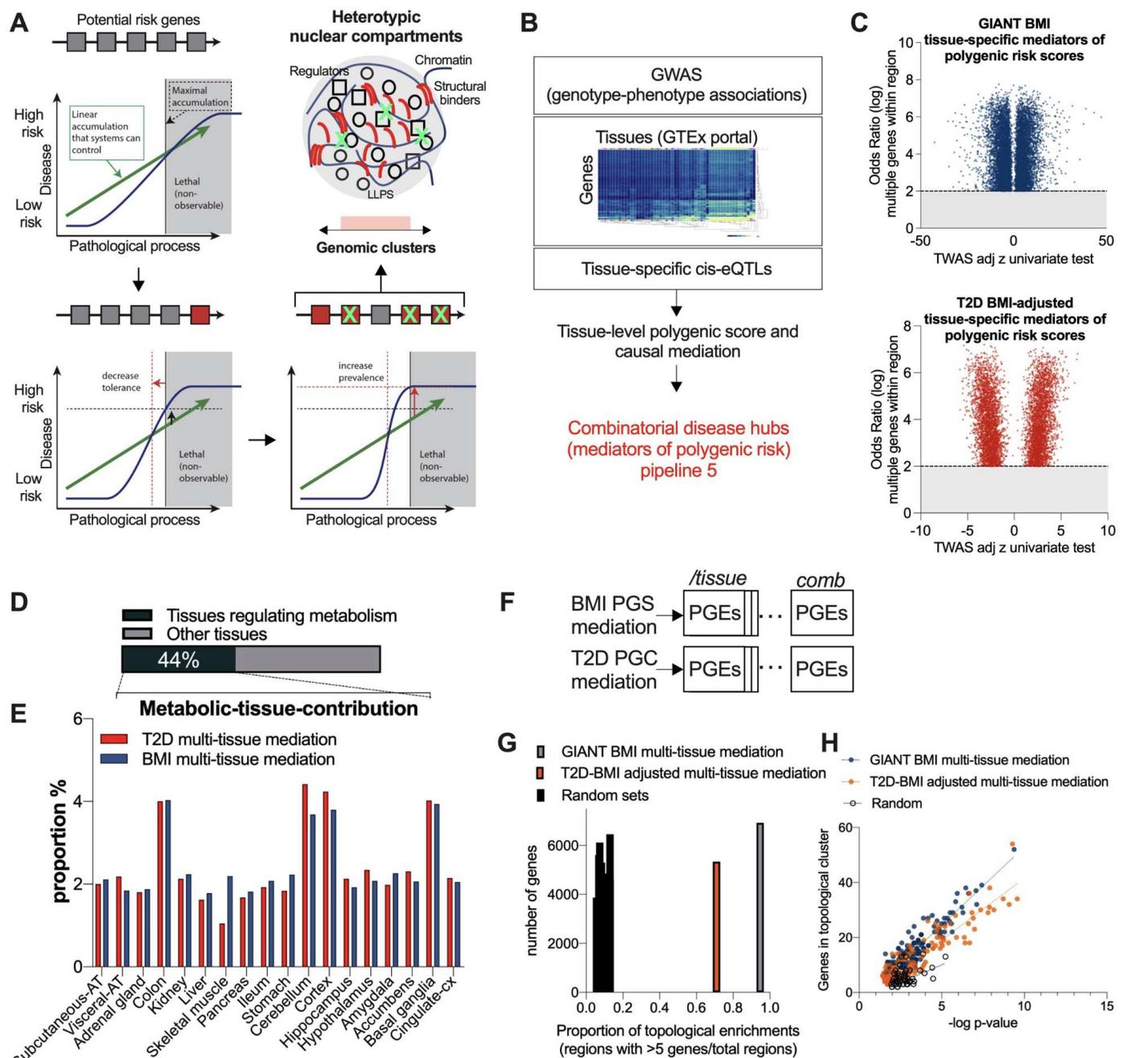


**Fig. S20. Single cell decomposition of spatial transcriptional hubs (related to Fig. 4).** (A) Chromosomal landscape of hub-positions for PGEs regulated by diet intervention in different adipose depots. With an asterisk highlighted top positional shifts caused by HFD. (B) t-SNE cluster plots for mice adipose tissue scRNA-seq from Taboula Muris consortia. (C) Proportion of hub-overlap between scRNAseq transcriptional hubs and chow-diet RNA-seq hubs compared to random hotspots. (D) Chromosomal landscape of hub-transcriptional positions for scRNA-hotspots and chow- and HFD mediated hubs. Proportion of overlap between top hubs for HFD and cell type.

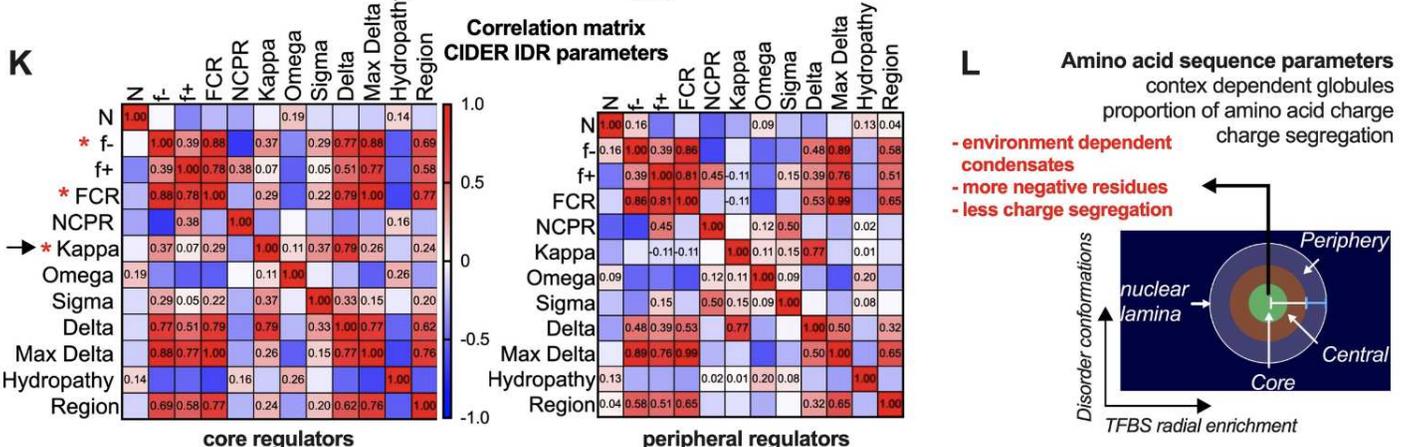
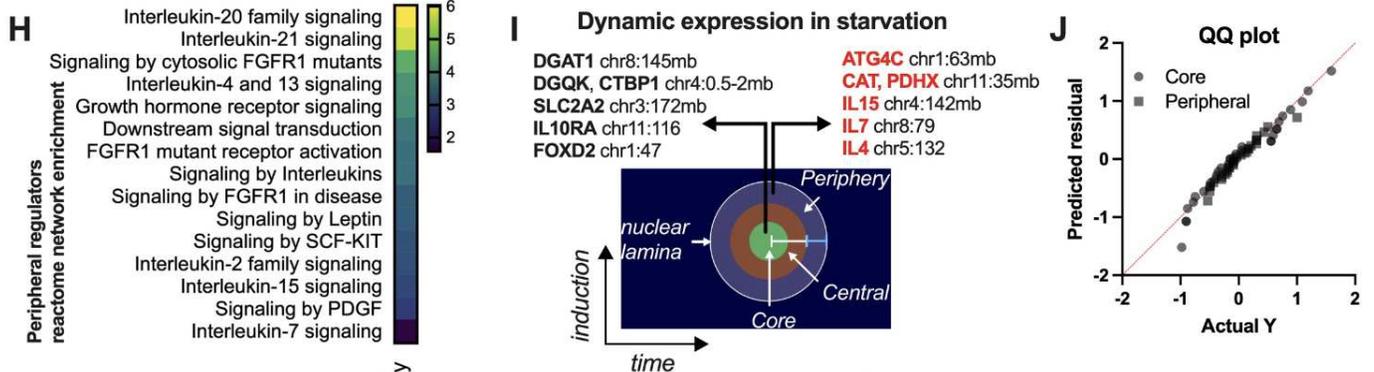
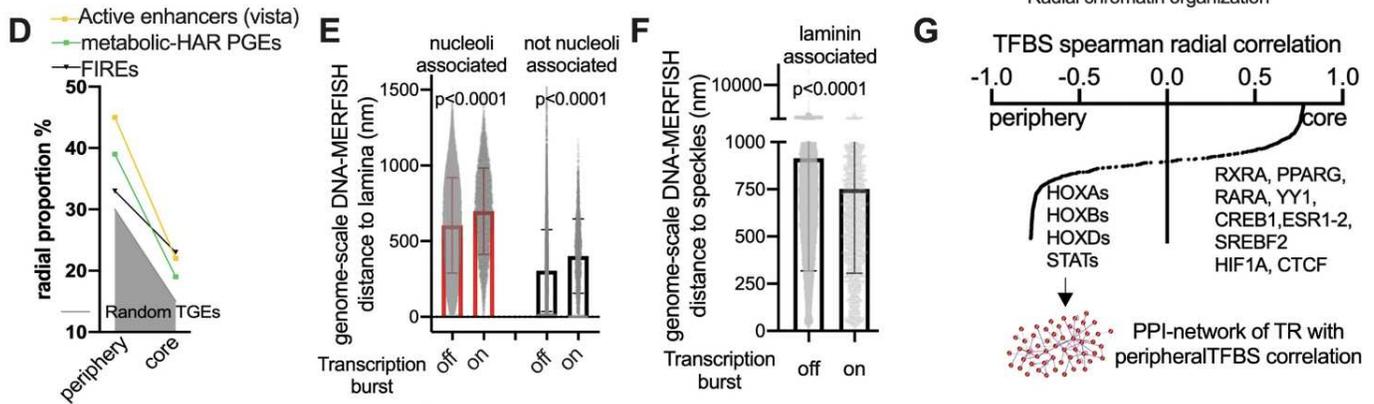
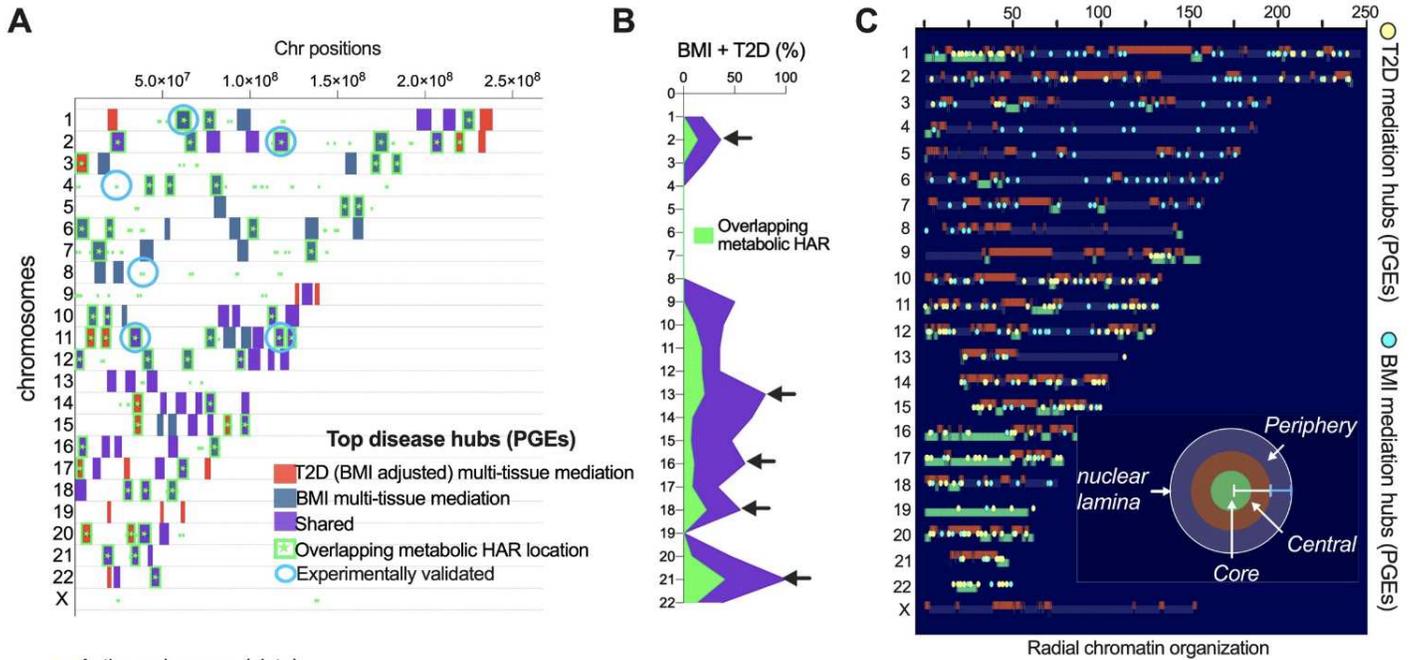


**Fig. S21. Mammalian tuning of spatial genome architecture** (related to Fig. 4). **(A)** Chromosome 2 hotspot regulated by diet and overlapping synteny metabolic-HAR regions. With an asterisk experimentally validated HAR. In lower panels, 3DMax reconstruction of homotypic distances from HiC data displaying diet-specific hotspot and positional shifts caused

