Literature Lab™ analysis of key associations, and comparisons between existing and novel gene sets linked to Multiple Sclerosis

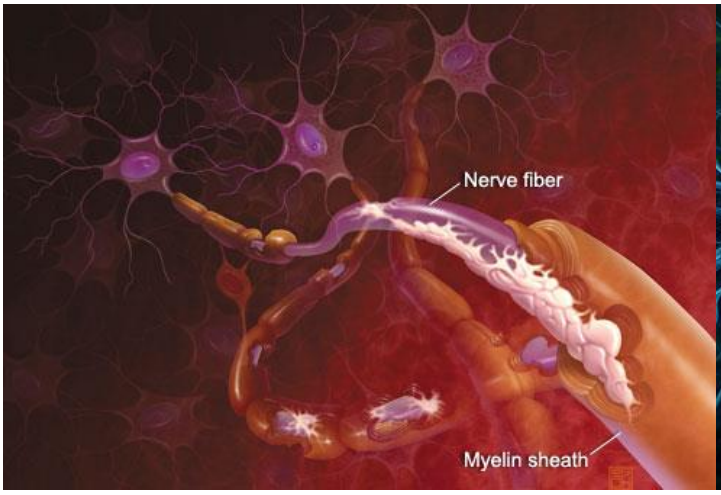
We analyzed three gene sets identified in the study: Known non-MHC variants (49), Novel non-MHC variants (48), and Variants from the 8 regions with consistent high-resolution fine mapping. All subject domains were explored and areas of focus included diseases, tissues, cell types, and pathways. The first objective was to identify the strength of the associations of the three gene sets to terms within the individual focus areas or domains. The second objective was to identify differences between the novel gene set and the existing gene set. Finally, the third objective was to explore the association of key critical genes to pathways and diseases. This work demonstrates the power of Literature Lab™ to interrogate both existing and novel genes linked to MS and uncover important associations and differences between MS gene sets and other diseases within PubMed.

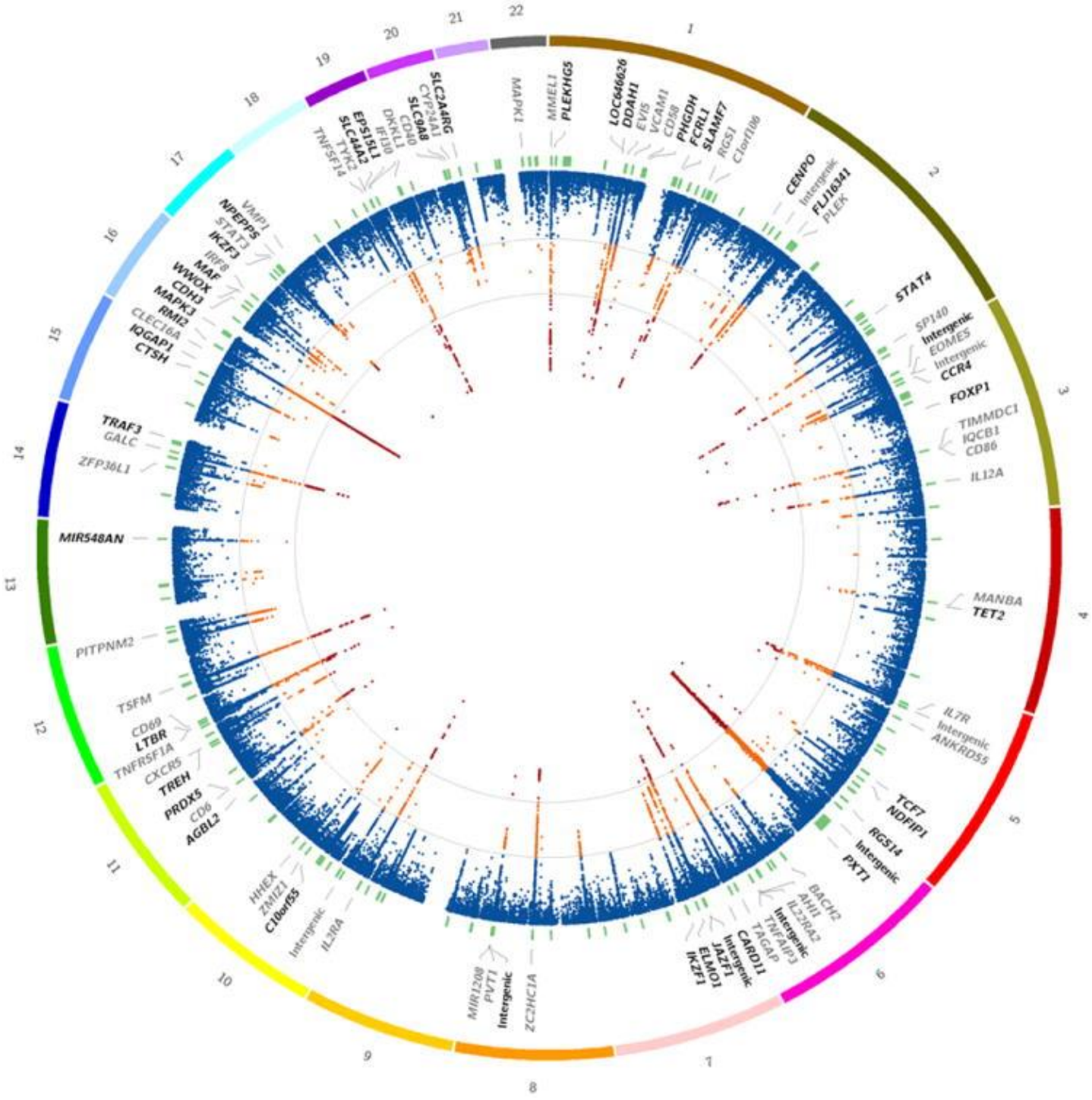
Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis

International Multiple Sclerosis Genetics Consortium (IMSGC)

