White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/184460/

Version: Published Version

Article:
Rosenberger, A., Muttray, N., Hung, R.J. et al. (2022) Gene–gene interaction of AhRwith and within the Wntcascade affects susceptibility to lung cancer. European Journal of Medical Research, 27. 14. ISSN: 2047-783X

https://doi.org/10.1186/s40001-022-00638-7

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Rosenberger *et al.*
European Journal of Medical Research (2022) 27:14
https://doi.org/10.1186/s40001-022-00638-7

European Journal of Medical Research

RESEARCH Open Access

Gene–gene interaction of *AhR*with and within the *Wnt*cascade affects susceptibility to lung cancer

Albert Rosenberger[1,35*], Nils Muttray[1], Rayjean J. Hung[2,3], David C. Christiani[4], Neil E. Caporaso[5], Geoffrey Liu[6,7], Stig E. Bojesen[8,9,10], Loic Le Marchand[11], Demetrios Albanes[5], Melinda C. Aldrich[12], Adonina Tardon[13], Guillermo Fernández-Tardón[13], Gad Rennert[14], John K. Field[15], Michael P. A. Davies[15], Triantafillos Liloglou[15], Lambertus A. Kiemeney[16], Philip Lazarus[17], Bernadette Wendel[1], Aage Haugen[18], Shanbeh Zienolddiny[18], Stephen Lam[19], Matthew B. Schabath[20], Angeline S. Andrew[21], Eric J. Duell[22], Susanne M. Arnold[23], Gary E. Goodman[24], Chu Chen[25], Jennifer A. Doherty[26], Fiona Taylor[27], Angela Cox[27], Penella J. Woll[27], Angela Risch[28], Thomas R. Muley[29,30], Mikael Johansson[31], Paul Brennan[32], Maria Teresa Landi[5], Sanjay S. Shete[33], Christopher I. Amos[34], Heike Bickeböller[1] and The INTEGRAL-ILCCO Consortium

Abstract

Background: Aberrant *Wnt* signalling, regulating cell development and stemness, influences the development of many cancer types. The Aryl hydrocarbon receptor (*AhR*) mediates tumorigenesis of environmental pollutants. Complex interaction patterns of genes assigned to *AhR/Wnt*-signalling were recently associated with lung cancer susceptibility.

Aim: To assess the association and predictive ability of *AhR/Wnt*-genes with lung cancer in cases and controls of European descent.

Methods: Odds ratios (OR) were estimated for genomic variants assigned to the Wnt agonist and the antagonistic genes *DKK2*, *DKK3*, *DKK4*, *FRZB*, *SFRP4* and *Axin2*. Logistic regression models with variable selection were trained, validated and tested to predict lung cancer, at which other previously identified SNPs that have been robustly associated with lung cancer risk could also enter the model. Furthermore, decision trees were created to investigate variant × variant interaction. All analyses were performed for overall lung cancer and for subgroups.

Results: No genome-wide significant association of *AhR/Wnt*-genes with overall lung cancer was observed, but within the subgroups of ever smokers (e.g., maker rs2722278 *SFRP4*; OR = 1.20; 95% CI 1.13–1.27; $p = 5.6 \times 10^{-10}$) and never smokers (e.g., maker rs1133683 *Axin2*; OR = 1.27; 95% CI 1.19–1.35; $p = 1.0 \times 10^{-12}$). Although predictability is poor, *AhR/Wnt-variants* are unexpectedly overrepresented in optimized prediction scores for overall lung cancer and for small cell lung cancer. Remarkably, the score for never-smokers contained solely two *AhR/Wnt-variants*. The optimal decision tree for never smokers consists of 7 *AhR/Wnt-variants* and only two lung cancer variants.

*Correspondence: arosenb@gwdg.de
[35] Institut Für Genetische Epidemiologie, Universitätsmedizin Göttingen, Humboldtallee 32, 37073 Göttingen, Germany
Full list of author information is available at the end of the article

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: The role of variants belonging to *Wnt/AhR*-pathways in lung cancer susceptibility may be underrated in main-effects association analysis. Complex interaction patterns in individuals of European descent have moderate predictive capacity for lung cancer or subgroups thereof, especially in never smokers.

Keywords: Susceptibility, Association, Gene–gene integration, Prediction, Polygenic risk score, Decision trees, Never smoker, Small cell lung cancer

Background

Lung cancer (LC) is the most common cancer worldwide since 1985. It is the leading cause of cancer related death around the world [1]. It was estimated for 2020, that globally 2.2 million new LC-cases were diagnosed, which are 11.4% of all new cancer cases. In the same year 1.8 million LC-cases died, which are 18% of all cancer related deaths [2]. The lifetime risk of developing a clinical manifest lung cancer (from birth to age 74) is higher in men (3.78%) than in women (1.77%).

The *Wnt* signalling pathway is a multi-regulator of, e.g., cell proliferation, differentiation, genetic stability, and much more. It is crucial in the development of embryos and in the dynamic balance of adult tissues, so also that of the lung. With respect to LC, changes of the *Wnt* signalling pathway have been observed for *Wnt* ligands, frizzled, TCF/LEF (T cell factor/lymphoid enhancer factor)-dependent transcription, and *Wnt* inhibitor silencing [3].

Genome-wide association studies (GWAS) have identified dozens of susceptibility loci throughout the genome that are associated with the susceptibility to lung cancer or one of its histological subtypes [4–11]. Genes related to *Wnt* signalling, one of the key pathway regulating cell development and stemness, were not detected as being associated to LC susceptibility in individuals of European descent so far, unlike *TERT* (5p15.33) that was one of the first for which a robust association was observed [12]. Aberrant *Wnt* signalling is often observed in expression profiles of many cancers, but to date no association of *Wnt*/*Ahr* genes with susceptibility to cancer of any type has been observed [13–15]. Administration of RNAi against *Wnt* was shown to reduce tumour burden in lung adenocarcinoma (adenoLC) [16]. In non-small cell lung cancer (NSCLC), overexpressed *miR-582-3p* maintains stemness features by negatively targeting the regulators of *Wnt* signalling *Axin2*, *DKK3* and *SRP1* for degradation, thereby increasing β-catenin mediated *Wnt* activity [17]. *TERT* expression was found to be directly enhanced by binding of β-catenin to its promoter region and thereby links telomerase activity to *Wnt* signalling [13]. This is as much important, as *TERT* is one of the first and most robust susceptibility genes for LC identified by GWAS [18, 19]. The tight regulatory machinery of the *Wnt* pathway has several major antagonists, such as Secreted Frizzled related protein (*sFRP*), Dickopff 5 (*DKK*) protein and *Axin2* protein [20]. Evidence also exists for a crosstalk between *AhR* and *Wnt* signalling [21].

