Published in final edited form as:

Ann Hum Genet. 2009 September ; 73(Pt 5): 514–519. doi:10.1111/j.1469-1809.2009.00533.x.

Genomic convergence of genome-wide investigations for complex traits

G. Kitsios^{1,2} and E. Zintzaras^{1,2}

¹Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Tufts University School of Medicine, 800 Washington Street, Tufts MC #63, Boston, MA 02111, USA

²Department of Biomathematics, University of Thessaly School of Medicine, Larissa, Greece

SUMMARY

Genome-wide investigations for identifying the genes for complex traits are considered to be agnostic in terms of prior assumptions for the responsible DNA alterations. The agreement of genome-wide association studies (GWAS) and genome-wide linkage scans (GWLS) has not been explored to date. In this study, a genomic convergence approach of GWAS and GWLS was implemented for the first time in order to identify genomic loci supported by both methods. A database with 376 GWLS and 102 GWAS for 19 complex traits was created. Data regarding the location and statistical significance for each genetic marker were extracted from articles or web-based databases. Convergence was quantified as the proportion of significant GWAS markers located within linked regions. Convergence was variable (0–73.3%) and was found to be significantly higher than expected by chance only for two of the 19 phenotypes. Seventy five loci of interest were identified, which being supported by independent lines of evidence, could merit prioritization in future investigations. Although convergence is supportive of genuine effects, lack of agreement between GWLS and GWAS is also indicative that these studies are designed to answer different questions and are not equally well suited for deciphering the genetics of complex traits.

Keywords

genomic convergence; linkage; association; complex traits

Correspondence: Prof. Elias Zintzaras, Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, 800 Washington Street, Tufts MC #63, Boston, MA 02111, tel: 617-636-1692, fax: 617-636-8628 ezintzaras@tuftsmedicalcenter.org.

Web Resources The URLs for data presented herein are as follows:

A Catalog of Published Genome-Wide Association Studies, <http://www.genome.gov/26525384>

dbSNP, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp>

HapMap, <http://www.hapmap.org/>

HuGE Navigator database, <http://hugenavigator.net>

T1Dbase, <http://T1DBase.org>

Alzgene, www.alzgene.org

Human Obesity Gene Map, <http://obesitygene.pbrb.edu>

Slep database, <https://slep.unc.edu/evidence>

Marshfield genetic map, <http://research.marshfieldclinic.org/genetics/GeneticResearch/compMaps.asp>

UniSTS database, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=unists>

INTRODUCTION

Despite extensive efforts, the genetic basis of common, multifactorial disorders remains poorly understood. Few genetic regions and genes can reliably be considered as true positives. The majority of findings from the two basic tools for deciphering the genetics of complex traits (genome-wide linkage scans (GWLS) and candidate gene studies) have shown inconsistency and non-reproducibility [Altmuller et al., 2001; Lohmuller et al., 2003, Ioannidis 2007; Morgan et al., 2007, Lasky-Su et al., 2008]. The advent of the first wave of genome-wide association studies (GWAS) provided substantial advances in uncovering DNA variations influencing the risk of common diseases and clinical phenotypes. More than 100 loci for 40 traits have been identified [Pearson & Manolio, 2008]. Although concerns relevant to the analysis, interpretation and translation of GWAS results have been expressed [Hardy & Singleton 2009; Shriner et al. 2007; McCarthy et al., 2008], publication trends for GWAS across several fields are rapidly increasing [Yu et al., 2008].

The multifactorial etiology of common disorders involving complex epistatic and gene-environment interactions reduces the likelihood that a single method could provide definitive answers. Alternatively, an integration approach combining different kinds of genetic data analysis ('genomic convergence') has been proposed for the prioritization of promising results from the bulk of exponentially accumulating data [Hauser et al., 2003]. Evidence for the implication of a genetic locus can be considered as more valid when supported by independent research methodologies, such as linkage, association or expression profiling studies.

Since the majority of genetic associations claimed by hypothesis-driven candidate-gene studies are eventually refuted [Lohmuller et al., 2003, Morgan et al., 2007], the 'hypothesis-free' methods of genetic analysis represent an area of major interest. The GWAS and GWLS can be considered as the only agnostic approaches, free from underlying assumptions regarding the location, number or functionality of causal genetic variants [Altmuller et al., 2001; Pearson & Manolio, 2008; Hunter et al., 2008]. Although GWAS have revealed many loci and variants of unprecedented biological implication [Pearson & Manolio, 2008; Hunter et al., 2008], no previous study has examined the agreement of these association findings with previous linkage analysis results. By comparing the independent findings stemming from 'hypothesis-free' methods, GWLS and GWAS can be considered as complementary strategies and not as concurrent [Bourgain et al., 2007]. In this study, we combined for the first time the strengths of GWLS and GWAS to identify genetic loci for complex traits supported by both type of studies, using a systematic literature search and data mining from web-based genetic databases.