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

Using the ImmunoChip custom genotyping array, we analysed 14,498 multiple sclerosis subjects and 24,091 healthy controls for 161,311 autosomal variants and identified 135 potentially associated regions (p-value < 1.0×10^{-4}). In a replication phase, we combined these data with previous genome-wide association study (GWAS) data from an independent 14,802 multiple sclerosis subjects and 26,703 healthy controls. In these 80,094 individuals of European ancestry we identified 48 new susceptibility variants (p-value < 5.0×10^{-8}); three found after conditioning on previously identified variants. Thus, there are now 110 established multiple sclerosis risk variants in 103 discrete loci outside of the Major Histocompatibility Complex. With high resolution Bayesian fine-mapping, we identified five regions where one variant accounted for more than 50% of the posterior probability of association. This study enhances the catalogue of multiple sclerosis risk variants and illustrates the value of fine-mapping in the resolution of GWAS signals.



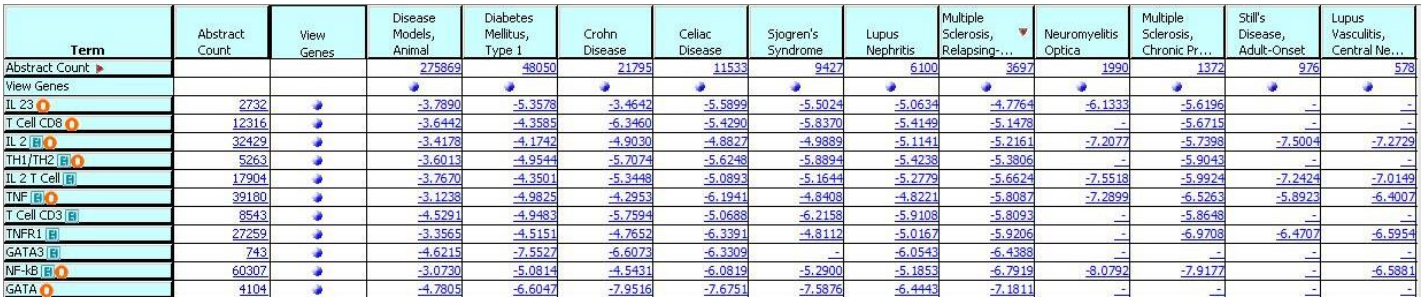


Circos plot showing primary association analysis in the discovery phase. The outer most track shows the numbered autosomal chromosomes. **The second track indicates the gene closest to the most associated SNP meeting all replication criteria. Previous identified associations are indicated in grey while novel genes are emboldened.** The third track indicates the physical position of the 184 fine mapping intervals. The inner most track indicates –log (P) which truncates the signal in several regions.

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

Literature Lab™ works by first searching the PubMed database for published abstracts that combine a gene from the user inputted gene set with a term, which can be associated with a user selected domain such as a disease, pathway, tissue, cell type, etc. It calculates a numeric score that measures the strength of the association between a target gene set and each term. This score is the Log of Product Frequency and the closer the value to zero the stronger the association. Literature Lab™ qualifies these scores against 1000 random gene sets to establish significance or P-value. It lists the relative contributions of individual genes to the product frequency and it visually illuminates the connections through clustering analysis and heat maps.

The Literature Lab™ basic platform allows the search for co-occurrences by comparing two domains such as diseases versus pathways.



Term	Abstract Count	View Genes	Disease Models, Animal	Diabetes Mellitus, Type 1	Crohn Disease	Celiac Disease	Sjogren's Syndrome	Lupus Nephritis	Multiple Sclerosis, Relapsing-...	Neuromyelitis Optica	Multiple Sclerosis, Chronic Pr...	Still's Disease, Adult-Onset	Lupus Vasculitis, Central Ne...
Abstract Count			275869	48050	21795	11533	9427	6100	3697	1990	1372	976	578
View Genes													
IL 23	2732		-3.7890	-5.3578	-3.4642	-5.5899	-5.5024	-5.0634	-4.7764	-6.1333	-5.6196	-	-
T Cell CD8	12316		-3.6442	-4.3585	-6.3460	-5.4290	-5.8370	-5.4149	-5.1478	-	-5.6715	-	-
IL 2	32429		-3.4178	-4.1742	-4.9030	-4.8827	-4.9889	-5.1141	-5.2161	-7.2077	-5.7398	-7.5004	-7.2729
TH1/TH2	5263		-3.6013	-4.9544	-5.7074	-5.6248	-5.8894	-5.4238	-5.3806	-	-5.9043	-	-
IL 2 T Cell	17904		-3.7670	-4.3501	-5.3448	-5.0893	-5.1644	-5.2779	-5.6624	-7.5518	-5.9924	-7.2424	-7.0149
TNF	39180		-3.1238	-4.9825	-4.2953	-6.1941	-4.8408	-4.8221	-5.8087	-7.2899	-6.5263	-5.8923	-6.4007
T Cell CD3	8543		-4.5291	-4.9483	-5.7594	-5.0688	-6.2158	-5.9108	-5.8093	-	-5.8648	-	-
TNFR1	27259		-3.3565	-4.5151	-4.7652	-6.3391	-4.8112	-5.0167	-5.9206	-	-6.9708	-6.4707	-6.5954
GATA3	743		-4.6215	-7.5527	-6.6073	-6.3309	-	-6.0543	-6.4388	-	-	-	-
NF-kB	60307		-3.0730	-5.0814	-4.5431	-6.0819	-5.2900	-5.1853	-6.7919	-8.0792	-7.9177	-	-6.5881
GATA	4104		-4.7805	-6.6047	-7.9516	-7.6751	-7.5876	-6.4443	-7.1811	-	-	-	-

The abstracts involved in each pathway and disease are hyperlinked. Also hyperlinked are the genes in the literature associated with each pathway or disease. The Literature Lab™ basic application also permits co-occurrences such as the search for comorbidities by comparing disease versus disease, with the option of highlighting the context of a gene set.

