

OPINION ARTICLE

Understanding Non-Mendelian Genetic Risk

Gerhard A. Coetzee^{1,*}¹Center for Neurodegenerative Science, Van Andel Institute, Grand Rapids, MI, USA

Abstract: This opinion paper highlights strategies for a better understanding of non-Mendelian genetic risk that was revealed by genome-wide association studies (GWAS) of complex diseases. The genetic risk resides predominantly in non-coding regulatory DNA, such as in enhancers. The identification of mechanisms, the causal variants (mainly SNPs), and their target genes are, however, not always apparent but are likely involved in a network of risk determinants; the identification presents a bottle-neck in the full understanding of the genetics of complex phenotypes. Here, we propose strategies to identify functional SNPs and link risk enhancers with their target genes. The strategies are 1) identifying fine-mapped SNPs that break/form response elements within chromatin bio-features in relevant cell types 2) considering the nearest gene on linear DNA, 3) analyzing eQTLs, 4) mapping differential DNA methylation regions and relating them to gene expression, 5) employing genomic editing with CRISPR/cas9 and 6) identifying topological associated chromatin domains using chromatin conformation capture.

ARTICLE HISTORY

Received: September 30, 2019

Revised: October 07, 2019

Accepted: October 07, 2019

DOI:

10.2174/1389202920666191018085511

Keywords: GWAS functionality, chromatin, genomics, SNP, genes, non-mendelian genetic risk.

1. INTRODUCTION

Non-Mendelian genomic risk is a relatively new field in understanding genetic diseases, such as cancer. Non-Mendelian genomic risk contrasts with Mendelian-inherited mutations; the latter can be followed in families *via* linkage analyses and have been known and studied for some time [1]. The contrast was revealed by genome-wide association studies (GWAS) of many complex traits, which identified non-Mendelian risk loci containing polymorphic variants mostly occurring in non-coding DNA; functional analyses of such risk have largely lagged the original GWAS signal identification. Until 2016, out of 3,836 successful GWAS studies, only 84 revealed some (but not complete) mechanistic understanding [2]. In most studies, risk SNPs have only been associated with disease (and not causing it). The reasons are that a GWAS signal at a particular locus has several surrogates in linkage disequilibrium and these, in turn, are linked to the functions of closely mapped (nearest) genes of interest [2]. Therefore, two main questions remain: Which SNP or SNPs are functional/causal and which genes functionally translate the risk signal.

2. FUNCTIONAL/CAUSAL RISK SNPs

SNP alleles are of various degrees in linkage disequilibrium (LD) in different racial-ethnic groups and at different loci [3]. Fine mapping can be achieved by direct genotyping and/or imputation of 1000 genomes [4]; SNPs with the lowest p-values and greatest effect sizes are the most likely causal

ones. The functionality of SNPs in cell types can be gleaned from nucleosome positioning (DNaseI-hypersensitivity and ATAC-seq) and surrounding histone modifications (h3K27ac and H3K1me) [5], SNPs at these sites are likely to be functionally involved, especially if they break/form transcription factor motifs [6].

3. IDENTIFICATION OF GENES THAT FUNCTIONALLY TRANSLATE THE RISK SIGNAL

Genes near GWAS risk signals may not necessarily impose risk. This is because many/most GWAS risk signals reside in regulatory DNA, such as enhancers. Matching enhancers with genes has revealed wide-spread intergenic and intragenic (intronic) interactions, which in turn may control gene expression at some genomic distance [7]. There are several examples of risk enhancers controlling genes at a distance, but a striking one is an obesity- and type 2 diabetes-associated non-coding sequence within the intron of gene FTO, which was shown to be functionally connected with gene IRX3, a megabase distant [8]. Even at very short map distances, more than 40% of enhancers skip over the nearest gene and interact with distant ones [9]. Furthermore, some enhancers regulate multiple genes and several enhancers interact with a given gene [10, 11]. Enhancers, containing risk SNPs (risk enhancers), may influence phenotypes (both normal and pathological) *via* complex mechanisms [12]. The problem of understanding non-Mendelian genetic risk can be formulated, in the first instance, as which are the causal SNPs and how best to match risk enhancers with promoters of the genes they regulate, thus revealing risk mechanisms.

