

Supplementary Materials: "A Network-Based Kernel Machine Test for the Identification of Risk Pathways in Genome-Wide Association Studies"

Saskia Freytag^{1,*,‡}, Juliane Manitz^{2,‡}, Martin Schlather³, Thomas Kneib^{2,4}, Christopher I. Amos⁵, Angela Risch^{6,7}, Jenny Chang-Claude⁸, Joachim Heinrich⁹, Heike Bickeböller^{1,4}

1 Institute of Genetic Epidemiology, Medical School, Georg-August University Göttingen, Göttingen, Germany

2 Department of Statistics and Econometrics, Georg-August University Göttingen, Göttingen, Germany

3 Institute for Mathematics, University of Mannheim, Mannheim, Germany

4 Center for Statistics, Georg-August University Göttingen, Göttingen, Germany

5 Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States of America

6 Division of Epigenomics and Cancer Risk Factors, Translational Lung Research Centre Heidelberg, German Cancer Research Center (DKFZ), Heidelberg, Germany

7 Translational Lung Research Center Heidelberg (TLRC-H), Member of the German Center for Lung Research (DZL), Heidelberg, Germany

8 Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

9 Institute of Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany

*** E-mail: Corresponding saskia.freytag@med.uni-goettingen.de**

‡ shared first authorship

Figures

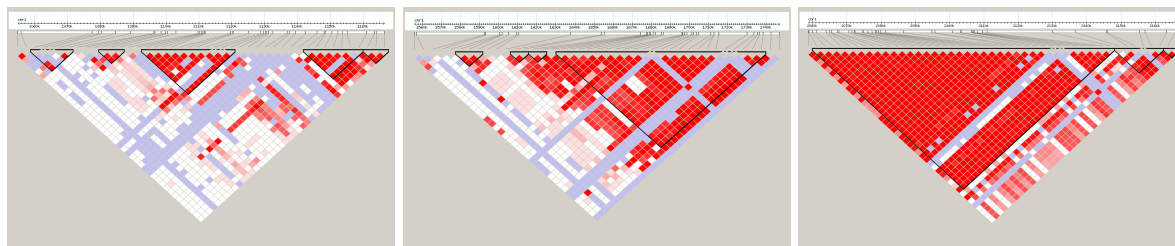


Figure 1. LD plots for the three causal 'proxy genes'. All 'proxy genes' exhibit major LD blocks. Figures were produced using HaploView [36].

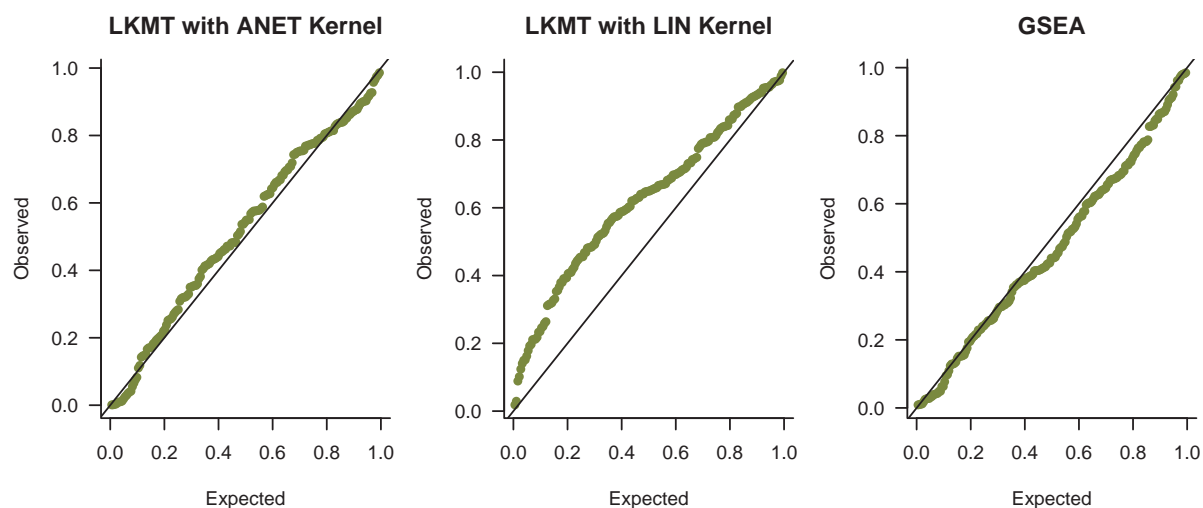


Figure 2. QQ-Plots for p-value distributions obtained from different pathway-based analysis methods for LC GWAS. We compare distributions obtained with LKMT using ANET or LIN kernel as well as distributions obtained by application of GSEA to the uniform distribution. Note, that in our analysis the pathway signals may be correlated, so that we expect deviations from the ideal case of an uniform distribution.