Aryl hydrocarbon receptor (7p21.1; *AhR*) is a ligand induced transcription factor, which is translocated into the nucleus. It is known to mediate the toxicity and tumorigenesis of a variety of environmental pollutants, including for NSCLC. *AhR* upregulates the enzyme CYP1A1 when cells are exposed to carcinogenic metabolites, such as some polycyclic aromatic hydrocarbons (PAHs) found in cigarette smoke. The CYP1A1 coding gene is discussed as a susceptibility gene for LC. *AhR* is a major determinant in the process of smoking driven LC [22–24]. The complexity of both the *AhR* signalling pathway and the *Wnt* signalling cascade is reflected by interaction effects of genomic variants within genes, which control their function [25]. Recently, the association of the *Wnt*-genes *DKK4* (8p11.21), *DKK3* (11p15.3), *DKK2* (4q25), *FRZB* (2q32.1, also known as *sFRP*3), *SFRP4* (7p14.1), *Axin2* (17q24.1) and a potential interaction with *AhR* was investigated with respect to the susceptibility to LC in a sample of 600 subjects from North India [25, 26]. A notable association with LC, e.g., for the *SFRP4* variant rs1802073 (OR = 3.19; 95% CI 1.81–5.63), was reported. Classification And Regression Tree (CART) analysis revealed an interaction of *DKK2* and *SFRP4* polymorphisms to be the best (off all investigated) predictors for LC; especially within smokers. They also reported to have identified several high-risk subgroups in smokers, e.g., characterised by *DKK2* (rs17037102/rs419558) and *Axin2* (rs9915936). A similar picture was observed in a sample of 270 subjects from Istanbul, Turkey [27]. A two-way interaction between *DKK*3 (rs3206824) and *SFRP4* (rs1802074) was found to be predictive of LC.

We aimed to assess a possible association of *AhR* pathway and *Wnt* signalling cascade with LC within the large-scale series of cases and controls of European descent hold by the International Lung Cancer Consortium (ILCCO)/Integrative analysis of Lung Cancer Etiology and Risk (INTEGRAL). To do this, we also evaluated the contribution of these genes to genetic prediction of LC as a complement to known LC-related markers.

Methods

The work presented has been reviewed and approved by the ILCCO Steering Committee.

Cases and controls

Phenotype and genotype data of 58,181 entries of the data repository of ILCCO were extracted. Details of the repository is described previously [4, 28]. QC control samples, individuals without information on smoking status or age, and samples of poor genotyping quality or sex discrepancies, were excluded. To avoid population stratification, this analysis is focused on European-ancestry population (defined as more than 95% probability of being of European descent). Fourteen thousand sixty-eight incident LC-cases and 12,390 cancer-free controls of European descent remained for analysis. Those genotyped with other genome-wide array in addition to OncoArray were separated to form an independent validation set (2nd validation set) of size ($n = 4359$, including 2360 LC-cases and 1999 controls).

Selected markers

For this investigation we extracted the genotypes of 113 genomic variants (markers) assigned to 58 genes, previously associated with the risk for LC in European decent people or one of its histological subtypes through a wide variety of approaches [4–11] or proxies thereof (called *LC-marker*), and 296 markers assigned to 7 genes involved in *Wnt* signalling and listed in Bahl et al. [25, 26] and Yilmaz et al. [27] (called *AhR*/*Wnt-marker*). Thus, we focused this analysis to genes previously investigated with respect to LC. Fifty of these 409 markers were eliminated before analysis due to a MAF $<1\%$ (minor allele frequency), or departure from HWE (Hardy–Weinberg equilibrium) in genotypes (unaffected $p < 10^{-7}$, affected $p < 10^{-12}$), or low imputation accuracy (info <0.8). Seventy-eight of the remaining *LC-markers* were genotyped with the OncoArray (44 thereof are proxy SNPs identified using LDlink [29]) and 32 had to be imputed. Two hundred and twenty one of the remaining *AhR*/*Wnt-markers* were genotyped and 28 have been imputed. A list of these markers extracted from ILCCO OncoArray repository is given in the Additional file 1.

Association analysis

We first performed association analysis for each marker separately using the program PLINK [30, 31]. Crude (model 1) and adjusted odds ratios (ORs) were estimated along with 95%-confidence intervals within log-additive models. Sex, age and smoking status and the first 3 principal components (PCs) to adjust for population stratification (model 2); and in addition the 6 most significantly associated *LC-markers* (rs55781567, 15q25.1 *CHRNA5*; rs11780471, 8p21.2 *CHRNA2*; rs7705526, 5p15.33 *TERT*; rs56113850, 19q13.2 *CYP2A6*; rs71658797, 1p31.1 AK5; rs11571833, 13q13.1 *BRCA2*) (model 3) were included in adjusted models. ORs were estimated for overall LC, small cell LC (SCLC), squamous cell LC (SqCLC), adenocarcinoma LC (adenoLC), ever smokers, never smokers and individuals aged ≤ 55 years (early onset LC) as subgroups. We generated QQ-plots for the *AhR*/*Wnt-markers* and estimated the genomic inflation factor λ. To account for multiple testing, genome-wide statistical significance was considered to correspond to a p value of 10^{-7} or lower, suggestive significance to a p value between 10^{-5} and 10^{-7} and nominal significance to a p value between 0.05 and 10^{-5}.

Logistic regression—predicting models with model selection

We fitted logistic regression models with variable selection to find appropriate polygenic risk scores (PRS) to predict the disease (LC) status (affected or unaffected). Any *AhR*/*Wnt-marker* or the *LC-marker* could be included in the model without preference. To avoid multi-collinearity we removed one of two SNPs in LD to another ($R^2 > 0.8$, pruning). The remaining entered the models as potential predictors. We performed forward selection until the Bayesian information criterion (BIC, most stringent selection), the Akaike information criterion (AIC, less stringent selection, contains in general more predictors) or the sample size corrected AIC (AICC) indicate a best solution (and 10 more selection steps). The resulting PRSs are called BIC-, AIC- and AICC-scores. Note, that for the purpose of model building, the AIC-selection is asymptotically equivalent to cross-validation (CV) [32, 33]. To avoid overfitting, we assigned individuals to a training or a validation set (to build a score) and a testing set (to examine the score performance) with a 1/3 probability each. For comparison, we also generated a BIC^{LC}-score with at least one marker, only allowing *LC-markers* to enter the model building. To compare the importance for LC prediction of the sets g of *LC-maker*s and *AhR*/*Wnt-markers, respectively,* we contrasted the importance-values defined as $I_g = \sum_{m \exists g} |\beta_m| \bullet \text{MAF}_m$ for each score (MAF_m the minor allele frequency and β_m the logistic regression coefficient of marker m). The superiority of the AIC-scores over the BIC^{LC}-score and the BIC-score was tested applying the nonparametric test of DeLong et al. [34] (1-sided) on AUCs of ROC (area under the receiver operation

characteristic curve). In addition, a corresponding precision-recall plot was created for the SCLC.

