



NIH PUBLIC ACCESS

Author Manuscript

*Aging Cell*. Author manuscript; available in PMC 2013 October 01.

Published in final edited form as:

*Aging Cell*. 2012 October ; 11(5): 727–731. doi:10.1111/j.1474-9726.2012.00871.x.

## Review: A Meta-Analysis of GWAS Studies and Age-Associated Diseases

**William R. Jeck**<sup>1,2</sup>, **Alex P. Siebold, Ph.D.**<sup>1</sup>, and **Norman E. Sharpless, M.D.**<sup>1,2,\*</sup>William R. Jeck: [william.jeck@gmail.com](mailto:william.jeck@gmail.com); Alex P. Siebold: [alex.siebold@gmail.com](mailto:alex.siebold@gmail.com); Norman E. Sharpless: [NES@med.unc.edu](mailto:NES@med.unc.edu)<sup>1</sup>Department of Genetics, The Lineberger Comprehensive Cancer Center, The University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA<sup>2</sup>Department of Medicine, The Lineberger Comprehensive Cancer Center, The University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA

### Summary

Genome-Wide Association studies (GWAS) offer an unbiased means to understand the genetic basis of traits by identifying single nucleotide polymorphisms (SNPs) linked to causal variants of complex phenotypes. GWAS have identified a host of susceptibility SNPs associated with many important human diseases, including diseases associated with aging. In an effort to understand the genetics of broad resistance to age-associated diseases (i.e. ‘wellness’), we performed a meta-analysis of human GWAS. Toward that end, we compiled 372 GWAS that identified 1,775 susceptibility SNPs to 105 unique diseases and used these SNPs to create a genomic landscape of disease susceptibility. This map was constructed by partitioning the genome into 200 kb ‘bins’ and mapping the 1,775 susceptibility SNPs to bins based on their genomic location. Investigation of these data revealed significant heterogeneity of disease association within the genome, with 92% of bins devoid of disease-associated SNPs. In contrast, 10 bins (0.06%) were significantly ( $p < 0.05$ ) enriched for susceptibility to multiple diseases, 5 of which formed two highly significant peaks of disease association ( $p < 0.0001$ ). These peaks mapped to the Major Histocompatibility (MHC) locus on 6p21 and the *INK4/ARF (CDKN2a/b)* tumor suppressor locus on 9p21.3. Provocatively, all 10 significantly enriched bins contained genes linked to either inflammation or cellular senescence pathways, and SNPs near regulators of senescence were particularly associated with disease of aging (e.g. cancer, atherosclerosis, type 2 diabetes, glaucoma). This analysis suggests that germline genetic heterogeneity in the regulation of immunity and cellular senescence influences the human health span.

---

\*To whom correspondence should be addressed: Norman E. Sharpless, M.D. The Lineberger Comprehensive Cancer Center, CB #7295, Departments of Medicine and Genetics, The University of North Carolina School of Medicine, Chapel Hill, NC 27599-7295, USA, Tel.: (919) 966-1185, Fax: (919) 966-8212, [NES@med.unc.edu](mailto:NES@med.unc.edu).

#### Author Contributions

WRJ contributed to the study design, analysis and interpretation of data, preparation of figures and tables, and writing and editing the manuscript. APS contributed to the study design, collection and interpretation of data, preparation of figures and tables, and writing and editing the manuscript. NES conceived the study, contributed to the study design, interpretation of data, and writing and editing the manuscript. The authors have no conflict of interest to report.

#### Supporting Information

Supporting information associated with this manuscript is intended to be made available in the online addition and consists a single Microsoft excel file with three supplemental tables listed as follows: Supplemental Table 1 - List of clinically relevant diseases included in this analysis, Supplemental Table 2 - List of non-disease traits, congenital deformities, and medical conditions of limited morbidity excluded from this analysis, Supplemental Table 3 - Genomic coordinates of each bin and their associated number of unique disease hits.

