The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell

*William Schierding[1], Wayne S. Cutfield[1,2] and Justin M. O'Sullivan[1,2] ***

[1] Liggins Institute, University of Auckland, Auckland, New Zealand
[2] Gravida – National Centre for Growth and Development, Auckland, New Zealand

Genome wide association studies are central to the evolution of personalized medicine. However, the propensity for single nucleotide polymorphisms (SNPs) to fall outside of genes means that understanding how these polymorphisms alter cellular function requires an expanded view of human genetics. Integrating the study of genome structure (chromosome conformation capture) into its function opens up new avenues of exploration. Changes in the epigenome associated with SNPs in gene deserts will allow us to define complex diseases in a much clearer manner, and usher in a new era of disease pathway exploration.

Keywords: GWAS, epigenetics, DNA folding, chromosomal folding, gene deserts
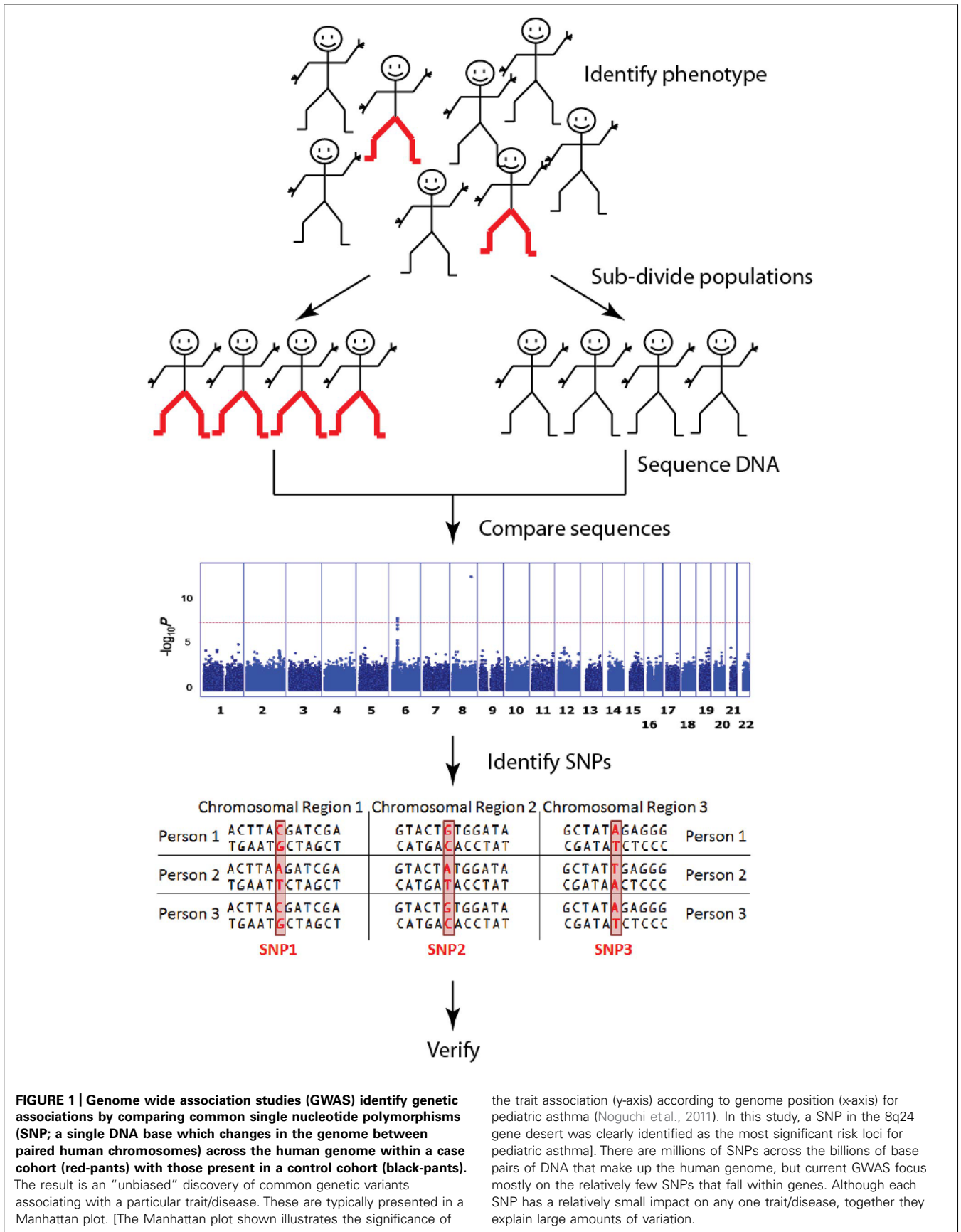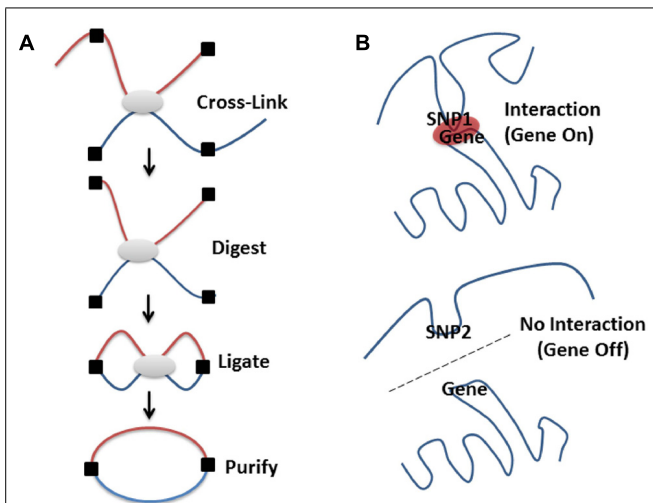
INTRODUCTION