MATERIALS AND METHODS

Selection of Studies

Applying a genomic convergence approach, data from GWLS and GWAS for complex traits were compared and examined for agreement. Eligible studies were retrieved by a systematic search of the PubMed database from its inception through August 2008, for all available GWLS and GWAS in humans. Combinations of search terms as 'genome-wide' OR 'linkage' OR 'association' OR 'gene' OR 'polymorphism' were used. Eligible phenotypes included disorders and traits for which multifactorial etiology has been proposed. To ensure accumulation of sufficient analyzable information and examine the most extensively studied phenotypes, only phenotypes with at least three published GWAS and at least three published GWLS were considered. Finally, a database including 376 GWLS and 102 GWAS for 19 phenotypes was created, available at our website (<http://biomath.med.uth.gr>).

Data Extraction

From each study, the following information was extracted: publication details (author, year of study, journal); phenotypic traits; details of study population (ethnic background, sample-size); genotyping methods (type and number of markers, type of genotyping platform); statistical methods; results obtained (marker accession numbers, genetic and physical location, genetic annotation, statistical significance).

Within each GWAS, the statistically significant Single Nucleotide Polymorphisms (SNPs) were recorded. In order to maximize the number of potentially promising findings from each study, a critical p-value at the 10^{-6} level was initially used [Hindorff et al., 2008], instead of the commonly used but more stringent 0.5×10^{-8} level [McCarthy et al., 2008]. For each SNP, the physical location and genetic annotation were identified through the dbSNP database [Sherry et al., 2001]. For SNPs exhibiting strong linkage disequilibrium ($r^2 > 0.8$) on the basis of HapMap data B35 [The International HapMap Project. 2003] within a single study, only the most significant representative marker of the locus was considered. The extracted GWAS data were cross-validated with the available GWAS data in HapMap B36, the Catalog of Published Genome-Wide Association Studies [Hindorff et al., 2008] and the HuGE Navigator database [Yu et al., 2008].

From the concordant to phenotypes available GWLS, the markers reaching genome-wide significant or suggestive linkage score (i.e. LOD score > 2) in their respective main analyses were recorded [Sawcer et al., 1997]. For each marker, the respective linkage confidence intervals were determined either by the 1.0-LOD-unit-down method when LOD score curves and markers were available, or by using the ± 15 centiMorgan interval around the peak marker in the absence of the above information, as previously described [Chen et al., 2007; Zintzaras et al., 2007]. These data were supplemented by linkage intervals identified in GWLS meta-analyses, and for five phenotypes (Diabetes Mellitus Type I, Obesity, Bipolar Disorder, Schizophrenia, Alzheimer's Disease) by their respective online linkage databases [Hulbert et al., 2007; Bertram et al., 2007; Rankinen et al., 2006; Konneker et al., 2008]. The extent of genomic coverage by linkage intervals for each phenotype was calculated by adding the non-overlapping intervals over the Marshfield genetic map [Broman et al., 1998] (Table 1). Comparisons of GWLS and GWAS results were mediated by translations of genetic and physical map distances using the MapOMat software [Kong et al., 2004], the Marshfield genetic map [Broman et al., 1998] and the Uni-STS database.

Data Analysis

Genomic convergence for each phenotype was defined as the proportion of statistically significant SNPs that fell within the areas of linkage intervals. Convergence was estimated when at least one significant SNP in at least one GWAS was detected for the specific phenotype. To examine whether the observed convergence exceeded or fell behind the convergence expected by chance, we used a two-sided, one-sample z-test of proportions at the $\alpha=0.05$ level of significance. The convergence expected by chance for each phenotype was set to be equal to the proportion of genomic coverage by linkage intervals. In example, for a phenotype where 40% of the genotype is covered by linkage intervals, each SNP has a 40% probability to fall within a linkage interval and produce a convergent result. Exact binomial test probabilities were calculated to test for agreement with the z-test. Correlation of the convergence proportions with the proportions of linkage coverage was also examined.

