Genomic Context Required

Full Chromosomal Platitude in the Study of Genetic Function

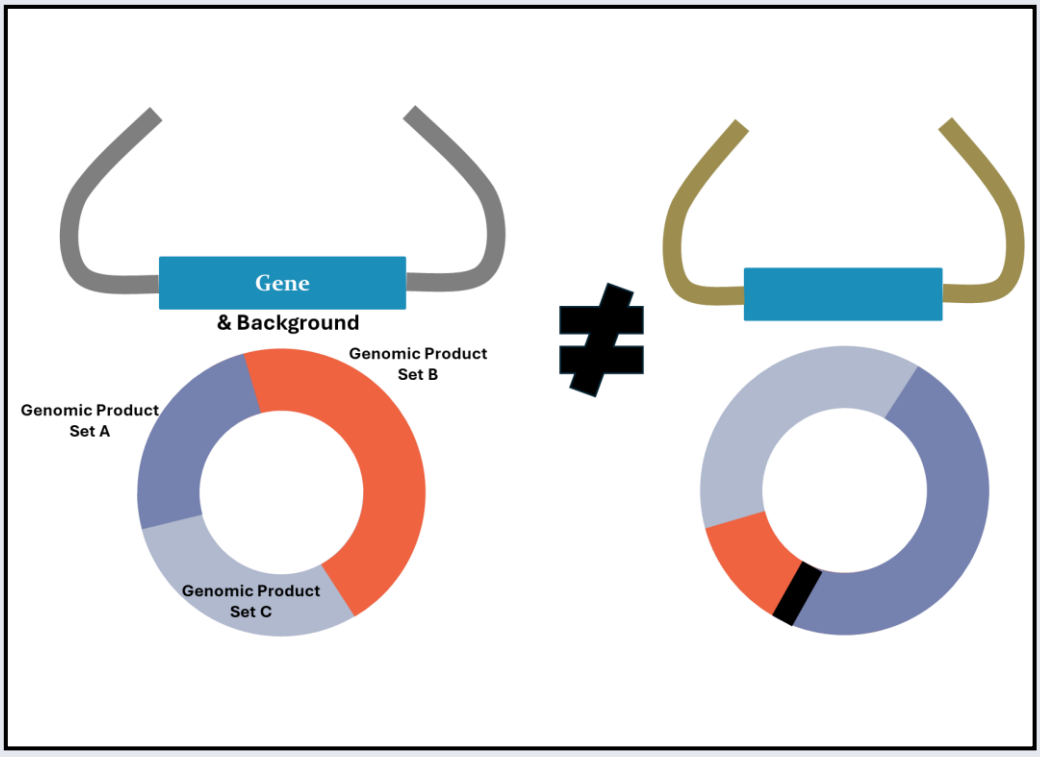
Josef Grey

Genomic Context Required

The Pursuit for Chromosomal Platitude in the Study of Genetic Function

ABSTRACT

Consideration of the full genomic context should be applied in genomic product study, disease modeling, and therapeutic development. Biological study on simplified or artificial systems results in mixed-relevancy of data. The makeup of genetic products alters significantly by artificial circularized platforms, non-native noncoding regions, and modified codon sequence. The entire spatial and temporal interactome and functome are altered from experimental to clinical environment, without requisite acknowledgement and study.



RELEVANCY

Research & Development, Drug Development

Genomic stability is a concept oft relegated to the domains of cancer, developmental disabilities, or DNA disrepair, but deserves focus across every stage of biological and medical science. Many common cell lines used to study gene product effects are done in notoriously unstable backgrounds. From (un)stable integration, expression of condensed genes for de novo effects, or genome-wide suppression studies – scientists have to balance experimental results with inconsistency in mechanism of action. The full interactome for each genetic product from genomic background to post-translational shuttling should be considered throughout the full therapeutic developmental process.