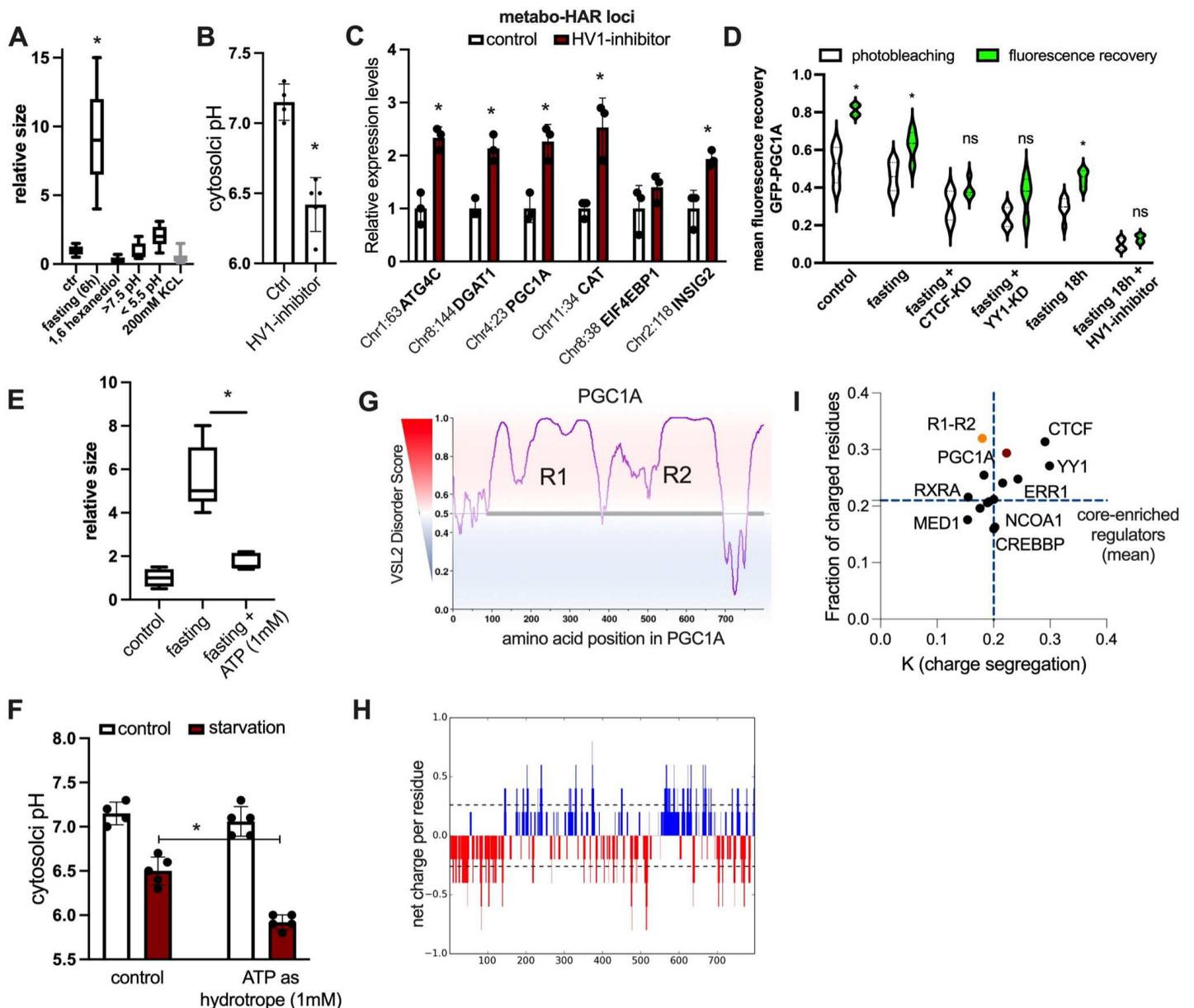
by HFD. **(B)** Representative HiC contact frequency map for chromosome 2 from mESCs (132). Blue squares represent chow-diet specific hotspots while red HFD hubs. Green arrow shows syntenic metabolic-HAR domain locations and blue arrows show single cell hubs located in homotypic interacting chromosome regions, and preferentially enriched in chow-diet hotspots. c-h homotypic interactions with MSCs hotspots and b-h with endothelial hotspots **(C)** Diet-induced shift in hub locations from trans-interacting in chow-diet to neighboring non interacting chromatin regions in HFD. Green arrow shows syntenic metabolic-HAR domain locations. **(D)** Free-fatty acid assay in stromal-vascular fraction cell cultures from mice fat tissue after induction of differentiation with PGC1A perturbation and glycolysis inhibition. **(E)** Top panel shows 3 experimentally validated metabolic-HAR regions in humans, connected to their syntenic mice domains in grey if it is fully preserved, and in red if it is split in mice. Lower panel shows CTCF-mediated long-range interactions for nuclear compartment in mice cell cultures during starvation. **(F)** PGC1A-binding by ChIP-qPCR in selected HAR elements and in mouse orthologue positions for top metabolic-HAR regions ATG4C and DGAT1. **(G)** Monte Carlo burst models from dynamic gene expression in starvation for selected genes DGAT and ATG4C from mice and human cell culture systems. Showing best-fit-values from discriminatory outcomes for amplitude, baseline and frequency. **(H)** PGC1A-binding by ChIP-qPCR in selected HAR elements and in mouse orthologue positions for top metabolic-HAR regions PGC1A domain. **(I)** Monte Carlo burst models from dynamic gene expression in starvation for selected gene PGC1A from mice and human cell culture systems. Showing best-fit-values from discriminatory outcomes for amplitude. **(J)** Functional assays for cytosolic pH in cell cultures during starvation at 12h. **(K)** Illustration showing nested model transcriptional variations between human and mice, and the transcriptional optimization and stability offered by cis-HAR anchoring of chromatin conformations. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.



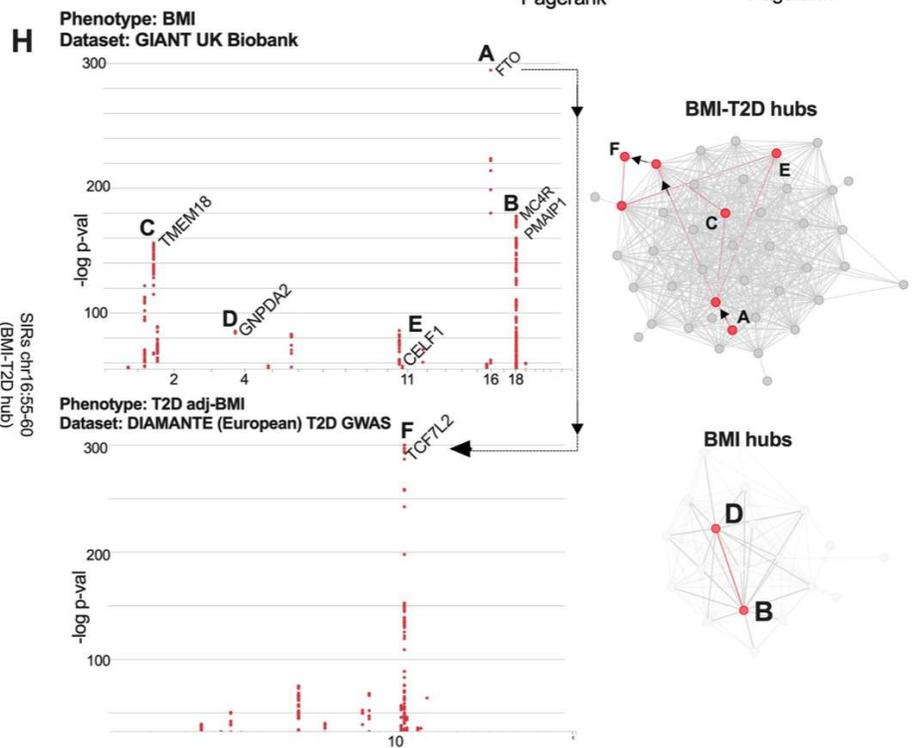
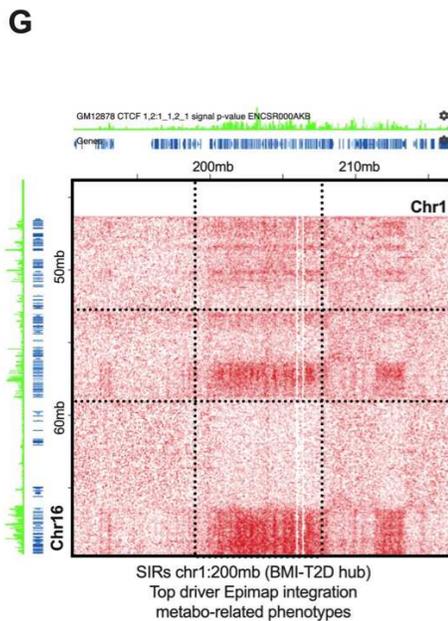
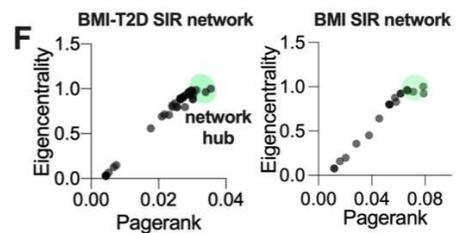
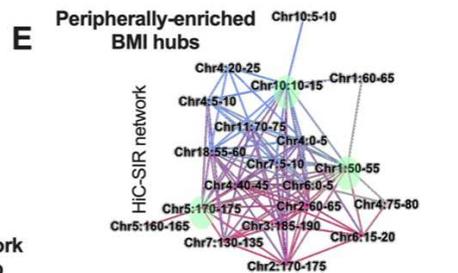
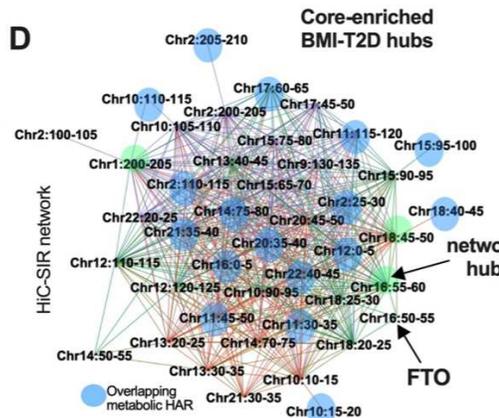
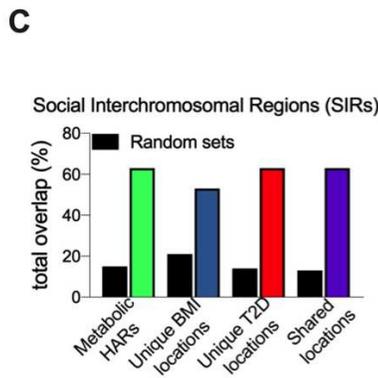
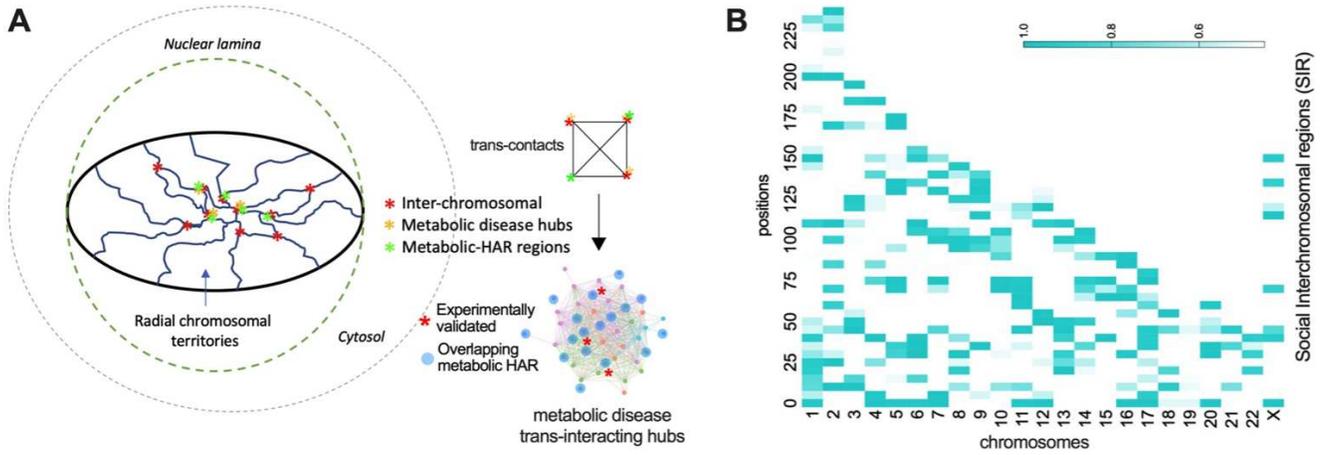
**Fig. S22. Genomic clusters of mediators of polygenic risk for metabolic disease (related to Fig. 5).** (A) Polygenic risk model illustration showing the cumulative effect of genetic variation on disease risk prediction. Here we exploit the cumulative effect of related variations and genes affected thereby, to find genomic risk clusters of co-regulated transcriptional compartments. (B) Illustration of mediation pipeline that integrates GWAS-association with tissue-specific cis-eQTLs and tissue-specific gene expression profiles from GTEx portal. Causal mediators of trait-risk are used as input for integrative topological genomic pipelines described in this paper. (C) Volcano plots of tissue-specific mediator genes of polygenic risk for BMI and T2D disease risk. (D) Proportion of mediator enrichments in metabolic tissues. (E) Metabolic tissue contribution to polygenicity for T2D and BMI traits. (F) Illustration showing individual and combinatorial positional hubs for gene mediators of polygenic risk for T2D and BMI. (G) Proportion of genomic clusters with > 5 genes per hotspot from combined topological enrichment for T2D and BMI mediators. (A) Positional hub enrichment p-values and number of genes composing each cluster.



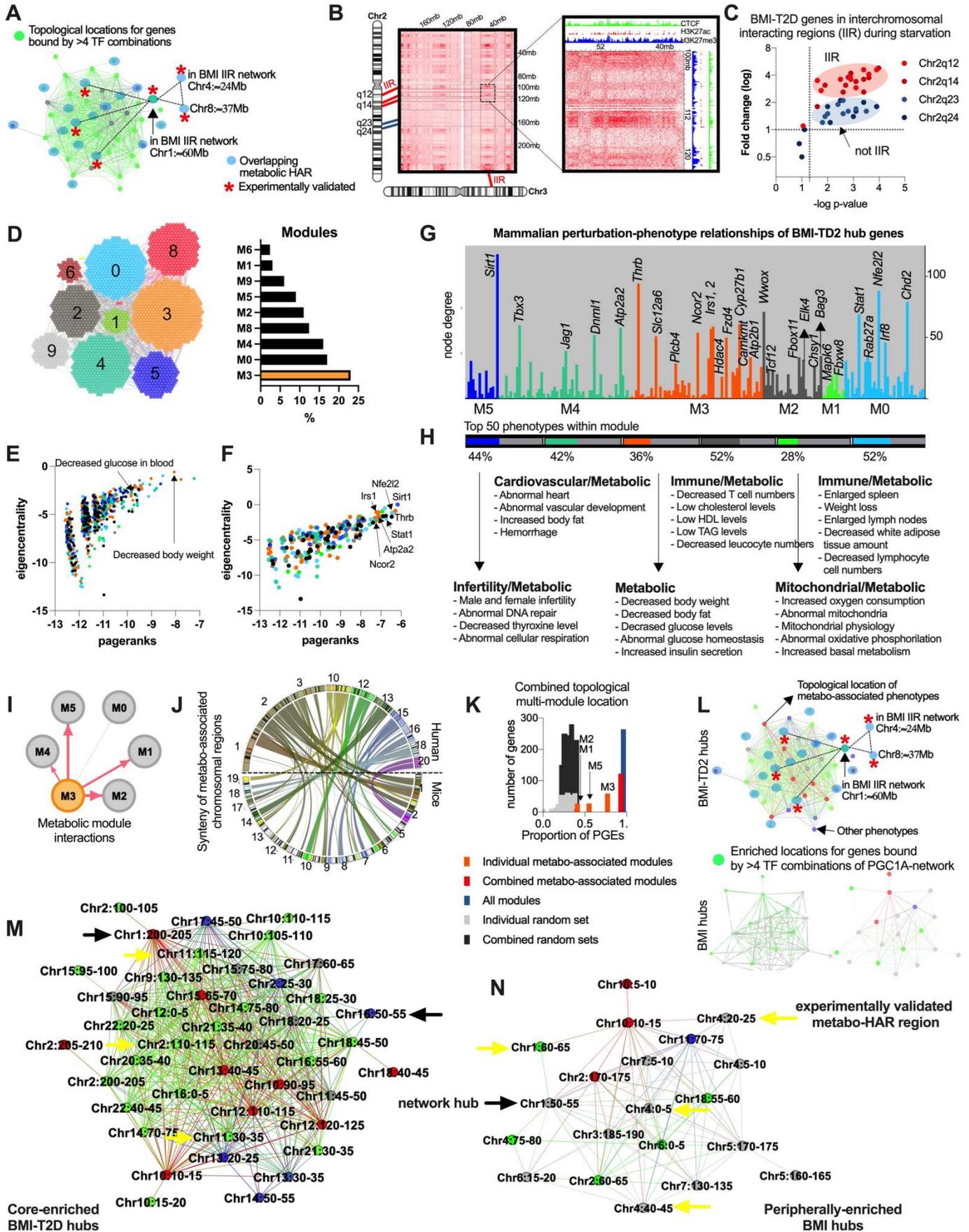
**Fig. S23. Radial divergent nuclear compartmentalization: impact to disease, and transcriptional control** (related to Fig. 5). **(A)** Chromosomal landscape of top disease hubs, trait-specific and shared, as well as overlapping metabolic-HAR regions and experimentally validated domains. **(B)** Chromosomal proportion of shared disease hubs from multi-tissue polygenic scores and causal mediation for BMI and T2D disease variants. **(C)** Radial chromosome organization using GPseq-data (80) and PGEs location for disease variant genes from **A**. In different colors radial nuclear location for chromatin loci. **(D)** Radial proportion enrichment for vista active enhancers, metabolic-HAR regions, frequently interacting regions (FIREs) and random gene set derived PGEs. **(E-F)** DNA-MERFISH data (77) from transcriptionally active loci (on or off), and their relation to different organelles as well as their distance to nuclear lamina. **(G)** TFBS radial correlation (80) with HOX family of transcription factors as top peripherally correlated. Lower panel showing PPI-network of peripherally enriched regulators, and their functional annotation by Reactome **(H)**. **(I)** Illustration showing selected genes from metabolic-HAR regions with specific radial enrichments for experimental validation in starvation. **(J)** QQ plot or normality test from experimental validation of dynamic gene expression comparison for radial enriched loci, peripheral vs core locations. **(K)** CIDER Correlation matrix parameters based on amino acid sequences (59) for ensemble disorder features from peripheral and core enriched regulators. With an asterisk, charged-mediated variations, and with an arrow segregation of charge within disorder regions. **(L)** Illustration of top parameter variation for core-enriched transcriptional regulators. Bars show mean values and error bars indicate SD. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups.