## Keywords

p16<sup>INK4a</sup>, p14<sup>ARF</sup>; ANRIL; TERT; longevity

A central tenet of gerontology is that common pathogenic mechanisms cause age-related phenotypes in disparate organs and tissues. For example, telomere dysfunction in the liver, bone marrow, and lung has been linked to age-associated, tissue-specific diseases such as cirrhosis, aplastic anemia and pulmonary fibrosis, respectively (Armanios 2011). Several broad pathways have been suggested as candidate global modifiers of human aging including sirtuins, insulin/IGF-1, ROS metabolism, inflammation, and cellular senescence. A prediction of the notion that common pathogenic pathways contribute to aging of distinct tissues is that there should be genes whose expression modulates these pathways, and heterogeneous expression of such genes within a population should be associated with multiple, seemingly distinct tissue-specific diseases.

High density single nucleotide polymorphism (SNP) arrays have provided population geneticists a high throughput method for identifying polymorphisms associated with the onset of complex phenotypes (e.g. physiological traits and/or markers, congenital abnormalities, and disease susceptibility/resistances). Large scale, population based studies that utilize SNP arrays to gain insights into gene(s) that may promote/cause a complex phenotype consist of Candidate Gene Association Studies (CGAS) and Genome Wide Association Studies (GWAS). The key difference between these two epidemiological study methods are that CGAS take a hypothesis driven approach whereas GWAS are performed in a non-biased manner (see review by (Jorgensen *et al.* 2009) for more detailed discussion of advantages/disadvantages of these methodologies). Moreover, modern pedigree studies (linkage analyses) can be performed using SNP arrays to perform genome-wide searches to identify variants associated with complex diseases, such as Alzheimer's disease (Zuchner *et al.* 2008), but variants identified from such efforts may be limited to small numbers of actual cases (i.e. individual families).

GWAS have been successfully employed to identify common polymorphic variations that contribute to several complex phenotypes. The value of GWAS are underscored by the ready identification of risk alleles that have been replicated in independent populations, which have identified both novel and known modulators of disease pathogenesis, as well as revealed new therapeutic targets (Altshuler *et al.* 2008). Moreover, the National Human Genome Research Institute (NHGRI) maintains a catalog of published GWAS that currently houses ~1000 studies which have identified >4,500 SNPs to >500 phenotypes (Hindorf *et al.* 2011). In an effort to understand what GWAS tell us about disease of human aging, we performed a meta-analysis of the NHGRI GWAS catalog. In particular, we used this resource to ask in an unbiased, genome-wide manner if there are 'hotspot' loci associated with multiple disease susceptibility/resistance phenotypes. Toward that end, we filtered this NHGRI dataset to only include studies that focused on clinically relevant human diseases. To better visualize chromosomal loci and candidate genes associated with multiple, distinct human diseases; especially age-associated diseases, we summed the frequency of disease associated SNPs in 200 kb bins spanning the whole genome. While clearly 'age-related disease' is not the same thing as 'aging', we elected to focus this analysis on disease susceptibility given the tractability of many well-delineated diseases to GWAS, as opposed to the mixed results obtained for less discrete endpoints (e.g. longevity, frailty, etc). We believe this approach is still of interest to gerontologists given that freedom from disease (wellness) is an essential determinant of health span.

In order to compile and filter GWAS that identified SNPs specific to human disease resistance/susceptibility, the complete 6/29/11 release of the NHGRI GWAS database was downloaded from the NHGRI GWAS website (Hindorff *et al.* 2011). This release contained 932 published GWAS that identified 4,558 SNPs in 511 phenotypes, with each SNP achieving a combined p-value of  $< 1.0 * 10^{-5}$ . Studies included in the catalog are also required to include at least 100,000 SNPs to permit a truly genome-wide analysis. Our analysis did not distinguish between ‘susceptibility’ SNPs and ‘protective’ SNPs, as each ‘susceptibility’ allele implies an alternative ‘protective’ allele at the same location. This dataset was filtered to exclude small GWAS ( $< 300$  cases) as well as those that investigated non-disease traits, congenital deformities, and medical conditions of limited morbidity (e.g. restless leg syndrome). In rare instances ( $n=14$ ) where disease vs. non-disease classification of a GWAS was not obvious, classification was performed with blinding to GWAS results (see supplemental Table 1 and Table 2 for included and excluded ‘diseases’). The inclusion/exclusion of these borderline conditions did not affect the analysis’ conclusions. The filtered GWAS dataset consisted of 372 studies that identified 1,775 SNPs associated with susceptibility/resistance to 105 unique human diseases. The total number of patients from these 372 studies totaled more than 2.3 million individuals from diverse ethnic backgrounds.

