Proceedings of The National Conference
On Undergraduate Research (NCUR) 2017
University of Memphis, TN
Memphis Tennessee
April 6-8, 2017

Determination of Biomarkers for or against Autoimmune Diseases In DNA Replication and Repair and Galectin-3 Enzymes.

Elena J. Hunsanger
Department of Chemistry
Wayne State University
42 West Warren Avenue
Detroit, Michigan 48202

Faculty Advisor: Dr. G. Andres Cisneros

Abstract

Systematic Lupus Erythematosus (SLE) and celiac disease are autoimmune diseases that are caused by the body's immune system targeting its own tissues, causing chronic inflammation and damage. This study searches for single nucleotide polymorphism (SNP) biomarkers associated with these diseases, to potentially open up new avenues of prevention, diagnosis, and treatment for them. The sample size for the SLE database was (approximately) 1150, and 2300 for the celiac disease database. A subset of SNPs in genes coding for the following proteins were examined using the program HyDn-SNP-S: Galectin-3, Mismatch Repair proteins, Human AlkB homologs, and the DNA polymerases. Two genes were found to contain SNPs in exonic regions: two exonic SNPs were found in mismatch repair protein MutS Homolog 3 (MSH3) and one in DNA Polymerase Alpha 2 (POLA2). Analysis of SLE population data showed that the two MSH3 SNPs are unlikely to have a significant association with SLE, but that the POLA2 SNP may decrease the incidence of the disease in carriers versus the general population. Previous studies show that POLA2 malfunction in T cells is connected with immunosuppression, indicating a potential mechanism for the protective effect.

Keywords: Single Nucleotide Polymorphisms, Autoimmune Disease, Biomarkers

1. Introduction

Autoimmune diseases present a prevalent challenge to our modern medical world: a group of 24 most-detectable autoimmune affected approximately 1 out of every 31 Americans in 1997.[1] The body misidentifying as antigens friendly tissues and cells is the underlying cause of these diseases. Autoimmune diseases can have a drastic impact the lives of those affected, with symptoms from chronic pain to digestive problems to neurological issues.[1] With such a broad range of diseases and presentations of symptoms, breaking down the field into smaller sections is helpful for purposes of study, diagnosis, and treatment.

The two particular autoimmune diseases that this study examines are Systematic Lupus Erythematosus (SLE) and celiac disease. SLE presents with a constellation of symptoms and a variety of intensities, ranging from bothersome to debilitating. The disease presents with a characteristic butterfly rash on the face, and other neurological and respiratory symptoms.[2] Celiac disease presents mainly with chronic inflammation of the gastrointestinal tract upon exposure to gluten, a chemical common in barley, wheat, and rye.[3] This can lead to malnutrition and poor quality of life.[1]

The etiology of these diseases is still unknown, though there are many investigations of the various possible causes. However, these diseases can still be diagnosed, treated, and even possibly prevented by testing and detection. Yet at this time there are not a large number of biomarkers known for these phenotypes.[4]

One useful subset of biomarkers consists of the single nucleotide polymorphisms (SNPs). SNPs are a frequently occurring type of genetic variation in human populations.[5] A SNP is a change of a single nucleotide in a genetic sequence. These mutations can be located in non-coding regions of DNA, or in regions that code for proteins. Of particular interest are SNPs located in the exons of genes; these regions contain the genetic code for proteins. A synonymous SNP in an exonic region will not alter the amino acid sequence, but a non-synonymous SNP will change the final amino acid sequence. This change can have significant– and potentially devastating– effects on the body's cellular processes.[5]

Deoxyribonucleic acid (DNA) replication and repair are essential to the cell cycle and cellular processes.[6] Any alteration or damage to these systems can have wide-reaching consequences for the health of those affected. Damage can come from a range of physical and chemical sources.[7] These can be mistakes made in the process of normal cell growth and function such as insertions, deletions, and mismatches of bases in DNA replication, or external agents as ionizing radiation inducing double-stranded breaks in the DNA or alkylating agents that cause damaging alkylation of the DNA molecule.[7]