The concept of a disease state is simple – an altered physiological condition driven by the change or perturbation of a genomic product(s). A genomic product can be a classical protein translated from mRNA, short interfering RNAs, long or short non coding RNAs acting enzymatically or antagonistically, tRNAs, even the genomic DNA itself as they act to modulate transcriptional levels. In the simplest of genetic diseases, a singular gene in a transcriptionally isolated region with one uninterrupted exon is transcribed and translated into a singular protein of one basic function. Repair or replacement of the protein or functional product will alleviate the disease, and screening for therapeutic hits can be done in a high-throughput linear assay and cell line. Very few of these diseases exist, and true-to-hypothesis, drug discovery has been *relatively* simple. Attempted juxtaposition of more complex disease states into simplified models has left therapeutic inadequacies.

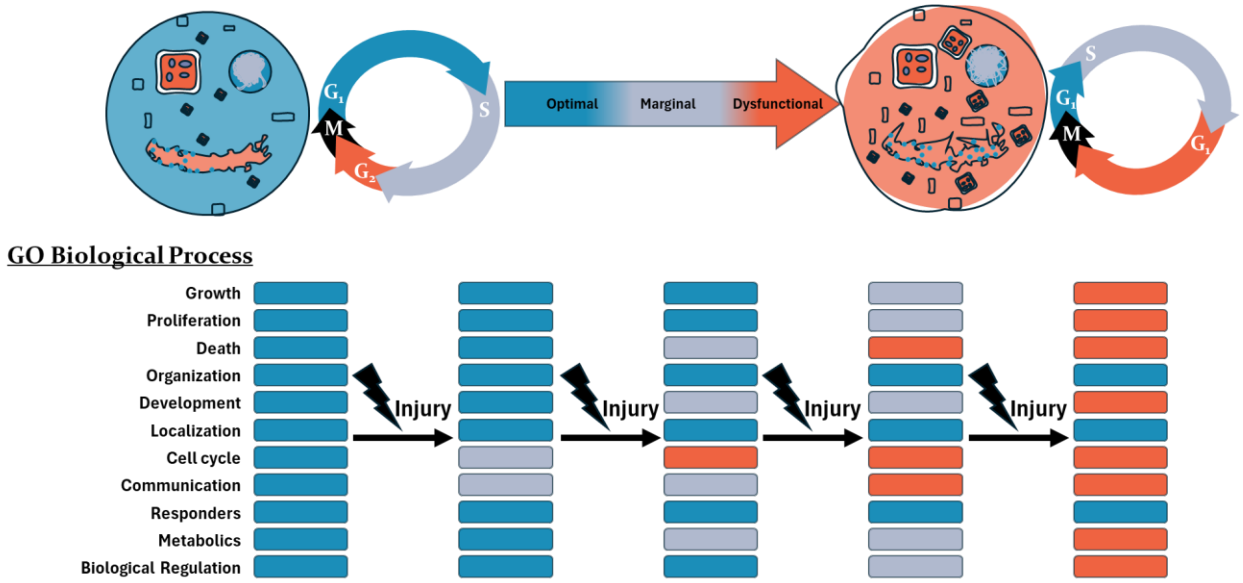


Figure 1 The transformation from healthy to diseased can be plotted across biological process axes, rarely with singular injury to disease state events. Gene Ontological biological processes represent critical pathways that need to be altered for disease progression.

Consider homo sapiens, with 23 pairs of chromosomes [22 + 1 sex-linked (XX/XY)]. Each pair, minus sex-linked, allows for the carry and access of duplicate genes, and the evolution-iteration process to continue with limited terminal mutation or combination events. Chromosomal access, and genomic transcription, is heavily controlled throughout the entire cell cycle. A majority of genomic DNA is tightly wound into inaccessible balls around histones, with specified sections of chromosomal DNA brought towards the edge of the nucleus for transcription in controlled periods of the cellular cycle. Transcription is conducted by various factors, ranging from proteins, RNAs, the same chromosome or another entirely – transcription factors ensure appropriate cellular functions are maintained as each cell has internal and external responsibilities that change according to triggers, time of day, metabolic status, consciousness, emotion, etc. The general rate and total output of transcription and translation are specific to each genomic product based on internal features to ensure appropriate action; cellular production and function is a digital scale, not analog. Synthetic systems meant to emulate a diseased state often rely on over-production of uniform genomic products on an (oft-forgotten) artificial promoter, 3' and 5' Untranslated Region (UTR) – changing everything from expression timing, levels, isoforms, and post-translational modifications.

