**Supplementary discussion: Meta-Analysis**

A series of twelve meta-analyses of the three individual AIDs with the four individual NHL subtypes demonstrated seven regions which passed a genome-wide threshold of significance in the meta-analysis, which would not have been discovered in analysis of the individual diseases (Table 3). Effect sizes are modest and total risk explained is low, however, the genes in these regions are worth discussion.

From the CLL and MS meta-analysis, rs140522 is in the 5' untranslated region of *ODF3B*, and rs6793295 is in the *LRRC34* gene, which were identified in recently published meta-analyses of multiple CLL GWAS (Law, et al., 2017) and MS GWAS (International Multiple Sclerosis Genetics Consortium, 2011) (Patsopoulos, et al., 2017). (Note: in that 2011 study (2), the SNP reached a p-value of 10-6, but additional validation targeting this SNP was performed in that study.

From the DLBCL and MS meta-analysis, rs2425752 is in the *NCOA5* gene, a nuclear receptor coactivator. This SNP has also been reported in the latest MS meta-analysis (Patsopoulos, et al., 2017). The gene is reported in solid cancers such as hepatocellular carcinoma (Liu, et al., 2017), luminal breast cancer (Ye, Huang, & Luo, 2017), squamous cell carcinoma (Chen, et al., 2014), and in the Hodgkin lymphoma versus MS meta analysis (Khankhanian, et al., 2016).

From the CLL and RA meta-analysis, rs3731714 lies on an intron of *CASP10*, a member of the cysteine-aspartic acid protease family, a cascade of enzymes which play a central role in the execution-phase of apoptosis. An ENCODE query (ENCODE 2017 -- https://www.encodeproject.org/) revealed significant DNase enrichment in multiple blood and small intestinal cell lines and evidence for altered motifs, including multiple forkhead box binding motifs. CHiP-seq data from ENCODE demonstrated binding of *GATA1*, a hematopoiesis-associated transcription factor, in peripheral blood-derived erythroblasts. Further, rs3731714 has been associated with *CASP10* expression in esophagus mucosa (Consortium, 2015). Mutations in *CASP10* have been associated with autoimmune lymphoproliferative syndrome (ALPS) type II(Wang, et al., 1999). Previous candidate gene studies have suggested that common variants in *CASP10* may be associated with NHL (Lan, et al., 2007) and gastric cancer (Hyland, et al., 2014). Other genes in the region include *PPIL3* and *CFLAR*. *PPIL3* (Peptidylprolyl isomerase like 3) encodes a member of the cyclophilin family. Cyclophilins catalyze the cis-trans isomerization of peptidylprolyl imide bonds in oligopeptides and are thought to act either as catalysts or as molecular chaperones in protein-folding events. *CFLAR* (CASP8 and FADD like apoptosis regulator) encodes a protein that is a regulator of apoptosis and is structurally like *CASP8*; however, the encoded protein lacks caspase activity. Rs3731714 has been associated with *PPIL3* expression in lymphoblastoid cell lines (Lappalainen, et al., 2013), peripheral blood monocytes (Zeller, et al., 2010) and with *CFLAR* expression in blood (Fehrmann, et al., 2011). Thus, the observed association between rs3731714 and the CLL-RA phenotype raises an intriguing possibility for a single allele to be associated with multi-genic consequences resulting in failure of apoptosis of leukocytes.

From MZL and RA, rs16947122 is in the intronic region of *FBXW8* (F-box and WD repeat domain containing 8), which belongs to the Fbw class in the f-box protein family. The F-box proteins function in phosphorylation-dependent ubiquitination and antigen processing. *FBXW8* has also been reported to play an essential role in cancer cell proliferation through proteolysis of cyclin D1(Okabe, et al., 2006). Association with disease endpoints have been reported with this SNP; however, *FBXW8* along with its adjacent genes, *HRK* and *TESC*, were recently suggested to be related to cognition and psychoses (Bulayeva, et al., 2015).

Also from MZL and RA, a group of three SNPs (rs1364229, rs7192064, rs2131402) are found in a gene desert, 500K base pairs upstream from the *CDH8* gene, and over 2M base pairs from the next closest documented human gene. These do not seem to be on any obvious regulatory region (ENCODE 2017 -- https://www.encodeproject.org/ ; and UCSC genome browser, regulatory tracks, 2017 -- https://genome.ucsc.edu/). A PubMed search returned no results for these SNPs (PubMed 2018 -- https://www.ncbi.nlm.nih.gov/pubmed/).