There are many cellular mechanisms and systems in place to repair the damage done to DNA by the environment.[6,8] In eukaryotic cells, a many genes are involved in the complex processes of DNA replication and repair.[6,8,9,10,11] This study will focus on a selected subset of genes coding for the following human replication and repair proteins: Galectin-3, mismatch repair proteins, Human AlkB homologs, and the DNA polymerases. If the genes are damaged in a way that causes alteration in their function, it can have a cascade of devastating effects on numerous cellular processes.[8,9,10,11] The exact function of Galectin-3 is not known at this time, but it is connected to cell growth and the cell cycle.[12] Mismatch repair proteins 2 and 3 function together to fix base mismatch errors, for example a G/T pairing, made during DNA replication.[11] AlkBH2 and AlkBH3 are homologs of the bacterial protein AlkB, and serve the function of repairing alkylation damage done to DNA by alkylating agents.[10] The DNA polymerases are essential to DNA replication and repair. DNA Polymerase Alpha 1 and 2 initiate DNA replication, Polymerase Beta repairs small single-strand DNA gaps, and Polymerase Delta 1, a subunit of the Polymerase Delta complex, is involved in DNA replication and repair. [9]

2. Materials and Methods

The data used for this study were found from studies in the Database of Genotypes and Phenotypes (dbGAP) maintained by the National Center for Biotechnology Information. This database contains information on SNPs and numerous phenotypes, ranging from the autoimmune diseases investigated here, to phenomena and conditions such as aging and heart disease. This data is submitted to dbGAP for the use of other research studies, and as such this study utilizes the genotype and phenotype data found from these previous studies to search for potential disease biomarkers.

The SNP data, downloaded from dbGAP reports, was then examined using Hypothesis Driven Single Nucleotide Polymorphism Search (HyDn-SNP-S), a program developed by the Cisneros Research Group, then at Wayne State University, now of the University of North Texas.[5] This program takes SNP data downloaded from dbGAP and scans within genomic parameters delineated by the investigator.[5] These parameters were chosen based on the known locations of start or end points of the genes in question, along with their chromosome location. This program selected, from genome-wide sets of SNP data, the SNP data only from the relevant genes under investigation for associations with the autoimmune diseases in question. This data was output into files labeled by disorder and gene name.

Subsequently, the SNP data was annotated using the Database of Single Nucleotide Polymorphisms (dbSNP), also maintained by the National Center for Biotechnology Information. Information about intronic or exonic character was found for each of the SNPs collected from the genomic regions under study. Information was also gathered as to whether the mutation was synonymous to the wild-type allele or non-synonymous– and thus potentially altering resultant protein function. If the SNP was found to be exonic in character, additional information to characterize the SNP was recorded on SNP contig position, allele change, position in the chromosome, residue change if non-synonymous, and position in the resulting altered protein.

Once a SNP was determined to be exonic and non-synonymous, the complete reports of phenotype and genotype information for each study that the SNP data had been downloaded from were accessed and downloaded from dbGAP. The phenotype and genotype information from the study on Systematic Lupus Erythematosus were annotated with case or control status and allele composition– homozygous wild type, heterozygous for the two possible versions of the SNP, and heterozygous for the less frequent version of the allele. The phenotype and genotype information for

celiac disease were not annotated in the same manner, and consequently could not be put through statistical analysis in the timeframe of the study.

Once the population phenotype data were obtained for the exonic non-synonymous SNPs, biostatistical analysis was run using the application R, which can be utilized to perform mathematical operations quickly and efficiently.[13] A chi-squared test was utilized to determine whether the null hypothesis of the SNP having no effect on the frequency of the disease could be rejected. Furthermore, the percent composition of case and control populations was analyzed to examine whether the different allele combinations caused by the SNP were correlated with different frequencies of the disease in the populations in the study.

The population numbers for cases and controls, and the results of the statistical analysis, were then tabulated.