Decision trees

Decision trees were created to examine marker × marker interaction with respect to the LC prediction. Any *AhR*/*Wnt-marker* or the *LC-marker* could be included in a tree without preference. This was accomplished in the entire sample and in all subgroups defined above. The R packages rpart and DescTools were used [35, 36]. To avoid trees being formed by spurious epistasis we removed one of two SNPs in LD to another (R^2 >0.8, pruning). Since overfitting is a point of concern when building decision trees, the complexity parameter was first optimized applying ten-fold cross-validation, grading the performance on the validation set by Somers' D (concordance of true and predicted LC-status). The ability of the optimal trees to predict the LC-status was then tested within the independent sample of 4359 cases and controls. True positive (TP) and true negative (TN) rates are given.

All statistical analyses were performed with SAS® 9.4, PLINK 1.90 and 2.0 or R 4.0.2.

Gene expression

We extracted information on gene expression from the Human Protein Atlas [37, 38] and LungGENS [39, 40].

Results

Sample description

The analysed sample consists of 14,068 LC-cases and 12,390 controls with median age of 63. Sixty-three percent were male, 52% of cases and 28% of controls were current smokers. The most frequent histological subtype is adenocarcinoma (38%), followed by squamous cell carcinoma (SqCLC) (26%) and small cell lung cancer (SCLC) (10%). The proportion of never-smokers was largest within the subgroup of adenocarcinoma cases (14%), but almost the same between those cases aged ≤55 years (10%) and aged >55 years (9%). Details on smoking status and histological subtypes are presented in Table 1.

Association analysis

We first performed association analysis for each *Wnt*/*AhR-marker* separately. The *p* values for an association of *AhR*/*Wnt-markers* with LC range from 0.005 (rs12115174; 8p11.21 *DKK4*; OR =0.9211) to 1 (model 2; adjusted for sex, age, smoking status and population stratification); with a negligible genomic inflation (λ =1.02). A nominally significant association ($10^{-5} < p \leq 0.05$) was observed for only 8 of the 249 markers (~3%). The corresponding point estimates of OR range from 0.88 (rs1053070054; 8p11.21 *DKK4*; p =0.007) to 1.12 (rs74596148; 7p14.1 *SFRP4*; p =0.25). A QQ-plot indicates that achieved p values almost perfectly agree with the expectation of no associated marker (see Fig. 1). p values and OR are in moderate agreement between the

Table 1 Smoking by LC status and subgroups

		Never smoker		Ever smoker					
	Total	Never		Former		Current		Ever[a]	
	n	*n*	%	*n*	%	*n*	%	*n*	%
Control									
Age ≤55 years	2762	951	34	698	25	896	32	217	8
Age >55 years	9628	2960	31	3572	37	2568	27	528	5
All	12,390	3911	32	4270	34	3464	28	745	6
Case									
SqCLC	3692	138	4	1257	34	2158	58	139	4
SCLC	1450	48	3	383	26	965	67	54	4
Other LC	3629	405	11	1200	33	1820	50	204	6
AdenoLC	5297	740	14	1989	38	2401	45	167	3
Age ≤55 years	2765	281	10	452	16	1945	70	87	3
Age >55 years	11,303	1050	9	4377	39	5399	48	477	4
All	14,068	1331	9	4829	34	7344	52	564	4
Total	26,458	5242	20	9099	34	10,808	41	1309	5

SCLC small cell lung cancer; *SqCLC* squamous cell lung cancer; *AdenoLC* adenocarcinoma of the lung; *other LC* other histological subtypes

[a] As recorded

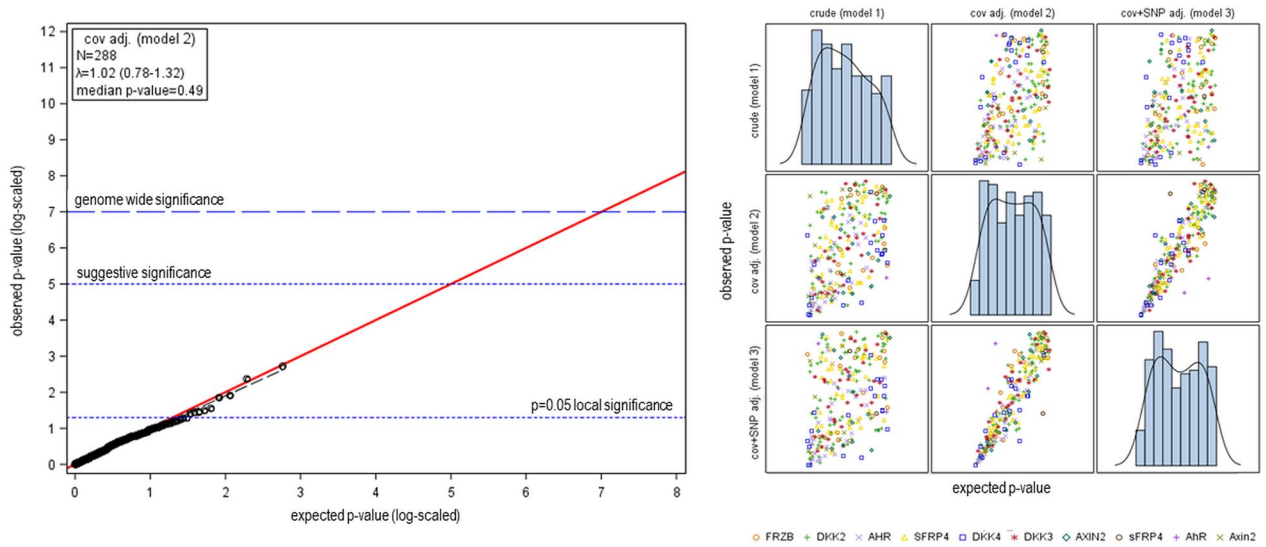

Fig. 1 Association of *AhR/Wnt-marker.* Left panel: QQ-Plot for model 2 (adjusted for sex, age and smoking status and the first three principal components); right panel: matrix of *p* values generated by model 1 (crude), model 2 (adjusted for sex, age and smoking status and the first three principal components) and model 3 (additionally adjusted for 6 selected *LC-markers*), genome-wide significance: *p* value $\leq 10^{-7}$, suggestive significance: $10^{-7} < p$ value $\leq 10^{-5}$, nominal significance: $10^{-5} < p$ value ≤ 0.05

models (e.g., model 2–3; additionally adjusted by *LC-markers*: Kendall's $\text{rho}_{\text{p}} = 0.75$, $\text{rho}_{\text{OR}} = 0.78$).