**Fig. S24. Transition from liquid-like to gel-like condensates controls fasting endurance** (related to Fig. 5). **(A)** GFP-PGC1A transcriptional condensate size in microscopy of living cells, after starvation and blocked by 1,6 hexanediol, extreme pH and KCL variations. **(A)** Cytosolic pH variation after proton pump HV1 inhibitor treatment. **(C)** Expression of top regulated genes within metabolic-HAR regions after HV1 treatment. **(C)** Fluorescence recovery mean-variation after refractory period of 6 seconds once bleaching was induced. Comparison in human cells in starvation and with transcriptional regulator perturbations as well as extreme starvation and Hv1 Co-treatments. **(E)** GFP-PGC1A transcriptional condensates relative size in living cells microscopy after starvation and blocked by ATP excess. **(F)** Cytosolic pH variation after fasting and ATP excess. **(G)** PONDR VSL2 disorder score of PGC1A amino acid sequence showing the R1 and R2 consensus disorder regions. **(H)** Net charge per residue parameter for PGC1A sequence. **(I)** Fraction of charge residues and charge segregation across full length PGC1A, PGC1A disorder regions R1 and R2, PGC1A-interacting regulators and mean nuclear core-enriched transcriptional regulators. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.

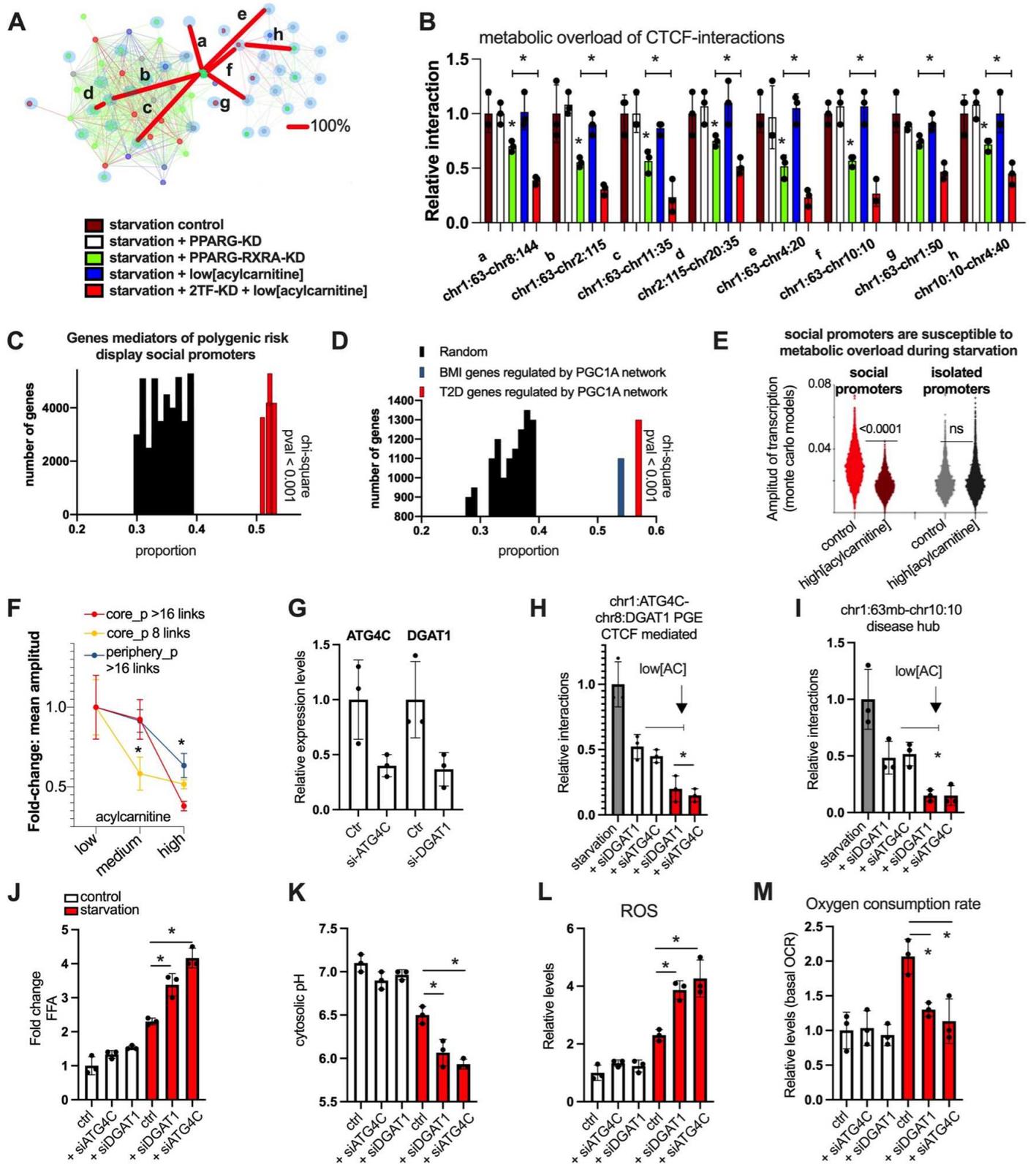


**Fig. S25. Topological systems genetics** (related to Fig. 6). **(A)** Illustration showing genome organization and clusters of co-localization between interacting chromosomal regions, metabolic disease hubs, and metabolic-HAR domains. In the right panel representations of loci-loci interactions between trans-domains and related metabolic-HAR regions. **(B)** Chromosomal landscape displaying social interchromosomal regions (SIR), used to define trans-domains for co-localization with disease hubs and HAR domains. **(C)** Proportion of colocalization between social interchromosomal regions, disease hubs and metabolic-HAR domains. **(D-E)** SIR Network reconstruction of shared T2D-BMI top hubs, unique BMI hubs, and metabolic-HAR domains. **(F)** Network-based analysis of SIR-network. In green showing top hubs in D and E. **(G)** Representative conservation of trans-interacting domains in HiC contact frequency maps (GM12878) for disease hub mediators. Displaying interaction between top hub in T2D-BMI SIR network FTO domain and top driver region in EpiMap integration for metabolic-related phenotypes. **(H)** Intra- and inter-trait relationships among conventional GWAS association studies. Top GWAS with their associated genes and their topological relationships in SIR-disease hub networks.



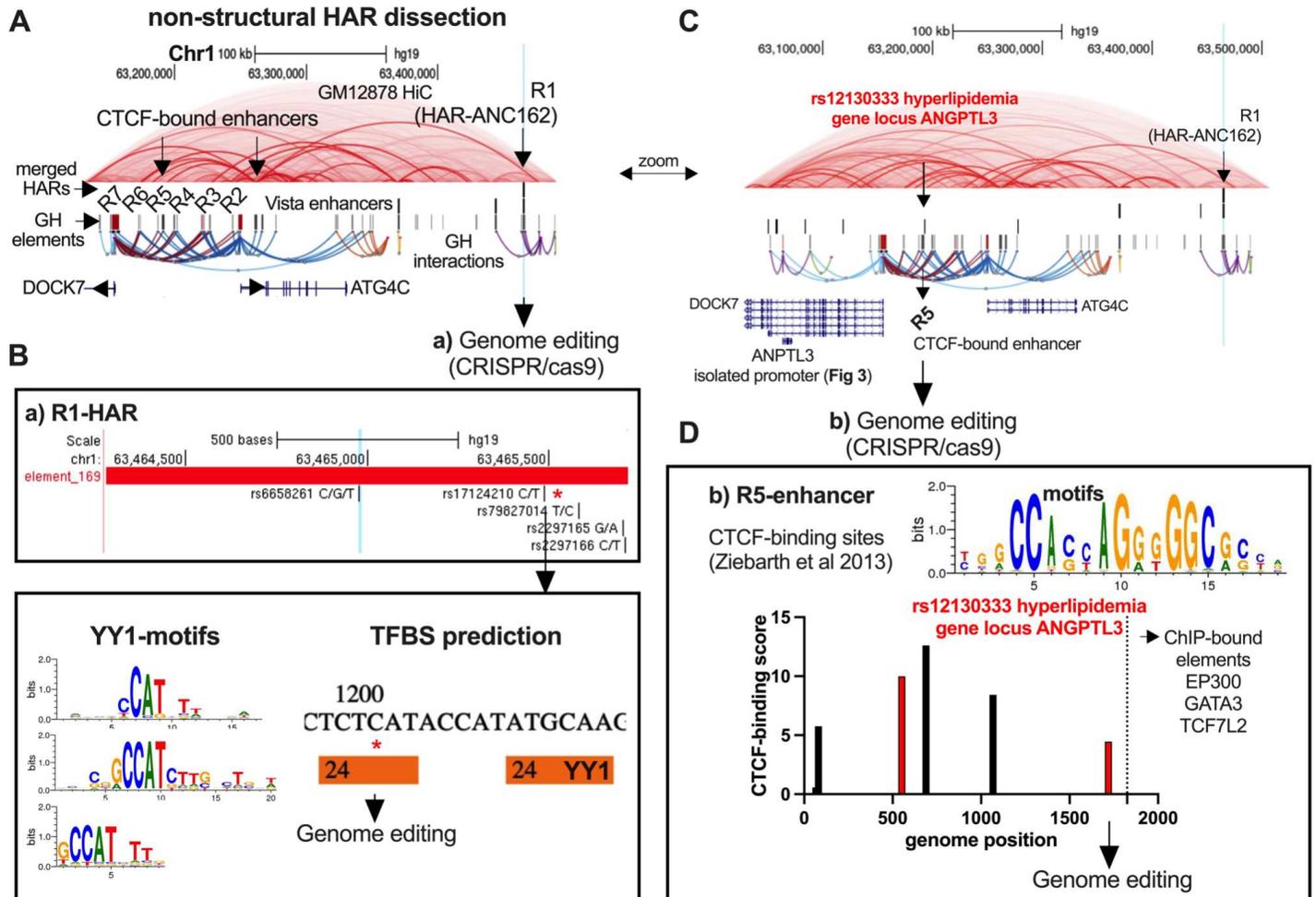
**Fig. S26. Topological systems genetics and integration of mammalian syntenic domains with metabolic phenotype associations (related to Fig. 6).** (A) SIR-disease hub network, highlighting in green location of disease-hub

genes regulated by combination of PGC1A-interacting regulators. With an asterisk experimentally validated metabolic-HAR regions. **(B)** Trans-interacting chromatin regions in HiC contact frequency maps, showing SIR-domains, and neighboring isolated regions, both harbouring metabolic-HAR and PGC1A regulated genes. **(C)** Experimental validation for regions in B comparing their fold-change variation in starvation. **(D)** Module layout of network integration of mammalian perturbation phenotypes for metabolic disease hubs, mediator of polygenic risk. In orange, top module with overall enrichment for metabolic phenotypes. **(E-F)** Network-based prioritization of phenotype and gene drivers for D. **(G)** Node degree for gene drivers in each module and phenotype enrichments **(H)**. **(I)** Inter-module relationships for metabolic phenotype module M3. **(J)** Syntenic conservation in mice for positional gene enrichment from human disease genes driving top interacting modules. **(K)** Combinatorial transcriptional hub enrichments for genes driving phenotypes in interacting modules and located in syntenic regions. **(L)** Colocalization of syntenic regions displaying metabolic phenotype associations in SIR-network. In green, regions regulated by a combination of PGC1A-interacting regulators. **(M-N)** Full SIR-network showing loci with phenotype associations and experimentally validated metabolic-HAR regions.