Subgroup analyses by a) study sample size and b) genomic location of SNPs were performed. In the first analysis, we explored convergence only for GWLS and GWAS with large sample sizes. Large sample sizes were defined as those exceeding the median sample

size of the group of studies investigating the same phenotype. In the second analysis, convergence was examined for intergenic and genic SNPs separately (Table 1).

In order to place the convergent findings in our database on exact chromosomal locations and define unique convergent loci, the linkage intervals harboring GWAS SNPs and showing extensive overlapping between them (>80%) were merged. Additionally, from SNPs across different GWAS that showed strong linkage disequilibrium or were located in physical proximity (less than $\pm 100\text{kb}$ apart), only the most significant 'hit' was recorded. The physical distance between the most highly associated SNP and the linkage peak marker in the interval was also recorded. In the final step, more stringent criteria for genome-wide linkage (LOD score > 3.2) and association ($p\text{-value} < 0.5 \times 10^{-8}$) were applied in this dataset of unique convergent findings, to identify signals with the most compelling evidence of convergence (Supplementary Table 1).

RESULTS

The number of linkage intervals, the proportion of genomic coverage by linkage intervals and the number of significant GWAS SNPs for each phenotype are shown in Table 1. Overall, 1221 linkage intervals were constructed and 387 (269 genic /118 intergenic) SNPs were identified. Convergence estimates ranged from 0% in the case of amyotrophic lateral sclerosis and colorectal cancer to 73.3 % for BMI-obesity (median 30.8, interquartile range (IQR) 12.0). The genomic convergence proportions correlated with the proportions of linkage genomic coverage (Spearman correlation coefficient $r=0.67$). However, convergence was significantly higher than expected by chance for only two phenotypes (LDL cholesterol concentration and restless leg syndrome).

In the subgroup analysis for studies with large sample sizes, the convergence ranged from 0% to 75.0% (median 22.2, IQR 28.6). Convergence was now statistically significant for six out of the 19 phenotypes (Alzheimer's disease, BMI-obesity, Crohn's disease, LDL cholesterol concentration, restless leg syndrome and rheumatoid arthritis).

When intergenic SNPs were considered separately, the observed convergence did not exceed the expected by chance for any of the phenotypes (median 10.0, IQR 33.3). On the contrary, the convergence for rheumatoid arthritis was significantly less than expected (convergence=0%, $p<0.05$). For SNPs located within genes, the convergence estimates ranged from 0–100% (median 35.0, IQR 35.7) and it was significantly higher than expected for four phenotypes (Alzheimer's disease, diabetes mellitus type I, LDL cholesterol concentration and rheumatoid arthritis).

A detailed description of convergent SNPs and their harbouring linkage intervals is provided in Supplementary Table 1. Seventy five unique convergent loci were identified. A graphical depiction of convergent findings is provided on a chromosomal map, available at our website (<http://biomath.med.uth.gr>). The stringent criteria for genome-wide linkage and association were met for 16 out of these 75 unique loci (21.3%). When considering all 389 SNPs initially identified from GWAS, 35 of them (9.0%) provided strong evidence of convergence (Supplementary Table 1).

DISCUSSION

Convergent findings can be considered as landmarks of loci and variants meriting priority in replication studies, in an effort to limit the plethora of false positive results in genetic epidemiology [NCI-NHGRI Working Group on Replication in Association Studies et al., 2007; Zintzaras & Lau., 2008]. By comparing the results from two agnostic methodologies

for unraveling the genetics of complex traits, we identified a subset of genomic loci with concomitant evidence of implication.

In our main analysis, genomic convergence was variable and was found to be significantly higher than the convergence expected by chance for only two of the 19 phenotypes. In the subgroup analysis for studies with large sample sizes, the number of phenotypes with statistically significant convergence increased to six. This pattern of results implied that studies with small sample sizes may have caused inflation in the inconsistency between GWLS and GWAS.

Consistently replicated GWAS findings that uncovered previously unsuspected pathophysiology, such as the 8q24 locus for prostate cancer or the 9p21.3 locus for coronary artery disease, have shown convergence in our database (Supplementary Table 1). Such signals located in “gene deserts” (intergenic regions), would never have been considered in a candidate gene study. The fact that intergenic signals had been previously implicated by linkage analysis and remained largely unnoticed until the emergence of GWAS is of particular importance. However, our subgroup analysis for convergence of intergenic SNPs did not identify any phenotype with significantly increased convergence. In contrast, convergence of genic SNPs showed significant results for four phenotypes.