Subgroup analysis

When dividing the cases according to histological subtypes (SCLC; SqCLC and adenoLC) the observation of no detectable association for *WNT/AhR-markers* remains. Merely the number of nominally significant association ($10^{-5} < p \leq 0.05$) increases to 12 (5%) or 21 (8%) of the 249 markers for SqCLC and SCLC, respectively, hence close to the expected type 1 error. (Additional file 1: Table S2). When dividing the cases and controls according to their smoking behaviour (ever and never smokers), genome-wide significance ($p \leq 10^{-7}$) was achieved for 7 and 8 markers, respectively. Another 12 and 3 markers, respectively, were found suggestively significant ($10^{-7} < p \leq 10^{-5}$) (see Additional file 1: Figure S1) for ever and never smokers. Those markers found associated among ever smokers have mainly been directly genotyped and are assigned to *SFRP4* and *DKK4*. For example, for marker rs2722278 we estimated an OR = 1.20 (95% CI 1.13–1.27), yielding a *p* value of 5.6×10^{-10}. Those markers found associated among never smokers have mainly been imputed and are mostly assigned to *Axin2,* but also to *AHR*, *FRZB* and *DKK2*. Marker rs17037102, assigned to *DKK2*, was the only one found associated with LC by Bahl et al. and in this analysis (see Table 2 and Additional file 1: Table S3). Interestingly, the ORs of these markers estimated by model 3 (additionally adjusted for selected *LC-marker*) differ from that estimated by model 2. They are closer to one and no more significant. For example, for rs1133683 (*Axin2*) we observe an OR = 1.27 (95% CI 1.19–1.35, $p = 1 \times 10^{-12}$) fitting model 2, but OR = 0.95 (95% CI 0.86–1.06, $p = 0.3586$) fitting model 3.

Logistic regression—predicting models with model selection

We further fit logistic regression models with variable selection to evaluate the contribution of *AhR/Wnt-markers* to a polygenic risk scores (PRS), but without postulating the usefulness of the score as such. Eight *LC-markers* from only eight *LC-genes* (*CYP2A6*, *CHRNA5*, *TERT*, *AMICA1*, *CHRNA3*, COPS2, *HCG4* and *CHRNA2*) were selected for the BIC-score (most stringent selection) to predict overall LC. Hence, the BIC-score and the BIC^{LC}-score are identical. In contrast, the AIC-score (for overall LC identical to the AICC-score) includes 20 *LC-markers* and remarkable 17 *AhR/Wnt-markers*, with *LC-markers* being more important than the *AhR/Wnt-markers* (importance ratio 0.56: 0.34) (see Fig. 2, Additional file 1: Figure S3 and Table S4). The ability to distinguish cases and controls from susceptibility genes only was, as expected, poor for each of the scores (see Additional file 1: Table S5). In the training set the performance of the AIC/AICC-score (AUC = 0.607) exceeded those of the BIC/BIC^{LC}-score (AUC = 0.582) significantly ($p < 0.001$). Within the test set (AUCs: 0.577 and 0.576) and the 2nd validation set (AUCs: 0.553 and 0.548), the higher complexity with additional *AhR/Wnt-markers* did

Table 2 Significantly associated *AhR/Wnt-markers* within never and ever smokers

SNP	Cyto band	MAF (%)	Gene	Model 2			Model 1	Model 3
				p value	OR	95% CI	OR	OR
Never smoker								
Imputed								
rs202198518[a]	7p21.1	14	*AHR*	3.4×10^{-13}	0.72	0.66–0.79	0.71	0.90^{ns}
Imputed								
rs2237297[a]		14		9.9×10^{-14}	0.71	0.65–0.78	0.71	0.90^{ns}
Imputed								
rs1133683	17q24.1	42	*Axin2*	1.0×10^{-12}	1.27	1.19–1.35	1.27	0.95^{ns}
Imputed								
rs2240307		5		7.7×10^{-24}	0.41	0.34–0.49	0.40	0.62^{ns}
Imputed								
rs35285779[a]		9		3.2×10^{-22}	0.58	0.52–0.65	0.58	1.10^{ns}
Imputed								
rs35415678[a]		9		3.7×10^{-19}	0.62	0.56–0.69	0.62	1.10^{ns}
Imputed								
rs288326	2q32.1	10	*FRZB*	2.5×10^{-8}	1.42	1.25–1.60	1.41	0.98^{ns}
Imputed								
rs17037102	4q25	15	*DKK2*	7.4×10^{-15}	0.69	0.63–0.76	0.69	1.09^{ns}
Ever smoker								
Genotyped								
rs12532321	7p14.1	45	*SFRP4*	1.3×10^{-9}	1.14	1.09–1.19	1.15	1.13^{ss}
Genotyped								
rs7811872		36		1.3×10^{-8}	0.88	0.84–0.92	0.88	0.88^{gws}
Genotyped								
rs10226308		42		1.8×10^{-8}	0.88	0.85–0.92	0.89	0.89^{gws}
Genotyped								
rs10488617		42		1.6×10^{-8}	0.88	0.85–0.92	0.89	0.89^{gws}
Genotyped								
rs2722278		16		5.6×10^{-10}	1.20	1.13–1.27	1.16	1.20^{gws}
Genotyped								
rs2722279		11		9.0×10^{-9}	1.22	1.14–1.31	1.17	1.23^{gws}
Genotyped								
rs7811420		43		7.9×10^{-8}	0.89	0.85–0.93	0.89	0.89^{gws}
Imputed								
rs2073664	8p11.21	9	*DKK4*	9.4×10^{-11}	1.20	1.14–1.27	1.15	1.08^{ss}

Model 1: crude odds ratio (OR); model 2: adjusted for sex, age and smoking status and the first three principal components; model 3: OR additional adjusted for 6 selected *LC-markers*. Only markers are listed for which genome-wide significance (p value $\leq 10^{-7}$) was achieved

MAF minor allele frequency; ^{gws} genome-wide significant (p value $\leq 10^{-7}$); ^{ss} suggestive significant ($10^{-7} < p$ value $\leq 10^{-5}$); ^{ns} not significant ($p > 0.05$)

[a] Pair of markers in LD ($R^2 > 0.8$ in populations of European decent)

not improve discriminability for overall LC ($p = 0.87$ and $p = 0.35$).

Similar score composition and performance was observed for most subgroups. The BIC-scores in the subgroups adenoLC (involved marker LC:*AhR*/*Wnt* = 6:–), SCLC (3:–) and smokers (7:–) contained *LC-markers* only, whereas *AhR*/*Wnt-markers* are included even under this stringent variable selection in the subgroups SqCLC (5:1) and Early onset LC (2:2). However, between 14 and 31 *AhR*/*Wnt-markers* entered these subgroup's AIC-scores. For these subgroups, the importance of the *LC-markers* for the AIC-score is higher than that of the included *AhR*/*Wnt-markers*.