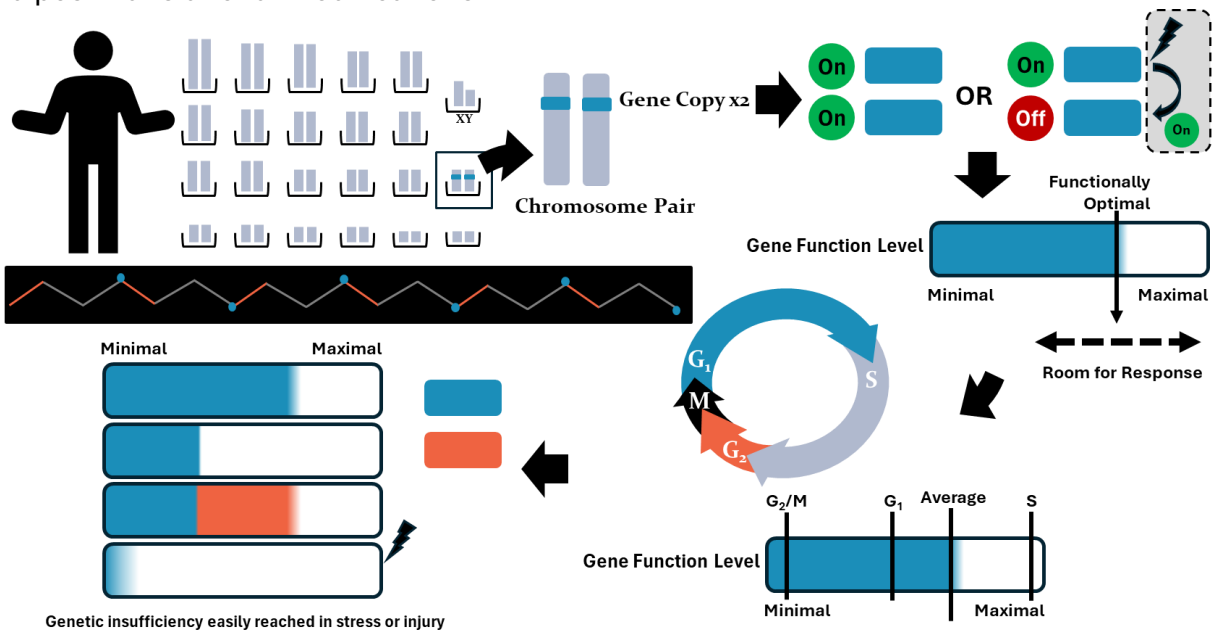


Figure 2 While at least two copies exist for each gene, most genes use only one for functional expression. While some genes are expressed consistently across the cell cycle, most fluctuate in response to changing demands. Many diseases, or disease effects, are only felt through specific periods or events where the affected function is lost. Similarly, some "benign" mutations become deleterious in specific triggers or scenarios, such as BRCA2 and alcohol intolerance.

On the macro, all widespread cell lines for biological research have been modified in significant manners to survive outside the primary organism, and to survive near

indefinitely. Cell lines such as HEK293 and HeLa were developed by taking human tissue samples, breaking them into as small clumps as possible, and just letting them run in the hopes of survival. HEK293 and HeLa are two lines that made the jump; HEK293 likely retaining transcriptional fluidity off of fetal pluripotency. HeLa already stemming from a ferociously malignant cancer. Both are pivotal to scientific discovery and medical therapeutic development. Both quickly devolve in genomic integrity through passages, resulting in extreme chromosomal configurations. Several labs have published on the variance of cell lines across institutions, with some notable headlines of collaborators working on the same "cell line" but with vastly different meta-omic (genomic, transcriptomic, proteomic, etc.) backgrounds. Studies involving genetic dosing or manipulation of any number of pathways present critical issues in translation and reproducibility across labs, cell lines, and tissues.