The use of prior linkage information in the analysis of GWAS data has been proposed as a strategy to increase statistical power [Roeder et al., 2006]. By applying a false-discovery rate that involves weighting the hypotheses on the basis of prior, informative linkage data, the power of the GWAS is expected to improve considerably. On the other hand, methodologies for enriching linkage results on the basis of the commonality of functional annotation have been proposed, in order to identify candidate genes expected to participate in a common pathway or process [Shriner et al., 2008]. Additionally, gene-prioritization bioinformatic tools have been developed, which exploit linkage and other types of prior information (such as gene expression profiles or protein-protein interactions) [Hutz et al., 2008]. These promising methodologies could help prioritize for further research those genes and SNPs that provide evidence of implication across different venues of research.

Nevertheless, our findings warrant cautious interpretation in the light of inherent weaknesses of genomic epidemiology investigations. Our study was based on the assumption that variants detectable by linkage designs are the same as those detectable by association, and vice versa. However, susceptibility loci in the multiplex families used in GWLS might not overlap with susceptibility loci in population-based samples, as the ones used in GWAS [Roeder et al. 2006].

The imprecision and limited power to detect multiple loci with modest contributions are well-recognized limitations of GWLS [Altmuller et al., 2001]. The current design of GWAS makes them more suitable in the context of the common disease-common variant hypothesis [Hunter et al., 2008; McCarthy et al., 2008]. Consequently, significant SNPs located outside linkage regions can still mark true positive associations, since the corresponding GWLS possibly lacked the power to uncover linkage in the surrounding region. Some of the most convincing associations that have emerged in the GWAS era, like the *IL23R* gene for Crohn's disease or the *KCNJ11* gene with type 2 Diabetes, were not supported by linkage data. Furthermore, the imprecision of GWLS results in outputs of extended linked regions containing hundreds of genes. Inability to narrow down the linkage interval might have resulted in some spurious convergent findings in our database.

Apart from the limitation of GWLS to detect low penetrance genes, GWAS do not capture the effect of SNPs with minor allele frequencies of less than 5%. However, according to an alternative hypothesis (common disease-rare variant hypothesis), complex traits are caused

collectively by multiple rare variants with moderate to high penetrance [Iyengar & Elston, 2007]. These rare variants, with frequencies lying somewhere between the limits of deleterious mutations and polymorphic variations (i.e. 0.1–1%), would be detectable only after extensive resequencing of carefully chosen candidate genes in selected cases [Bodmer & Bonilla, 2008].

In conclusion, genomic convergence techniques do not provide direct validation of findings, nor represent a substitute for high-quality replication studies. Interpreted with caution, evidence from convergence could provide an additional piece of information (together with strong statistical evidence or biological plausibility) [McCarthy et al., 2008; NCI-NHGRI Working Group on Replication in Association Studies et al., 2007] in determining the design of replication studies. The convergent findings that emerged from two exploratory techniques methodologies for complex traits are supportive of genuine effects that could merit prioritization in future studies. Lack of convergence is not indicative of false positive findings or methodological flaws; genome-wide linkage and association studies are probably designed to answer different questions and are not equally well suited for deciphering the genetics of complex diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Georgios Kitsios is Pfizer-Tufts Medical Center Research Fellow in Clinical Research.

Scientific support for this project was provided through the Tufts Clinical and Translational Science Institute (Tufts CTSI) under funding from the National Institute of Health/National Center for Research Resources (UL1 RR025752). Points of view or opinions in this paper are those of the authors and do not necessarily represent the official position or policies of the Tufts CTSI.

REFERENCES

- Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am. J. Hum. Genet.* 2001; 69:936–950. [PubMed: 11565063]
- Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat. Genet.* 2007; 39:17–23. [PubMed: 17192785]
- Bourgain C, Génin E, Cox N, Clerget-Darpoux F. Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases? *Eur. J. Hum. Genet.* 2007; 15:260–263. [PubMed: 17164797]
- Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* 2008; 40:695–701. [PubMed: 18509313]
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* 1998; 63:861–869. [PubMed: 9718341]
- Chen Y, Rollins J, Paigen B, Wang X. Genetic and genomic insights into the molecular basis of atherosclerosis. *Cell. Metab.* 2007; 6:164–179. [PubMed: 17767904]
- Hauser MA, Li YJ, Takeuchi S, Walters R, Noureddine M, Maready M, Darden T, Hulette C, Martin E, Hauser E, Xu H, Schmechel D, Stenger JE, Dietrich F, Vance J. Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. *Hum. Mol. Genet.* 2003; 12:671–677. [PubMed: 12620972]
- Hardy J, Singleton A. Genomewide association studies and human disease. *N. Engl. J. Med.* 2009; 360:1759–1768. [PubMed: 19369657]