Literature Lab™ vs Literature Lab™ PLUS

- The Literature Lab™ platform (above) allows the user to search for co-occurrences between two domains, for example diseases and pathways, with or without the context of a gene list.

- The LPFs are calculated and links are provided to the literature.

- The Literature Lab™ PLUS platform provides a significantly more rigorous analysis with significance using the 1000 random gene sets; LPF, P-value, individual gene contributions, clustering analysis, and many other features.

- While Literature Lab™ can point out interesting co-occurrences, PLUS can drill down to provide statistically significant associations and unexpected results

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

The Literature Lab™ PLUS platform

Gene Set - Known non-MHC variants

Domain – Diseases

Positive Associations – Multiple Sclerosis, MS relapsing-remitting, etc.

Moderate Associations – Celiac Disease, etc.

Diseases-MeSH (Subset)	Count	Celiac Disease	Malabsorption Syndromes	Myasthenia Gravis, Autoimmune...	Multiple Sclerosis	Multiple Sclerosis, Relapsing-...	Demyelinating Autoimmune Diseases, CNS	Whipple Disease	Autoimmune Diseases of the Nervou...	Nervous System Autoimmune...	Encephalom... Autoimmune, Experimental	Myasthenia Gravis	Neuritis, Autoimmune, Experimental
View All Genes For Term													
Association		Moderate	Positive	Positive	Positive	Positive							
Gene X Term Abstracts		123	130	28	1218	119	1209	2	1398	598	548	128	21
Term Abstracts		11533	19341	227	47435	3697	40162	916	60169	7188	6528	8049	386
Nonzero Genes		19	21	5	35	12	33	1	33	20	20	15	6
LPF		-3.39	-3.75	-3.99	-3.01	-3.94	-3.00	-5.50	-3.08	-3.05	-3.09	-4.23	-4.49
Random Sets	Count	1000	1000	760	1000	995	1000	528	1000	1000	1000	999	811
Experiment Set	Term rank	36	67	56	65	65	87	58.5	97	104	97	97	134
	Score	2.18	1.53	1.47	1.41	1.31	1.36	1.29	1.27	1.21	1.20	1.10	0.91
	P-Value	0.0146	0.0633	0.0710	0.0796	0.0944	0.0868	0.0994	0.1022	0.1123	0.1152	0.1352	0.1817
	Score rank	26	92	103	112	138	123	145	153	175	182	233	354
Symbol/OrigID	Max PF	66.80%	73.99%	66.76%	24.75%	58.03%	29.55%	100%	34.21%	45.26%	43.87%	68.11%	35.49%
IL2RA	16605	4.95%	5.13%	66.76%	20.16%	58.03%	29.55%	-	34.21%	45.26%	43.87%	68.11%	23.60%
MAPK1	14372	-	0.02%	-	0.05%	-	0.01%	-	0.02%	0.04%	0.05%	0.24%	0.56%
STAT3	13876	0.10%	0.13%	-	1.80%	1.09%	2.10%	-	1.82%	6.13%	7.48%	0.06%	-
CD40	12385	0.08%	0.08%	8.74%	4.85%	4.86%	4.79%	-	5.04%	9.06%	8.84%	3.86%	10.33%
VCAM1	9579	0.04%	1.61%	-	4.91%	3.54%	5.49%	-	5.73%	7.06%	7.61%	0.35%	13.36%
CD86	8113	0.59%	0.81%	19.22%	4.82%	7.42%	8.80%	-	9.93%	22.74%	21.54%	6.70%	35.49%
TNFRSF1A	4322	0.00%	0.11%	1.00%	4.45%	0.49%	3.61%	-	3.69%	2.62%	2.32%	0.44%	16.66%
CD69	4049	0.19%	0.26%	4.28%	0.82%	0.93%	1.31%	-	1.40%	1.69%	1.69%	2.57%	-
PLEK	2346	-	0.01%	-	0.02%	-	0.00%	-	0.00%	-	-	0.09%	-
IL7R	2302	0.04%	0.11%	-	12.30%	20.03%	11.08%	-	9.65%	1.52%	1.85%	0.83%	-
CD58	1366	-	0.09%	-	1.16%	1.55%	1.05%	100%	1.61%	-	-	7.61%	-
TYK2	1249	-	-	-	1.27%	0.19%	0.88%	-	0.71%	1.24%	1.52%	-	-
MANBA	804	-	-	-	0.00%	-	0.00%	-	0.00%	-	-	-	-
CYP24A1	777	-	-	-	0.47%	1.21%	0.26%	-	0.26%	0.08%	0.10%	-	-
TNFAIP3	645	1.62%	0.41%	-	0.27%	-	0.14%	-	0.38%	0.10%	0.12%	5.26%	-
GALC	577	-	-	-	1.66%	-	0.73%	-	1.01%	0.43%	0.53%	-	-

Domains & Association Counts

	Strong	Moderate	Positive
Pathways	28	32	38
Diseases-MeSH	0	18	53
PathologicalCond-MeSH	0	2	4
Psychology-MeSH	0	2	9
Physiology-MeSH	2	9	9
CellPhysiology-MeSH	9	6	4
Metabolism-MeSH	4	1	0
ChemicalsDrugs-MeSH	60	117	208
ChemicalActions-MeSH	0	8	11
Anatomy-MeSH	13	27	25
TissueTypes-MeSH	1	8	2
CellTypes-MeSH	11	20	12
CellStructures-MeSH	0	0	2
Biogenetics-MeSH	4	5	11
OtherBiology-MeSH	27	36	24
Organisms-MeSH	3	28	32