Five enhancer/promoter matching strategies can be considered to shed light on the above conundrum.

*Address correspondence to this author at the Center for Neurodegenerative Science, Van Andel Institute, Grand Rapids, MI, USA; Tel: +1(616)238-4903; E-mail: gerry.coetzee@vai.org

4. STRATEGIES TO MATCH ENHANCERS WITH GENE PROMOTER

4.1. Nearest Gene

This strategy is used most often in gene identification of GWAS loci. In some cases, (but not in all), this makes perfect sense, especially if the nearest gene also happens to carry Mendelian-inherited mutations. Germline mutations and SNPs at genes such as *TERT*, *p53*, and *BRCA1/2* indicate that these genes are involved in cancer etiology [13]. However, as stated above, in most cases, the nearest genes are not involved in complex disease risk.

4.2. eQTL Analyses

Expression-quantitative-trait-loci (eQTL) analyses are based on the correlation between variant genotypes (homozygous, heterozygous, homozygous alternate allele) and gene expression among a large number of samples [14]. In this approach, a priori candidates must be identified to see which genes in Cis are the most likely ones since genome-wide analyses suffer from multiple-hypotheses restrictions of significance. eQTL analyses suffer from lack of power (thus resulting in false negatives), cell type heterogeneity and false positives due to stochastic variation and abundant association [15].

4.3. Epigenetic Traits

By comparing DNA methylation with gene expression levels, one can correlate increased DNA methylation at enhancers with gene expression inhibition in multiple samples. The method is based on the negative correlation between CpG methylation at enhancers and their gene regulatory activities. This method was recently employed by comparing tumor with normal tissues [9, 16]. The advent of genome-wide bisulfate sequencing (to detect all methylated DNA sites), will in the future, reveal active and inactive enhancers in many cell types. A more recent software update tool has been published, called ELMER2 [17]. It is important to note that this type of analysis is only correlative and thus cannot be used to understand precise mechanisms of direct interactions.

4.4. Genomic Editing

Genome editing using CRISPR/cas9 technology has gained prominence due to the amazing precision by which this can be done and the potential benefits that can be achieved both *in vitro* (experimental systems) and *in vivo* (correcting genetic defects) [18]. This powerful technology may be employed to understand the enhancer/target gene pairs. Enhancers containing risk variants can be edited using CRISPR/cas9 by direct deletion or allelic replacement [19]. Both manipulation types can be followed by RNA-seq to determine changes in gene expression because of the manipulation. Risk enhancers and insolated sites (CTCF binding) containing risk SNPs can be manipulated by bringing different enhancer/promoter matches into play [20]. A major concern is a possibility that the guide-RNAs used to target the locus in question, may also bind to irrelevant sites and create off-target artifacts; this can be addressed by using different non-overlapping guide-RNAs on the same locus but

this is expensive and labor-intensive. It is important to note that this approach does not distinguish between direct and indirect effects as mediated by intermediate genes.

4.5. Chromatin Conformation Capture

Several versions of this approach have been developed. At the basis of these approaches is the crosslinking of intact chromatin, followed by restriction enzyme digestion and subsequent ligation. This covalently links DNA fragments, which in linear DNA are at a distance, revealing looping and what has been described at topological-associated domains (TADs). Gene-enhancer pairs most likely function within such TADs. Several versions of this strategy exist. Conformation chromatin capture (3C) between one to one interactions, chromatin conformation capture using circular DNA (4C) between a locus (aka viewpoint) and genome-wide targets, chromatin conformation capture using carbon-copy techniques (5C) between many interactions at a locus and finally chromatin conformation capture genome-wide (Hi-C) between many viewpoints and targets. Many reviews have been written on this topic; here are two reviews [21, 22]. These powerful methods directly assess enhancer/promoter interactions and are not dependent on gene intermediates. False-positive interactions may result from fortuitous non-relevant interactions.