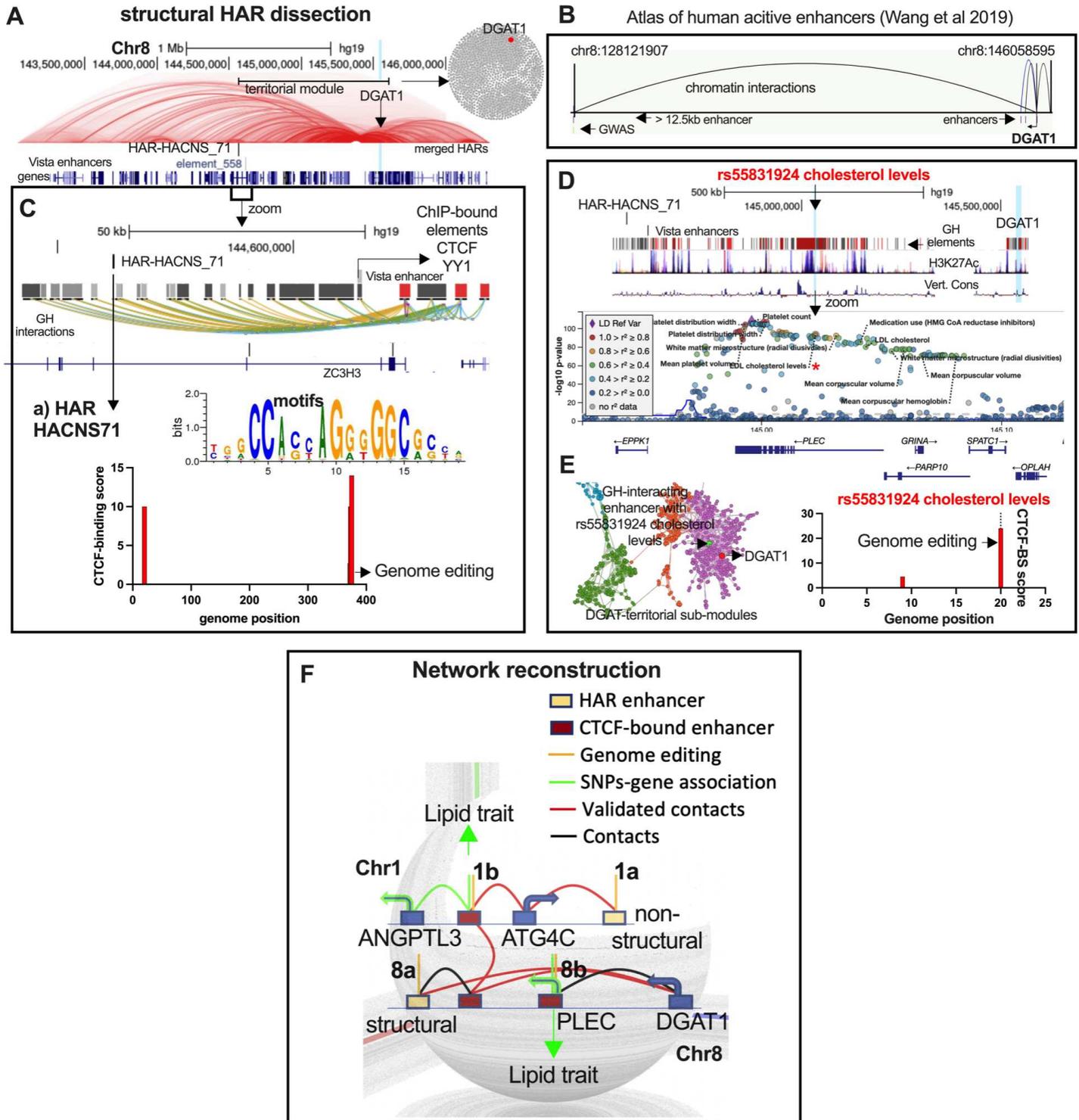


**Fig. S27. Combinatorial binding PGC1A-network to SIR-PRS network** (related to Fig. 6). (A) BMI-T2D SIR- and BMI SIR-networks connected through top-bridging chromatin loci Chr1:60–65mb, harboring metabolic-HAR ATG4C region. Red lines showing starvation-mediated regulation of long-range interactions among diverse loci. (B) CTCF-mediated long-range interactions among loci displayed in A and with combinatorial transcriptional perturbation as well as metabolic overload wti low concentration of acylcarnitine (1 $\mu$ M). (C) Proportion of mediators of polygenic risk displaying social promoters as defined by their links in regulatory interaction networks in Fig 2. (D) Proportion of mediators of polygenic risk displaying social promoters as defined by their links in regulatory interaction networks in Fig 2, and being combinatorially

bound by PGC1A-interacting regulators. **(E)** Amplitude for transcriptional burst models by Monte Carlo for genes within metabolic-HAR domains with either social and isolated promoters during starvation and with metabolic overload (acylcarnitine excess 20uM). **(F)** Mean amplitude of transcriptional models of core and peripheral genes with social promoters (>16 enhancer-promoter interactions) and mid-social promoters (8 interactions) in human adipocytes after starvation and treated with different concentrations of acylcarnitine (low=1uM; medium=10uM; high=20uM). **(G)** Gene expression levels after targeted perturbation with siRNA. **(H)** metabolic-HAR region interactions in starvation and after targeted perturbation of top regulated genes ATG4C and DGAT1, and with low acylcarnitine. **(I)** Long-range interactions for different disease loci hubs. **(J)** Functional assays in the same conditions as in H. Free-fatty acids **(J)**, cytosolic pH **(K)**, oxidative stress **(L)** and oxygen consumption rate by seahorse **(M)**. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.

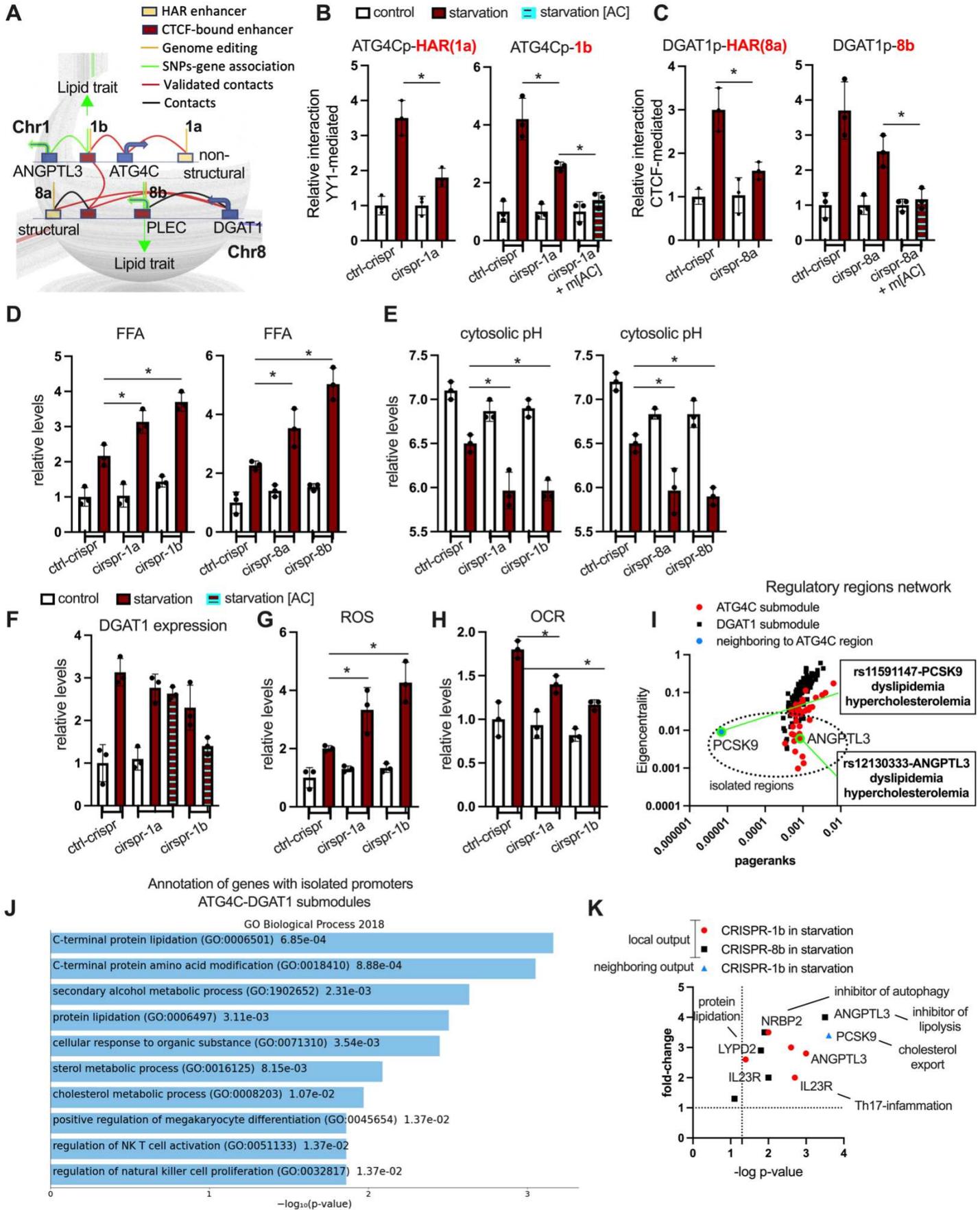


**Fig. S28. Disease circuit dissection and genome editing of metabolic-HAR ATG4C-domain (related to Fig. 6).** **(A)** Human genome track showing top targets within metabolic-HAR domain in chr1:62-63mb: nCHARs (ANC162), regulatory elements and interactions, GM12878-Hi-C data, CTCF-bound elements, and promoter locations for ATG4C and DOCK7. **(B)** HAR-ANC162 (element\_169) displaying variants within the element. With an arrow, location of TFBS for YY1 transcription factor, which was selected for genome editing. **(C)** Human genome track as in A, showing with arrows selected enhancers, in which R5 element (CTCF-bound) harbours a SNP related to hyperlipidemia (rs12130333). **(D)** R5 element was scanned for CTCF-binding sites as described by Ziebarth et al (101). A CTCF-binding site in close proximity to the variant rs12130333 was selected for genome editing.

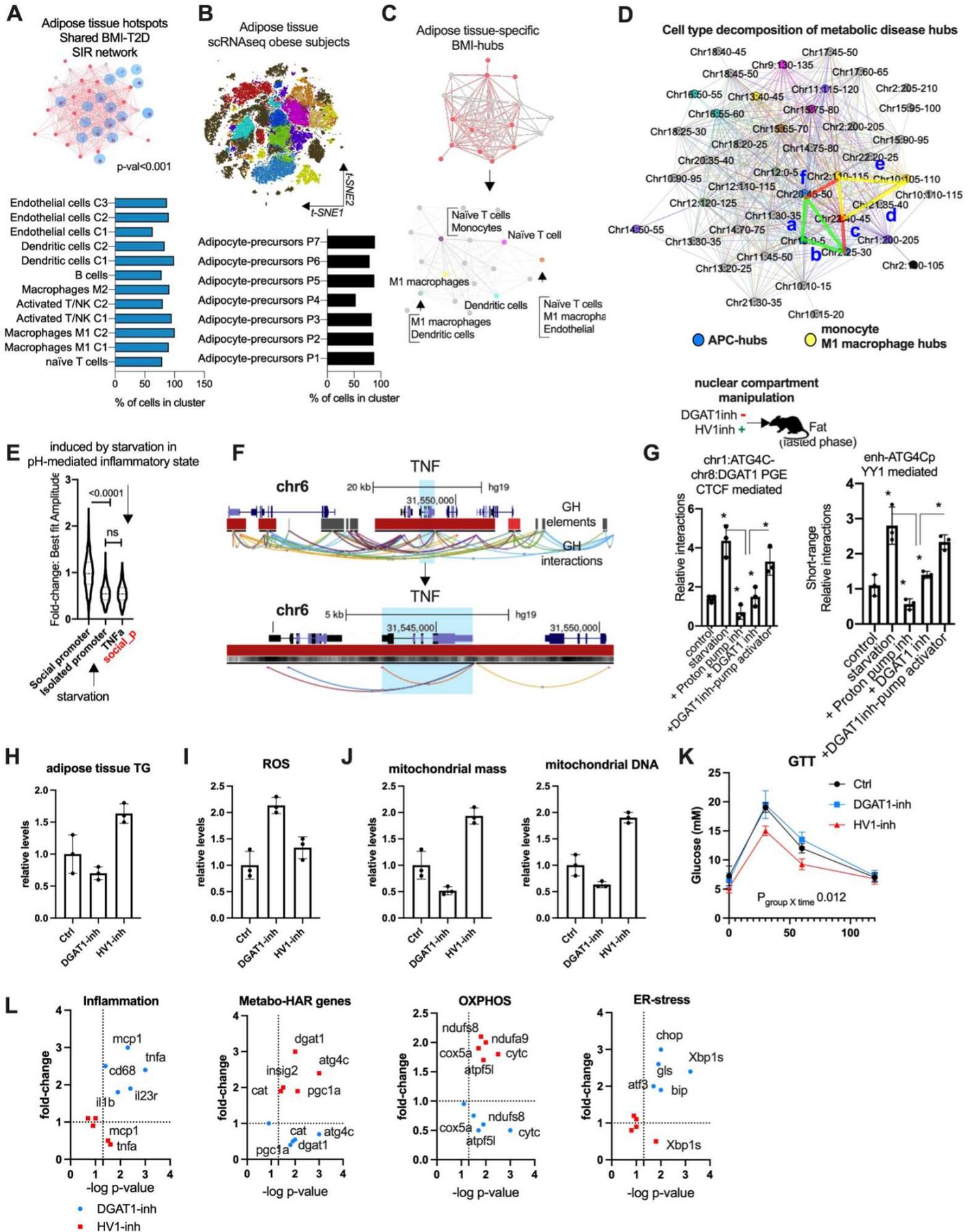


**Fig. S29. Disease circuit dissection and genome editing of metabolic-HAR DGAT1-domain** (related to Fig. 6). (A) Human genome track showing top targets within metabolic-HAR domain in chr8:144-146mb: nCHARs (HACNS\_71), vista regulatory element 558, GM12878-Hi-C data, DGAT1-promoter location, and DGAT1-module reconstruction from regulatory elements interaction network. (B) Enhancer long-range regulation of DGAT1 from the atlas of human active enhancers database (147). (C) HAR-HACNS\_71, vista regulatory element 558, regulatory elements and interactions by GeneHancer. With an arrow to the lowest panel showing CTCF-binding sites in the HAR element, one of which was selected for genome editing. (D) Human genome track as in A, showing with an arrow selected variant related to cholesterol levels in regulatory element, followed in middle panel by all SNPs within region by HugAmp metabolic database and with an asterisk the lipid-related variant rs55831924. (E) Force atlas layout of DGAT1 module and submodules from genome-wide regulatory region interactions. Highlighted in green the lipid-variant and in red DGAT1 promoter element. In the right panel, CTCF-binding sites for the element harbouring the lipid-variant. (F) Full trans-network reconstruction Chr1-Chr8 coregulated metabolic HAR domains. In green arrow, lipid-related variants within

structural enhancers, HAR-locations, target genes, chromatin contacts as well as notation for genome editing interventions.