Most important, we observed a significantly higher predictive accuracy (larger AUCs) of the *AhR*/*Wnt-markers* enriched AIC-scores compared to BIC^{LC}–score in the subgroup of SCLC patients ($p = 0.019$; $AUC_{AIC} = 0.577$ $AUC_{BIC} = 0.546$) within the test set (see Additional file 1:

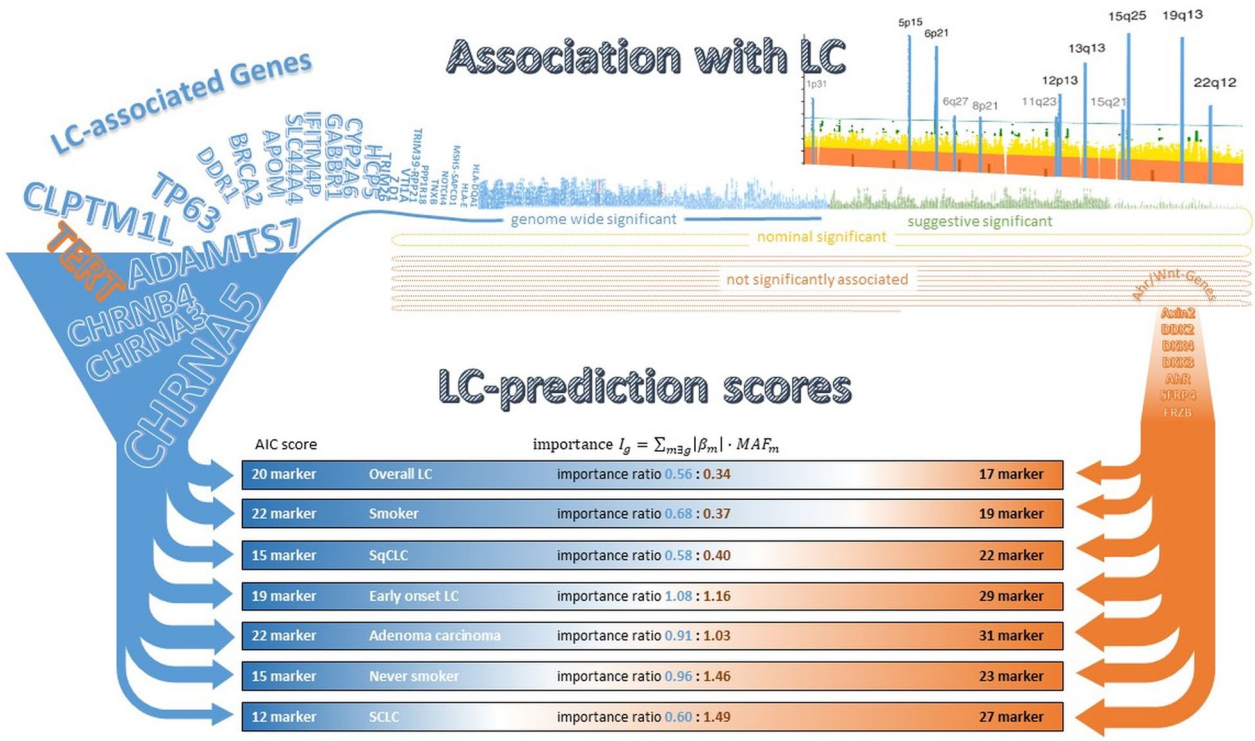

Fig. 2 Comparison of score composition. *LC* lung cancer; *AIC score* score of a logistic regression model with variant selection according to the Akaike information criterion (AIC); MAF_m minor allele frequency of variant (marker) m; β_m regression parameter of variant m; *LC-associated genes* previously reported as associated to LC or one of its histological subtypes; *AhR/Wnt*-genes selected genes assigned to Wnt-signalling, including AhR; *Smoker* ever, former and current smoker; *SCLC* small cell lung cancer, *SqCLC* squamous cell lung cancer, Early onset LC: aged $\leq$ 55 years; TERT is framed in orange, because telomerase activity is related to Wnt signalling

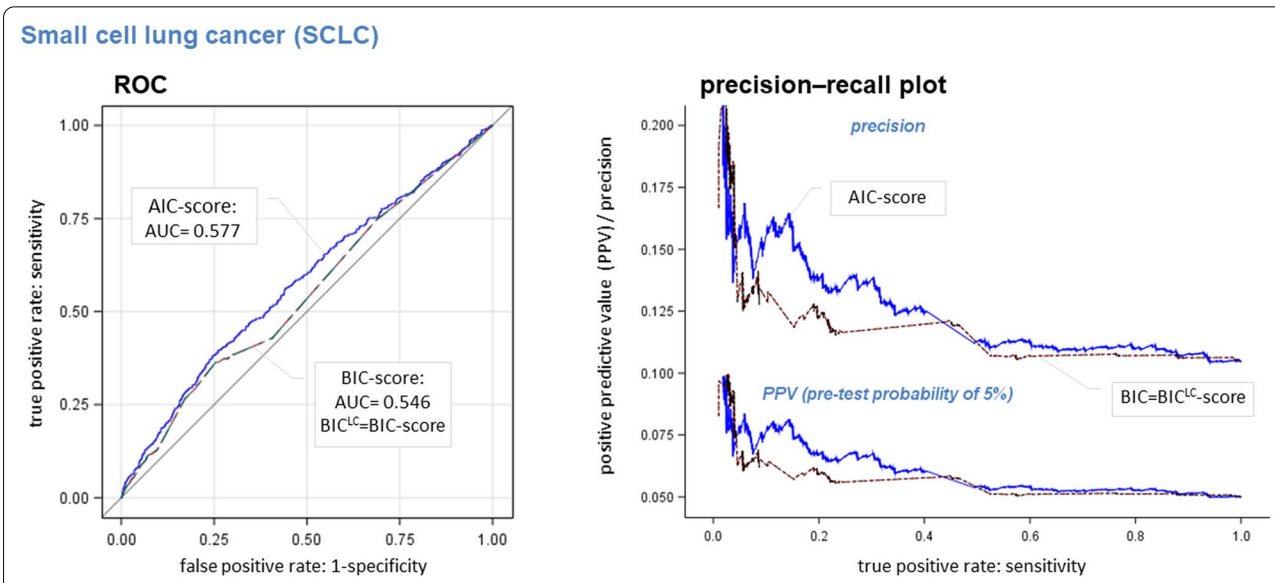

Fig. 3 ROC and precision-recall-plot: SCLC. The diagnostic performance of the AIC-score compared to the BIC/BIC^{LC}-score in the test-set is presented. Left panel: *ROC* receiver operation characteristics; right panel: corresponding precision-recall plot; precision = (true positive cases)/(true positive cases + false positive controls), positive predictive value (PPV) = (sensitivity × pre-test-probability)/[(sensitivity × pre-test-probability) + (1 − specificity × 1 − pre-test-probability)] for a pre-test-probability of 5%

Figure S4). For this subgroup, the selected *AhR*/*Wnt-markers* contribute to the AIC-score more than twice as much as the *LC-markers* (importance ratio 0.60:1.49). The precision-recall plot of Fig. 3 indicates that a positive SCLC prediction based on the AIC-score can be trusted more than that based on *LC-markers* alone (BIC^{LC}-score). In the 2nd validation set the score-specific AUCs were similar but no more significantly different ($p = 0.08$; $AUC_{AIC} = 0.564$ vs. $AUC_{BIC} = 0.531$). The AIC-score of this SCLC-subgroup is composed of 12 *LC-markers* (assigned to *CHRNA5*, *HCG4*, *DNAJB4* (4× each), *CYP2A6*, *CHRNA3*, *CHRNA2*, *AMICA1*, *KCNJ4*, *AS1*, *BRCA2*, *EGFL8* and *WNK1* (2 × each) and 27 *AhR*/*Wnt-markers* (assigned to all *AhR*/*Wnt*-genes except *DKK3*). However, only one LC patient in the test set ($n = 434$) and one in the 2nd validation set ($n = 164$) was recognized as a patient at a threshold of 50% case probability.