This dataset was then used to construct the genome-wide disease susceptibility map (Figure 1). The hg19 release of the human genome was divided into 15,157 bins with each bin containing 200 kb of genomic sequence (see supplemental Table 3 for genomic coordinates and hits associated with each bin). The analysis was also not sensitive to choice of bin size. SNPs from the filtered GWAS dataset were mapped to the binned genome with redundant hits of the same disease to the same bin counted as a single hit. As an example of this, when GWAS SNPs mapped to the same bin of a disease that encompassed more specific disease states within it (e.g. inflammatory bowel disease (IBD), Crohn’s disease, and ulcerative colitis), it was counted as a single disease hit for that bin. Alternatively, if only specific forms of IBD mapped to the same bin, they were counted as individual disease hits for that bin. This approach allowed studies that identified distinct effects on disease subtypes to be included, without over representing studies focused on identifying SNPs associated with disease categories. Distinguishing disease subtypes (i.e. IBD) had minimal impact on our findings, and only shifted one locus, 17q12, above the significance threshold. The number of unique disease associations per bin was then graphed versus chromosomal location in a ‘Manhattan plot’ (Figure 1) and a 10,000 iteration permutation analysis was performed on the mapped SNPs to estimate statistical significance.

We elected to use a permutation test to estimate statistical significance, as this approach accounts for variation in the number of SNPs tested, our method of counting diseases in shared categories, and multiple comparisons of assessing each bin for significance. Permutation testing is the gold standard for determining significance, provided that it is computationally tractable (Johnson *et al.* 2010). In each iteration of this test, all SNPs were randomly and independently assigned to the 15,157 bins that represent the whole genome, and the bin with the maximum number of randomly assigned SNPs was identified. Bins containing more than four unique disease-associated SNPs occurred in less than 5% (i.e.  $p<0.05$ ) of the 10,000 iterations performed, setting this as our threshold for significance (indicated by a dashed line on figure 1). Although it is possible that not all of the 15,000+ bins are assayed equally well by GWAS, the inclusion criterion of the NHGRI requiring at least 100,000 mapping SNPs indicates that the large majority of the genome is covered in these analyses, and the major conclusions of the study remain significant even if the permutation analysis is restricted to a small fraction of the genome.

This analysis revealed substantial heterogeneity in the human genome with regard to disease susceptibility. The majority of bins (13,900 of 15,157; 92%) did not contain any disease-

associated SNPs. In contrast, only ten bins (2 Mb or 0.06% of the genome) showed statistically significant enrichment ( $p < 0.05$ ) for disease association, with two strong ‘peaks’ ( $p < 0.0001$ ) of multi-disease association. The largest peak spanned four neighboring bins (800 kb) that contain the gene-rich MHC locus on chromosome 6p21 (Figure 1). SNPs in this bin were linked to 24 unique diseases, most of which were autoimmune in nature (e.g. asthma, inflammatory bowel disease, lupus, Hodgkins Disease, Table 1), and not classical disease of aging. This finding confirms the well-established pathogenic role of MHC polymorphisms in the development of diverse autoimmune diseases (Fernando *et al.* 2008; Rioux *et al.* 2009). Therefore, while the association of the MHC locus with autoimmune diseases is not surprising, this finding serves as a positive control for the analysis.

The second highest disease susceptibility association peak mapped to a gene-poor bin on chromosome 9p21.3. This bin contains only four transcripts emanating from the *INK4/ARF* (or *CDKN2a/b*) locus, which harbors three related protein-encoding transcripts ( $p15^{INK4b}$ ,  $p16^{INK4a}$ ,  $p14^{ARF}$ ) as well as a long non-coding RNA (*ANRIL*) that is anti-sense to  $p15^{INK4b}$ . The *INK4/ARF* locus is a key mediator of cellular senescence which inhibits cell cycle progression from G<sub>1</sub> to S phase in response to various forms of cellular stress (Sharpless & DePinho 2007). The 9p21.3 bin was linked to 10 unique diseases, almost all of which are age-associated: cancers (e.g. breast, glioblastoma), type 2 diabetes mellitus (T2DM), glaucoma and several atherosclerotic diseases (e.g. stroke, aortic aneurysm, myocardial infarction) (Table 1). It is worth noting the considerable size of these two peaks: the 6p21 and 9p21.3 disease susceptibility hotspots represent 0.03% of the genome, but combined were associated with nearly a third (34 of 105) of the unique diseases analyzed by GWAS.