**Fig. S30. Disease circuit dissection and genome editing of trans-interacting metabolic-HAR domains** (related to Fig. 6). **(A)** Full trans-network reconstruction Chr1-Chr8 coregulated metabolic HAR domains. In green arrow, lipid-related variants within structural enhancers, HAR-locations, target genes, chromatin contacts as well as notation for genome editing interventions. **(B)** Left panel shows relative YY1-mediated loops of HAR-ATG4C\_promoter in control mutation and perturbation of HAR YY1 binding site during starvation. Right panel shows enhancer 1b-ATG4C\_promoter loops in the same conditions (control mutation and HAR-mutation) and with metabolic overload with acylcarnitine (medium concentration of 10uM). **(C)** Left panel shows relative CTCF-mediated loops of HAR-DGAT1\_promoter in control mutation and perturbation of HAR CTCF-binding site during starvation. Right panel shows enhancer 8b-DGAT1\_promoter loops in the same conditions (control mutation and HAR-mutation) and with metabolic overload with acylcarnitine (medium concentration of 10uM). **(D)** Free-fatty acids functional assays in the same conditions comparing control mutations, HAR mutations (a) and mutation in key-structural enhancers (b). **(E)** Cytosolic pH functional assays in the same conditions comparing control mutations, HAR mutations (a) and mutation in key-structural enhancers (b). **(F)** Trans-regulatory effects on chr8 DGAT1 expression from chr1-mutations in starvation and with acylcarnitine excess (10uM). **(G)** Oxidative stress and **(H)** oxygen consumption rate functional assays in the same conditions and in response to mutations in chr1 (a and b). **(I)** Node-prioritization scores for ATG4C, DGAT1 and ATG4C-neighboring regulatory submodules. Highlighted in dotted circle promoter regions with few connections and defined as isolated in Fig. 2, and in green those regions sharing strong genetic association to lipid-related traits. **(J)** Functional annotation of isolated promoters in both submodules. **(K)** Gene expression of top isolated genes in the same conditions as in B comparing control and key-structural mutations. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.



**Fig. S31. Single cell decomposition of metabolic disease hubs in human fat tissue and mammalian in-vivo manipulation of nuclear compartment formation** (related to Fig. 6). **(A)** Social interchromosomal region (SIR) network of trans-interacting chromatin regions and associated metabolic-HAR domains (blue nodes). In red, nodes harbouring fat-specific transcriptional hubs from mediation of polygenic risk for T2D and BMI in fat tissue. Overall modularity enrichment by z-score and p-value <0.05. **(B)** Cluster specific t-SNE plots from human adipose tissue scRNA seq and percentage of cell types within. **(C)** SIR-network and cell-specific decomposition of transcriptional hotspots found in differential expression scRNA seq by positional enrichments. **(D)** Full SIR-network coordinates with different cell-specific hubs. Highlighted in blue and connected by green lines adipose precursor cell hubs. In yellow lines monocyte and M1 macrophages cells specific hubs. **(E)** Mean amplitude from Monte Carlo models of dynamic expression profiles in adipocytes during starvation. Comparing mean amplitude for social promoter gene DOCK7, isolated promoter gene ANGPTL3, and TNF burst kinetics. **(F)** Human genome track picture showing chr6 region harboring TNF gene, enhancer elements and interactions. In the lower panel, TNF gene span with local interactions. **(G)** Top panel shows illustration for acute in vivo intervention by intraperitoneal injections of fasting mice with either DGAT1-inhibitor or HV1-inhibitor (2mg/kg). In lower panels, CTCF-long range and YY1 short-range relative interactions in the same conditions. **(H)** Adipose tissue functional assays after chronic manipulation of nuclear compartment formation in mice treated daily with either inhibitor as in G during 7 days: **(H)** triglycerides, **(I)** oxidative stress, **(J)** mitochondrial mass and mitochondrial DNA. **(K)** Glucose tolerance test in the same conditions as in H. **(L)** Gene expression profile in adipose tissue in the same conditions as in H. Bars show mean values and error bars indicate SEM. Unpaired, two-tailed student's t-test was used when two groups were compared, and ANOVA followed by fisher's least significant difference (LSD) test for post hoc comparisons for multiple groups. \* p-value <0.05.

## Supplementary Tables

### Table S1. Metabolic phenotype comparison across mammals

- A. Metabolic parameters and fasting endurance score in nonprimate mammals
- B. Metabolic parameters and fasting endurance score in primate mammals
- C. Metabolic parameters and fasting endurance score in hominoids
- D. Tissue mass and metabolic parameters across mammals

Link: <https://www.dropbox.com/s/7czlr2dkukutsxl/Table%20S1.xlsx?dl=0>

### Table S2. Differential human specific molecular signatures in fat and brain

- A. GREAT analysis from ATACseq comparison between human-primate in fat: human increase
- B. GREAT analysis from ATACseq comparison between human-primate in fat: human decrease
- C. Fat branch-specific acceleration and positive selection from ATACseq comparison between humans and primates
- D. Accelerated evolution signature genes in brain

Link: <https://www.dropbox.com/s/h0bva3hjq0jx37c/Table%20S2.xlsx?dl=0>

### Table S3. Human-accelerated regulatory regions

- A. List of human accelerated regions coordinates
- B. HAR-associated genes discovery from Capra et al. 2013
- C. Full list of HAR-genes compiled for this study
- D. Functional annotation of HAR genes in brain-specific networks
- E. Functional annotation of HAR genes in fat-specific networks

Link: <https://www.dropbox.com/s/j89lhmskpxoo8x7/Table%20S3.xlsx?dl=0>

### Table S4. Transcriptional regulatory networks of HAR-genes

- A. ChIP-seq enrichment ChIP-atlas of HAR-genes
- B. Adipocytes ChIP-seq enrichment ChIP-atlas of HAR-genes
- C. Neural ChIP-seq enrichment ChIP-atlas of HAR-genes
- D. Stem cells ChIP-seq enrichment ChIP-atlas of HAR-genes
- E. Blood ChIP-seq enrichment ChIP-atlas of HAR-genes
- F. Muscle ChIP-seq enrichment ChIP-atlas of HAR-genes
- G. Liver ChIP-seq enrichment ChIP-atlas of HAR-genes
- H. TFBS enrichment from MSDB
- I. TFBS enrichment from genome browser
- J. ENCODE and ChEA consensus enrichment

<https://www.dropbox.com/s/1sqdr5741pqq6t0/Table%20S4.xlsx?dl=0>

### Table S5. Transcriptional regulatory networks, and PPI reconstitution

- A. Node + 1 network reconstruction of PPIs for HAR-genes regulators
- B. Network topological prioritization of network in A
- C. Full PGC1A-network reconstruction of protein-protein interactions
- D. Network topological prioritization of full PGC1A-network
- E. Reactome functional annotation of HAR cooperative regulators
- F. Gene-ontology functional annotation of HAR cooperative regulators
- G. Fat and brain tissue-specific PPIs for PGC1A-network

<https://www.dropbox.com/s/brp2olmem6o5zgy/Table%20S5.xlsx?dl=0>

### Table S6. Intrinsically disordered regions of HAR-gene regulators

- A. Consensus disordered regions in PGC1A protein
- B. Hydropathy boundary ordered and disordered proteins
- C. Multivalent pi-pi disordered score human proteins
- D. DisProt disordered score human proteins

<https://www.dropbox.com/s/idqxtlylhl1lg24/Table%20S6.xlsx?dl=0>

### Table S7. Phylogenetic conservation of co-activator protein residues

- A. ConSurf amino acid variation of PGC1A residues
- B. Amino acid conservation scores for PGC1A co-activator

<https://www.dropbox.com/s/57k3ql68793bwql/Table%20S7.xlsx?dl=0>

**Table S8. Combinatorial transcriptional regulation of HAR-genes**

- A. Functional interacting transcriptional regulators co-regulated genes by promoter ChIP-seq enrichment
- B. Combinatorial co-regulated HAR-genes (metabolic-HAR genes)

<https://www.dropbox.com/s/0zq6kakrtw5q773/Table%20S8.xlsx?dl=0>

**Table S9. Functional topological loci decomposition of metabolic-HAR regions**

- A. Positional gene enrichments of metabolic-HAR genes
- B. Cytogenetic band enrichments of metabolic-HAR genes
- C. Functional annotation of metabolic-HAR PGE-domains

<https://www.dropbox.com/s/10wlpjohnu4n83u/Table%20S9.xlsx?dl=0>

**Table S10. Compendia of HiC data**

- A. HiC chromatin interacting regions
- B. Information on HiC compendia from Schmitt et al. 2016
- C. Frequently interacting regions FIREs
- D. Conserved TADs

<https://www.dropbox.com/s/uejs9r1h405i5n2/Table%20S10.xlsx?dl=0>

**Table S11. hESC interacting interchromosomal regions**

- A. Network from hESC-HiC interacting chromatin segments and metabolic-HAR regions overlap
- B. Network ID of interacting segments in A
- C. Topological prioritization and network analysis of hESC-HiC interacting chromatin segments

<https://www.dropbox.com/s/4hl6jvb7l87enbb/Table%20S11.xlsx?dl=0>

**Table S12. Genome-wide short-range enhancer-promoter relationships**

- A. GeneHancer database of enhancer-promoter associations.

<https://www.dropbox.com/s/ih9mof2smzyc6os/Table%20S12.xlsx?dl=0>

**Table S13. Genome-wide short-range regulatory elements network**

- A. Regulatory elements associations
- B. Network analyses of regulatory elements associations
- C. Module-based and degree prioritization
- D. Degree-based prioritization of enhancers
- E. Social promoter genes and genomic cluster positions
- F. Nonsocial promoter genes and genomic cluster positions

<https://www.dropbox.com/s/4b4u8hmns6u3el9/Table%20S13.xlsx?dl=0>

**Table S14. metabolic-HAR domains short-range regulatory subnetwork from reconstituted nuclear compartment**

- A. ATG4C-submodule short-range relationships
- B. DGAT1-submodule short-range relationships
- C. PGC1A-submodule short-range relationships
- D. CAT-submodule short-range relationships
- E. ATG4C-submodule territorial ChIP-seq enrichment for social enhancer and promoters
- F. DGAT-submodule territorial ChIP-seq enrichment for social enhancer and promoters
- G. PGC1A-submodule territorial ChIP-seq enrichment for social enhancer and promoters
- H. CAT-submodule short-range relationships
- I. PPI network reconstruction of territorial ChIP-seq enrichment of regulators within ATG4C submodule

<https://www.dropbox.com/s/dcy05kn2hjbcaz2/Table%20S14.xlsx?dl=0>

**Table S15. Active enhancer elements**

- A. Vista active enhancer elements human and mouse coordinates and associated genes
- B. Enhancer identity and human coordinates

<https://www.dropbox.com/s/2tpw0kos0xqyndv/Table%20S15.xlsx?dl=0>

**Table S16. Structural variants, genomic range function and genetics of metabolic-HAR compartments**

- A. Common structural variants with SNVs
- B. Number of SVs in reconstituted nuclear compartment with metabolic-HAR domains
- C. ATG4C metabolic-HAR domain genomic range genetics (chr1:62mb)
- D. INSIG2 metabolic-HAR domain genomic range genetics (chr2:110mb)
- E. PGC1A metabolic-HAR domain genomic range genetics (chr4:25mb)
- F. CAT metabolic-HAR domain genomic range genetics (chr8:38mb)
- G. APOA5 metabolic-HAR domain genomic range genetics (chr11:117mb)
- H. GREAT analysis metabolic-HAR genomic coordinates

<https://www.dropbox.com/s/gun54eshm93xl3u/Table%20S16.xlsx?dl=0>

**Table S17. Epimap network modeling of metabolic-HAR regions**

- A. Epimap all SNPs-enhancer intersections
- B. Epimap all GWAS-SNPs associations
- C. Tissue and trait network topological dependencies
- D. Chromatin loci network topological dependencies
- E. Transcriptional hubs from Epimap-modular dependencies

<https://www.dropbox.com/s/jth6erg3syhkf4c/Table%20S17.xlsx?dl=0>

**Table S18. ChIP-MS structuring candidates and PGC1A-network**

- A. List of proteins identified by chromatin proteomic profiling in mESC dual enhancer-promoter
- B. Interrogation of PPIs for PGC1A-network and ChIP-MS structuring candidates
- C. IDs node summary annotation for B
- D. Functional network annotation
- E. Network topological dependencies

<https://www.dropbox.com/s/c8helx3ck1engfa/Table%20S18.xlsx?dl=0>

**Table S19. Tissue- and cell-specific enhancer activity**

- A. Tissue- and cell-specific enhancer activity
- B. Global mean and sum of activity for each enhancer coordinate
- C. Adipocyte-specific enhancer

<https://www.dropbox.com/s/10ojb9escn0h9zm/Table%20S19.xlsx?dl=0>

**Table S20. Human active enhancers from GRO-seq, PRO-seq and CAGE data**

<https://www.dropbox.com/s/whsch6czar35mrz/Table%20S20.txt?dl=0>

**Table S21. Human adipocyte precursors profiling and brain cortex snRNA**

- A. Cell-specific genes in snRNAseq from human cortex
- B. Human adipocyte precursors cell profiling

<https://www.dropbox.com/s/dnw1anv8jqzkn9e/Table%20S21.xlsx?dl=0>

**Table S22. Synteny of metabolic-HAR regions and mammalian perturbation networks**

- A. Murine transcriptional hubs of metabolic-HAR regions
- B. Metabolic-HAR gene perturbation-phenotype association
- C. Gene network topological dependencies
- D. Phenotype network topological dependencies

<https://www.dropbox.com/s/7aq1nhhnwsbp5fq/Table%20S22.xlsx?dl=0>

**Table S23. Differentially expressed genes in diet-induced obesity across 19 organs**

- A. Mesenteric adipose tissue. B. Brown adipose tissue. C. Visceral adipose tissue. D. Retroperitoneal adipose tissue. E. Subcutaneous adipose tissue. F. Hypothalamus. G. Ileum. H. Jejunum. I. Kidney. J. Liver. K. Stomach. L. Duodenum. M. Soleus. N. Colon. O. Cortex. P. Adrenal gland. Q. Cerebellum. R. Midbrain stem.

<https://www.dropbox.com/s/lmd3k8k8xaiveth/Table%20S23.xlsx?dl=0>

**Table S25. Transcriptional hubs controlled by diet**

- A. Mesenteric adipose tissue. B. Brown adipose tissue. C. Visceral adipose tissue. D. Retroperitoneal adipose

tissue. E. Subcutaneous adipose tissue. F. Hypothalamus. G. Ileum. H. Jejunum. I. Kidney. J. Liver. K. Stomach. L. Duodenum. M. Soleus. N. Colon. O. Cortex. P. Adrenal gland. Q. Cerebellum. R. Midbrain stem.