3. Results and Discussion

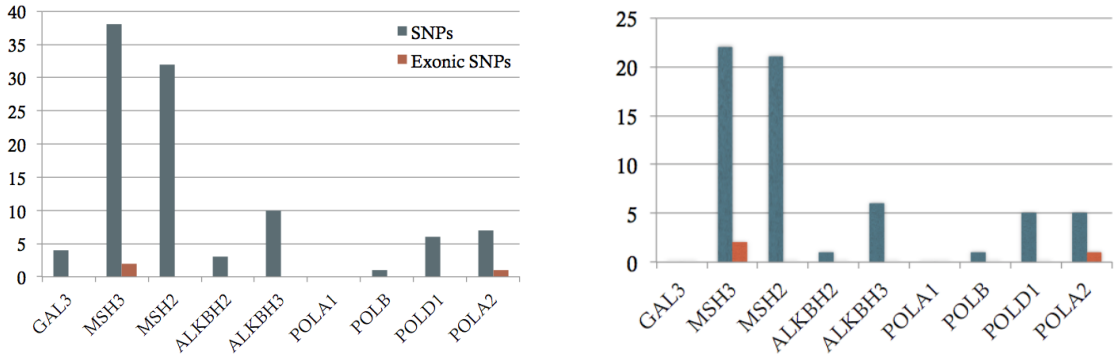

Figure 1: Number of SNPs found in celiac disease (left) and SLE (right) data, both total SNPs (blue bar) and the number out of those which were exonic (red bar).

In the SLE dataset, zero SNPs were found in GAL3, twenty-two in MSH3, twenty-one in MSH2, one in ALKBH2, six in ALKBH3, zero in POLA1, one in POLB, five in POLD1, and five in POLA2 Figure 1.

GAL3, MSH2, ALKBH2, ALKBH3, POLA1, POLB, and POLD1 were not found to contain any exonic non-synonymous SNPs in the datasets for either disease studied. However, in MSH3 two SNPs were found both in the celiac disease data and in the SLE data: rs184967 and rs26279. Another exonic non-synonymous SNP was found in POLA2: rs487989. Rs184967 is a base change from A to G, and results in an amino acid residue change of glutamine to arginine. Rs26279 is the inverse change, from G to A, and leads to a residue change of alanine to threonine. Rs487989 is another G to A change, and results in the substitution of arginine for glycine. Interestingly, these three SNPs were all present in both of the datasets.

Table 1: SLE population data for each exonic non-synonymous SNP, assorted by phenotype and genotype, with associated P values. G11 is the homozygous minor allele, G12 is the heterozygote, and G22 is the homozygous major allele.

SNP	G11 count (case)	G12 count (case)	G22 count (case)	G11 count (control)	G12 count (control)	G22 count (control	P value
rs184967	30 GG	353 AG	927 AA	100 GG	939 AG	2295 AA	0.2599
rs26279	669 AA	527 GA	114 GG	1640 AA	1390 GA	310 GG	0.468
rs487989	19 AA	308 GA	975 GG	84 AA	893 GA	2298 GG	0.001846

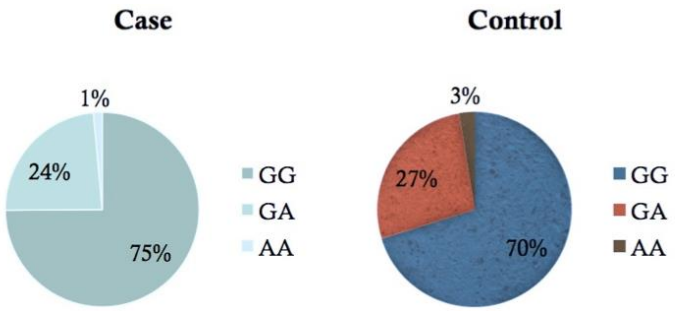

Figure 2: Population composition data, showing the relative frequencies of each set of alleles in the SNP by rounded percent.

Upon statistical analysis of the SLE population phenotype and genotype data, P values were found for rs184967 and rs26279. The P value associated with rs184967 was 0.2599, and 0.468 for rs26279 [Table 1]. As the P value of 0.05 is the accepted standard required to reject the null hypothesis, the null hypothesis could not be rejected with regard to the two exonic non-synonymous SNPs present in MSH3. However, for rs487989 in POLA2, the P value for the SNP was found to be 0.001846. This value is far smaller than the level required to reject the null hypothesis.

The population composition analysis performed on the SLE phenotype and genotype data for this SNP gave interesting results. The observed relative frequency of GG (major homozygote), AG (heterozygote) and AA (minor homozygote) was 0.749, 0.237, and 0.015, respectively, for the cases. The frequencies were 0.702, 0.273, and 0.026, respectively, for the controls. Notably, this data shows a statistically significant difference in the frequency of the SNP between case and control populations [Figure 2]. The SNP has a higher frequency in the control population.

Considering the resulting pattern of data, this SNP does not appear to have clinical value as a biomarker for SLE. However, the statistically significant change in disease frequency that it correlates with indicates an interesting potential effect on the disease mechanisms of action. The residue change resulting from this SNP– a change from a small, nonpolar glycine residue [Figure 3] to a larger, polar basic arginine residue– is a significant change in protein side chain properties.