Studies of human variation and its links to disease promise to usher in an era of personalized medicine, where the genetic code of an individual is assessed and used to guide clinical practice. This advance toward individualized medicine is being made possible through the cataloging of common genetic variants and their associations with complex traits and disease.

Genome wide association studies (GWAS) have a central role in this human genetics revolution. The primary purpose of GWAS is to identify single nucleotide polymorphisms (SNPs) that are associated with phenotypic traits, typically those associated with a particular disease (**Figure 1**). In a handful of cases, particularly in cancer, the move to genome wide SNP detection has led to clinically relevant predictions (Jostins and Barrett, 2011). However, despite these best-case-scenarios, nearly half of the disease-associated SNPs from published GWAS are not located in or near genes (Visel et al., 2009; Hindorff et al., 2013). Therefore, despite the fact that significant associations are often found between complex traits and SNPs in gene deserts (i.e., genomic regions of > 500kb that lack annotated genes or protein-coding sequences; Venter et al., 2001; Libioulle et al., 2007; Grant et al., 2009), their location within gene deserts means they have no readily annotated gene function and cannot be assigned to a specific biological pathway. Explanations abound for why apparently significant SNPs are located in gene deserts (Visel et al., 2009; Uddin et al., 2011; Zhang et al., 2012). However, typically these sorts of results only make their way into publications as part of supplementary tables of findings, often without any attempt to explain the association. Current trends are now shifting to focus on SNPs within gene deserts as potentially contributing to diseases by regulating gene function from a distance.

How is it possible for SNPs within gene deserts to regulate unlinked genes? The answer lies within the twists and turns that form when 3 m of human DNA (chromosomes) is packaged within a roughly spherical nucleus that is only approximately 10 µm in diameter. Within the hierarchy of folding necessary to package the genome within the eukaryotic nucleus, regions of each chromosome contact other chromosomes to form an intricate 3-dimensional DNA network. Therefore, while two regions of DNA (loci) may be distant on a linear scale, DNA folding provides a mechanism for these two loci to become spatially close together. Implicit in this concept is the idea that all genetic functions (regulation, reading, repair, and replication) are influenced by this 3-dimensional architecture, generating the cell's morphology and function (Misteli, 2001). Intra-cellular DNA structure cannot be divorced from its functions.

Methods to study intra-cellular DNA organization, e.g., chromosomal conformation capture, 3C, or related methodologies (**Figure 2**) have confirmed that loci on separate chromosomes (Ling et al., 2006; Lomvardas et al., 2006) or on the same chromosome but separated by large intervening sequences (Carter et al., 2002; Tolhuis et al., 2002; Sotelo et al., 2010; French et al., 2013) can interact in space to regulate the expression of multiple genes (Schoenfelder et al., 2010). Therefore, it is possible that intergenic SNPs associated with diseases are indeed involved in the regulation of genes and pathways through spatial associations with different genes. In effect, intergenic SNPs represent sequence alterations that affect the ability of these regions to interact with other loci or recruit the proteins necessary for gene regulation at a distance (Sotelo et al., 2010; French et al., 2013). This proposal gains some support from the finding that 76% of GWAS SNPs are near (or in linkage within a haplotype block) DNaseI hypersensitive sites,