From CLL and SLE, rs1439112 resides on an intron of the *MGAT5* gene which encodes the enzyme mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase V. This enzyme catalyzes the decisive step in the generation of tetra-antennary N-glycans attached to the cell surface, and of secreted glycoproteins. These are key molecules in innate as well as adaptive immunity. For example, some are implicated in the discrimination between “self” and “non-self” (Rudd, 2001). Thus, genetic variability of this enzyme could perturb glycosylation processes instrumental for maintaining immune system tolerance and recognition of foreign antigens such as cancer cells. Of interest, the glycosylation pattern of immunoglobulins contributes to their affinities for Fc and other immune receptors, linking these molecules to the pathogenesis of immune-based diseases. In a mouse model study, loss of *MGAT5* expression lowered the threshold needed for T cell activation, and *MGAT5* deficient mice presented a variety of autoimmune phenotypes (Demetriou, Granovsky, Quaggin, & Dennis, 2001). In humans, GWAS identified variants in *MGAT5* as susceptibility factors for psoriasis (Aterido, et al., 2016), and variants in this gene were associated with the plasma N-glycome (Huffman, et al., 2011). The psoriasis-associated variant impacted N-glycosylation levels as well as activity of CD8+ and CD4+ T cells (Aterido, et al., 2016). In the cancer context, *MGAT5* mRNA and glycan products are upregulated in tumor cells and correlate with poor prognosis (Taniguchi & Kizuka, 2015). *MGAT5* knockout mice showed reduced cancer growth and metastasis (Granovsky, et al., 2000). Blocking expression of MGAT5-modified glycans in breast cancer cell lines led to suppression of tumor progression with activation of Th1 cytokine production and macrophages (Li, et al., 2008). Additionally, rs1439112 affects binding of several transcription factors including *GATA3*(Enciso-Mora, et al.), a regulator of T-helper-Type 2 (Th2) cell development, and EP300, a histone acetyltransferase frequently mutated in lymphoma subtypes (Pasqualucci, et al., 2011). Further, the SNP is predicted to affect an enhancer region in primary B-cells as well as primary T helper memory cells.

A group four SNPs (rs10936599, rs1317082, rs13069553, and rs7621631) at the *TERC-ACTRT3-MYNN* locus were also associated in the meta-analysis of SLE versus CLL. These four SNPs are in high LD with one another (r2 ≥ 0.88 for all pairwise comparisons). This locus was associated with the risk of bladder cancer (Figueroa, et al., 2014), colorectal cancer (Houlston, et al., 2010), multiple myeloma (Chubb, et al., 2013), CLL (Speedy, et al., 2014), MGUS (Weinhold, et al., 2014), pulmonary fibrosis (Fingerlin, et al., 2013), and telomere length (Codd, et al., 2013) at genome-wide significance. Additionally, suggestive associations have been observed for multiple sclerosis (International Multiple Sclerosis Genetics Consortium, 2011) and celiac disease (Dubois, et al., 2010). Several of these variants are located with the *MYNN* (myoneurin) gene, which encodes a member of the *BTB/POZ* and zinc finger domain-containing protein family that is involved in the control of gene expression. Other genes in this region include *ACTRT3* (actin related protein, formerly *ARPM1*), *TERC* (telomerase RNA component), and *LRRC34* (leucine rich repeat containing 34). *TERC* is an attractive biological candidate due to its role in telomere maintenance. Telomerase is a ribonucleoprotein polymerase that maintains telomere ends by the addition of the telomere repeat TTAGGG. The enzyme consists of a protein component with reverse transcriptase activity and a RNA component, encoded by *TERC*, that serves as a template for the telomere repeat. Telomerase expression plays a role in cellular senescence, as it is normally repressed in postnatal somatic cells resulting in progressive shortening of telomeres. Deregulation of telomerase expression in somatic cells may be involved in oncogenesis. Studies in mice suggest that telomerase also participates in chromosomal repair, since *de novo* synthesis of telomere repeats may occur at double-stranded breaks. The C allele at rs10936599 was associated with increased telomere length in genome-wide association studies (from SNPedia July 2018 -- https://www.snpedia.com/), while the alternate allele was found to be the risk allele in our study. Interestingly, patients with SLE had shorter leukocyte telomere lengths than controls in cross-sectional studies (Haque, et al., 2013).

The final SNP of interest in the meta-analysis of CLL versus SLE was rs10069690, an intronic variant in the *TERT* gene (telomerase reverse transcriptase). *TERT* encodes the reverse transcriptase subunit of telomerase, hTERT. The SNP variant has been associated with CLL (Berndt, et al., 2016) (Speedy, et al., 2014) and multiple solid cancers, including breast (Haiman, et al., 2011), testicular (Turnbull, et al., 2010), ovarian (Kuchenbaecker, et al., 2015) and prostate cancer(Kote-Jarai, et al., 2011). Other variants in this region have been associated with other cancer types (Wang, et al., 2014), underscoring the importance of this region in cancer risk. The A (minor) allele of rs10069690 has been shown to result in a splice site variant of hTERT which acts as a dominant negative inhibitor of telomerase (Killedar, et al., 2015). Specifically, the alternatively spliced, INS1b, transcript does not encode the reverse transcriptase domain required for telomerase enzyme activity, but retains its ability to bind to the telomerase RNA subunit, hTR, resulting in decreased telomerase activity and telomere shortening (Killedar, et al., 2015).