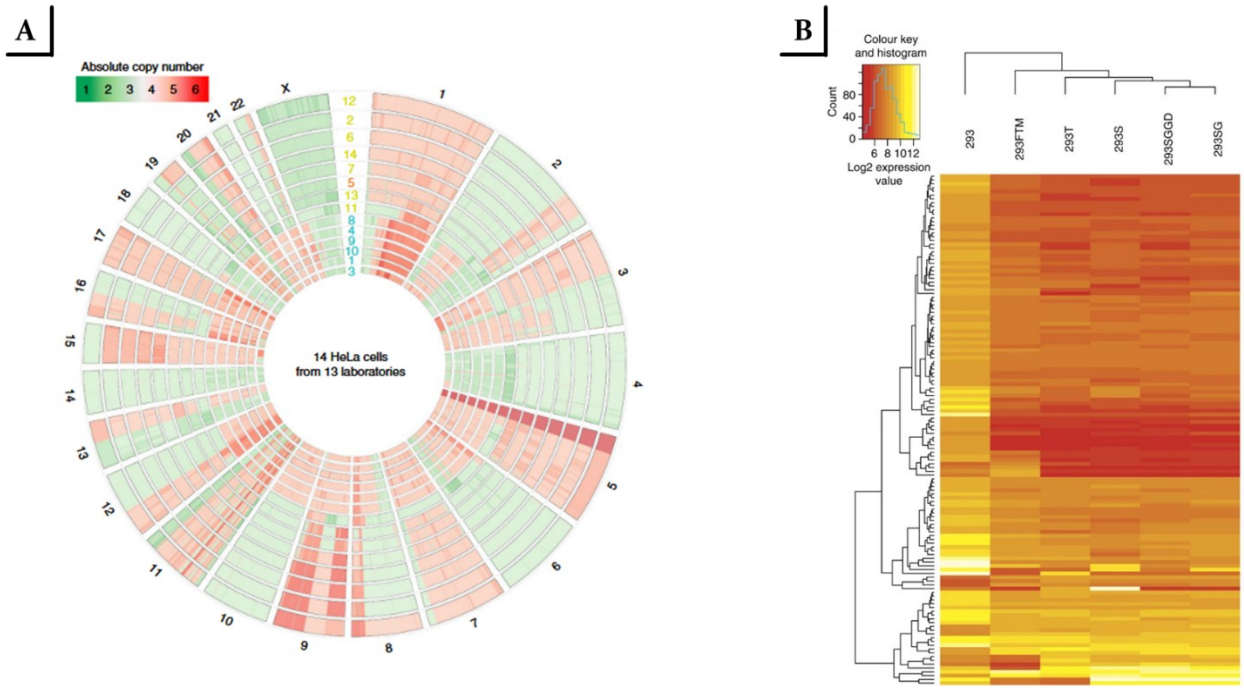


Figure 3 Figures representing genomic diversity in HeLa & HEK293 cells across collaborating laboratories. While both lines show common deletions and insertions, significant chromosomal regions alter in expression levels, leaving genetic dosing and product makeup questions. A. Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E. G., Van Drogen, A., Borel, C., Frank, M., Germain, P. L., Bludau, I., Mehnert, M., Seifert, M., Emmenlauer, M., Sorg, I., Bezrukov, F., Bena, F. S., Zhou, H., Dehio, C., Testa, G., Saez-Rodriguez, J., ... Aebersold, R. (2019). Multi-omic measurements of heterogeneity in HeLa cells across laboratories. Nature biotechnology, 37(3), 314–322. https://doi.org/10.1038/s41587-019-0037-y
B. Lin, Y. C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van de Peer, Y., Tavernier, J., & Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nature communications, 5, 4767. https://doi.org/10.1038/ncomms5767