FIGURE 1 | Genome wide association studies (GWAS) identify genetic associations by comparing common single nucleotide polymorphisms (SNP; a single DNA base which changes in the genome between paired human chromosomes) across the human genome within a case cohort (red-pants) with those present in a control cohort (black-pants). The result is an "unbiased" discovery of common genetic variants associating with a particular trait/disease. These are typically presented in a Manhattan plot. [The Manhattan plot shown illustrates the significance of the trait association (y-axis) according to genome position (x-axis) for pediatric asthma (Noguchi et al., 2011). In this study, a SNP in the 8q24 gene desert was clearly identified as the most significant risk loci for pediatric asthma]. There are millions of SNPs across the billions of base pairs of DNA that make up the human genome, but current GWAS focus mostly on the relatively few SNPs that fall within genes. Although each SNP has a relatively small impact on any one trait/disease, together they explain large amounts of variation.

FIGURE 2 | High-resolution molecular techniques for studying the spatial organization of chromosomes. Various methods exist to capture when two genomic regions are in close spatial proximity. Microscopy is a powerful tool for visualizing structure, especially when combined with FISH methodologies (Branco and Pombo, 2006; Rouquette et al., 2010). However, even the super resolution microscopes only resolve 15–20 nm structures, which limits the ability to visualize the lowest levels of DNA packing (Baker, 2011). By contrast, proximity based ligation technologies [e.g., 3C (Dekker et al., 2002), 4C (Zhao et al., 2006), GCC (Rodley et al., 2009), ChIA-PET (Li et al., 2013 n.d.), or 5C (Dostie et al., 2006)] enable the high-resolution identification of which DNA sequences are contacting each other, directly or indirectly, in 3D space at a given moment in time. **(A)** Proximity based ligation methodologies currently incorporate the same basic steps. The structure of the chromatin is captured by cross-linking which ensures that regions that are in contact are chemically held together. The chromatin structure is then broken into pieces with restriction enzymes (digested), pieces that are held together by cross-links are then enzymatically joined together (ligated), purified away from the cellular debris and detected by various methods (for a more in depth discussion see Grand et al., 2011). The method used for detection of these ligated products (i.e., PCR, low or high through-put sequencing) depends on whether specific, close range or global organization is being investigated. By coupling 3C technologies to next-generation sequencing (NGS) it is possible to create unbiased low or high-resolution 3-dimensional maps of whole genomes (Lieberman-Aiden et al., 2009; Rodley et al., 2009). This has led to discoveries of interesting interactions in regions of the DNA that otherwise wouldn't be thought to be part of a particular disease process (Rodley et al., 2012). **(B)** Genetic variation (SNPs, see **Figure 1A**) can affect chromatin structure by altering folding patterns. These changes result in differential spatial relationships, leading to a gain or loss of function through altered associations between enhancers, promoters, LCRs, silencers, or imprinting control regions.

which are often locus control regions (LCRs) – regions associated with enhancers (Maurano et al., 2012; Malin et al., 2013). Despite the fact that "our understanding of higher-order genomic structure is coarse, fragmented and incomplete," (Dixon et al., 2012) there are certain things that have been learned by a decade of research on the topic.

HOW DO LONG-DISTANCE INTERACTIONS AFFECT GENE EXPRESSION?

The "dog-on-a-lead" model has been proposed to account for the coordination of positioning of chromosomes and the formation of interactions within and between chromosomes. In this model, chromosome positioning and folding dominates the

3-dimensional organization of the nucleus, while genes, domains, and enhancers are largely limited to form contacts within the chromosomal context within which they are located (Krijger and de Laat, 2013). This model allows for the occurrence of cell-specific interactions within similar chromosome domains as a cause of variegated gene expression amongst otherwise identical cells. (de Wit et al., 2013).