Moderate Assocations for Diseases-MeSH

Term	Score	P-Value
Hyper-IgM Immunodeficiency Syndrome	2.66	0.0040
Celiac Disease	2.18	0.0146
Hypergammaglobulinemia	2.17	0.0150
Graft vs Host Disease	2.04	0.0206
Lymphocytic Choriomeningitis	1.99	0.0234
Kidney Diseases, Cystic	1.98	0.0241
HTLV-I Infections	1.96	0.0251
Deltaretrovirus Infections	1.95	0.0259
Dermatitis, Contact	1.94	0.0260
Leukemia, Hairy Cell	1.93	0.0267
Dysgammaglobulinemia	1.90	0.0286
Arenaviridae Infections	1.87	0.0310
Agammaglobulinemia	1.83	0.0338
Listeriosis	1.81	0.0353
Liver Cirrhosis, Biliary	1.77	0.0382
Lymphopenia	1.76	0.0389
Meningitis, Viral	1.75	0.0397
Mastocytosis	1.66	0.0481

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

Gene Set - Novel non-MHC variants

Domain – Diseases

Strong Associations – Lymphoma, blood neoplasms, Rheumatoid Arthritis, etc.

Moderate Associations – Leukemia, blood diseases, lymphoma, etc.

Diseases-MeSH (Subset)	Count	Lymphoma, B-Cell, Marginal Zone	Hematologic Neoplasms	Lymphoma, Large B-Cell, Diffuse	Lymphoma, B-Cell	Arthritis, Rheumatoid	Leukemia-L... Adult T-Cell	Precursor Cell Lymphoblas...	Multiple Myeloma	Lymphoma, Non-Hodgkin	Myeloproli... Disorders	Paraprotei...	Lymphoma, T-Cell, Cutaneous
View All Genes For Term													
Association		Strong	Strong	Strong	Strong	Strong	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Gene X Term Abstracts		152	69	144	403	371	78	208	255	585	241	206	53
Term Abstracts		3728	10169	8912	29735	55125	3332	20878	26664	61252	27181	26830	5323
Nonzero Genes		9	12	14	21	17	9	14	18	21	12	19	7
LPF		-2.02	-3.52	-2.79	-2.56	-2.86	-3.21	-2.90	-3.31	-2.79	-3.06	-3.47	-3.69
Random Sets	Count	888	999	999	1000	1000	989	1000	1000	1000	1000	1000	960
Experiment Set	Term rank	6	10	17	21	16	30	45	33	33	26	41	41
	Score	3.16	2.86	2.70	2.55	2.54	2.22	2.16	2.13	2.08	2.06	1.97	1.92
	P-Value	0.0008	0.0021	0.0035	0.0054	0.0055	0.0131	0.0155	0.0165	0.0188	0.0196	0.0243	0.0274
	Score rank	1	7	16	22	23	32	33	34	38	40	45	47
Symbol/OrigID	Max PF	92.81%	94.68%	69.28%	60.95%	87.32%	92.98%	98.33%	43.28%	56.01%	93.94%	43.64%	94.35%
MAPK3	28176	0.00%	0.09%	0.04%	0.08%	0.24%	0.04%	0.11%	3.63%	0.19%	1.31%	3.87%	0.03%
PLEKHG5	3538	0.00%	0.15%	0.00%	0.00%	0.00%	-	-	0.00%	0.00%	0.00%	0.00%	-
TCF7	1765	-	0.07%	0.02%	0.04%	0.00%	1.34%	0.31%	0.02%	0.21%	0.01%	0.01%	0.05%
CCR4(GID:1233)/**CC	1568	0.01%	0.02%	0.63%	0.25%	0.61%	92.98%	0.02%	0.04%	4.41%	-	0.01%	94.35%
STAT4	1180	-	0.03%	-	0.01%	9.15%	-	0.00%	0.01%	0.12%	0.00%	0.01%	3.84%
CDH3(GID:1001)/**CDH3	875	-	-	-	0.00%	-	-	-	-	0.00%	0.00%	-	-
IKZF1	847	-	3.10%	0.20%	0.37%	0.01%	3.64%	98.33%	2.02%	1.21%	4.22%	0.82%	-
CXCR5	720	0.02%	0.05%	0.35%	1.73%	1.97%	-	0.02%	0.26%	3.21%	-	0.14%	1.16%
LTBR	585	-	0.06%	-	0.03%	0.44%	-	-	0.01%	0.08%	-	0.02%	-
TET2	548	-	94.68%	0.01%	0.01%	-	-	0.01%	-	0.26%	93.94%	-	-
TRAF3	510	0.01%	-	0.01%	0.47%	0.01%	-	-	5.38%	0.79%	0.03%	3.62%	-
BCL10	506	92.81%	-	10.70%	60.95%	0.00%	-	0.12%	0.96%	56.01%	0.08%	1.06%	-
MAF	475	-	-	0.01%	0.02%	-	0.10%	0.01%	43.28%	0.08%	0.22%	42.47%	-
CTSH	471	-	-	-	-	0.14%	-	-	-	-	-	-	-
DDAH1	458	-	-	-	-	0.01%	-	-	-	-	-	-	-
IQGAP1	416	-	-	-	-	-	0.12%	-	0.07%	-	-	0.03%	-

Domains & Association Counts

	Strong	Moderate	Positive
Pathways	8	27	52
Diseases-MeSH	7	18	47
PathologicalCond-MeSH	0	0	7
Psychology-MeSH	1	6	14
Physiology-MeSH	2	7	8
CellPhysiology-MeSH	4	7	4
Metabolism-MeSH	0	2	3
ChemicalsDrugs-MeSH	27	103	208
ChemicalActions-MeSH	2	7	14
Anatomy-MeSH	9	9	17
TissueTypes-MeSH	3	4	3
CellTypes-MeSH	2	12	14
CellStructures-MeSH	0	1	5
Biogenetics-MeSH	0	8	9
OtherBiology-MeSH	8	22	22
Organisms-MeSH	0	1	21

Strong Associations for Diseases-MeSH

Term	Score	P-Value
Lymphoma, B-Cell, Marginal Zone	3.16	0.0008
Hematologic Neoplasms	2.86	0.0021
Acute Lung Injury	2.81	0.0025
Endometrial Neoplasms	2.79	0.0026
Lymphoma, Large B-Cell, Diffuse	2.70	0.0035
Lymphoma, B-Cell	2.55	0.0054
Arthritis, Rheumatoid	2.54	0.0055

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

Gene Set – Variants from the 8 regions with consistent high resolution fine mapping

Domain - Diseases

Moderate Associations – Nervous system autoimmune, demyelinating autoimmune, Multiple Sclerosis, MS relapsing-remitting, Diabetes Mellitus - Type 1, Rheumatoid Arthritis, Lupus, Herpes Simplex infections, etc.