- Hindorff, LA.; Jenkins, HA.; Mehta, JP.; Manolio, TA. A Catalog of Published Genome-Wide Association Studies. 2008 [Accessed at 11-15-2008]. Available at: www.genome.gov/26525384.
- Hulbert EM, Smink LJ, Adlem EC, Allen JE, Burdick DB, Burren OS, Cassen VM, Cavnar CC, Dolman GE, Flamez D, Friery KF, Healy BC, Killcoyne SA, Kutlu B, Schuilenburg H, Walker NM, Mychaleckyj J, Eizirik DL, Wicker LS, Todd JA, Goodman N. T1DBase: integration and presentation of complex data for type 1 diabetes research. *Nucleic. Acids. Res.* 2007; 35:D742–746. [PubMed: 17169983]
- Hunter DJ, Altshuler D, Rader DJ. From Darwin's finches to canaries in the coal mine-mining the genome for new biology. *N. Engl. J. Med.* 2008; 358:2760–2763. [PubMed: 18579810]
- Hutz JE, Kraja AT, McLeod HL, Province MA. CANDID: a flexible method for prioritizing candidate genes for complex human traits. *Genet. Epidemiol.* 2008; 32:779–790. [PubMed: 18613097]
- Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum. Hered.* 2007; 64:203–213. [PubMed: 17551261]
- Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or “common gene, common disease”? *Methods. Mol. Biol.* 2007; 376:71–84. [PubMed: 17984539]
- Konneker T, Barnes T, Furberg H, Losh M, Bulik CM, Sullivan PF. A searchable database of genetic evidence for psychiatric disorders. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 2008; 147B: 671–675. [PubMed: 18548508]
- Kong X, Murphy K, Raj T, He C, White PS, Matice TC. A combined linkage-physical map of the human genome. *Am. J. Hum. Genet.* 2004; 75:1143–1148. [PubMed: 15486828]
- Lasky-Su J, Lyon HN, Emilsson V, Heid IM, Molony C, Raby BA, Lazarus R, Klanderman B, Soto-Quiros ME, Avila L, Silverman EK, Thorleifsson G, Thorsteinsdottir U, Kronenberg F, Vollmert C, Illig T, Fox CS, Levy D, Laird N, Ding X, McQueen MB, Butler J, Ardlie K, Papoutsakis C, Dedoussis G, O'Donnell CJ, Wichmann HE, Celedón JC, Schadt E, Hirschhorn J, Weiss ST, Stefansson K, Lange C. On the replication of genetic associations: timing can be everything! *Am. J. Hum. Genet.* 2008; 82:849–858. [PubMed: 18387595]
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* 2003; 33:177–182. [PubMed: 12524541]
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 2008; 9:356–369. [PubMed: 18398418]
- Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *JAMA.* 2007; 297:1551–1561. [PubMed: 17426274]
- NCI-NHGRI Working Group on Replication in Association Studies. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. *Nature.* 2007; 447:655–660. [PubMed: 17554299]
- Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA.* 2008; 299:1335–1344. [PubMed: 18349094]
- Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, Argyropoulos G, Walts B, Pérusse L, Bouchard C. The human obesity gene map: the 2005 update. *Obesity(Silver.Spring).* 2006; 14:529–644. [PubMed: 16741264]
- Roeder K, Bacanu SA, Wasserman L, Devlin B. Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* 2006; 78:243–252. [PubMed: 16400608]
- Sawcer S, Jones HB, Judge D, Visser F, Compston A, Goodfellow PN, Clayton D. Empirical genomewide significance levels established by whole genome simulations. *Genet. Epidemiol.* 1997; 14:223–229. [PubMed: 9181352]
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic. Acids. Res.* 2001; 29:308–311. [PubMed: 11125122]

- Shriner D, Baye TM, Padilla MA, Zhang S, Vaughan LK, Loraine AE. Commonality of functional annotation: a method for prioritization of candidate genes from genome-wide linkage studies. *Nucleic. Acids. Res.* 2008; 36:e26. [PubMed: 18263617]
- Shriner D, Vaughan LK, Padilla MA, Tiwari HK. Problems with genome-wide association studies. *Science.* 2007; 316:1840–1842. [PubMed: 17600199]
- The International HapMap Project. *Nature.* 2003; 426:789–796. [PubMed: 14685227]
- Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat. Genet.* 2008; 40:124–125. [PubMed: 18227866]
- Zintzaras E, Kitsios G, Kent D, Camp NJ, Atwood L, Hopkins PN, Hunt SC. Genome-wide scans meta-analysis for pulse pressure. *Hypertension.* 2007; 50:557–564. [PubMed: 17635856]
- Zintzaras E, Lau J. Synthesis of genetic association studies for pertinent gene-disease associations requires appropriate methodological and statistical approaches. *J. Clin. Epidemiol.* 2008; 61:634–645. [PubMed: 18538260]

Table 1
Results from genomic convergence approach for the most extensively studied complex traits.