The dog-on-a-lead model for the spatial organization of eukaryotic genomes encompasses the formation of short and long-distance interactions, between enhancers and promoters, which form an essential component of the regulatory systems for eukaryotic gene expression (Sotelo et al., 2010; French et al., 2013). These interactions are hypothesized to function to bring genes and regulatory regions to spatial domains which contain high concentrations of the relevant enzymes, proteins and raw materials required for the production of messenger RNA. In effect, loci co-localize at factories that promote transcription (Cook, 2002). For example, activation of the HoxB and uPA genes is accompanied by alterations in their spatial organization that include the association of these genes into a factory that promotes their transcription (Chambeyron and Bickmore, 2004; Marenduzzo et al., 2007; Ferrai et al., 2010). Thus, inter- and intra-chromosomal organization is a reflection of the cell machinery integrating spatial relationships into higher-order epigenomic regulation (Baker, 2011).

However, as clear as these models have become, technological challenges have left questions about the existence, formation and maintenance of transcription factories that require chromosome-mixing and long-range interactions. Firstly, the 3C based methods that have become popular for the study of the 3-dimensional chromosome networks can potentially bias the results (Razin et al., 2010; Gavrilov et al., 2013). Moreover, these methods are probabilistic (O'Sullivan et al., 2013) and thus, interactions identified by 3C are representative of what is happening in a population of cells. Any one cell in the population is likely to display only a small proportion of the interactions that were captured. Therefore, the "factory" may not actually be present in any one cell. Secondly, the amount of mixing (Branco and Pombo, 2006) that occurs between spatially adjacent chromosomes within the nucleus remains controversial with some contending that chromosomes are self-contained with little mixing (Olivares-Chauvet et al., 2011).

IS GENOME ORGANIZATION STABLE?

Evidence exists for organization of chromosomes into topological domains in pluripotent stem cells that then remain stable between cell types (de Wit et al., 2013). Reproducible spatial associations between fragile loci may be reflected in translocation hot-spots in somatic cells (Roix et al., 2003). For example, Burkitt's Lymphoma is often characterized by a translocation involving the IgH locus, located on chromosome 14, and the Myc gene promoter, located on chromosome 8 (Roix et al., 2003). Critically, the Myc and IgH loci co-localized during transcription within the nucleus of primary B cells (but not other cell types; Osborne et al., 2007; Wang et al., 2009) supporting the idea that inter-chromosomal interactions can promote disease-associated translocations. Overall, it is evident that genome organization has

many common factors across cell type, but is unlikely to be a singular structure. Rather, the spatial organization is dynamic over time and space, helping to regulate the cell's current needs through cycles of aging, type (function), and current needs (cell-cycle dependent).

DOES THE SPATIAL ORGANIZATION OF A GENOME ALTER WITH AGE?

DNA damage repair, histone modifications, and chromatin remodeling are all highly affected by the aging process (Burgess et al., 2012). Moreover, cellular aging has been shown to affect the shape of the nucleus and nuclear lamina organization in cell cultures of cells from Hutchinson–Gilford progeria syndrome patients (Bridger and Kill, 2004). The aging process also plays a great role in changes in gene expression, a process shown across species. (Zhan et al., 2007; Brink et al., 2009; de Magalhaes et al., 2009; Park, 2011). Epigenomics is generally considered to be highly dependent on the interplay between the spatial organization of chromosomes and nuclear functions. Therefore, it is reasonable to hypothesize that aging-associated alterations to short- and long-distance regulatory interactions within the spatial organization of the genome affect the development-related expression of critical genes.

It has been shown that in some circumstances similar genes come together in 3D space, regardless of the age of the individual (Gandhi et al., 2006). However, this result is controversial due to its low resolution and observations of the developmental-dependent regulated formation of long-distance interactions associated with the expression of fetal and adult hemoglobin (Tolhuis et al., 2002; Bank, 2006). Specifically, the erythroid-specific genes Hbb-b1 and Eraf, separated by over 20 Mb on the same chromosome, co-localize into a transcriptional factory when actively transcribed (Osborne et al., 2004). Similarly, mice Hbb has been shown to form transcription-associated preferential inter-chromosomal connections with 359 erythroid genes from different genomic locations (Schoenfelder et al., 2010).

IS IT IMPORTANT TO INCORPORATE THE CHROMOSOME INTERACTION NETWORK INTO FUTURE GWAS/SNP STUDIES?