<https://www.dropbox.com/s/8e2w8wrzxtvxsws/Table%20S24.xlsx?dl=0>

#### **Table S25. Fat-specific transcriptional hubs**

A. Fat-specific transcriptional hubs controlled by diet

<https://www.dropbox.com/s/bq1fe09wox9krkr/Table%20S25.xlsx?dl=0>

#### **Table S26. Single cell topological decomposition of fat transcriptional hubs**

- A. Cell-specific genes in murine adipose from Tabula Muris consortium
- B. Natural killer cell-specific transcriptional hubs in murine adipose tissue
- C. T cell-specific transcriptional hubs in murine adipose tissue
- D. Myeloid cell-specific transcriptional hubs in murine adipose tissue
- E. B cell-specific transcriptional hubs in murine adipose tissue
- F. Mesenchymal cell-specific transcriptional hubs in murine adipose tissue
- G. Endothelial cell-specific transcriptional hubs in murine adipose tissue

<https://www.dropbox.com/s/oesjv972mu0dna/Table%20S26.xlsx?dl=0>

#### **Table S27. Multi-tissue mediators of polygenic risk scores for BMI and T2D**

- A. Tissue-specific mediators of polygenic risk for T2D (BMI-adjusted)
- B. Tissue-specific mediators of polygenic risk for BMI (GIANT)
- C. Genomic hubs for BMI mediators of polygenic risk
- D. Genomic hubs for T2D mediators of polygenic risk

<https://www.dropbox.com/s/efr9d3oq1ihikh4/Table%20S27.xlsx?dl=0>

#### **Table S28. Radial chromatin organization**

- A. GPseq radial chromatin loci coordinates
- B. DNA-MERFISH chromatin organization
- C. Ontology of core and nuclear speckles associated genes with transcriptional burst on from B
- D. Biological enrichment of core and nuclear speckles associated genes with transcriptional burst on

<https://www.dropbox.com/s/n6iayzilav1x8if/Table%20S28.xlsx?dl=0>

#### **Table S29. Radial transcriptional regulators**

- A. TFBS density and GPseq score radial correlations
- B. Functional interacting regulators with peripheral enrichment
- C. Node + 1 network reconstruction of B
- D. Functional interacting regulators with core enrichment
- E. Node + 1 network reconstruction of D

<https://www.dropbox.com/s/ymx5kshsev9kpn1/Table%20S29.xlsx?dl=0>

#### **Table S30. Radial transcriptional condensate features**

- A. Amino acid sequences for core enriched regulators
- B. Amino acid sequences for peripherally enriched regulators
- C. Context-dependent condensate features for core enriched regulators
- D. Context-dependent condensate features for peripherally enriched regulators

<https://www.dropbox.com/s/v69f5yb6p7ap73y/Table%20S30.xlsx?dl=0>

#### **Table S31. Social interchromosomal network: interacting metabolic disease hubs and metabolic-HAR regions**

- A. Social interchromosomal regions: 5mb range interacting scores
- B. Interacting metabolic disease hubs for shared hotspots for BMI and T2D polygenic risk
- C. Interacting metabolic disease hubs coordinates for BMI mediators of polygenic risk
- D. Interacting metabolic disease hubs coordinates for T2D mediators of polygenic risk
- E. Mammalian phenotype integration of gene perturbations for shared mediators of BMI and T2D polygenic risk
- F. Network topological dependencies for mediators
- G. Network topological dependencies for phenotypes

<https://www.dropbox.com/s/47fnm8cqlijs5ki/Table%20S31.xlsx?dl=0>

**Table S32. Interacting metabolic disease hubs regulated by PGC1A-network**

- A. Combinatorial regulation of T2D hotspots by PGC1A-network
- B. Combinatorial regulation of BMI hotspots by PGC1A-network
- C. Gene list > 4 combinatorial PGC1A-network for T2D-BMI mediators of polygenic risk
- D. Genomic clusters by cytogenetic band enrichments
- E. Genomic clusters by positional gene enrichments

<https://www.dropbox.com/s/0b95i0r8x3ng4n4/Table%20S32.xlsx?dl=0>

**Table S33. Single cell topological decomposition of interacting metabolic disease hubs**

- A. Subcutaneous progenitors genes in obese fat scRNAseq
- B. Visceral progenitors genes in obese fat scRNA
- C. Immune cell-specific genes in obese fat scRNAseq
- D. Endothelial cell-specific genes in obese fat scRNAseq
- E. Genomic hubs from fat progenitors scRNAseq genes
- F. Genomic hubs from fat immune scRNAseq genes

<https://www.dropbox.com/s/ml1zqsonmgez16h/Table%20S33.xlsx?dl=0>

## References and notes

1. E. H. Finn, T. Misteli, Molecular basis and biological function of variability in spatial genome organization. *Science*. **365** (2019), doi:10.1126/science.aaw9498.
2. A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, G. H. Karpen, Phase separation drives heterochromatin domain formation. *Nature*. **547**, 241–245 (2017).
3. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A Phase Separation Model for Transcriptional Control. *Cell*. **169**, 13–23 (2017).
4. J. A. Riback, L. Zhu, M. C. Ferrolino, M. Tolbert, D. M. Mitrea, D. W. Sanders, M.-T. Wei, R. W. Kriwacki, C. P. Brangwynne, Composition-dependent thermodynamics of intracellular phase separation. *Nature*. **581**, 209–214 (2020).
5. T. Misteli, The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*. **183**, 28–45 (2020).
6. B. R. Sabari, A. Dall’Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. **361** (2018), doi:10.1126/science.aar3958.
7. A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, C. Zechner, Phase separation provides a mechanism to reduce noise in cells. *Science*. **367**, 464–468 (2020).
8. J. A. Riback, C. D. Katanski, J. L. Kear-Scott, E. V. Pilipenko, A. E. Rojek, T. R. Sosnick, D. A. Drummond, Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell*. **168**, 1028–1040.e19 (2017).
9. A. W. Fritsch, A. F. Diaz-Delgado, O. Adame-Arana, C. Hoegge, M. Mittasch, M. Kreysing, M. Leaver, A. A. Hyman, F. Jülicher, C. A. Weber, Local thermodynamics governs the formation and dissolution of protein condensates in living cells, , doi:10.1101/2021.02.11.430794.
10. A. Tantos, P. Friedrich, P. Tompa, Cold stability of intrinsically disordered proteins. *FEBS Lett*. **583**, 465–469 (2009).
11. R. P. Joyner, J. H. Tang, J. Helenius, E. Dultz, C. Brune, L. J. Holt, S. Huet, D. J. Müller, K. Weis, A glucose-starvation response regulates the diffusion of macromolecules. *Elife*. **5** (2016), doi:10.7554/eLife.09376.
12. M. A. Mourão, J. B. Hakim, S. Schnell, Connecting the dots: the effects of macromolecular crowding on cell physiology. *Biophys. J*. **107**, 2761–2766 (2014).
13. L. B. Persson, V. S. Ambati, O. Brandman, Cellular Control of Viscosity Counters Changes in Temperature and Energy Availability. *Cell*. **183**, 1572–1585.e16 (2020).
14. H. Mostafavi, S. Thiyagarajan, B. S. Stratton, E. Karatekin, J. M. Warner, J. E. Rothman, B. O’Shaughnessy, Entropic forces drive self-organization and membrane fusion by SNARE proteins. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5455–5460 (2017).
15. R. Hancock, K. W. Jeon, *New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals* (Academic Press, 2013).
16. L. Jawerth, E. Fischer-Friedrich, S. Saha, J. Wang, T. Franzmann, X. Zhang, J. Sachweh, M. Ruer, M. Ijavi, S. Saha, J. Mahamid, A. A. Hyman, F. Jülicher, Protein condensates as aging Maxwell fluids. *Science*. **370**, 1317–1323 (2020).
17. A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, J. P. Taylor, Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell*. **163**, 123–133 (2015).
18. G. Hultqvist, E. Åberg, C. Camilloni, G. N. Sundell, E. Andersson, J. Dogan, C. N. Chi, M. Vendruscolo, P. Jemth,

Emergence and evolution of an interaction between intrinsically disordered proteins. *Elife*. **6** (2017), doi:10.7554/eLife.16059.

19. L. Shenhav, D. Zeevi, Resource conservation manifests in the genetic code. *Science*. **370**, 683–687 (2020).
20. J. A. Capra, G. D. Erwin, G. McKinsey, J. L. R. Rubenstein, K. S. Pollard, Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130025 (2013).
21. I. E. Eres, K. Luo, C. J. Hsiao, L. E. Blake, Y. Gilad, Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet.* **15**, e1008278 (2019).
22. H. Pontzer, D. A. Raichlen, A. D. Gordon, K. K. Schroepfer-Walker, B. Hare, M. C. O'Neill, K. M. Muldoon, H. M. Dunsworth, B. M. Wood, K. Isler, J. Burkart, M. Irwin, R. W. Shumaker, E. V. Lonsdorf, S. R. Ross, Primate energy expenditure and life history. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1433–1437 (2014).
23. H. Pontzer, M. H. Brown, D. A. Raichlen, H. Dunsworth, B. Hare, K. Walker, A. Luke, L. R. Dugas, R. Durazo-Arvizu, D. Schoeller, J. Plange-Rhule, P. Bovet, T. E. Forrester, E. V. Lambert, M. E. Thompson, R. W. Shumaker, S. R. Ross, Metabolic acceleration and the evolution of human brain size and life history. *Nature*. **533**, 390–392 (2016).
24. A. Navarrete, C. P. van Schaik, K. Isler, Energetics and the evolution of human brain size. *Nature*. **480**, 91–93 (2011).
25. S. L. Lindstedt, M. S. Boyce, Seasonality, Fasting Endurance, and Body Size in Mammals. *The American Naturalist*. **125** (1985), pp. 873–878.
26. D. Swain-Lenz, A. Berrio, A. Safi, G. E. Crawford, G. A. Wray, Comparative Analyses of Chromatin Landscape in White Adipose Tissue Suggest Humans May Have Less Beigeing Potential than Other Primates. *Genome Biol. Evol.* **11**, 1997–2008 (2019).
27. S. Berto, I. Mendizabal, N. Usui, K. Toriumi, P. Chatterjee, C. Douglas, C. A. Tamminga, T. M. Preuss, S. V. Yi, G. Konopka, Accelerated evolution of oligodendrocytes in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24334–24342 (2019).
28. C. S. Greene, A. Krishnan, A. K. Wong, E. Ricciotti, R. A. Zelaya, D. S. Himmelstein, R. Zhang, B. M. Hartmann, E. Zaslavsky, S. C. Sealfon, D. I. Chasman, G. A. FitzGerald, K. Dolinski, T. Grosser, O. G. Troyanskaya, Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47**, 569–576 (2015).
29. K. S. Pollard, S. R. Salama, N. Lambert, M.-A. Lambot, S. Coppens, J. S. Pedersen, S. Katzman, B. King, C. Onodera, A. Siepel, A. D. Kern, C. Dehay, H. Igel, M. Ares Jr, P. Vanderhaeghen, D. Haussler, An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. **443**, 167–172 (2006).
30. S. Uebbing, J. Gockley, S. K. Reilly, A. A. Kocher, E. Geller, N. Gandotra, C. Scharfe, J. Cotney, J. P. Noonan, Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2007049118.
31. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science*. **357** (2017), doi:10.1126/science.aaf4382.
32. V. Martínez-Redondo, A. T. Pettersson, J. L. Ruas, The hitchhiker's guide to PGC-1 $\alpha$  isoform structure and biological functions. *Diabetologia*. **58**, 1969–1977 (2015).
33. P. Puigserver, Z. Wu, C. W. Park, R. Graves, M. Wright, B. M. Spiegelman, A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell*. **92**, 829–839 (1998).
34. H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–50 (2016).
35. K. You, Q. Huang, C. Yu, B. Shen, C. Sevilla, M. Shi, H. Hermjakob, Y. Chen, T. Li, PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.* **48**, D354–D359 (2020).
36. S. Oki, T. Ohta, G. Shioi, H. Hatanaka, O. Ogasawara, Y. Okuda, H. Kawaji, R. Nakaki, J. Sese, C. Meno, ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19** (2018), doi:10.15252/embr.201846255.
37. O. Argüello-Miranda, Y. Liu, N. E. Wood, P. Kositangool, A. Doncic, Integration of Multiple Metabolic Signals

- Determines Cell Fate Prior to Commitment. *Mol. Cell.* **71**, 733–744.e11 (2018).
38. H. A. Shaban, A. Seeber, Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Res.* **48**, 3423–3434 (2020).
  39. K. De Preter, R. Barriot, F. Speleman, J. Vandesompele, Y. Moreau, Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. *Nucleic Acids Res.* **36**, e43 (2008).
  40. A. D. Schmitt, M. Hu, I. Jung, Z. Xu, Y. Qiu, C. L. Tan, Y. Li, S. Lin, Y. Lin, C. L. Barr, B. Ren, A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
  41. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* **485**, 376–380 (2012).
  42. S. Kaufmann, C. Fuchs, M. Gonik, E. E. Khrameeva, A. A. Mironov, D. Frishman, Inter-chromosomal contact networks provide insights into Mammalian chromatin organization. *PLoS One.* **10**, e0126125 (2015).
  43. S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, D. Cohen, GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database.* **2017** (2017), doi:10.1093/database/bax028.
  44. C. H. Eskiw, N. F. Cope, I. Clay, S. Schoenfelder, T. Nagano, P. Fraser, Transcription factories and nuclear organization of the genome. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 501–506 (2010).
  45. D. J. Watts, S. H. Strogatz, Collective dynamics of “small-world” networks. *Nature.* **393**, 440–442 (1998).
  46. Z. Zhang, J. Zhang, A big world inside small-world networks. *PLoS One.* **4**, e5686 (2009).
  47. A.-L. Barabási, *Network Science* (Cambridge University Press, 2016).
  48. G.-S. Peng, S.-Y. Tan, J. Wu, P. Holme, Trade-offs between robustness and small-world effect in complex networks. *Sci. Rep.* **6**, 37317 (2016).
  49. G. Lawyer, Understanding the influence of all nodes in a network. *Sci. Rep.* **5**, 8665 (2015).
  50. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–92 (2007).
  51. Z. Tang, O. J. Luo, X. Li, M. Zheng, J. J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Ruszczycycki, P. Michalski, E. Piecuch, P. Wang, D. Wang, S. Z. Tian, M. Penrad-Mobayed, L. M. Sachs, X. Ruan, C.-L. Wei, E. T. Liu, G. M. Wilczynski, D. Plewczynski, G. Li, Y. Ruan, CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell.* **163**, 1611–1627 (2015).
  52. M. Wlasnowolski, M. Sadowski, T. Czarnota, K. Jodkowska, P. Szalaj, Z. Tang, Y. Ruan, D. Plewczynski, 3D-GNOME 2.0: a three-dimensional genome modeling engine for predicting structural variation-driven alterations of chromatin spatial structure in the human genome. *Nucleic Acids Res.* **48**, W170–W176 (2020).
  53. J. Weischenfeldt, O. Symmons, F. Spitz, J. O. Korb, Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* **14**, 125–138 (2013).
  54. G. H. Perry, F. Yang, T. Marques-Bonet, C. Murphy, T. Fitzgerald, A. S. Lee, C. Hyland, A. C. Stone, M. E. Hurles, C. Tyler-Smith, E. E. Eichler, N. P. Carter, C. Lee, R. Redon, Copy number variation and evolution in humans and chimpanzees. *Genome Res.* **18**, 1698–1710 (2008).
  55. R. L. Collins, H. Brand, K. J. Karczewski, X. Zhao, J. Alföldi, L. C. Francioli, A. V. Khera, C. Lowther, L. D. Gauthier, H. Wang, N. A. Watts, M. Solomonson, A. O’Donnell-Luria, A. Baumann, R. Munshi, M. Walker, C. W. Whelan, Y. Huang, T. Brookings, T. Sharpe, M. R. Stone, E. Valkanas, J. Fu, G. Tiao, K. M. Laricchia, V. Ruano-Rubio, C. Stevens, N. Gupta, C. Cusick, L. Margolin, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, K. D. Taylor, H. J. Lin, S. S. Rich, W. S. Post, Y.-D. I. Chen, J. I. Rotter, C. Nusbaum, A. Philippakis, E. Lander, S. Gabriel, B. M. Neale, S. Kathiresan, M. J. Daly, E. Banks, D. G. MacArthur, M. E. Talkowski, A structural variation reference for medical and population genetics. *Nature.* **581**, 444–451 (2020).
  56. C. J. Willer, Y. Li, G. R. Abecasis, METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* **26**, 2190–2191 (2010).