Strong Associations – Deltaretrovirus infections, HTLV-1 infections, Leukemia T-cell, Leukemia-Lymphoma adult T-cell.

Diseases-MeSH (Subset)	Count	Deltaretro... Infections	HTLV-I Infections	Lymphopenia	Leukemia-L... Adult T-Cell	Leukemia, T-Cell	Herpes Simplex	Paraparesis, Tropical Spastic	Paraparesis	Lymphoma, B-Cell, Marginal Zone	Sezary Syndrome	Nervous System Autoimmune...	Vaccinia
View All Genes For Term													
Association		Strong	Strong	Strong	Strong	Strong	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Gene X Term Abstracts		113	96	185	131	226	103	29	58	31	52	291	80
Term Abstracts		5288	4141	5912	3332	5961	28420	1403	5050	3728	1290	7188	9220
Nonzero Genes		4	3	6	4	5	6	3	3	3	5	7	6
LPF		-3.89	-3.90	-3.52	-3.56	-3.35	-3.73	-4.46	-4.41	-3.86	-3.94	-3.24	-3.99
Random Sets	Count	869	846	952	932	979	999	598	847	717	682	986	989
Experiment Set	Term rank	5	7	19	25	24	26	15	26	37	29	38	39
	Score	2.60	2.60	2.11	2.10	2.04	2.27	2.09	2.09	1.96	1.89	1.86	1.85
	P-Value	0.0047	0.0047	0.0174	0.0179	0.0207	0.0115	0.0183	0.0183	0.0249	0.0295	0.0311	0.0321
	Score rank	6	7	19	20	27	12	21	22	29	32	34	36
Symbol/OrigID	Max PF	96.33%	96.42%	97.82%	97.27%	95.08%	96.63%	98.00%	96.51%	98.26%	75.82%	70.15%	89.76%
IL2RA	16605	96.33%	96.42%	97.82%	97.27%	95.08%	2.50%	98.00%	96.51%	1.70%	75.82%	70.15%	5.75%
TNFRSF1A	4322	0.55%	0.04%	0.83%	0.23%	0.43%	0.35%	0.48%	0.12%	0.05%	0.16%	4.06%	3.53%
CD58	1366	2.71%	3.54%	0.04%	1.27%	2.75%	0.05%	1.52%	3.37%	-	-	-	89.76%
STAT4	1180	-	-	0.43%	-	0.03%	0.25%	-	-	-	20.77%	23.58%	0.36%
TNFAIP3	645	-	-	-	-	-	-	-	-	98.26%	1.06%	0.15%	-
IL12A	353	-	-	-	-	-	0.21%	-	-	-	-	1.09%	0.30%
CD6	352	0.42%	-	0.16%	1.23%	1.71%	-	-	-	-	-	0.27%	0.30%
TNFSF14	311	-	-	0.72%	-	-	96.63%	-	-	-	2.19%	0.70%	-

Diseases-MeSH (Subset)	Count	Encephalom... Autoimmune, Experimental	Lymphocytosis	Myasthenia Gravis	Lupus Erythemato... Systemic	Multiple Sclerosis, Relapsing-...	Demyelinating Autoimmune Diseases, CNS	Arthritis, Rheumatoid	Sjogren's Syndrome	Autoimmune Diseases of the Nervou...	Lymphoma, T-Cell, Cutaneous	Neuromuscular Junction Diseases	Myelitis
View All Genes For Term													
Association		Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Gene X Term Abstracts		262	85	88	350	72	601	541	102	716	97	74	38
Term Abstracts		6528	3351	8049	30259	3697	40162	55125	9427	60169	5323	8549	9656
Nonzero Genes		7	4	6	7	5	8	8	7	8	6	5	4
LPF		-3.28	-3.93	-4.32	-3.46	-4.15	-3.39	-3.43	-4.07	-3.43	-4.06	-4.48	-5.11
Random Sets	Count	985	760	902	999	819	998	1000	972	999	869	864	849
Experiment Set	Term rank	38	49	27	28	30	36	49	31	38	45	33	45
	Score	1.85	1.84	1.83	1.80	1.80	1.79	1.78	1.78	1.73	1.70	1.70	1.69
	P-Value	0.0321	0.0332	0.0333	0.0358	0.0359	0.0370	0.0373	0.0375	0.0421	0.0444	0.0450	0.0460
	Score rank	37	40	41	45	47	50	51	52	60	64	65	66
Symbol/OrigID	Max PF	67.05%	98.29%	82.85%	50.58%	94.31%	72.89%	34.01%	46.80%	75.27%	86.45%	85.03%	92.50%
TNFAIP3	645	0.18%	-	6.40%	19.21%	-	0.34%	28.89%	12.43%	0.83%	0.33%	4.96%	-
CD6	352	0.33%	0.72%	-	-	-	3.43%	0.50%	8.89%	2.47%	-	-	-
TNFSF14	311	0.84%	-	-	0.03%	-	0.50%	1.00%	-	0.36%	2.75%	-	-
IL12A	353	1.31%	-	0.73%	0.98%	1.08%	2.11%	0.12%	1.42%	1.81%	-	-	3.76%
TNFRSF1A	4322	3.55%	0.24%	0.54%	1.07%	0.80%	8.91%	11.67%	3.51%	8.12%	0.20%	0.33%	2.77%
STAT4	1180	26.75%	-	0.22%	50.58%	1.30%	9.22%	34.01%	46.80%	7.61%	8.87%	0.30%	-
IL2RA	16605	67.05%	98.29%	82.85%	28.13%	94.31%	72.89%	22.40%	23.65%	75.27%	86.45%	85.03%	92.50%
CD58	1366	-	0.75%	9.26%	0.01%	2.52%	2.60%	1.42%	3.30%	3.53%	1.41%	9.38%	0.97%

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

Literate Lab™ allows the user to compare experimental sets and identify critical differences among domains and individual gene contributions. The magnitude of these differences are based on the calculated strength of the associations of the gene sets to domains. The user can also toggle to similarities between gene sets.