Interestingly the BIC-score for never smokers was built by only two *AhR*/*Wnt-markers* (assigned to *Axin2* and *SFRP4*) but not a single *LC-marker*. Furthermore, the *LC-markers* are the minority in the composite of the AIC-score (15:23). They also contribute less to the AIC-score than the *AhR*/*Wnt-markers* (importance ratio of 0.96:1.46). The median predicted case probability, in the test set (24.8%) and 2nd validation set (25.6%), exceeds that of controls by 1–2%-points. However, AUC differed neither in the test set ($p = 0.13$; $AUC_{AIC} = 0.540$ $AUC_{BIC} = 0.514$) nor in the 2nd validation set ($p = 0.36$; $AUC_{AIC} = 0.535$ $AUC_{BIC} = 0.526$) significantly. Nevertheless, this observation highlights the value of the *AhR*/*Wnt-markers* in the subgroup of never smokers.

Decision trees

Finally, we generated decision trees to evaluate the contribution of *AhR*/*Wnt-markers* to LC prediction that allow for a complex interaction structure, but without postulating the usefulness of the trees as such. The decision tree for overall LC (whole sample) consists off solely a single decision node (rs55781567 assigned to *CHRNA5*), achieving a Somers' concordance index $D = 0.0565$ in the 2nd validation set (see Additional file 1: Table S6 and Figure S2). A single-node decision-tree was also found optimal for participants aged ≤ 55 years (split: rs1051730 assigned to *CHRNA3*), achieving a Somers' concordance index $D = 0.096$. These two, unsophisticated trees are characterised by balanced TP- (about 62%) and TN-rates (about 44%).

The decision trees for ever smokers, SCLC and SqCLC were more complex achieving Somers' concordance indexes D of 0.007, − 0.0005 and 0.0126, respectively. The trees for SCLC and SqCLC are characterised by an extreme TP-rate < 5% and TN-rate > 99%; the tree for Ever Smokers by a TP-rate > 99% and TN-rate < 5%. Remarkably, a marker assigned to *CHRNA5* was always chosen as the first and most important split for the trees for ever smokers, for SCC and SqCLC. However, markers assigned to *AhR*/*Wnt-genes* (smoker: *DKK2*; *SCLC*: *FRZB*; *SqCLC*; *DKK2* and *DKK3*) appear at lower-level decision-nodes (Additional file 1: Figures S5–S8). With the same program settings, no decision tree could be created for adenocarcinoma.

Most notable is the optimal decision tree for the 5242 never smokers (75% LC-cases, 25% controls), the only one that does not contain a marker belonging to the CHRN (Cholinergic receptors nicotinic subunits) gene group (see Fig. 4). The tree is built from only two *LC-markers* but 7 *AhR*/*Wnt-markers*, achieving a Somers' concordance index $D = -0.002$. One can make out three branches of this tree. Branch I covers two thirds of individuals ($n = 754$, 66% of 1141 in the 2nd validation set): all of these are graded as "unaffected" based on only the two *LC-markers*: first decision node (rs885518 assigned to MTAP) and second decision node (rs7705526 assigned to TERT that links telomerase activity to Wnt signalling). For branch II an additional node (rs17214897 assigned to DKK2) is taken into account, covering a further tenth (9.9%) of never smokers. In this branch, very few subjects of the training set (1.7% within branch II eq. 0.17% of all never smokers) are graded "affected". However, one in four individuals of the 2nd validation set belonging to both branches, I and II, is truly "affected" but has not been detected (TP-rate = 0%, TN-rate = 100%). Rated as "affected" appears in the test set only in the third branch III, covering the remaining fourth of never smokers ($n = 284$ of the 2nd validation set). This third branch requires genotypes of several *AhR*/*Wnt-markers* assigned to *AHR*, *Axin2*, *DKK2* and/or *SFRP4*. Herein, one in three ($n = 97$ of the 2nd validation set) is truly "affected" and is given a chance to be correctly identified, which appears in 8 LC-cases (TP-rate = 9%, TN-rate = 88%). We also noted that the histological subtypes are equally distributed between the branches (see Additional file 1: Table S7).

Gene expression

AHR, *Axin2*, *DKK3* are ubiquitously expressed, with RNA expression detected in many tissues and evidence for protein expression. *Axin2* and *DKK3* are moderately to highly expressed in normal lung tissues according to the Human Protein Atlas [37]. *AhR* is expressed at low levels in macrophage cells of the lung. No expression is reported for other *Wnt*/*AhR*-genes (see Additional file 1: Figure S9 and Table S9). Significant differential expression is listed in LungGENS for *AhR*, *Axin2 DKK2*, *DKK3* and *SFRP4* [39] (see Additional file 1: Table S8). Furthermore, AhR is reported to be abundantly expressed in