Extensive efforts have been made to catalog human variation. The most recent versions of dbSNP and the human gene mutation database contain 38,072,522 validated variants (Sherry et al., 2001) and ∼100,000 mutations in nuclear genes (Stenson et al., 2009) that are associated with complex human traits, respectively. However, the associations between common-variants (SNPs) and phenotypic traits or diseases held in these databases, and others like them, only describe a small fraction of the overall heritability of complex disease traits (Frazer et al., 2009). Thus, our ability to elucidate functional pathways related to these SNPs has been limited. Part of the reason for missing heritability has been proposed to be bias toward results focused on the coding regions, which comprise only 1.5% of the genome (Consortium et al., 2007). This ignores the rest of the ∼2.5–15% of the genome that is estimated to be functionally constrained, yet outside of coding regions (Vernot et al., 2012). Thus, given the fact that enhancers can be located in gene deserts (Harismendy et al., 2011) and can

control multiple genes through physical interactions, it is important to determine if SNPs located outside of genes contribute to disease phenotypes through alterations to spatial regulatory interactions.

One caveat to the study of SNPs within non-genic regions is that, while it is known that common SNPs explain a substantial portion of heritability, not all SNPs contribute equally to the heritability of a trait. SNPs in genes explain the most heritability, while those near genes (or in areas regulating them) explain some, and those in non-genic regions (SNP deserts) explain little of the heritability (Smith et al., 2011; Yang et al., 2011; Schork et al., 2013). Despite this, it remains possible that SNPs located outside of coding regions represent a new class of regulatory SNPs that make an important contribution toward explaining heritability.

ARE THERE ANY CURRENT METHODS FOR ASSOCIATING SNPs AND 3D FUNCTION?

Methods that try to explain the roles of these SNPs in the context of 3D structure have recently begun to be developed. For example, a recently developed database provides functional annotations of SNPs using actual long-range interaction datasets (Wang et al., 2012; Li et al., 2013). By going beyond conservation information and incorporating information from multiple different sources (e.g., HapMap, ENCODE), the GWAS3D database has branded itself as an "efficient solution to interpret the regulatory role of genetic variation in the non-coding regions," associating SNPs with 3D structure changes. This database brings 3D structure out of LD blocks, but it does not provide a mechanism through which gene deserts would have a functional role within the cell.

EXAMPLES OF GWAS HITS IN GENE DESERTS PLAYING A FUNCTIONAL ROLE THROUGH 3D INTERACTIONS

Recent advances in the theoretical and experimental methods used to study DNA packaging within cells make it possible to elucidate the biological function and pathways to which SNPs located within gene deserts can contribute. This has been shown in a number of gene deserts, most notably: SNPs within a 1.2 Mb region on chromosome 8q24, a known gene desert, have been implicated in cancer-type-specific interactions with Myc, a highly potent cancer gene > 300 kb away (Amundadottir et al., 2006; Ghoussaini et al., 2008; Ahmadiyeh et al., 2010; Wasserman et al., 2010). This same region has also been implicated in pediatric asthma in an Asian population (Noguchi et al., 2011) and non-syndromic cleft lip in pediatric patients (Grant et al., 2009).

In examples reflecting the complex nature of the metabolic syndrome, several studies have found long-range interactions that regulate metabolic disease pathways. 3D genome interactions within the 9p21 locus were shown to play a functional role for GWAS-significant SNPs associated with coronary artery disease and type 2 diabetes; enhancers in this region with significant STAT-1 binding have the ability to impair interferon-γ signaling response (Harismendy et al., 2011). Li et al. (2013) using the GWAS3D database, showed that SNPs associated with diabetes mellitus have a number of cell-type-specific interactions across the genome, specifically the 1p13 region. This group also explored the regulation of IRS1, active in type 2 diabetes and coronary artery

disease, discussing the possible regulatory role of GWAS SNPs in sites 600 Kb and 1 Mb downstream from the IRS1 gene promoter.

Beyond cancer and metabolic syndromes, long-range interactions can have a role in human development. A study on the transcription of Sonic hedgehog (SHH), an important regulator of human development, has shown that its expression can be altered by a SNP in an intron in a gene 1Mb away which acts on a *cis*-acting regulator of SHH (2° of separation; Lettice et al., 2002). This interaction plays a role in the development of preaxial polydactyly.

CONCLUSION

Since the GWAS era began over a decade ago, the understanding of the genetics of human disease has undergone many significant breakthroughs. Despite this, the likelihood that individual SNPs explain any single heritable disease predisposition decreases with the complexity of the cellular pathways and processes that ultimately contribute to the disease. The "low hanging fruit" from the common-disease, common-variant hypothesis is likely gone, prompting the drive to find new breakthroughs and move the pursuit beyond simply cataloging the deleterious SNPs within genes (Li et al., 2013).