Comparison of Novel Non-MHC genes to Existing Non-MHC genes

Domain – Pathways

Strong Association Differences – Beta-catenin, E-cadherin Signaling, PITX2

Moderate Association Differences – Adherins Junction, WNT, etc.

Pathways	Count	Beta-Catenin	E-cadherin Signaling	PITX2	Adherens Junction	N-cadherin Signaling	WNT	Methylation	Endometrial Cancer	AHR	D4-GDI	GATA	Classical Complement
View All Genes For Term													
Association		Strong 1^	Strong 1^	Strong 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Positive 1^	Positive 1^	Positive 1^
Gene X Term Abstracts		903/344	396/234	286/125	79/37	117/62	672/280	362/385	109/79	224/240	21/15	190/168	27/34
Term Abstracts		16211/16211	7560/7560	5050/5050	2659/2659	2012/2012	17668/17668	31325/31325	4177/4177	11348/11348	883/883	4101/4101	3589/3589
Nonzero Genes		18/22	11/17	9/16	8/7	7/7	20/21	26/25	9/10	15/18	6/4	15/20	4/13
LPF		-1.94/-3.60	-2.51/-3.79	-2.48/-3.90	-3.21/-4.79	-2.93/-4.21	-2.12/-3.56	-3.10/-3.91	-3.63/-4.35	-3.27/-4.11	-2.79/-4.66	-2.78/-3.82	-3.31/-5.05
Random Sets	Count	1000/1000	1000/1000	1000/1000	1000/1000	999/1000	1000/1000	1000/1000	1000/1000	1000/1000	971/999	1000/1000	991/1000
Experiment Set	Term rank	43/609	38/555	34/594	71/806	35/499	103/560	183/805	71/572	143/694	32/589	55/418	131/516
	Score	2.12/-0.45	2.25/-0.30	2.01/-0.39	1.48/-0.86	2.16/-0.17	1.56/-0.39	0.89/-0.89	1.47/-0.29	1.05/-0.56	2.03/-0.26	1.88/-0.04	1.44/-0.25
	P-Value	0.0172/0.3279	0.0123/0.3832	0.0224/0.3481	0.0690/0.1947	0.0153/0.4330	0.0592/0.3480	0.1874/0.1861	0.0711/0.3854	0.1476/0.2873	0.0212/0.3979	0.0301/0.4857	0.0750/0.3997
	Score rank	38/518	32/486	44/503	101/601	35/456	86/504	221/605	107/483	184/551	41/477	53/419	115/476
Symbol/OrigID	Max PF	92.68/59.50%	73.95/66.18%	92.03/35.77%	76.78/46.96%	90.34/67.74%	98.62/29.88%	82.87/46.44%	67.97/56.31%	88.21/33.35%	99.14/79.02%	41.99/27.87%	99.22/49.80%
MAPK3	28176	0.92%	5.20%	1.55%	2.19%	6.41%	0.39%	1.46%	26.47%	3.85%	0.12%	0.83%	0.05%
IL2RA	16605	0.09%	0.18%	0.09%	-	-	0.06%	4.20%	0.29%	15.03%	-	12.57%	4.74%
MAPK1	14372	7.39%	12.56%	8.63%	46.96%	18.33%	1.63%	3.32%	56.31%	9.63%	17.56%	2.87%	0.88%
STAT3	13876	59.50%	66.18%	35.77%	8.25%	67.74%	26.29%	46.44%	30.06%	33.35%	-	27.87%	0.91%
CD40	12385	0.05%	3.19%	0.05%	-	-	0.06%	1.41%	0.04%	2.64%	-	1.05%	49.80%
VCAM1	9579	0.43%	2.18%	0.41%	11.95%	4.16%	0.26%	0.10%	-	5.71%	2.15%	7.42%	1.31%
CD86	8113	0.08%	5.33%	0.18%	-	-	0.02%	0.46%	-	6.74%	-	2.40%	13.96%
TNFRSF1A	4322	0.56%	0.17%	0.15%	0.54%	0.19%	1.06%	0.05%	0.49%	7.56%	-	0.60%	0.73%
CD69	4049	0.01%	0.18%	-	-	-	-	0.16%	-	5.47%	1.27%	1.43%	0.78%
PLEKHG5	3538	-	-	-	-	-	-	0.00%	-	0.00%	-	-	-
PLEK	2346	0.51%	1.25%	0.27%	15.93%	-	0.14%	0.40%	0.91%	0.43%	79.02%	0.07%	-
IL7R	2302	0.52%	0.14%	2.48%	-	0.35%	0.22%	2.52%	-	0.44%	-	7.00%	1.37%
TCF7	1765	92.68%	3.88%	92.03%	0.55%	0.22%	98.62%	1.11%	1.46%	88.21%	-	4.75%	-
CCR4(GID:1233)/**CC	1568	-	0.00%	-	-	-	0.00%	0.13%	-	0.52%	-	0.46%	0.04%