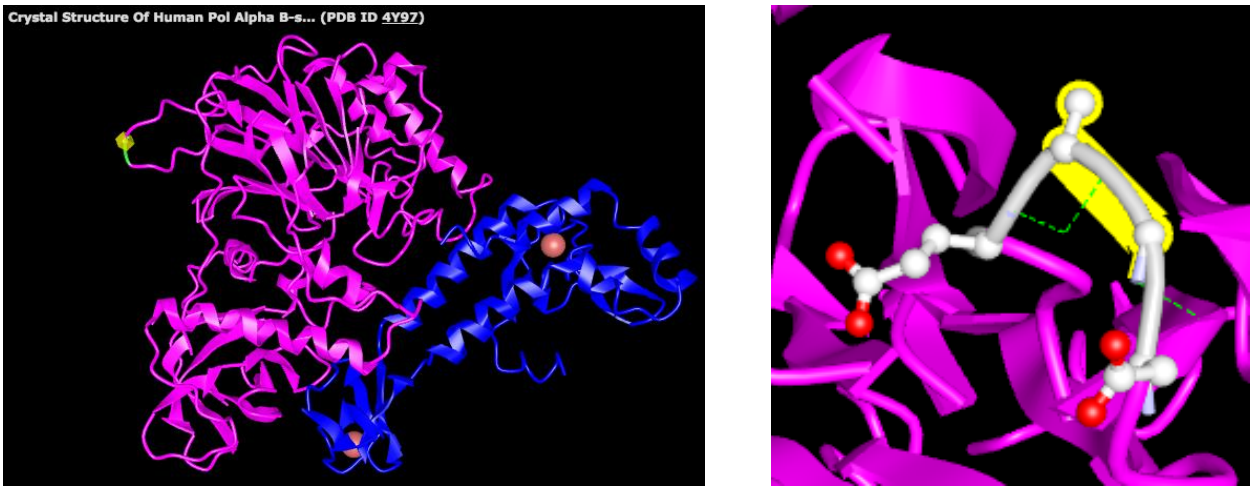

Figure 3: Left– crystal structure of DNA polymerase alpha 2 (fuchsia) complexed with the c-terminal of the catalytic subunit (blue). The glycine residue at position 583, where the SNP locus occurs, in the POLA2 subunit is highlighted in green and yellow. Right– a closer view of the glycine residue at position 583 (highlighted in yellow), two carboxylate residues near it in the protein, and hydrogen bonding (green).

Structure taken from “Crystal Structure of the human pol alpha b subunit in complex with the c-terminal domain of the catalytic subunit. Suwa Y et. al., *J.Biol.Chem.* (2015) 290 p. 14328”

Molecular dynamics simulations may assist in determining what type of structural and functional changes this amino acid residue alteration causes in the resultant protein. However, previous research suggests a possible method of action for this protective effect. Dysfunction, malfunction, or other inhibition of DNA polymerase alpha has been shown to have immunosuppressive effects. A study was published in the Journal Of Immunology on immunosuppression due to factors released by cancer cells. This study proposed that the immunosuppressive effects resulted from the direct inhibition of DNA polymerase alpha.[14] Furthermore, another study from the Proceedings of the National Academy of Sciences found a link between the inhibition of T cell proliferation and suppression of DNA polymerase alpha.[15] These studies indicate a set of connections between immunosuppression, T cell regulation, and POLA function.

This SNP, which changes an amino acid residue in the POLA subunit POLA2 from glycine to arginine, could possibly interfere with the function of POLA. If so, then the connection found in previous research provides a potential path for protective effect this SNP displays against SLE. This disease is characterized by an overactive immune system attacking the body’s own tissues. Therefore, it is possible a mutation that suppresses or otherwise alters the function of POLA may help guard against it, considering the immunosuppressive effect alteration to POLA function has been found to have.

Future work to examine the phenomenon associated with this SNP may include molecular dynamics studies, to confirm or reject the hypothesis that this SNP and the amino acid residue change it leads to cause functional alteration to the POLA2 protein.

4. Acknowledgments

The author would like to thank Dr. G. Andres Cisneros at the University of North Texas and Dr. Gregory Dyson at Wayne State University for the guidance, assistance, and support both provided through the course of this research study. Additionally, this work was supported and made possible by a grant from the Undergraduate Research Opportunities Program at Wayne State University.