Alterations to the genome can occur to the actual DNA sequence and also at the level of the spatial organization. Individually, or collectively, these changes contribute to diseases in a number of different ways. Therefore, while intergenic SNPs do not affect the coding sequence of any known genes, it remains likely that they contribute to disease phenotypes by changing the 3-dimensional organization of the genome.

While the functional characterization of intergenic SNPs is no easy feat, it is critical that new research identifies and functionally characterizes all variants and variant combinations that make a significant contribution to disease etiology. Expressed quantitative trait loci (eQTL) studies have attempted to link SNPs, changes in gene expression, and phenotype. However, the mechanism by which non-coding SNPs affect expression remains unclear, particularly for trans-eQTLs. Integrating spatial organization, eQTL, and SNP data may provide evidence for direct linkages that explain some of these correlations.

Integrating clinical, computational and molecular approaches to identify changes in pathways mediated by spatial associations with intergenic SNPs will open up avenues of exploration that cannot be otherwise elucidated, defining complex disease in a much clearer manner, and ushering in a new era of disease pathway exploration.

ETHICAL STANDARDS

All experiments comply with the current laws of the country in which they were performed.

AUTHOR CONTRIBUTIONS

William Schierding primary author; Wayne S. Cutfield and Justin M. O'Sullivan project leaders (all authors contributed significantly to the writing of this paper).

ACKNOWLEDGMENTS

REFERENCES

Ahmadiyeh, N., Pomerantz, M. M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., et al. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with Myc. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9742–9746. doi: 10.1073/pnas.0910668107

Amundadottir, L. T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B. A., et al. (2006). A common variant associated with prostate cancer in European and African populations. *Nat. Genet.* 38, 652–658. doi: 10.1038/ng1808

Baker, M. (2011). Genomics: genomes in three dimensions. *Nature* 470, 289–294. doi: 10.1038/470289a

Bank, A. (2006). Regulation of human fetal hemoglobin: new players, new complexities. *Blood* 107, 435–443. doi: 10.1182/blood-2005-05-2113

Branco, M. R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* 4:e138. doi: 10.1371/journal.pbio.0040138

Bridger, J. M., and Kill, I. R. (2004). Aging of Hutchinson-Gilford progeria syndrome fibroblasts is characterised by hyperproliferation and increased apoptosis. *Exp. Gerontol.* 39, 717–724. doi: 10.1016/j.exger.2004.02.002

Brink, T. C., Demetrius, L., Lehrach, H., and Adjaye, J. (2009). Age-related transcriptional changes in gene expression in different organs of mice support the metabolic stability theory of aging. *Biogerontology* 10, 549–564. doi: 10.1007/s10522-008-9197-8

Burgess, R. C., Misteli, T., and Oberdoerffer, P. (2012). DNA damage, chromatin, and transcription: the trinity of aging. *Curr. Opin. Cell Biol.* 24, 724–730. doi: 10.1016/j.ceb.2012.07.005

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* 32, 623–626. doi: 10.1038/ng1051

Chambeyron, S., and Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* 18, 1119–1130. doi: 10.1101/gad.292104

Consortium, E. P., Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874

Cook, P. R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* 32, 347–352. doi: 10.1038/ng1102-347

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799

de Magalhaes, J. P., Curado, J., and Church, G. M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* 25, 875–881. doi: 10.1093/bioinformatics/btp073

de Wit, E., Bouwman, B. A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M. J., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 501, 227–231. doi: 10.1038/nature12420

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309. doi: 10.1101/gr.5571506

Ferrai, C., Xie, S. Q., Luraghi, P., Munari, D., Ramirez, F., Branco, M. R., et al. (2010). Poised transcription factories prime silent uPA gene prior to activation. *PLoS Biol.* 8:e1000270. doi: 10.1371/journal.pbio.1000270

Frazer, K. A., Murray, S. S., Schork, N. J., and Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.* 10, 241–251. doi: 10.1038/nrg2554

French, J. D., Ghoussaini, M., Edwards, S. L., Meyer, K. B., Michailidou, K., Ahmed, S., et al. (2013). Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am. J. Hum. Genet.* 92, 489–503. doi: 10.1016/j.ajhg.2013.01.002

Gandhi, M., Medvedovic, M., Stringer, J. R., and Nikiforov, Y. E. (2006). Interphase chromosome folding determines spatial proximity of genes participating in carcinogenic RET/PTC rearrangements. *Oncogene* 25, 2360–2366. doi: 10.1038/sj.onc.1209268