Disorder	GWLS (n)	Linkage Intervals from GWLS (n)	% coverage of the genome ^a	GWAS (n)	Significant SNPs from GWAS (Genic / Intergenic)	Convergence		Convergence for intergenic SNPs ^b		Convergence for genic SNPs ^c		Convergence for studies with large sample sizes ^d	
						%	p-value	%	p-value	%	p-value	%	p-value
Alzheimer's Disease	19	73	30.1	6	6 (5/1)	66.7	0.05	0	0.51	80.0	<0.05	75.0	<0.05
Adult Height	18	65	34.9	5	42 (33/0)	28.6	0.39	0	n/a	36.3	0.86	22.2	0.93
Anyotrophic Lateral Sclerosis	11	12	5.2	5	8 (5/3)	0	0.50	0	0.69	0	0.60	0	0.61
Bipolar Disorder	21	57	24.5	3	8 (4/4)	20.0	0.77	25.0	0.98	25.0	0.98	0	0.35
BMI-Obesity	65	215	52.5	7	15 (11/4)	73.3	0.11	75.0	0.37	72.7	0.17	75.0	<0.05
Bone Mineral Density / Osteoporosis	23	75	41.7	3	16 (12/4)	46.6	0.69	75.0	0.18	33.3	0.56	30.0	0.44
Breast Cancer	5	14	8.1	6	26 (14/12)	3.5	0.39	0	0.30	7.1	0.89	7.1	0.85
Colorectal Cancer	3	13	6.7	5	9 (7/2)	0	0.42	0	0.70	0	0.48	0	0.72
Coronary Artery Disease	12	40	28.6	6	16 (3/13)	25.0	0.75	30.8	0.86	0	0.27	16.7	0.87
Crohn's Disease	14	48	25.3	8	55 (29/26)	30.9	0.34	19.2	0.47	41.4	0.05	34.5	<0.05
Diabetes Mellitus Type 1	14	36	18.8	3	26 (16/10)	30.8	0.12	10.0	0.48	43.8	<0.05	20.0	0.14
Diabetes Mellitus Type 2	24	55	32.3	12	48 (42/6)	27.1	0.44	33.3	0.96	23.8	0.24	12.5	0.74
LDL cholesterol	30	55	30.5	4	26 (25/1)	57.7	<0.05	0	0.51	60.0	<0.05	63.6	<0.05
Nicotine Dependence	20	49	26.5	3	3 (2/1)	33.3	0.79	0	0.55	50.0	0.45	0	0.70
Prostate Cancer	23	67	32.7	8	27 (16/11)	29.6	0.73	36.4	0.79	25.0	0.51	21.4	0.82
Restless Leg Syndrome	6	17	6.2	3	7 (7/0)	28.6	<0.05	0	n/a	14.3	0.37	50.0	<0.05
Rheumatoid Arthritis	10	139	35.6	4	16 (5/11)	31.3	0.72	0	<0.05	100.0	<0.05	30.7	<0.05
Schizophrenia	36	113	35.5	6	6 (4/2)	33.3	0.91	50.0	0.67	25.0	0.67	25.0	0.61
Systemic Lupus Erythematosus	13	78	34.9	5	27 (20/7)	29.6	0.56	28.6	0.73	35.0	0.99	21.1	0.68
Total	367	1221		102	387 (269/118)								

^a % coverage of a 3487 cM genome (Marshfield Map) by linkage intervals.

^b Convergence defined as the ratio: (Number of significant intergenic SNPs in GWAS located in linked regions)/(Total number of significant intergenic SNPs in GWAS)

^c Convergence defined as the ratio: (Number of significant genic SNPs in GWAS located in linked regions)/(Total number of significant genic SNPs in GWAS)

d_c Convergence defined as the ratio: (Number of significant SNPs in GWAS with large sample sizes, located in linked regions identified from GWLS with large sample sizes)/(Total number of significant SNPs in GWAS with large sample sizes).