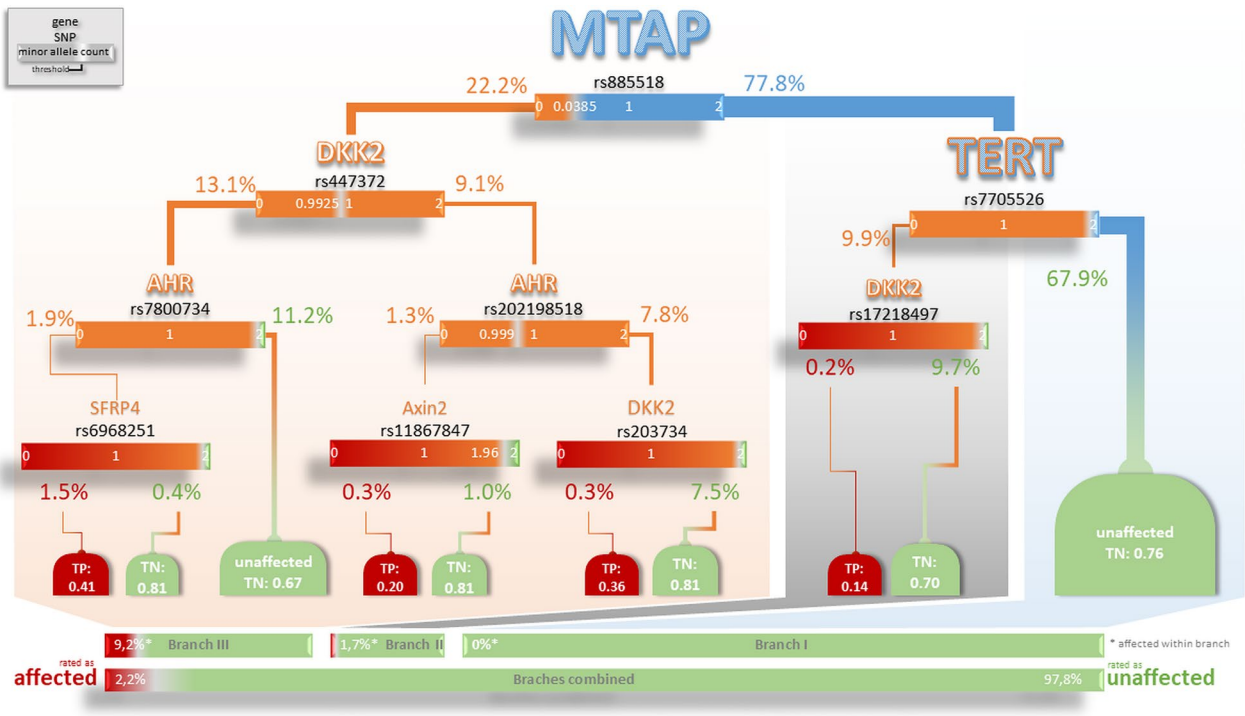

Fig. 4 Decision tree for never smoker. Node information: gene name, marker; split information below the node: threshold for minor allele count; blue split nodes: LC-genes, orange split nodes: *AhR/Wnt*-genes,; TERT is framed in orange, because telomerase activity is related to Wnt signalling; decision nodes and bars: green for unaffected; red for affected, *TN* true negative rate, *TP* true positive red; the size of gene names, lines and decision notes is proportional to the size of the respective (sub)sample

solid lung tumours, especially in adenocarcinomas. AhR overexpression was associated with upregulation of IL-6 secretion, which is critical for lung cancer initiation [41]. Detailed information on gene expression is given in the Additional file 1. In addition, the *DKK*1 serum level was seen as significantly lower in NSCLC and SCLC patients compared to healthy controls [42]. Significant upregulation of *DKK2* expression was found in APC (adenomatous polyposis coli)-mutated non-SCLC lung cancers [43].

Discussion

This investigation was intended to discover association of the *Wnt*-genes *DKK4* (8p11.21), *DKK3* (11p15.3), *DKK2* (4q25), *FRZB* (2q32.1, also known as *sFRP*3), *SFRP4* (7p14.1), *Axin2* (17q24.1) and a potential interaction with *AhR-genes,* to LC in a large sample of 26,458 individuals of European descent. No marginal association of *AhR*/*Wnt-markers* with overall LC was observed. Interestingly, an accumulation of associated markers was observed splitting the sample by smoking status, where respective markers in ever smokers are assigned to *SFRP4*. On the other hand, association analysis in never smokers reflects complex gene–gene interactions, as markers of several *AhR*/*Wnt-genes* were found to be genome-wide associated with LC. This complexity is also visible through the decision tree analysis.

Our results are in line with findings from northern India [25, 26] and from Istanbul, Turkey [27], both of which are based on much smaller samples (approx. 600 and 270 people, respectively). In these investigations, the interaction of DKK2 and DKK3 with SFRP4 and Axin2 polymorphisms turned out to be the best (of all examined) predictors of LC, especially in smokers. Axin2, but also AHR, FRZB and DKK2, were observed to be complex associated in never smokers. Our analysis agrees with both previous studies that complex interaction patterns between the examined genes contribute to overall LC susceptibility or within certain subgroups. However, we have not been able to replicate reported single marker associations directly.

To discover patterns of *AhR*/*Wnt-genes* involved in LC genesis we further changed the focus from significance of association to inclusion in prediction models, and followed two approaches: first, we searched for polygenic risk scores (PRS). Doing so, we add up marker main effects to construct multidimensional scores, optimising model fit (instead of marker preselection by p-value

below some threshold), to discriminate cases from controls in a somehow ideal way. Complex gene × gene (G × G) interactions are not modelled.

Nevertheless, the proportion of *AhR*/*Wnt-genes* entering some of the predictive models was remarkable large, given that these markers are not, all other candidates, however, genome-wide significantly associated to LC. This was particularly noticeable for SCLC, since *AhR*/*Wnt-markers* contribute more than twice as much to the prediction score as *LC-markers*. It is known, that within current smokers, tobacco consumption is strongest associated to SCLC [44]. Moreover, within never smokers, a stringed defined score is made up from only two *AhR*/*Wnt-markers*, assigned to *Axin2* and *SFRP4*. However, the discriminative ability of PRSs for LC, contributing markers with significance for main effect at different levels, is in general poor. The AUC of the BIC^{LC} score for overall LC (0.58 in the test set and 0.55 in the 2nd validation set) corresponds to the AUC = 0.54 based on four top *LC-genes* in a simulated population, as given by the GWAS-ROCS Database (https://gwasrocs.ca/). This may be due to other overpowering risk factors, since models including, e.g., age, sex and smoking variables achieve higher AUCs (0.62–0.79) [45].

Recently two polygenic risk scores (PRSs) for overall-LC had been developed, validated and assessed with respect to improving eligibility to low-dose computed tomography (LDCT) as the only recommended screening test for lung cancer. Jia et al. [46, 47] build a PRS on 19 genome-wide associated SNPs ($p < 0.5 \times 10^{-8}$). Hung et al. [48], integrated their PRS on 128 SNPs (35 "known" LC-related loci, 93 suggestive associated loci selected by LASSO-regression model) into the $PLCO_{all2014}$ risk model. Both approaches have been validated using data from the UK Biobank. For both scores, the mean PRS differed only slightly between LC cases and cancer-free controls (Jia: effect size ~0.19; Hung: effect size ~0.22). For both scores, no substantial increase in discriminability of cases from controls is reported, when adding the PRS to existing risk models (Jia: family history—AUC = 0.589, family history + PRS – AUC = 0.615; Hung: $PLCO_{all2014}$–AUC = 0.828, $PLCO_{all2014}$ + PRS – AUC = 0.832). However, both were able to show that the age at which a smoker crosses the recommended screening threshold of 1.5% for the 5-year LC risk depends on the genetic background, which is sufficiently quantified by the PRS examined. Some smokers will be eligible by < 50 years of age, others by > 60 years of age. Hence, constructing reliable PRS, even with small discriminability, may help to improve the performance of LDCT.