CONCLUSION

The strategies outlined above are complementary and should be used in combination to ensure viable mechanistic insight into non-Mendelian genetic risk. Such understanding will yield novel insight (and likely therapeutic targets) in the genetic etiology of complex diseases.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

Funding is from the NIH (R01CA136924-09 and R01CA190182-04) and the Van Andel Institute, USA.

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

I appreciate the input from members of my lab, Steven Pierce, Trevor Tyson, JC van der Schans, Alix Booms, and Jordan Prahl.

REFERENCES

- [1] Rice, J.P.; Saccone, N.L.; Corbett, J. The lod score method. *Adv. Genet.*, **2001**, *42*, 99-113. [[http://dx.doi.org/10.1016/S0065-2660\(01\)42017-7](http://dx.doi.org/10.1016/S0065-2660(01)42017-7)] [PMID: 11037316]
- [2] Gallagher, M.D.; Chen-Plotkin, A.S. The Post-GWAS era: From association to function. *Am. J. Hum. Genet.*, **2018**, *102*(5), 717-730. [<http://dx.doi.org/10.1016/j.ajhg.2018.04.002>] [PMID: 29727686]
- [3] Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.*, **2008**,

- 9(6), 477-485.
[<http://dx.doi.org/10.1038/nrg2361>] [PMID: 18427557]
- [4] Marchini, J.; Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.*, **2010**, *11*(7), 499-511. [<http://dx.doi.org/10.1038/nrg2796>] [PMID: 20517342]
- [5] Buffry, A.D.; Mendes, C.C.; McGregor, A.P. The functionality and evolution of eukaryotic transcriptional enhancers. *Adv. Genet.*, **2016**, *96*, 143-206. [<http://dx.doi.org/10.1016/bs.adgen.2016.08.004>] [PMID: 27968730]
- [6] Coetzee, S.G.; Coetzee, G.A.; Hazelett, D.J. motifbreakR: An R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics*, **2015**, *31*(23), 3847-3849. [<http://dx.doi.org/10.1093/bioinformatics/btv470>] [PMID: 26272984]
- [7] Li, G.; Ruan, X.; Auerbach, R.K.; Sandhu, K.S.; Zheng, M.; Wang, P.; Poh, H.M.; Goh, Y.; Lim, J.; Zhang, J.; Sim, H.S.; Peh, S.Q.; Mulawadi, F.H.; Ong, C.T.; Orlov, Y.L.; Hong, S.; Zhang, Z.; Landt, S.; Raha, D.; Euskirchen, G.; Wei, C.L.; Ge, W.; Wang, H.; Davis, C.; Fisher-Aylor, K.I.; Mortazavi, A.; Gerstein, M.; Gingeras, T.; Wold, B.; Sun, Y.; Fullwood, M.J.; Cheung, E.; Liu, E.; Sung, W.K.; Snyder, M.; Ruan, Y. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, **2012**, *148*(1-2), 84-98. [<http://dx.doi.org/10.1016/j.cell.2011.12.014>] [PMID: 22265404]
- [8] Smemo, S.; Tena, J.J.; Kim, K.H.; Gamazon, E.R.; Sakabe, N.J.; Gómez-Marín, C.; Aneas, I.; Credidio, F.L.; Sobreira, D.R.; Wasserman, N.F.; Lee, J.H.; Puvion, V.; Tam, D.; Shen, M.; Son, J.E.; Vakili, N.A.; Sung, H.K.; Naranjo, S.; Acemel, R.D.; Manzanares, M.; Nagy, A.; Cox, N.J.; Hui, C.C.; Gomez-Skarmeta, J.L.; Nóbrega, M.A. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, **2014**, *507*(7492), 371-375. [<http://dx.doi.org/10.1038/nature13138>] [PMID: 24646999]
- [9] Yao, L.; Shen, H.; Laird, P.W.; Farnham, P.J.; Berman, B.P. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.*, **2015**, *16*, 105. [<http://dx.doi.org/10.1186/s13059-015-0668-3>] [PMID: 25994056]