Gavrilov, A. A., Gushchanskaya, E. S., Strelkova, O., Zhironkina, O., Kireev, I. I., Iarovaia, O. V., et al. (2013). Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res.* 41, 3563–3575. doi: 10.1093/nar/gkt067

Ghoussaini, M., Song, H., Koessler, T., Al Olama, A. A., Kote-Jarai, Z., Driver, K. E., et al. (2008). Multiple loci with different cancer specificities within the 8q24 gene desert. *J. Natl. Cancer Inst.* 100, 962–966. doi: 10.1093/jnci/djn190

Grand, R. S., Gehlen, L. R., and O'Sullivan, J. M. (2011). "Methods for the investigation of chromosome organization," in Vol. 5, *Advances in Genetics Research* ed. K. V. Urbano (New York: NOVA), 111–130.

Grant, S. F., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C. E., et al. (2009). A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J. Pediatr.* 155, 909–913. doi: 10.1016/j.jpeds.2009.06.020

Harismendy, O., Notani, D., Song, X., Rahim, N. G., Tanasa, B., Heintzman, N., et al. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* 470, 264–268. doi: 10.1038/nature09753

Hindorff, L. A., MacArthur, J., Morales, J., Junkins, H. A., Hall, P. N., Klemm, A. K., et al. (2013). *A Catalog of Published Genome-Wide Association Studies.* Available at: www.genome.gov/gwastudies [Accessed July 31, 2013].

Jostins, L., and Barrett, J. C. (2011). Genetic risk prediction in complex disease. *Hum. Mol. Genet.* 20, R182–R188. doi: 10.1093/hmg/ddr378

Krijger, P. H. L., and de Laat, W. (2013). Identical cells with different 3D genomes; cause and consequences? *Curr. Opin. Genet. Dev.* 23, 191–196. doi: 10.1016/j.gde.2012.12.010

Lettice, L. A., Horikoshi, T., Heaney, S. J., Van Baren, M. J., Van Der Linde, H. C., Breedveld, G. J., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7548–7553. doi: 10.1073/pnas.112212199

Li, M. J., Wang, L. Y., Xia, Z., Sham, P. C., and Wang, J. (2013). GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Res.* 41, W150–W158. doi: 10.1093/nar/gkt456

Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., et al. (2007). Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* 3:e58. doi: 10.1371/journal.pgen.0030058

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369

Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., et al. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* 312, 269–272. doi: 10.1126/science.1123191

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* 126, 403–413. doi: 10.1016/j.cell.2006.06.035

Malin, J., Aniba, M. R., and Hannenhalli, S. (2013). Enhancer networks revealed by correlated DNAse hypersensitivity states of enhancers. *Nucleic Acids Res.* 41, 6828–6838. doi: 10.1093/nar/gkt374

Marenduzzo, D., Faro-Trindade, I., and Cook, P. R. (2007). What are the molecular ties that maintain genomic loops? *Trends Genet.* 23, 126–133. doi: 10.1016/j.tig.2007.01.007

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195. doi: 10.1126/science.1222794

Misteli, T. (2001). The concept of self-organization in cellular architecture. *J. Cell Biol.* 155, 181–185. doi: 10.1083/jcb.200108110

Noguchi, E., Sakamoto, H., Hirota, T., Ochiai, K., Imoto, Y., Sakashita, M., et al. (2011). Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet.* 7:e1002170. doi: 10.1371/journal.pgen.1002171

Olivares-Chauvet, P., Fennessy, D., Jackson, D. A., and Maya-Mendoza, A. (2011). Innate structure of DNA foci restricts the mixing of DNA from different chromosome territories. *PLoS ONE* 6:e27527. doi: 10.1371/journal.pone.0027527

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36, 1065–1071. doi: 10.1038/ng1423

Osborne, C. S., Chakalova, L., Mitchell, J. A., Horton, A., Wood, A. L., Bolland, D. J., et al. (2007). Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* 5:e192. doi: 10.1371/journal.pbio.0050192

O'Sullivan, J., Hendy, M., Pichugina, T., Wake, G., and Langowski, J. (2013). The statistical-mechanics of chromosome conformation capture. *Nucleus* 4, 390–398. doi: 10.4161/nucl.26513

Park, S. K. (2011). Genomic approaches for the understanding of aging in model organisms. *BMB Rep.* 44, 291–297. doi: 10.5483/BMBRep.2011.44.5.291