The remaining five bins (1p31.3, 2p16.1, 5p15.33, 7q32.1, 17q12) that were significantly enriched for disease-associations ( $p < 0.05$ ) were also directly linked to either immunity/inflammation or cellular senescence pathways. The 1p31 and 2p16 bins contain *IL23R* and *REL* respectively, which modulate immunity and lymphocyte biology, and these bins were predominantly associated with autoimmune disease (Table 1). The 5p15.33 bin includes *TERT*, a critical subunit of telomerase, which is associated with cellular senescence by modulating telomere length (Martinez & Blasco 2011). Disease susceptibilities mapping to the 5p15.33 bin were mainly comprised of cancers, consistent with the association between telomere length and cancer susceptibility (Table 1) (Hills & Lansdorp 2009; Willeit *et al.* 2010). The 5p15.33 bin was also associated with idiopathic pulmonary fibrosis (IPF), consistent with the finding of increased IPF in patients with congenital telomerase deficiency (Armanios *et al.* 2007; Tsakiri *et al.* 2007). Candidate genes in the 7q32 and 17q12 bins are less obvious, but these loci were also solely associated with autoimmune or inflammatory diseases (Table 1), suggesting these bins harbor modulators of the immune response. In general, the 5 loci associated with immunity and inflammation were mostly associated with autoimmune diseases (e.g. T1DM, asthma, IBD, Hodgkins disease) and were not as strongly linked to age-associated diseases as the two bins associated with senescence (i.e. cancers, atherosclerosis, T2DM, glaucoma, pulmonary fibrosis).

Although the finding that all loci associated with broad disease resistance appeared related to effects on immunity or senescence, there are limitations to this analysis. First, cis-regulatory elements can act over a large genomic scale (e.g. several Mb’s); for example, 9p21.3 variants have been suggested to influence expression both of the nearby tumor suppressor proteins of the *INK4a/ARF* locus, as well as *IFNa-21*, a more distant (~1 Mb) regulator of inflammation (Liu *et al.* 2009; Harismendy *et al.* 2011). Likewise, another gene in the 5p15.33 bin, *CLPTMIL*, has also been postulated to contribute to cancer progression (McKay *et al.* 2008). Therefore, the true causal variant located near GWAS-identified SNPs may influence expression of one or more local transcripts, some or all of which may not be

located in the same bin. Moreover, an ascertainment bias exists in that certain well-demarcated disease states (e.g. autoimmune diseases) appear more tractable to GWAS than less clinically distinct entities (e.g. community acquired pneumonia). Therefore, not all morbid conditions of aging are tractable to GWAS.

Importantly, the prevalence and morbidity of each disease was not weighted in this study. For example, scleroderma (rare) and myocardial infarction (common) were each counted as a single, unique disease per genomic bin, despite differing greatly in their total contribution to human morbidity. Additionally, this analysis does not account for SNP prevalence or scale of their effect. Future work could incorporate these factors to estimate the multi-disease population attributable risk associated with certain SNP genotypes. As a result of these limitations, this analysis may overestimate the importance of the MHC locus, which is strongly associated with several rare diseases. By contrast it may underestimate the relevance of the 9p21.3 bin, which is associated with common, highly morbid diseases (Table 1).