- [10] Jin, F.; Li, Y.; Dixon, J.R.; Selvaraj, S.; Ye, Z.; Lee, A.Y.; Yen, C.A.; Schmitt, A.D.; Espinoza, C.A.; Ren, B. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, **2013**, *503*(7475), 290-294. [<http://dx.doi.org/10.1038/nature12644>] [PMID: 24141950]
- [11] Pennacchio, L.A.; Bickmore, W.; Dean, A.; Nobrega, M.A.; Bejerano, G. Enhancers: Five essential questions. *Nat. Rev. Genet.*, **2013**, *14*(4), 288-295. [<http://dx.doi.org/10.1038/nrg3458>] [PMID: 23503198]
- [12] Coetzee, G.A.; Pierce, S. The five dimensions of parkinson's disease genetic risk. *J. Parkinsons Dis.*, **2017**. [<http://dx.doi.org/10.3233/JPD-171256>] [PMID: 29254107]
- [13] Parry, E.M.; Gable, D.L.; Stanley, S.E.; Khalil, S.E.; Antonescu, V.; Florea, L.; Armanios, M. Germline mutations in DNA repair genes in lung adenocarcinoma. *J. Thorac. Oncol.*, **2017**, *12*(11), 1673-1678. [<http://dx.doi.org/10.1016/j.jtho.2017.08.011>] [PMID: 28843361]
- [14] Brandt, M.; Lappalainen, T. SnapShot: Discovering genetic regulatory variants by QTL analysis. *Cell*, **2017**, *171*, 980. [<http://dx.doi.org/10.1016/j.cell.2017.10.031>]
- [15] Liu, B.; Gloudemans, M.J.; Rao, A.S.; Ingelsson, E.; Montgomery, S.B. Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.*, **2019**, *51*(5), 768-769. [<http://dx.doi.org/10.1038/s41588-019-0404-0>] [PMID: 31043754]
- [16] Rhie, S.K.; Guo, Y.; Tak, Y.G.; Yao, L.; Shen, H.; Coetzee, G.A.; Laird, P.W.; Farnham, P.J. Identification of activated enhancers and linked transcription factors in breast, prostate, and kidney tumors by tracing enhancer networks using epigenetic traits. *Epigenetics Chromatin*, **2016**, *9*, 50. [<http://dx.doi.org/10.1186/s13072-016-0102-4>] [PMID: 27833659]
- [17] Silva, T.C.; Coetzee, S.G.; Gull, N.; Yao, L.; Hazelett, D.J.; Noush-mehr, H.; Lin, D.C.; Berman, B.P. ELMER v.2: An R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics*, **2019**, *35*(11), 1974-1977. [<http://dx.doi.org/10.1093/bioinformatics/bty902>] [PMID: 30364927]
- [18] Bak, R.O.; Gomez-Ospina, N.; Porteus, M.H. Gene editing on center stage. *Trends Genet.*, **2018**, *34*(8), 600-611. [<http://dx.doi.org/10.1016/j.tig.2018.05.004>] [PMID: 29908711]
- [19] Soldner, F.; Stelzer, Y.; Shivalila, C.S.; Abraham, B.J.; Latourelle, J.C.; Barrasa, M.I.; Goldmann, J.; Myers, R.H.; Young, R.A.; Jaenisch, R. Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature*, **2016**, *533*(7601), 95-99. [<http://dx.doi.org/10.1038/nature17939>] [PMID: 27096366]
- [20] Guo, Y.; Perez, A.A.; Hazelett, D.J.; Coetzee, G.A.; Rhie, S.K.; Farnham, P.J. CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops. *Genome Biol.*, **2018**, *19*(1), 160. [<http://dx.doi.org/10.1186/s13059-018-1531-0>] [PMID: 30296942]
- [21] Dekker, J.; Misteli, T. Long-range chromatin interactions. *Cold Spring Harb. Perspect. Biol.*, **2015**, *7*(10), a019356. [<http://dx.doi.org/10.1101/cshperspect.a019356>] [PMID: 26430217]
- [22] Dekker, J.; Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell*, **2016**, *164*(6), 1110-1121. [<http://dx.doi.org/10.1016/j.cell.2016.02.007>] [PMID: 26967279]