Razin, S. V., Gavrilov, A. A., and Yarovaya, O. V. (2010). Transcription factories and spatial organization of eukaryotic genomes. *Biochemistry (Mosc.)* 75, 1307–1315. doi: 10.1134/S0006297910110015

Rodley, C. D., Bertels, F., Jones, B., and O'sullivan, J. M. (2009). Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genet. Biol.* 46, 879–886. doi: 10.1016/j.fgb.2009.07.006

Rodley, C. D. M., Grand, R. S., Gehlen, L. R., Greyling, G., Jones, M. B., and O'Sullivan, J. M. (2012). Mitochondrial-nuclear DNA interactions contribute to the regulation of nuclear transcript levels as part of the inter-organelle communication system. *PLoS ONE* 7:e30943. doi: 10.1371/journal.pone.0030943

Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A., and Misteli, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat. Genet.* 34, 287–291. doi: 10.1038/ng1177

Rouquette, J., Cremer, C., Cremer, T., and Fakan, S. (2010). Functional nuclear architecture studied by microscopy: present and future. *Int. Rev. Cell Mol. Biol.* 282, 1–90. doi: 10.1016/S1937-6448(10)82001-5

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42, 53–61. doi: 10.1038/ng.496

Schork, A. J., Thompson, W. K., Pham, P., Torkamani, A., Roddey, J. C., Sullivan, P. F., et al. (2013). All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* 9:e1003449. doi: 10.1371/journal.pgen.1003449

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311.

Smith, E. N., Koller, D. L., Panganiban, C., Szelinger, S., Zhang, P., Badner, J. A., et al. (2011). Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes. *PLoS Genet.* 7:e1002134. doi: 10.1371/journal.pgen.1002134

Sotelo, J., Esposito, D., Duhagon, M. A., Banfield, K., Mehalko, J., Liao, H., et al. (2010). Long-range enhancers on 8q24 regulate c-Myc. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3001–3005. doi: 10.1073/pnas.0906067107

Stenson, P. D., Mort, M., Ball, E. V., Howells, K., Phillips, A. D., Thomas, N. S., et al. (2009). The human gene mutation database: 2008 update. *Genome Med.* 1, 13. doi: 10.1186/gm13

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., and De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465. doi: 10.1016/S1097-2765(02)00781-5

Uddin, M., Sturge, M., Peddle, L., O'rielly, D. D., and Rahman, P. (2011). Genome-wide signatures of "rearrangement hotspots" within segmental duplications in humans. *PLoS ONE* 6:e28853. doi: 10.1371/journal.pone.0028853

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351. doi: 10.1126/science.1058040

Vernot, B., Stergachis, A. B., Maurano, M. T., Vierstra, J., Neph, S., Thurman, R. E., et al. (2012). Personal and population genomics of human regulatory variation. *Genome Res.* 22, 1689–1697. doi: 10.1101/gr.134890.111

Visel, A., Rubin, E. M., and Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature* 461, 199–205. doi: 10.1038/nature08451

Wang, J. H., Gostissa, M., Yan, C. T., Goff, P., Hickernell, T., Hansen, E., et al. (2009). Mechanisms promoting translocations in editing and switching peripheral B cells. *Nature* 460, 231–236. doi: 10.1038/nature08159

Wang, M. C., Chen, F. C., Chen, Y. Z., Huang, Y. T., and Chuang, T. J. (2012). LDGIdb: a database of gene interactions inferred from long-range strong linkage disequilibrium between pairs of SNPs. *BMC Res. Notes* 5:212. doi: 10.1186/1756-0500-5-212

Wasserman, N. F., Aneas, I., and Nobrega, M. A. (2010). An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a Myc enhancer. *Genome Res.* 20, 1191–1197. doi: 10.1101/gr.105361.110

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43, 519–525. doi: 10.1038/ng.823

Zhan, M., Yamaza, H., Sun, Y., Sinclair, J., Li, H., and Zou, S. (2007). Temporal and spatial transcriptional profiles of aging in *Drosophila melanogaster*. *Genome Res.* 17, 1236–1243. doi: 10.1101/gr.6216607

Zhang, X., Cowper-Sal Lari, R., Bailey, S. D., Moore, J. H., and Lupien, M. (2012). Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res.* 22, 1437–1446. doi: 10.1101/gr.135665.111

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347. doi: 10.1038/ng1891

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.