57. C. A. Boix, B. T. James, Y. P. Park, W. Meuleman, M. Kellis, Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature*. **590**, 300–307 (2021).
58. H. Yoo, C. Triandafillou, D. A. Drummond, Cellular sensing by phase separation: Using the process, not just the products. *J. Biol. Chem.* **294**, 7151–7159 (2019).
59. A. S. Holehouse, J. Ahad, R. K. Das, R. V. Pappu, CIDER: Classification of Intrinsically Disordered Ensemble Regions. *Biophysical Journal*. **108** (2015), p. 228a.
60. N. Khanna, Y. Zhang, J. S. Lucas, O. K. Dudko, C. Murre, Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions. *Nat. Commun.* **10**, 2771 (2019).
61. X. Ji, D. B. Dadon, B. J. Abraham, T. I. Lee, R. Jaenisch, J. E. Bradner, R. A. Young, Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3841–3846 (2015).
62. A. S. Weintraub, C. H. Li, A. V. Zamudio, A. A. Sigova, N. M. Hannett, D. S. Day, B. J. Abraham, M. A. Cohen, B. Nabet, D. L. Buckley, Y. E. Guo, D. Hnisz, R. Jaenisch, J. E. Bradner, N. S. Gray, R. A. Young, YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*. **171**, 1573–1588.e28 (2017).
63. B. R. Sabari, A. Dall’Agnese, R. A. Young, Biomolecular Condensates in the Nucleus. *Trends Biochem. Sci.* **45**, 961–977 (2020).
64. F. Erdel, K. Rippe, Formation of Chromatin Subcompartments by Phase Separation. *Biophys. J.* **114**, 2262–2270 (2018).
65. M. Conte, L. Fiorillo, S. Bianco, A. M. Chiariello, A. Esposito, M. Nicodemi, Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nat. Commun.* **11**, 3289 (2020).
66. B. Levine, G. Kroemer, Biological Functions of Autophagy Genes: A Disease Perspective. *Cell*. **176**, 11–42 (2019).
67. T. B. Nguyen, S. M. Louie, J. R. Daniele, Q. Tran, A. Dillin, R. Zoncu, D. K. Nomura, J. A. Olzmann, DGAT1-Dependent Lipid Droplet Biogenesis Protects Mitochondrial Function during Starvation-Induced Autophagy. *Dev. Cell*. **42**, 9–21.e5 (2017).
68. C. Chitraju, N. Mejhert, J. T. Haas, L. G. Diaz-Ramirez, C. A. Grueter, J. E. Imbriglio, S. Pinto, S. K. Koliwad, T. C. Walther, R. V. Farese Jr, Triglyceride Synthesis by DGAT1 Protects Adipocytes from Lipid-Induced ER Stress during Lipolysis. *Cell Metab.* **26**, 407–418.e3 (2017).
69. C. S. McCoin, T. A. Knotts, S. H. Adams, Acylcarnitines--old actors auditioning for new roles in metabolic physiology. *Nat. Rev. Endocrinol.* **11**, 617–625 (2015).
70. K. Huang, Y. Li, A. R. Shim, R. K. A. Virk, V. Agrawal, A. Eshein, R. J. Nap, L. M. Almassalha, V. Backman, I. Szleifer, Physical and data structure of 3D genome. *Sci Adv.* **6**, eaay4055 (2020).
71. T. Gao, J. Qian, EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.* **48**, D58–D64 (2020).
72. Y. Guo, Q. Xu, D. Canzio, J. Shou, J. Li, D. U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M. Q. Zhang, B. Ren, A. R. Krainer, T. Maniatis, Q. Wu, CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*. **162**, 900–910 (2015).
73. A. J. M. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, Å. Segerstolpe, C. M. Rivera, B. Ren, R. Sandberg, Genomic encoding of transcriptional burst kinetics. *Nature*. **565**, 251–254 (2019).
74. T. Fukaya, B. Lim, M. Levine, Enhancer Control of Transcriptional Bursting. *Cell*. **166**, 358–368 (2016).
75. A. Raajendiran, G. Ooi, J. Bayliss, P. E. O’Brien, R. B. Schittenhelm, A. K. Clark, R. A. Taylor, M. S. Rodeheffer, P. R. Burton, M. J. Watt, Identification of Metabolically Distinct Adipocyte Progenitor Cells in Human Adipose Tissues. *Cell Rep.* **27**, 1528–1540.e7 (2019).
76. H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang, A. J. Martorell, R. M. Ransohoff, B. P. Hafler, D. A. Bennett, M. Kellis, L.-H. Tsai, Single-cell transcriptomic analysis of Alzheimer’s disease. *Nature*. **570**, 332–337 (2019).

77. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell*. **182**, 1641–1659.e26 (2020).
78. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. **562**, 367–372 (2018).
79. Y. Park, A. Sarkar, L. He, J. Davilla-Velderrain, P. L. De Jager, M. Kellis, Causal gene inference by multivariate mediation analysis in Alzheimer's disease. *bioRxiv* (2017), p. 219428.
80. G. Girelli, J. Custodio, T. Kallas, F. Agostini, E. Wernersson, B. Spanjaard, A. Mota, S. Kolbeinsdottir, E. Gelali, N. Crosetto, M. Bienko, GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* **38**, 1184–1193 (2020).
81. H. Hou, Y. Zhao, C. Li, M. Wang, X. Xu, Y. Jin, Single-cell pH imaging and detection for pH profiling and label-free rapid identification of cancer-cells. *Sci. Rep.* **7**, 1759 (2017).
82. P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q.-X. Jiang, B. T. Nixon, M. K. Rosen, Phase transitions in the assembly of multivalent signalling proteins. *Nature*. **483**, 336–340 (2012).
83. P. Chien, L. M. Gierasch, Challenges and dreams: physics of weak interactions essential to life. *Mol. Biol. Cell*. **25**, 3474–3477 (2014).
84. H. Pang, J. Li, H. Du, Y. Gao, J. Lv, Y. Liu, S. J. Li, Loss of voltage-gated proton channel Hv1 leads to diet-induced obesity in mice. *BMJ Open Diabetes Res Care*. **8** (2020), doi:10.1136/bmjdr-2019-000951.
85. J. Dekker, T. Misteli, Long-Range Chromatin Interactions. *Cold Spring Harb. Perspect. Biol.* **7**, a019356 (2015).
86. A. Belyaeva, S. Venkatachalapathy, M. Nagarajan, G. V. Shivashankar, C. Uhler, Network analysis identifies chromosome intermingling regions as regulatory hotspots for transcription. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 13714–13719 (2017).
87. S. A. Quinodoz, N. Ollikainen, B. Tabak, A. Palla, J. M. Schmidt, E. Detmar, M. M. Lai, A. A. Shishkin, P. Bhat, Y. Takei, V. Trinh, E. Aznauryan, P. Russell, C. Cheng, M. Jovanovic, A. Chow, L. Cai, P. McDonel, M. Garber, M. Guttman, Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*. **174**, 744–757.e24 (2018).
88. M. Abifadel, M. Varret, J.-P. Rabès, D. Allard, K. Ouguerram, M. Devillers, C. Cruaud, S. Benjannet, L. Wickham, D. Erlich, A. Derré, L. Villéger, M. Farnier, I. Beucler, E. Bruckert, J. Chambaz, B. Chanu, J.-M. Lecerf, G. Luc, P. Moulin, J. Weissenbach, A. Prat, M. Krempf, C. Junien, N. G. Seidah, C. Boileau, Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).
89. J. Wang, M. R. Ban, G. Y. Zou, H. Cao, T. Lin, B. A. Kennedy, S. Anand, S. Yusuf, M. W. Huff, R. L. Pollex, R. A. Hegele, Polygenic determinants of severe hypertriglyceridemia. *Hum. Mol. Genet.* **17**, 2894–2899 (2008).
90. J. Vijay, M.-F. Gauthier, R. L. Biswell, D. A. Louiselle, J. J. Johnston, W. A. Cheung, B. Belden, A. Pramatarova, L. Biertho, M. Gibson, M.-M. Simon, H. Djambazian, A. Staffa, G. Bourque, A. Laitinen, J. Nystedt, M.-C. Vohl, J. D. Fraser, T. Pastinen, A. Tchernof, E. Grundberg, Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. *Nat Metab.* **2**, 97–109 (2020).
91. D. Nicetto, G. Donahue, T. Jain, T. Peng, S. Sidoli, L. Sheng, T. Montavon, J. S. Becker, J. M. Grindheim, K. Blahnik, B. A. Garcia, K. Tan, R. Bonasio, T. Jenuwein, K. S. Zaret, H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science*. **363**, 294–297 (2019).
92. T. Krausgruber, N. Fortelny, V. Fife-Gernedl, M. Senekowitsch, L. C. Schuster, A. Lercher, A. Nemc, C. Schmidl, A. F. Rendeiro, A. Bergthaler, C. Bock, Structural cells are key regulators of organ-specific immune responses. *Nature*. **583**, 296–302 (2020).
93. G. S. Hotamisligil, N. S. Shargill, B. M. Spiegelman, Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science*. **259**, 87–91 (1993).
94. H.-C. Chang, L. Guarente, SIRT1 and other sirtuins in metabolism. *Trends Endocrinol. Metab.* **25**, 138–145 (2014).

95. F. He, X. Ru, T. Wen, NRF2, a Transcription Factor for Stress Response and Beyond. *Int. J. Mol. Sci.* **21** (2020), doi:10.3390/ijms21134777.
96. A. C. P. Thirone, C. Huang, A. Klip, Tissue-specific roles of IRS proteins in insulin signaling and glucose transport. *Trends Endocrinol. Metab.* **17**, 72–78 (2006).
97. R. Tuyeras, L. Z. Agudelo, S. P. Ram, A. Loon, B. Kutlu, K. Grove, M. Kellis, Cellular intelligence: reaching dynamic specialization through non-equilibrium multi-scale compartmentalization.
98. S. A. Murray, L. S. Dalbøge, K. Baquero, C. A. Sanford, A. Misquith, A. J. Mercer, T. H. Meek, M. Guldbrandt, B. Andersen, P. Kievit, K. L. Grove, B. Kutlu, Whole transcriptome analysis and validation of metabolic pathways in subcutaneous adipose tissues during FGF21-induced weight loss in non-human primates. *Sci. Rep.* **10**, 7287 (2020).
99. S. Kajimura, P. Seale, K. Kubota, E. Lunsford, J. V. Frangioni, S. P. Gygi, B. M. Spiegelman, Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex. *Nature.* **460**, 1154–1158 (2009).
100. J. L. Ruas, J. P. White, R. R. Rao, S. Kleiner, K. T. Brannan, B. C. Harrison, N. P. Greene, J. Wu, J. L. Estall, B. A. Irving, I. R. Lanza, K. A. Rasbach, M. Okutsu, K. S. Nair, Z. Yan, L. A. Leinwand, B. M. Spiegelman, A PGC-1 $\alpha$  isoform induced by resistance training regulates skeletal muscle hypertrophy. *Cell.* **151**, 1319–1331 (2012).
101. J. D. Ziebarth, A. Bhattacharya, Y. Cui, CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res.* **41**, D188–94 (2013).
102. M. Claussnitzer, S. N. Dankel, K.-H. Kim, G. Quon, W. Meuleman, C. Haugen, V. Glunk, I. S. Sousa, J. L. Beaudry, V. Puviindran, N. A. Abdennur, J. Liu, P.-A. Svensson, Y.-H. Hsu, D. J. Drucker, G. Mellgren, C.-C. Hui, H. Hauner, M. Kellis, FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
103. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).
104. K. Luck, D.-K. Kim, L. Lambourne, K. Spirohn, B. E. Begg, W. Bian, R. Brignall, T. Cafarelli, F. J. Campos-Laborie, B. Charlotiaux, D. Choi, A. G. Coté, M. Daley, S. Deimling, A. Desbuleux, A. Dricot, M. Gebbia, M. F. Hardy, N. Kishore, J. J. Knapp, I. A. Kovács, I. Lemmens, M. W. Mee, J. C. Mellor, C. Pollis, C. Pons, A. D. Richardson, S. Schlabach, B. Teeking, A. Yadav, M. Babor, D. Balcha, O. Basha, C. Bowman-Colin, S.-F. Chin, S. G. Choi, C. Colabella, G. Coppin, C. D'Amata, D. De Ridder, S. De Rouck, M. Duran-Frigola, H. Ennajdaoui, F. Goebels, L. Goehring, A. Gopal, G. Haddad, E. Hatchi, M. Helmy, Y. Jacob, Y. Kassa, S. Landini, R. Li, N. van Lieshout, A. MacWilliams, D. Markey, J. N. Paulson, S. Rangarajan, J. Rasla, A. Rayhan, T. Rolland, A. San-Miguel, Y. Shen, D. Sheykhkarimli, G. M. Sheynkman, E. Simonovsky, M. Taşan, A. Tejada, V. Tropepe, J.-C. Twizere, Y. Wang, R. J. Weatheritt, J. Weile, Y. Xia, X. Yang, E. Yeger-Lotem, Q. Zhong, P. Aloy, G. D. Bader, J. De Las Rivas, S. Gaudet, T. Hao, J. Rak, J. Tavernier, D. E. Hill, M. Vidal, F. P. Roth, M. A. Calderwood, A reference map of the human binary protein interactome. *Nature.* **580**, 402–408 (2020).
105. R. Oughtred, J. Rust, C. Chang, B.-J. Breitkreutz, C. Stark, A. Willems, L. Boucher, G. Leung, N. Kolas, F. Zhang, S. Dolma, J. Coulombe-Huntington, A. Chatr-Aryamontri, K. Dolinski, M. Tyers, The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200 (2021).
106. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
107. Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, Predicting intrinsic disorder from amino acid sequence. *Proteins.* **53 Suppl 6**, 566–572 (2003).
108. R. Pancsa, W. Vranken, B. Mészáros, Computational resources for identifying and describing proteins driving liquid-liquid phase separation. *Brief. Bioinform.* (2021), doi:10.1093/bib/bbaa408.
109. Z. Dosztányi, B. Mészáros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics.* **25**, 2745–2746 (2009).
110. B. Mészáros, G. Erdos, Z. Dosztányi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).