Tables

Table 1. List of top five highly ranked pathways for lung cancer and their respective p-values as identified with signed ANET.

Pathway	KEGG Name (Function)	Type	p-value
hsa00620	Pyruvate metabolism	Metabolism	$1.04 \cdot 10^{-3}$
hsa00240	Pyrimidine metabolism	Metabolism	$1.38 \cdot 10^{-3}$
hsa00250	Alanine, aspartate and glutamate metabolism	Metabolism	$2.68 \cdot 10^{-3}$
hsa00750	Vitamin B6 metabolism	Metabolism	$3.66 \cdot 10^{-3}$
hsa00630	Glyoxylate and dicarboxylate metabolism	Metabolism	$8.76 \cdot 10^{-3}$

Table 2. List of significantly rheumatoid arthritis associated pathways identified by LKMT with signed ANET.

Pathway	KEGG Name (Function)	Type	p-value
hsa04141	Protein processing in endoplasmic reticulum	Genetic Information Processing	$2.33 \cdot 10^{-122}$
hsa04330	Notch signaling pathway	Environmental Information Processing	$9.14 \cdot 10^{-92}$
hsa00140	Steroid hormone biosynthesis	Genetic Information Processing	$5.38 \cdot 10^{-47}$
hsa01100	Metabolic pathways	Metabolism	$1.35 \cdot 10^{-45}$
hsa03018	RNA degradation	Genetic Information Processing	$8.71 \cdot 10^{-28}$
hsa05150	Staphylococcus aureus infection	Human Disease	$4.57 \cdot 10^{-25}$
hsa04612	Antigen processing and presentation	Organismal Systems	$1.46 \cdot 10^{-17}$
hsa04650	Natural killer cell mediated cytotoxicity	Organismal Systems	$1.32 \cdot 10^{-16}$
hsa04060	Cytokine-cytokine receptor interaction	Environmental Information Processing	$1.47 \cdot 10^{-16}$
hsa04610	Complement and coagulation cascades	Organismal Systems	$4.24 \cdot 10^{-16}$
hsa05014	Amyotrophic lateral sclerosis	Human Disease	$7.98 \cdot 10^{-15}$
hsa05160	Hepatitis C	Human Diseases	$8.90 \cdot 10^{-14}$
hsa04210	Apoptosis	Cellular Processes	$1.25 \cdot 10^{-13}$
hsa04010	MAPK signaling pathway	Environmental Information Processing	$1.29 \cdot 10^{-13}$
hsa04920	Adipocytokine signaling pathway	Organismal Systems	$1.58 \cdot 10^{-12}$
hsa05145	Toxoplasmosis	Human Diseases	$4.44 \cdot 10^{-11}$
hsa05142	Chagas disease	Human Diseases	$6.89 \cdot 10^{-11}$
hsa04380	Osteoclast differentiation	Organismal Systems	$1.41 \cdot 10^{-9}$
hsa04620	Toll-like receptor signaling pathway	Organismal Systems	$1.69 \cdot 10^{-9}$
hsa00983	Drug metabolism - other enzymes	Metabolism	$5.46 \cdot 10^{-6}$
hsa04020	Calcium signaling pathway	Environmental Information Processing	$8.02 \cdot 10^{-6}$
hsa03015	mRNA surveillance pathway	Genetic Information Processing	$2.46 \cdot 10^{-5}$
hsa04660	T cell receptor signaling pathway	Organismal Systems	$3.02 \cdot 10^{-5}$
hsa03013	RNA transport	Genetic Information Processing	$3.88 \cdot 10^{-5}$
hsa04622	RIG-I-like receptor signaling pathway	Organismal Systems	$1.06 \cdot 10^{-4}$
hsa05200	Pathways in cancer	Human Diseases	$1.76 \cdot 10^{-4}$

Highlighted pathways do not include genes located in the HLA region or genes previously identified to be associated with RA in peer-reviewed scientific publications (for a list of these see [26]). 88 of the 180 pathways include previously identified genes or genes located in the HLA region.