Although the association of senescence regulators such as *TERT* and *p16<sup>INK4a</sup>* with cancer and the MHC locus with autoimmunity is not surprising, the finding that all identified hotspots of recurrent disease association map to bins linked to either inflammation/immunity or cellular senescence is striking. The diversity of age-related diseases associated with the 9p21.3 bin is particularly remarkable (Figure 2). Of the four principal causes of age-related morbidity (neoplasia, metabolic disease, atherosclerosis and neurodegeneration), three are recurrently associated with polymorphisms mapping near the *INK4/ARF* locus by GWAS (Figure 2). Recently, even the outlier, neurodegenerative disease, has been linked to this locus based on a genome-wide pedigree study of late onset Alzheimer's disease (Zuchner *et al.* 2008). While it remains unclear how modulating senescence may contribute to some diseases in the 9p21.3 bin, this finding is consistent with several recent murine studies showing an effect of modulating *p16<sup>INK4a</sup>* expression *in vivo* on many non-malignant, age-associated phenotypes including T2DM, atherosclerosis, T-cell function, cataracts, and sarcopenia (Krishnamurthy *et al.* 2006; Baker *et al.* 2011; Chen *et al.* 2011; Kuo *et al.* 2011; Liu *et al.* 2011).

It is also worth noting what was not associated with broad disease susceptibility: conserved pathways that modulate longevity in model organisms (e.g. Insulin/IGF-1 signaling, mTOR signaling, reactive oxygen species signaling, Sirtuins, etc). It is possible that regulation of these pathways is not variable among human populations or that these pathways do not modulate general disease resistance in humans, but we think more likely this observation reflects a lack of power of the GWAS meta-analysis approach. Accordingly, SNPs near *IGF1R*, *FOXO3A*, and *AKT1* have been associated with longevity in candidate studies and pedigree analyses (Suh *et al.* 2008; Pawlikowska *et al.* 2009; Sebastiani *et al.* 2012), suggesting an association of these loci with age-associated conditions may emerge in genome-wide analyses with further study. Nonetheless, this unbiased meta-analysis of results from ~2.3 million patients only identifies polymorphic regulation of cellular senescence and immunity as general determinants of genetic susceptibility to a host of human diseases, with in particular a striking association of senescence with *age-associated* disease. These genetic data support the therapeutic targeting of these specific pathways to promote broad disease resistance and augment the human health span.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



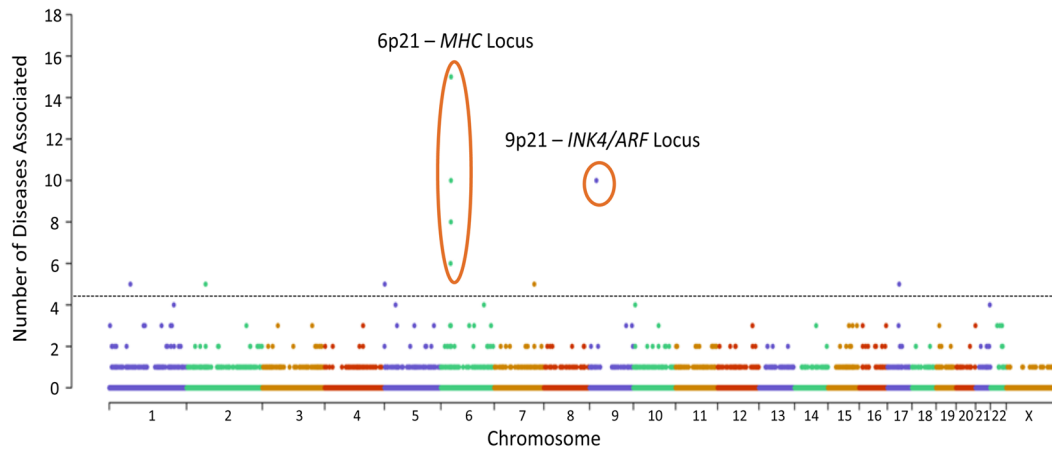
## Acknowledgments

This work was supported by grants from the Burroughs Wellcome Foundation and the NIA.