111. R. M. Vernon, P. A. Chong, B. Tsang, T. H. Kim, A. Bah, P. Farber, H. Lin, J. D. Forman-Kay, Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife*. **7** (2018), doi:10.7554/eLife.31486.
112. V. N. Uversky, J. R. Gillespie, A. L. Fink, Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins*. **41**, 415–427 (2000).
113. R. K. Das, R. V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13392–13397 (2013).
114. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
115. T. Pupko, I. Pe'er, M. Hasegawa, D. Graur, N. Friedman, A branch-and-bound algorithm for the inference of ancestral amino-acid sequences when the replacement rate varies among sites: Application to the evolution of five gene families. *Bioinformatics*. **18**, 1116–1123 (2002).
116. A. Biegert, J. Söding, Sequence context-specific profiles for homology searching. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3770–3775 (2009).
117. I. Mayrose, D. Graur, N. Ben-Tal, T. Pupko, Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol. Biol. Evol.* **21**, 1781–1791 (2004).
118. P. W. Rose, A. Prlić, C. Bi, W. F. Bluhm, C. H. Christie, S. Dutta, R. K. Green, D. S. Goodsell, J. D. Westbrook, J. Woo, J. Young, C. Zardecki, H. M. Berman, P. E. Bourne, S. K. Burley, The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.* **43**, D345–56 (2015).
119. M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*. **9**, 173–175 (2011).
120. A. Veidenberg, A. Medlar, A. Löytynoja, Wasabi: An Integrated Platform for Evolutionary Sequence Analysis and Data Visualization. *Mol. Biol. Evol.* **33**, 1126–1130 (2016).
121. M. Kateri, *Contingency Table Analysis: Methods and Implementation Using R* (Springer, 2014).
122. M. P. Creighton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
123. J. R. Dixon, D. U. Gorkin, B. Ren, Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell.* **62**, 668–680 (2016).
124. S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. **159**, 1665–1680 (2014).
125. N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L. Aiden, Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101 (2016).
126. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).
127. E. Crane, Q. Bian, R. P. McCord, B. R. Lajoie, B. S. Wheeler, E. J. Ralston, S. Uzawa, J. Dekker, B. J. Meyer, Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*. **523**, 240–244 (2015).
128. F. Ay, T. L. Bailey, W. S. Noble, Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
129. R. Fang, M. Yu, G. Li, S. Chee, T. Liu, A. D. Schmitt, B. Ren, Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* **26**, 1345–1348 (2016).
130. M. Hu, K. Deng, S. Selvaraj, Z. Qin, B. Ren, J. S. Liu, HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics*. **28**, 3131–3133 (2012).

131. M. Imakaev, G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker, L. A. Mirny, Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods*. **9**, 999–1003 (2012).
132. E. Yaffe, A. Tanay, Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* **43**, 1059–1065 (2011).
133. K. Kruse, S. Sewitz, M. M. Babu, A complex network framework for unbiased statistical analyses of DNA-DNA contact maps. *Nucleic Acids Res.* **41**, 701–710 (2013).
134. R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhashi, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A. R. R. Forrest, P. Carninci, M. Rehli, A. Sandelin, An atlas of active enhancers across human cell types and tissues. *Nature*. **507**, 455–461 (2014).
135. H.-J. Westra, M. J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, M. W. Christiansen, B. P. Fairfax, K. Schramm, J. E. Powell, A. Zhernakova, D. V. Zhernakova, J. H. Veldink, L. H. Van den Berg, J. Karjalainen, S. Withoff, A. G. Uitterlinden, A. Hofman, F. Rivadeneira, P. A. C. 't Hoen, E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez, M. A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Völker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dacey, S. A. Gharib, D. A. Enquobahrie, T. Lumley, G. W. Montgomery, S. Makino, H. Prokisch, C. Herder, M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R. C. Jansen, P. M. Visscher, J. C. Knight, B. M. Psaty, S. Ripatti, A. Teumer, T. M. Frayling, A. Metspalu, J. B. J. van Meurs, L. Franke, Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
136. B. Mifsud, F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder, L. Ferreira, S. W. Wingett, S. Andrews, W. Grey, P. A. Ewels, B. Herman, S. Happe, A. Higgs, E. LeProust, G. A. Follows, P. Fraser, N. M. Luscombe, C. S. Osborne, Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
137. W. J. Shim, E. Sinniah, J. Xu, B. Vitrinel, M. Alexanian, G. Andreatti, S. Shen, Y. Sun, B. Balderson, C. Boix, G. Peng, N. Jing, Y. Wang, M. Kellis, P. P. L. Tam, A. Smith, M. Piper, L. Christiaen, Q. Nguyen, M. Bodén, N. J. Palpant, Conserved Epigenetic Regulatory Logic Infers Genes Governing Cell Identity. *Cell Syst.* **11**, 625–639.e13 (2020).
138. O. Oluwadare, Y. Zhang, J. Cheng, A maximum likelihood algorithm for reconstructing 3D structures of human chromosomes from chromosomal contact data. *BMC Genomics.* **19**, 161 (2018).
139. O. Oluwadare, M. Highsmith, D. Turner, E. Lieberman Aiden, J. Cheng, GSDB: a database of 3D chromosome and genome structures reconstructed from Hi-C data. *BMC Mol Cell Biol.* **21**, 60 (2020).
140. T. Trieu, J. Cheng, 3D genome structure modeling by Lorentzian objective function. *Nucleic Acids Res.* **45**, 1049–1058 (2017).
141. M. J. Fullwood, C.-L. Wei, E. T. Liu, Y. Ruan, Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* **19**, 521–532 (2009).
142. J. M. Downen, Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham, L. N. Zhang, A. S. Weintraub, J. Schuijers, T. I. Lee, K. Zhao, R. A. Young, Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell.* **159**, 374–387 (2014).
143. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
144. G. Li, L. Cai, H. Chang, P. Hong, Q. Zhou, E. V. Kulakova, N. A. Kolchanov, Y. Ruan, Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics.* **15 Suppl 12**, S11 (2014).
145. G. Li, X. Ruan, R. K. Auerbach, K. S. Sandhu, M. Zheng, P. Wang, H. M. Poh, Y. Goh, J. Lim, J. Zhang, H. S. Sim, S. Q. Peh, F. H. Mulawadi, C. T. Ong, Y. L. Orlov, S. Hong, Z. Zhang, S. Landt, D. Raha, G. Euskirchen, C.-L. Wei, W. Ge, H. Wang, C. Davis, K. I. Fisher-Aylor, A. Mortazavi, M. Gerstein, T. Gingeras, B. Wold, Y. Sun, M. J. Fullwood, E. Cheung, E. Liu, W.-K. Sung, M. Snyder, Y. Ruan, Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell.* **148**, 84–98 (2012).

146. M. Sadowski, A. Kraft, P. Szalaj, M. Wlasnowolski, Z. Tang, Y. Ruan, D. Plewczynski, Spatial chromatin architecture alteration by structural variations in human genomes at the population scale. *Genome Biol.* **20**, 148 (2019).
147. J. Wang, X. Dai, L. D. Berry, J. D. Cogan, Q. Liu, Y. Shyr, HACER: an atlas of human active enhancers to interpret regulatory variants. *Nucleic Acids Res.* **47**, D106–D112 (2019).
148. D. R. Zerbino, S. P. Wilder, N. Johnson, T. Juettemann, P. R. Flicek, The ensembl regulatory build. *Genome Biol.* **16**, 56 (2015).
149. J. Ernst, M. Kellis, Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
150. L. Teng, B. He, J. Wang, K. Tan, 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics.* **32**, 2727 (2016).
151. J. MacArthur, E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon, A. Milano, J. Morales, Z. M. Pendlington, D. Welter, T. Burdett, L. Hindorf, P. Flicek, F. Cunningham, H. Parkinson, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
152. T. Gao, B. He, S. Liu, H. Zhu, K. Tan, J. Qian, EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics.* **32**, 3543–3551 (2016).
153. Y. Yang, J. Fear, J. Hu, I. Haecker, L. Zhou, R. Renne, D. Bloom, L. M. McIntyre, Leveraging biological replicates to improve analysis in ChIP-seq experiments. *Comput. Struct. Biotechnol. J.* **9**, e201401002 (2014).
154. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
155. T. E. P. Consortium, The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489** (2012), pp. 57–74.
156. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature.* **518**, 317–330 (2015).
157. M. L. Heltberg, S. Krishna, M. H. Jensen, On chaotic dynamics in transcription factors and the associated effects in differential gene regulation. *Nat. Commun.* **10**, 71 (2019).
158. J. E. Henninger, O. Oksuz, K. Shrinivas, I. Sagi, G. LeRoy, M. M. Zheng, J. O. Andrews, A. V. Zamudio, C. Lazaris, N. M. Hannett, T. I. Lee, P. A. Sharp, I. I. Cissé, A. K. Chakraborty, R. A. Young, RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell.* **184**, 207–225.e24 (2021).
159. H. J. Motulsky, R. E. Brown, Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics.* **7**, 123 (2006).
160. A. Christopoulos, Assessing the distribution of parameters in models of ligand-receptor interaction: to log or not to log. *Trends Pharmacol. Sci.* **19**, 351–357 (1998).
161. K. I. Tenekedjiev, N. D. Nikolova, K. Kolev, in *Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science*, C. J. Mode, Ed. (InTech, Rijeka (HR), 2017).
162. J. Lee, W.-Y. Hong, M. Cho, M. Sim, D. Lee, Y. Ko, J. Kim, Synteny Portal: a web-based application portal for synteny block analysis. *Nucleic Acids Res.* **44**, W35–40 (2016).

163. P. Chanda, E. Costa, J. Hu, S. Sukumar, J. Van Hemert, R. Walia, Information Theory in Computational Biology: Where We Stand Today. *Entropy* . **22** (2020), doi:10.3390/e22060627.
164. C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, 1998).
165. K. Watanabe, S. Stringer, O. Frei, M. Umićević Mirkov, C. de Leeuw, T. J. C. Polderman, S. van der Sluis, O. A. Andreassen, B. M. Neale, D. Posthuma, A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
166. T. Berisa, J. K. Pickrell, Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics.* **32**, 283–285 (2016).
167. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* **4**, 7 (2015).
168. E. Baker, K. M. Schmidt, R. Sims, M. C. O'Donovan, J. Williams, P. Holmans, V. Escott-Price, W. T. G. Consortium, POLARIS: Polygenic LD-adjusted risk score approach for set-based analysis of GWAS data. *Genet. Epidemiol.* **42**, 366–377 (2018).
169. A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. J. H. Penninx, R. Jansen, E. J. C. de Geus, D. I. Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimäki, E. Raitoharju, M. Kähönen, I. Seppälä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, B. Pasaniuc, Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
170. N. Mancuso, H. Shi, P. Goddard, G. Kichaev, A. Gusev, B. Pasaniuc, Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100**, 473–487 (2017).
171. G. D. Smith, S. Ebrahim, Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.* **33**, 30–42 (2004).
172. G. Davey Smith, G. Hemani, Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–98 (2014).
173. B. Schölkopf, D. W. Hogg, D. Wang, D. Foreman-Mackey, D. Janzing, C.-J. Simon-Gabriel, J. Peters, Modeling confounding by half-sibling regression. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7391–7398 (2016).
174. D. Risso, J. Ngai, T. P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
175. M. R. Branco, A. Pombo, Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, e138 (2006).
176. G. Cavalli, T. Misteli, Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290–299 (2013).
177. R. Dali, M. Blanchette, A critical assessment of topologically associating domain prediction tools. *Nucleic Acids Res.* **45**, 2994–3005 (2017).
178. L. Yengo, J. Sidorenko, K. E. Kemper, Z. Zheng, A. R. Wood, M. N. Weedon, T. M. Frayling, J. Hirschhorn, J. Yang, P. M. Visscher, GIANT Consortium, Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
179. A. Mahajan, D. Taliun, M. Thurner, N. R. Robertson, J. M. Torres, N. W. Rayner, A. J. Payne, V. Steinthorsdottir, R. A. Scott, N. Grarup, J. P. Cook, E. M. Schmidt, M. Wuttke, C. Sarnowski, R. Mägi, J. Nano, C. Gieger, S. Trompet, C. Lecoeur, M. H. Preuss, B. P. Prins, X. Guo, L. F. Bielak, J. E. Below, D. W. Bowden, J. C. Chambers, Y. J. Kim, M. C. Y. Ng, L. E. Petty, X. Sim, W. Zhang, A. J. Bennett, J. Bork-Jensen, C. M. Brummett, M. Canouil, K.-U. Ec Kardt, K. Fischer, S. L. R. Kardia, F. Kronenberg, K. Läll, C.-T. Liu, A. E. Locke, J. 'an Luan, I. Ntalla, V. Nylander, S. Schönherr, C. Schurmann, L. Yengo, E. P. Bottinger, I. Brandslund, C. Christensen, G. Dedoussis, J. C. Florez, I. Ford, O. H. Franco, T. M. Frayling, V. Giedraitis, S. Hackinger, A. T. Hattersley, C. Herder, M. A. Ikram, M. Ingelsson, M. E. Jørgensen, T. Jørgensen, J. Kriebel, J. Kuusisto, S. Ligthart, C. M. Lindgren, A. Linneberg, V. Lyssenko, V. Mamakou, T. Meitinger, K. L. Mohlke, A. D. Morris, G. Nadkarni, J. S. Pankow, A. Peters, N. Sattar, A. Stančáková, K. Strauch, K. D. Taylor, B. Thorand, G. Thorleifsson, U. Thorsteinsdottir, J. Tuomilehto, D. R. Witte, J. Dupuis, P. A. Peyser, E. Zeggini, R. J. F. Loos, P. Froguel, E. Ingelsson, L. Lind, L. Groop, M. Laakso, F. S. Collins, J. W. Jukema, C. N. A. Palmer, H. Grallert, A. Metspalu, A. Dehghan, A. Köttgen, G. R. Abecasis, J. B. Meigs, J. I. Rotter, J. Marchini, O. Pedersen, T. Hansen, C. Langenberg, N. J. Wareham, K. Stefansson, A. L. Gloyn, A. P. Morris, M.

- Boehnke, M. I. McCarthy, Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
180. H. Tanabe, S. Müller, M. Neusser, J. von Hase, E. Calcagno, M. Cremer, I. Solovei, C. Cremer, T. Cremer, Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4424–4429 (2002).
181. E. Gelali, G. Girelli, M. Matsumoto, E. Wernersson, J. Custodio, A. Mota, M. Schweitzer, K. Ferenc, X. Li, R. Mirzazadeh, F. Agostini, J. P. Schell, F. Lanner, N. Crosetto, M. Bienko, iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nat. Commun.* **10**, 1636 (2019).
182. A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Chèneby, S. R. Kulkarni, G. Tan, D. Baranasic, D. J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W. W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).
183. S. Wang, J.-H. Su, B. J. Beliveau, B. Bintu, J. R. Moffitt, C.-T. Wu, X. Zhuang, Spatial organization of chromatin domains and compartments in single chromosomes. *Science.* **353**, 598–602 (2016).
184. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science.* **348**, aaa6090 (2015).
185. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell.* **177**, 1888–1902.e21 (2019).
186. L. Wang, Z. Feng, X. Wang, X. Wang, X. Zhang, DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics.* **26**, 136–138 (2010).
187. R. T. Pillich, J. Chen, V. Rynkov, D. Welker, D. Pratt, NDEx: A Community Resource for Sharing and Publishing of Biological Networks. *Methods Mol. Biol.* **1558**, 271–301 (2017).