Microscopically, an active gene consists of three core elements, a promoter to initiate replication, an open reading frame (ORF) carrying the functional information, and a

terminator. The promoter sequence presents as a unique structural formation that will recruit various protein and/or nucleic acid factors to start transcription, minor variations open or close various loops or "arms" that are critical for highly-specific binding. The ORF consists of the sequence to be transcribed, which can contain exons and introns, to be spliced for product diversification. The terminator, or stop, is the softest requirement; some genes lack a proper terminator and rely on chromosomal DNA structure or a following gene's terminator, and specific polymerase-complexes have limited transcription run-time. Outside of these, there is a 5' and 3' Untranslated region (UTR) that modifies genomic products from transcription through post-translational modifications, mRNA half-life, etc. These UTRs can recruit genomic elements from entirely separate chromosomes, with some full genomic units requiring multiple chromosomes to be brought together for transcription initiation – hence repeated chromosomal "stitching" events across evolution and disease. Abrogation of any of these molecular pairings, of any binding motif, of any function sequence in the exon, any nucleotide within the intron; genomic effect is driven by an orchestra. Understanding effects of any genetic change is limited by the understanding of all forces acting upon the whole of the system, and is tethered to the sum.

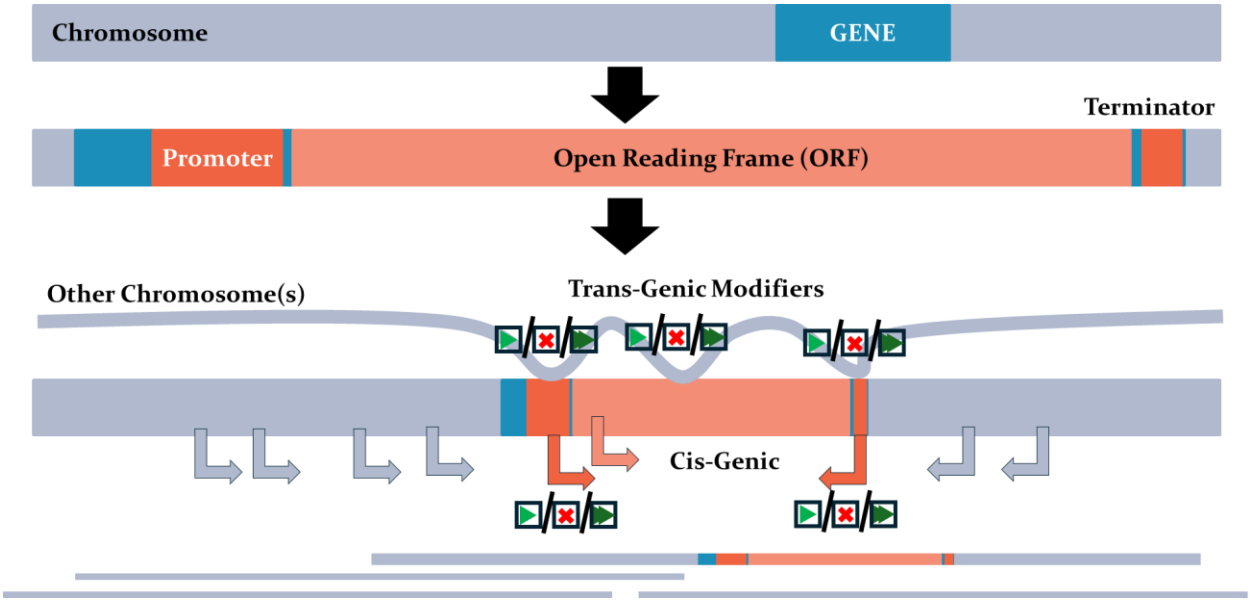


Figure 4 Affective genomic elements are not limited by sequential proximity to the gene, nor even spatial proximity – numerous DNA binding proteins orchestrate entirely separate chromosomes to speed up, slowdown, turn on/off, or alter the transcript products of a gene. Artificial systems with barebones native sequence, structure, or context for genes of interest lose important specificity in genetic function.

A gene's function is the sum of its genomic products. Coming a long way from the 'one gene-one protein' theory of olde, a gene can have any number of products. Primarily driven by the splicing of introns – either part or entire exons can be removed from mRNA with an intronic excision. In fact, many introns splice themselves out, requiring or allowing

no partner or dispute – meaning that there is an underlying mechanism of mRNA splicing that is resulting in an unknown variation in genomic products, even in genes without or very few modeled introns. Consider a short interfering RNA, hyper-specific to a sequence, it will bind to and prevent mRNA translation via DICER. Now consider it's targeting gene, containing multiple exons; only mRNA transcripts carrying the siRNA-target sequence will be silenced, while any transcripts without that exon will remain. Sequestered-mRNAs can undergo further modifications or processing, resulting in even more unknown. Barring critical analysis with holistic methods, any conclusion on genetic perturbation effects is questionable.