## References

- Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008; 322:881–888. [PubMed: 18988837]
- Armanios M. Telomerase and idiopathic pulmonary fibrosis. *Mutation research*. 2011
- Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, Lawson WE, Xie M, Vulto I, Phillips JA 3rd, Lansdorf PM, Greider CW, Loyd JE. Telomerase mutations in families with idiopathic pulmonary fibrosis. *The New England journal of medicine*. 2007; 356:1317–1326. [PubMed: 17392301]
- Baker DJ, Wijshake T, Tchkonian T, Lebrasseur NK, Childs BG, van de Sluis B, Kirkland JL, van Deursen JM. Clearance of p16(Ink4a)-positive senescent cells delays ageing-associated disorders. *Nature*. 2011
- Chen H, Gu X, Liu Y, Wang J, Wirt SE, Bottino R, Schorle H, Sage J, Kim SK. PDGF signalling controls age-dependent proliferation in pancreatic beta-cells. *Nature*. 2011; 478:349–355. [PubMed: 21993628]
- Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet*. 2008; 4:e1000024. [PubMed: 18437207]
- Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature*. 2011; 470:264–268. [PubMed: 21307941]
- Hills M, Lansdorf PM. Short telomeres resulting from heritable mutations in the telomerase reverse transcriptase gene predispose for a variety of malignancies. *Annals of the New York Academy of Sciences*. 2009; 1176:178–190. [PubMed: 19796246]
- Hindorf LA, Junkins HA, Hall PN, Mehta JP, Manolio TA. A Catalog of Published Genome-Wide Association Studies. 2011.
- Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, O'Brien SJ. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC genomics*. 2010; 11:724. [PubMed: 21176216]
- Jorgensen TJ, Ruczinski I, Kessing B, Smith MW, Shugart YY, Alberg AJ. Hypothesis-driven candidate gene association studies: practical design and analytical considerations. *American journal of epidemiology*. 2009; 170:986–993. [PubMed: 19762372]
- Krishnamurthy J, Ramsey MR, Ligon KL, Torrice C, Koh A, Bonner-Weir S, Sharpless NE. p16INK4a induces an age-dependent decline in islet regenerative potential. *Nature*. 2006; 443:453–457. [PubMed: 16957737]
- Kuo CL, Murphy AJ, Sayers S, Li R, Yvan-Charvet L, Davis JZ, Krishnamurthy J, Liu Y, Puig O, Sharpless NE, Tall AR, Welch CL. Cdkn2a Is an Atherosclerosis Modifier Locus That Regulates Monocyte/Macrophage Proliferation. *Arteriosclerosis, thrombosis, and vascular biology*. 2011
- Liu Y, Johnson SM, Fedoriw Y, Rogers AB, Yuan H, Krishnamurthy J, Sharpless NE. Expression of p16(INK4a) prevents cancer and promotes aging in lymphocytes. *Blood*. 2011; 117:3257–3267. [PubMed: 21245485]
- Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Mohlke KL, Ibrahim JG, Thomas NE, Sharpless NE. INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS One*. 2009; 4:e5027. [PubMed: 19343170]
- Martinez P, Blasco MA. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nature reviews. Cancer*. 2011; 11:161–176.
- McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njolstad I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, Lubinski J,