# References for supplementary discussion

Aterido, A., Julià, A., Ferrándiz, C., Puig, L., Fonseca, E., Fernández-Ló Pez, E., . . . Marsal, S. (2016). Genome-Wide Pathway Analysis Identifies Genetic Pathways Associated with Psoriasis. *Journal of Investigative Dermatology, 136*, 593-602.

Berndt, S., Camp, N., Skibola, C., Vijai, J., Wang, Z., Gu, J., . . . Slager, S. (2016). Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun, 7*, 1-9.

Bulayeva, K., Lesch, K., Bulayev, O., Walsh, C., Glatt, S., Gurgenova, F., . . . Thompson, P. (2015). Genomic structural variants are linked with intellectual disability. *Journal of Neural Transmission, 122*, 1289-1301.

Chen, G.-q., Tian, H., Yue, W.-m., Li, L., Li, S.-h., Qi, L., . . . Lu, M. (2014). NCOA5 low expression correlates with survival in esophageal squamous cell carcinoma. *Medical Oncology, 100*, 1835-1843.

Chubb, D., Weinhold, N., Broderick, P., Chen, B., Johnson, D., Försti, A., . . . Goldschmidt, H. (2013). Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nature Genetics, 45*, 1221-1225.

Codd, V., Nelson, C., Albrecht, E., Mangino, M., Deelen, J., Buxton, J., . . . Samani, N. (2013). Identification of seven loci affecting mean telomere length and their association with disease. *Nature Genetics, 45*, 422-427.

Consortium, G. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science, 348*, 648-660.

Demetriou, M., Granovsky, M., Quaggin, S., & Dennis, J. (2001). Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature, 409*, 733-739.

Dubois, P., Trynka, G., Franke, L., Hunt, K., Romanos, J., Curtotti, A., . . . van Heel, D. (2010). Multiple common variants for celiac disease influencing immune gene expression. *Nature Genetics, 42*, 295-302.

Enciso-Mora, V., Broderick, P., Ma, Y., Jarrett, R., Hjalgrim, H., Hemminki, K., . . . Houlston, R. (n.d.). A genome-wide association study of Hodgkin Lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21, and 10p14 (GATA3). *42*, 1126-1130.

Fehrmann, R., Jansen, R., Veldink, J., Westra, H., Arends, D., Bonder, M., . . . Franke, L. (2011). Trans-eqtls reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the hla. *PLoS Genetics, 7*, 1-14.

Figueroa, J., Ye, Y., Siddiq, A., Garcia-Closas, M., Chatterjee, N., Prokunina-Olsson, L., . . . Rothman, N. (2014). Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum Mol Genet, 23*, 1387-1398.

Fingerlin, T., Murphy, E., Zhang, W., Peljto, A., Brown, K., Steele, M., . . . Schwartz, D. (2013). Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nature Genetics, 45*, 613-620.

Granovsky, M., Fata, J., Pawling, J., Muller, W., Khokha, R., & Dennis, J. (2000). Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Natural Medicines, 6*, 306-312.

Haiman, C., Chen, G., Vachon, C., Canzian, F., Dunning, A., Millikan, R., . . . Peto, J. (2011). A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor–negative breast cancer. *Nature Genetics, 43*, 1210-1214.

Haque, S., Rakieh, C., Marriage, F., Ho, P., Gorodkin, R., Teh, L., . . . Bruce, I. (2013). Shortened telomere length in patients with systemic lupus erythematosus. *Arthritis Rheum, 65*, 1319-1323.

Houlston, R., Cheadle, J., Dobbins, S., Tenesa, A., Jones, A., Howarth, K., . . . Tomlinson, I. (2010). Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nature Genetics, 42*, 973-977.

Huffman, J., Knežević, A., Vitart, V., Kattla, J., Adamczyk, B., Novokmet, M., . . . Lauc, G. (2011). Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Human Molecular Genetics, 20*, 5000-5011.

Hyland, P., Lin, S.-W., Hu, N., Zhang, H., Wang, L., Su, H., . . . Taylor, P. (2014). Genetic variants in fas signaling pathway genes and risk of gastric cancer. *International Journal of Cancer, 134*, 822-831.

International Multiple Sclerosis Genetics Consortium, T. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature, 476*, 214-9.

Khankhanian, P., Cozen, W., Himmelstein, D., Madireddy, L., Din, L., Van Den Berg, A., . . . Hjalgrim, H. (2016). Meta-analysis of genome-wide association studies reveals genetic overlap between Hodgkin lymphoma and multiple sclerosis. *International Journal of Epidemiology, 45*, 728-740.