The data used for the analyses described in this paper were obtained from the database of Genotypes and Phenotypes (dbGaP), at http://www.ncbi.nlm.nih.gov/gap. Genotype and phenotype data for the Genome-wide Association Study of Celiac Disease (dbGaP accession number phs000274.v1.p1) were provided by Principal Investigators: Susan Neuhausen, PhD. Beckman Research Institute at the City of Hope, Duarte, CA, USA; Chad Garner, PhD. University of California, Irvine, CA, USA; Co-Investigators: Joseph Murray, MD. Mayo Clinic, Rochester, MN, USA; Alessio Fasano, MD. University of Maryland, Baltimore, MD, USA; Peter Green, MD. Columbia University, New York, New York, USA. Funding support for the original study was provided by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA and NIH contract "High throughput genotyping for studying the genetic contributions to human disease". National Institutes of Health, Bethesda, MD, USA.

The data used for the analyses described in this paper were obtained from the database of Genotypes and Phenotypes (dbGaP), at http://www.ncbi.nlm.nih.gov/gap. Genotype and phenotype data for the International Consortium on the Genetics of Systemic Lupus Erythematosus (SLEGEN) (dbGaP accession number phs000216.v1.p1) were provided by Carl D. Langefeld. Funding support for the original study was provided by the Alliance for Lupus Research, the National Institutes of Health, and other sources as detailed in International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD. Genomewide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet. 2008. 40(2):204-10.

5. References

1. Jacobson, Denise L., et al. "Epidemiology and estimated population burden of selected autoimmune diseases in the United States." Clinical immunology and immunopathology 84.3 (1997): 223-243.

2. Cervera, Ricard, et al. "Systemic lupus erythematosus: clinical and immunologic patterns of disease expression in a cohort of 1,000 patients." Medicine 72.2 (1993): 113-124.

3. Green, Peter HR, and Christophe Cellier. "Celiac disease." New England Journal of Medicine 357.17 (2007): 1731-1743.

4. Hueber, Wolfgang, and William H. Robinson. "Proteomic biomarkers for autoimmune disease." Proteomics 6.14 (2006): 4100-4105.

5. Swett, Rebecca J., et al. "Hypothesis driven single nucleotide polymorphism search (HyDn-SNP-S)." DNA repair 12.9 (2013): 733-740.

6. Alberts, Bruce, et al. "DNA replication, repair, and recombination." (2002).

7. Helleday, Thomas, Saeed Eshtad, and Serena Nik-Zainal. "Mechanisms underlying mutational signatures in human cancers." Nature Reviews Genetics 15.9 (2014): 585-598.

8. Sedgwick, Barbara, et al. "Repair of alkylated DNA: recent advances." DNA repair 6.4 (2007): 429-442.

9. Cotterill, Sue, and Kearsey, Stephen(Aug 2014) Eukaryotic DNA Polymerases. In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net. 13 Nov 2015.

10. Duncan, Tod, et al. "Reversal of DNA alkylation damage by two human dioxygenases." Proceedings of the National Academy of Sciences 99.26 (2002): 16660-16665.

11. Habraken, Yvette, et al. "Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3." Current Biology 6.9 (1996): 1185-1187.

12. Kim, Hyeong-Reh Choi, et al. "Cell cycle arrest and inhibition of anoikis by galectin-3 in human breast epithelial cells." Cancer research 59.16 (1999): 4148-4154.

13. Ihaka, Ross, and Robert Gentleman. "R: a language for data analysis and graphics." Journal of computational and graphical statistics 5.3 (1996): 299-314.

14. Venturelli, Donatella, Salvatore Travali, and Bruno Calabretta. "Inhibition of T-cell proliferation by a MYB antisense oligomer is accompanied by selective down-regulation of DNA polymerase alpha expression." Proceedings of the National Academy of Sciences 87.15 (1990): 5963-5967.

15. Oh, S. K., and DAVID P. Lapenson. "E receptor-related immunosuppressive factor in malignant pleural fluid and plasma: molecular mechanism of action on DNA-polymerase-alpha." The Journal of Immunology 135.1 (1985): 355-361.

16. Crystal Structure of the human pol alpha b subunit in complex with the c-terminal domain of the catalytic subunit. Suwa Y et. al., *J.Biol.Chem.* (2015) 290 p. 14328