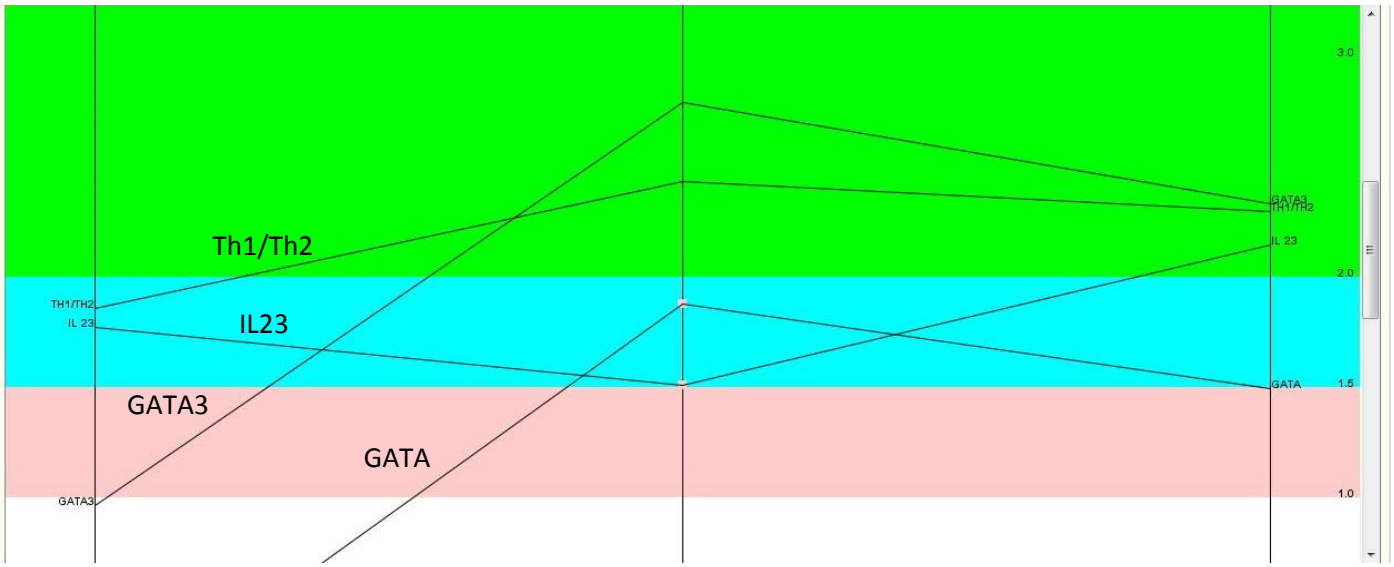
Domain – Diseases

Strong Association Differences – Neoplasms; Moderate Association Differences - Carcinomas

Diseases-MeSH (Subset)	Count	Endometrial Neoplasms	Uterine Neoplasms	Neoplasms, Complex and Mixed	Adenomatous Polyposis Coli	Neoplasms, Ductal, Lobular, a...	Genital Neoplasms, Female	Stomach Neoplasms	Carcinoma, Ductal, Breast	Neoplasms, Cystic, Mucinous, ...	Eye Neoplasms	Neoplasms, Basal Cell	Carcinoma, Ductal
View All Genes For Term													
Association		Strong 1^	Strong 1^	Strong 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^	Moderate 1^
Gene X Term Abstracts		122/78	221/260	70/87	104/26	128/145	477/654	284/355	52/57	34/40	61/45	30/42	93/119
Term Abstracts		15644/15644	66843/66843	20582/20582	6998/6998	25656/25656	119160/119160	47630/47630	12261/12261	15618/15618	19179/19179	10324/10324	17043/17043
Nonzero Genes		11/15	18/21	12/16	8/10	14/14	19/27	16/22	9/13	9/9	9/12	6/12	11/14
LPF		-3.21/-5.02	-3.78/-4.79	-3.60/-5.13	-3.31/-5.53	-4.14/-4.72	-3.86/-4.26	-3.93/-4.35	-4.21/-4.77	-4.98/-5.63	-4.57/-5.25	-4.49/-5.31	-4.24/-4.66
Random Sets	Count	1000/1000	1000/1000	1000/1000	999/1000	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000	999/1000	999/1000	1000/1000
Experiment Set	Term rank	15/826	29/860	37/831	56/923	143/780	109/732	130/683	97/547	390/962	163/676	119/602	163/706
	Score	2.79/-0.89	2.06/-1.00	1.93/-0.89	1.83/-1.08	1.03/-0.74	1.07/-0.67	1.15/-0.57	1.44/-0.25	0.11/-1.51	1.05/-0.52	1.18/-0.39	0.93/-0.62
	P-Value	0.0026/0.1880	0.0198/0.1591	0.0267/0.1870	0.0334/0.1391	0.1516/0.2288	0.1427/0.2508	0.1242/0.2828	0.0746/0.4003	0.4553/0.0656	0.1475/0.3007	0.1195/0.3495	0.1755/0.2674
	Score rank	10/2115	41/2215	46/2121	53/2308	227/1995	210/1926	173/1827	99/1430	987/2585	221/1763	164/1601	275/1870
Symbol/OrigID	Max PF	97.48/57.05%	89.84/33.70%	97.24/23.77%	85.10/41.94%	84.80/29.22%	60.45/34.29%	41.73/56.04%	94.86/68.69%	56.82/21.31%	65.14/51.56%	87.71/23.00%	84.19/38.68%
MAPK3	28176	1.60%	5.58%	0.25%	0.07%	9.70%	23.75%	15.37%	1.70%	3.68%	6.65%	0.86%	10.07%
IL2RA	16605	1.45%	10.33%	14.41%	-	8.33%	15.87%	13.61%	3.50%	19.88%	0.89%	2.96%	7.19%
MAPK1	14372	57.05%	32.54%	6.64%	1.34%	20.55%	25.35%	13.53%	6.56%	15.38%	18.58%	1.23%	16.95%
STAT3	13876	19.29%	33.70%	9.36%	41.94%	28.53%	34.29%	56.04%	5.86%	15.93%	0.60%	20.38%	21.25%
CD40	12385	0.22%	3.00%	6.47%	-	4.77%	4.77%	3.00%	0.97%	0.22%	2.69%	3.96%	5.60%
VCAM1	9579	1.75%	1.17%	0.28%	2.01%	1.37%	1.26%	0.84%	1.25%	2.56%	1.54%	3.28%	1.02%
CD86	8113	0.08%	3.30%	5.23%	2.37%	0.23%	3.67%	2.58%	-	-	1.82%	0.24%	0.03%
TNFRSF1A	4322	0.62%	2.15%	3.83%	4.45%	1.18%	2.99%	0.40%	0.45%	15.78%	-	0.45%	1.57%
CD69	4049	0.66%	1.12%	16.37%	-	0.20%	1.84%	0.29%	0.12%	-	0.23%	-	0.27%
PLEKHG5	3538	0.01%	0.31%	-	-	0.02%	0.39%	0.25%	-	-	0.06%	-	-
PLEK	2346	1.14%	0.36%	0.28%	18.45%	-	0.32%	0.32%	-	-	-	20.92%	-
IL7R	2302	-	0.04%	4.61%	-	0.36%	0.06%	0.52%	0.21%	1.19%	0.40%	-	0.47%
TCF7	1765	0.21%	0.33%	0.27%	85.10%	0.03%	0.50%	1.47%	0.08%	-	0.44%	-	0.06%
CCR4(GID:1233)/**CC	1568	-	0.01%	0.01%	-	0.14%	0.02%	0.29%	-	0.39%	-	-	0.06%