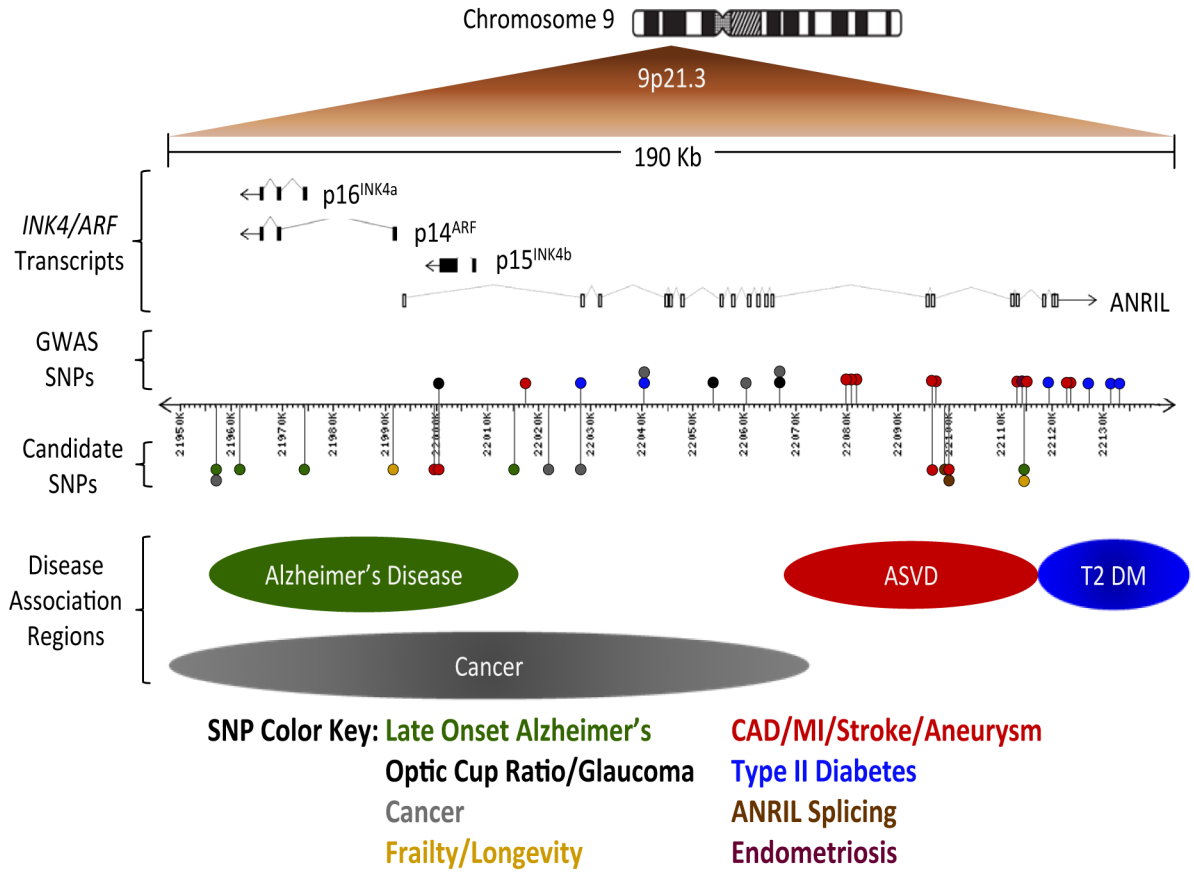
- Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T, Xinarianos G, Cassidy A, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, Gonzalez CA, Ramon Quiros J, Martinez C, Navarro C, Ardanaz E, Larranaga N, Kham KT, Key T, Bueno-de-Mesquita HB, Peeters PH, Trichopoulou A, Linseisen J, Boeing H, Hallmans G, Overvad K, Tjonneland A, Kumle M, Riboli E, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P. Lung cancer susceptibility locus at 5p15.33. *Nat Genet.* 2008; 40:1404–1406. [PubMed: 18978790]
- Pawlikowska L, Hu D, Huntsman S, Sung A, Chu C, Chen J, Joyner AH, Schork NJ, Hsueh WC, Reiner AP, Psaty BM, Atzmon G, Barzilai N, Cummings SR, Browner WS, Kwok PY, Ziv E. Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. *Aging Cell.* 2009; 8:460–472. [PubMed: 19489743]
- Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, Fernando MM, Green T, De Jager PL, Foisy S, Wang J, de Bakker PI, Leslie S, McVean G, Padyukov L, Alfredsson L, Annese V, Hafler DA, Pan-Hammarstrom Q, Matell R, Sawcer SJ, Compston AD, Cree BA, Mirel DB, Daly MJ, Behrens TW, Klareskog L, Gregersen PK, Oksenberg JR, Hauser SL. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc Natl Acad Sci U S A.* 2009; 106:18680–18685. [PubMed: 19846760]
- Sebastiani P, Solovieff N, Sun JX. Naive Bayesian Classifier and Genetic Risk Score for Genetic Risk Prediction of a Categorical Trait: Not so Different after all! *Frontiers in genetics.* 2012; 3:26. [PubMed: 22393331]
- Sharpless NE, DePinho RA. How stem cells age and why this makes us grow old. *Nature reviews. Molecular cell biology.* 2007; 8:703–713.
- Suh Y, Atzmon G, Cho MO, Hwang D, Liu B, Leahy DJ, Barzilai N, Cohen P. Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc Natl Acad Sci U S A.* 2008; 105:3438–3442. [PubMed: 18316725]
- Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, Rosenblatt RL, Shay JW, Garcia CK. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A.* 2007; 104:7552–7557. [PubMed: 17460043]
- Willeit P, Willeit J, Mayr A, Weger S, Oberhollenzer F, Brandstatter A, Kronenberg F, Kiechl S. Telomere length and risk of incident cancer and cancer mortality. *JAMA: the journal of the American Medical Association.* 2010; 304:69–75. [PubMed: 20606151]
- Zuchner S, Gilbert JR, Martin ER, Leon-Guerrero CR, Xu PT, Browning C, Bronson PG, Whitehead P, Schmechel DE, Haines JL, Pericak-Vance MA. Linkage and association study of late-onset Alzheimer disease families linked to 9p21.3. *Annals of human genetics.* 2008; 72:725–731. [PubMed: 18761660]