Killedar, A., Stutz, M., Sobinoff, A., Tomlinson, C., Bryan, T., Beesley, J., . . . Pickett, H. (2015). A Common Cancer Risk-Associated Allele in the hTERT Locus Encodes a Dominant Negative Inhibitor of Telomerase. *PLoS Genetics, 11*, 1-23.

Kote-Jarai, Z., Olama, A., Giles, G., Severi, G., Schleutker, J., Weischer, M., . . . Eeles, R. (2011). Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nature Genetics, 43* , 785-791.

Kuchenbaecker, K., Ramus, S., Tyrer, J., Lee, A., Shen, H., Beesley, J., . . . Chenevix-Trench, G. (2015). Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nature Genetics, 47*, 164-171.

Lan, Q., Zheng, T., Chanock, S., Zhang, Y., Shen, M., Wang, S., . . . Rothman, N. (2007). Genetic variants in caspase genes and susceptibility to non-Hodgkin lymphoma. *Carcinogenesis, 28*, 823-827.

Lappalainen, T., Sammeth, M., Friedländer, M., ‘t Hoen, P., Monlong, J., Rivas, M., . . . Dermitzakis, E. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature, 501*, 506-511.

Law, P., Berndt, S., Speedy, H., Camp, N., Sava, G., Skibola, C., . . . Slager, S. (2017). Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nature Communications, 8*, 14175.

Li, D., Li, Y., Wu, X., Li, Q., Yu, J., Gen, J., & Zhang, X. (2008). Knockdown of Mgat5 inhibits breast cancer cell growth with activation of CD4+ T cells and macrophages. *The Journal of Immunology, 180*, 3158-3165.

Liu, X., Liu, F., Gao, S., Reske, J., Li, A., Wu, C., . . . Xiao, H. (2017). A single non-synonymous NCOA5 variation in type 2 diabetic patients with hepatocellular carcinoma impairs the function of NCOA5 in cell cycle regulation. *Cancer Letters, 391*, 152-161.

Okabe, H., Lee, S., Puchareon, J., Albertson, D., McCormick, F., & Tetsu, O. (2006). A critical role for FBXW8 and MAPK in cyclin D1 degradation and cancer cell proliferation. *PLoS ONE, 1*, 1-12.

Pasqualucci, L., Dominguez-Sola, D., Chiarenza, A., Fabbri, G., Grunn, A., Trifonov, V., . . . Dalla-Favera, R. (2011). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature, 471*, 189-196.

Patsopoulos, N., Baranzini, S., Santaniello, A., Shoostari, P., Cotsapas, C., Wong, G., . . . De Jager, P. (2017). The multiple sclerosis genomic map: Role of peripheral immune cells and resident microglia in susceptibility. *bioRxiv, 1*, 1-43.

Rudd, P. (2001). Glycosylation and the Immune System. *Science, 291*, 2370-2376.

Speedy, H., Di Bernardo, M., Sava, G., Dyer, M., Holroyd, A., Wang, Y., . . . Houlston, R. (2014). A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat Genet, 46*, 56-60.

Taniguchi, N., & Kizuka, Y. (2015). Glycans and cancer: Role of N-Glycans in cancer biomarker, progression and metastasis, and therapeutics. *Advances in Cancer Research, 126*, 11-51.

Turnbull, C., Rapley, E., Seal, S., Pernet, D., Renwick, A., Hughes, D., . . . Rahman, N. (2010). Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nature Genetics, 42*, 604-607.

Wang, J., Zheng, L., Lobito, A., Chan, F., Dale, J., Sneller, M., . . . Lenardo, M. (1999). Inherited human caspase 10 mutations underlie defective lymphocyte and dendritic cell apoptosis in autoimmune lymphoproliferative syndrome type II. *Cell, 98*, 47-58.

Wang, Z., Zhu, B., Zhang, M., Parikh, H., Jia, J., Chung, C., . . . Amundadottir, L. (2014). Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Human Molecular Genetics, 23*, 6616-6633.

Weinhold, N., Johnson, D., Rawstron, A., Försti, A., Doughty, C., Vijayakrishnan, J., . . . Houlston, R. (2014). Inherited genetic susceptibility to monoclonal gammopathy of unknown significance. *Blood, 123*, 2513-2517.

Ye, X., Huang, D., & Luo, R. (2017). NCOA5 is correlated with progression and prognosis in luminal breast cancer. *Biochemical and Biophysical Research Communications, 482*, 253-6.

Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., . . . Cambien, F. (2010). Genetics and beyond - the transcriptome of human monocytes and disease susceptibility. *PLoS ONE, 5*, 1-15.