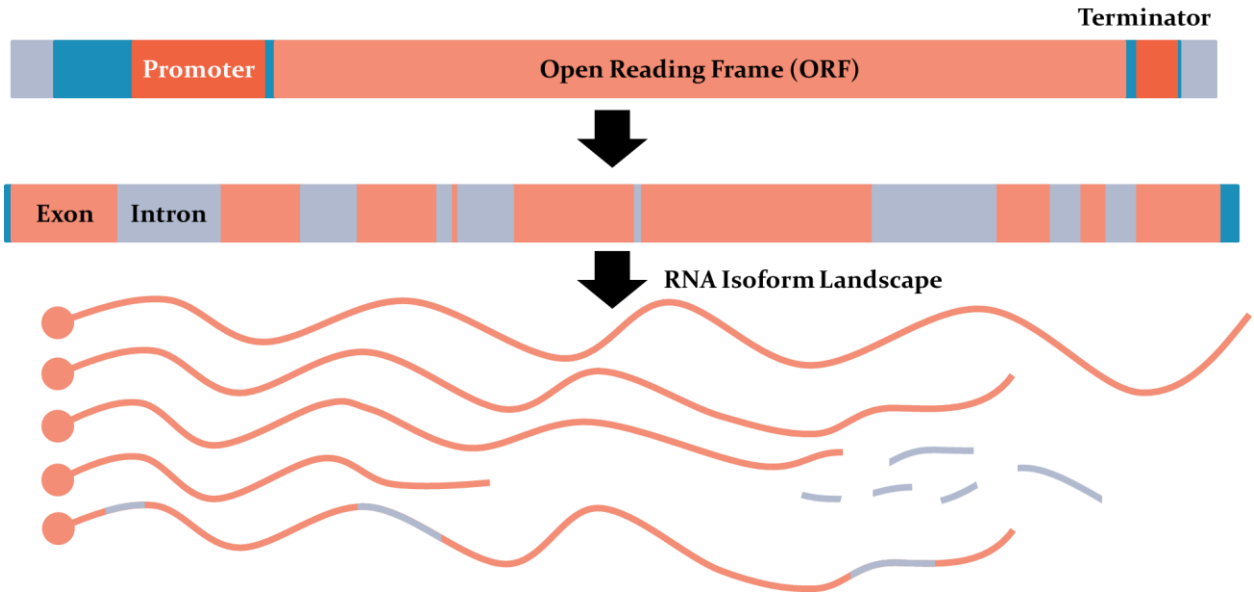


Figure 5 With potentially less than 5% of human genes being intron-less, the use of such systems to study genetic effects has left linear interactome and function maps. While not every gene may have multiple products, the potential has been relatively unstudied. Several disease-causing and impacting genes have functionally distinct isoform products, either taking on specific functions driven by less functional domains or from temporal co-expression of cycle-specific partners. Formally un-targetable genes could become accessible with isoform-specific therapeutics.

Despite the long-running understanding of cis and trans factors on genomic product expression and functionality, flawed cellular and genomic models have been used to establish and enforce scientific hegemony. Therapeutic development has maintained artificial experimental backgrounds until general population clinical trials and suffered for it. Additionally, a push for more artificial intelligence in the discovery process via computational modeling assembled from unnatural backgrounds threatens modern initiatives. Even the industry capitulation to 3D or organoid models risks failure without appropriate integration of diverse genomic backgrounds and holistic multi-omics analysis through tissues and time. Science and medicine require the continued push for diversity of thought, integration of competing and holistic analytical systems, and continued path of interdisciplinary cooperation. To quote a physicist, "everything should be made as simple as possible, but not simpler."

Genomic Context Required

Contact Advisory@Greytheorem.com

See more at Greytheorem.com