**Figure 1. The genetic landscape of human disease**

Manhattan plot depicting the number of unique human diseases per bin linked to disease susceptibility SNPs identified by GWAS. Each point represents a 200kb bin ordered by chromosomal location. The dotted line represents the cutoff for statistical significance as determined by a 10,000 iteration permutation test ( $p < 0.05$ ). The two highest peaks of disease association ( $p < 0.0001$ ) are circled.





**Figure 2. The *INK4/ARF* Locus is a Genomic Hotspot for Age-Associated Disease Susceptibility**  
 The primary transcripts within the human *INK4/ARF* locus including the long non-coding RNA, ANRIL, are depicted relative to chromosome 9. Disease susceptibility SNPs identified by GWAS that mapped to the *INK4/ARF* locus are represented by color coded circles above the genomic ruler. Susceptibility SNPs displayed below the ruler were identified by non-GWAS association studies such as genome-wide pedigree studies. Each color represents a specific human disease which is described in the SNP color key. Clusters of susceptibility SNPs associated to a specific disease are marked by large color coded ovals. Abbreviations: ASVD = Atherosclerotic vascular diseases and T2DM = Type 2 Diabetes Mellitus.

**Table 1**

Chromosomal loci of significantly enriched ('hotspot') bins.

Chromosome Region	# Unique Diseases/Bin	Candidate Gene(s)	Associated Disease Susceptibilities
1p31.3	5	IL23R	Immune: IBD <sup>A</sup> (x2) <sup>B</sup> , Behcet's disease, Psoriasis, Ankylosing spondylitis
2p16.1	5	REL	Immune: IBD, RA <sup>C</sup> , Psoriasis, Celiac Disease, Hodgkin's Lymphoma
5p15.33	5	TERT	Senescence: Cancers (x5), Idiopathic pulmonary fibrosis
6p21	26 across four bins	MHC, NOTCH4	Immune: Arthritis (x4), IBD (x2), Cancer (x5), Lupus, MS <sup>D</sup> , Scleroderma, Celiac disease, T1DM <sup>E</sup> , Asthma, Primary biliary cirrhosis, Psoriasis
7q32.1	5	IRF5, TNPO3	Immune: SLE <sup>F</sup> , IBD, RA, Primary biliary cirrhosis, Scleroderma
9p21.3	10	p15 <sup>INK4b</sup> , p16 <sup>INK4a</sup> , p14 <sup>ARF</sup> , ANRIL	Senescence: MI <sup>G</sup> , stroke, T2DM <sup>H</sup> , Glaucoma, Aortic aneurysm, Intracranial aneurysm, Cancers (x3), Endometriosis
17q12	5	IKZF3, GSDMA, GSDMB	Immune: IBD (x2), Asthma, RA, T1DM

<sup>A</sup> Inflammatory Bowel Syndrome.

<sup>B</sup> Numbers in parenthesis represent unique number of specific diseases within that disease family.

<sup>C</sup> Rheumatoid Arthritis.

<sup>D</sup> Multiple sclerosis,

<sup>E</sup> Type 1 Diabetes mellitus.

<sup>F</sup> Systemic Lupus Erythematosus.

<sup>G</sup> Myocardial Infarction,

<sup>H</sup> Type 2 Diabetes mellitus.