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

Literature Lab™ allows the user to track the strength of associations across different gene sets using the gene set comparison feature. The 1) existing non-MHC gene set, 2) novel non-MHC gene set, and 3) variants consistent with high resolution mapping. Panel 1: pathways involving IL23, Th1/Th2, GATA3, and GATA all show a strengthening association trend from the existing to the novel gene sets.



Summary

Literature Lab™ interrogates the PubMed database and produces quantitative data based on the strength of associations and the statistical significance between gene sets and term domains. The existing non-MHC, the novel non-MHC, and the variants identified through high-resolution mapping were analyzed and pathway and disease associations were explored.

- While the existing gene set was associated with MS and demyelinating diseases, it also showed moderate association with Celiac Disease, another autoimmune disorder. Autoimmune diseases are thought to involve overlapping genes and pathways which manifest in tissue and disease specific ways.

- In addition to the autoimmune disease Rheumatoid Arthritis, the novel gene set showed positive associations with a range of Lymphomas and various Neoplasms. It is evident that genes in this set are associated with the susceptibility and outgrowth of various cancerous states. This is not entirely surprising as current hypotheses speculate that heritability of certain genetic mutations implicated in cancer are also associated with susceptibility to MS and autoimmune diseases in general. Furthermore, Cancer and MS have been linked as comorbidities.

- The variant fine mapping gene set identified moderate associations with multiple MS disease states and other autoimmune diseases (*e.g.* Lupus, Type-1 Diabetes, and Rheumatoid Arthritis). Interestingly, strong associations with viral infections (*e.g.* Herpes Simplex and Deltaretrovirus) and Leukemias were found, supporting the hypotheses that infections and genetic abnormalities are liked to MS susceptibility and can be involved as co-morbidities.

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.

- The viral associations are especially interesting because again current hypotheses have begun to link viral infections (Herpes Simplex) to susceptibilities to MS and other autoimmune diseases. This is a very active area of recent investigation and important questions include whether viral infection increases the susceptibility to MS and the nature of the genetic underpinnings of this connection. Literature Lab™ was able to pick up these interesting and informative associations even though research in this area is in the early stages.

- Critical differences in associations between the gene sets and domains were identified. In the pathways domain, the novel gene set showed strong associations with Beta-catenin and moderate associations with E-cadherin signaling, Adherin Junction, whereas the existing gene set did not. In the disease domain, the novel gene set showed moderate associations with Neoplasms and Carcinomas in contrast to the existing gene set.

- Pathway association trends were tracked over the three gene sets using the gene set comparison feature. The novel and high resolution mapping gene sets showed moderate and strong associations with the IL23, Th1/Th2, and GATA pathways.

- Recent studies in experimental autoimmune encephalomyelitis, a mouse model of multiple sclerosis, showed that IL-23 was responsible for the inflammation observed, not IL-12 as previously thought. Subsequently, IL-23 was shown to facilitate development of inflammation in numerous other models of immune pathology where IL-12 had previously been implicated including models of arthritis, intestinal inflammation, and psoriasis.

- Proliferating helper T cells that develop into effector T cells differentiate into two major subtypes of cells known as T_h1 and T_h2 cells. T_h1 helper cells are the host immunity effectors against intracellular bacteria and protozoa. They are triggered by IL-12, IL-2 and their effector cytokine is IFN-γ. The key T_h2 transcription factors are STAT6 and GATAs.

- The GATA transcription factor GATA3 is an important regulator of T cell development which has been shown to promote the secretion of IL-4, IL-5, and IL-13 from Th2 cells, and to induce the differentiation of Th0 cells towards this T cell subtype. IL2, TNF, TNFR1, T cell CD3/8 were also strongly associated with the set.

- The pronounced associations of these genes and pathways in the novel fine mapping variant gene set points to their relevance in MS and may enhance their visibility as attractive druggable pathways.

Click here to arrange a live demo of gene set data or to arrange for Literature Lab™ analysis of your data:

http://www.acumenta.com/acumenta/product/Aug_31_Offer.php

Damon Anderson, PhD
Application Scientist
412-901-7785
danderson@acumenta.com
www.acumenta.com

Nat Genet. 2013 Nov; 45(11): 10.1038/ng.2770.