Two- and multiway G × G interaction can also contribute to LC susceptibility, rather than just markers with observed (marginal) main effects. G × G interaction is in general less commonly investigated, not only because this requires much larger samples. However, Li et al. [49] found RGL1:RAD51B in overall LC and non-SCLC, SYNE1:RNF43 in adenocarcinoma and FHIT:TSPAN8 in SqCLC to interactively contribute to LC susceptibility. As in the presented data analysis, the impact of these genes would also have been overlooked considering main effects only. Another reason could be that LC itself is just a generic term of several subcategories that differ in terms of LC initiation and require separate PRSs [45, 50]. A third reason of the poor performance may be due to the exclusively concentration on genetic effects, rather than modelling lifelong interaction with the environment as well. For example, G × E interaction effects for LC have been observed smoking [51], exposure to asbestos fibres [52, 53] and exposure to radon [54, 55].

With this in mind, the data analysis presented shows that the complex interaction of Wnt-related genes has the potential to be part of an adequate risk assessment for never-smokers or in relation to certain histological subtypes of LC.

As a second approach, we constructed decision trees, which mainly depict G × G interaction patterns. Although, the ability to discriminate cases from controls is again poor, CHRNA5 was in general the most important first node for overall LC and in many subgroups. *AhR*/*Wnt-genes* play a complex but important role in at least one quarter of never smokers, as seen before. Remarkably, *TERT*, which links telomerase activity to *Wnt* signalling, was central in that branch and important for the remaining three quarters of never smoker. This corresponds to a concentration of relevant genes for this subgroup in the *CLPTM1L-TERT* region on chromosome 5, as previously reported by Hung et al. [56]. Our observations confirm the suspicion, that LC in never smokers is a different entity, justified beforehand on differences in epidemiological, clinical and molecular characteristics [50].

We would like to emphasize that this study was not intended to provide a definitive and reliable risk assessment, but rather aimed to examine in depth the LC-relevant complex interaction pattern of *AhR*/*Wnt-genes* hypnotized by Bahl et al. Indeed, considering prediction instead of association provides weaker evidence for this, but is valid in view of the large amount of external evidence. The importance of the *Wnt*-signalling pathway and its antagonist's *sFRP*, *DKKs* and *Axin2* for cancer is outlined in the introduction. One can also assume a connection with the molecular functionality, since involved genes are expressed ubiquitously or in lung tissues.

Although the large-scale, thoroughly quality checked, and representative sample of genetically proven European descent individuals was used for the presented analysis,

some limitations must be noted. We used a rather narrow definition of *AhR*/*Wnt-genes* to limit the number of possible interactions. An extension to, e.g., EGRF, APC, FRAT2 or the CYP-family would also be justified. We further could have chosen the random forest method as a more contemporary and robust approach than decision trees, but we would not be able to present our results so illustrative. However, the sample size allowed subgroup analyses, whereby the special importance of *AhR*/*Wnt-genes* for SCLC and never smokers could be shown.

Conclusions

The role of markers belonging to *Wnt* signalling and the *AhR* pathway in LC susceptibility may be underrated in main-effects association analysis. Complex interaction patterns in individuals of European decent have moderate predictive capacity for LC or subsets thereof, especially in never smokers.

Abbreviations

AhR: Aryl hydrocarbon receptor; GWAS: Genome-wide association studies; LC: Lung cancer; NSCLC: Non-small cell lung cancer; SCLC: Small cell lung cancer; SqCLC: Squamous cell lung cancer; adenoLC: Adenocarcinoma lung cancer; OR: Odds ratio; CART: Classification and regression tree; AUCs of ROC: Area under the receiver operation characteristic curve; ILCCO: International Lung Cancer Consortium; INTEGRAL: Integrative analysis of lung cancer etiology and risk; PRS: Polygenic risk scores; BIC: Bayesian information criterion; AIC: Akaike information criterion; CV: Cross validation; MAF: Minor allele frequency; TP: True positive rate; TN: True negative rate; LDCT: Low-dose computed tomography.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40001-022-00638-7.

Additional file 1: Table S1. List of investigated SNPs (*AhR/Wnt-markers* and *LC-markers*). **Table S2.** Association of *AhR/Wnt-marker* within subgroups. **Table S3.** Association of markers reported elsewhere. **Table S4.** Score composition and importance ratio. **Table S5.** Discriminability of prediction scores. **Table S6.** Prediction accuracy of the decision trees. **Table S7.** Prediction accuracy in never smokers by histological subtypes. **Table S8.** Expression in normal tissue of the lung according LungGENS. **Table S9.** Expression in normal tissue of the lung according the Human Protein Atlas. **Figure S1.** Association of *AhR/Wnt-markers* within never and ever smokers. **Figure S2.** Decision tree for overall LC. **Figure S3.** LC-risk score: model selection and ROCs for overall LC. **Figure S4.** LC-risk score: model selection and ROCs for SCLC and Never smoker. **Figure S5.** Decision tree for early onset LC (age ≤55 years). **Figure S6.** Decision tree for SqCLC. **Figure S7.** Decision tree for SCLC. **Figure S8.** Decision tree for ever smoker. **Figure S9.** Expression profiles according to the Human Protein Atlas.

Acknowledgements

Not applicable.

Authors' contributions

AR designed the investigation, carried out parts of the formal analysis and wrote the main manuscript text. NM carried out parts of the formal analysis, prepared figures and critical reviewed and revised the manuscript. BW carried out parts of the formal analysis. RJH and CIA coordinate the research activity of the consortium, including data curation and funding acquisition. HB supervised the investigation, including funding acquisition. RJH, HB, LLM, CIA and LAK critical reviewed and revised the manuscript. AC, AH, AR, ASA, AT, CC, DA, DCC, EJD, FT, GEG, GF-T, GL, GR, HB, JAD, JKF, LAK, LLM, MBS, MCA, MJ, MPAD, MTL, NEC, PB, PJW, PL, RJH, SEB, SL, SMA, SSS, SZ, TL and TRM collected and provided study materials and data. All the authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The National Institutes of Health (7U19CA203654-02/ 397 114564-5111078 Integrative Analysis of Lung Cancer Etiology and Risk) supported this work. CARET is funded by the National Cancer Institute, National Institutes of Health through grants U01 CA063673, UM1 CA167462, R01 CA 111703, RO1 CA 151989, U01 CA167462 and funds from the Fred Hutchinson Cancer Research Center. The Boston Lung Cancer Study was funded by NCI grant 5U01CA209414. Other individual funding for participating studies and members of INTEGRAL-ILCCO are listed elsewhere [4, 10, 30]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We acknowledge support by the Open Access Publication Funds of the Göttingen University.

Availability of data and materials

The data that support the findings of this study are available from ILCCO/INTEGRAL but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of ILCCO/INTEGRAL.

Declarations

Ethics approval and consent to participate

All participants in this study signed an informed consent, approved by the local internal review board or ethics committee and administered by trained personnel. All consortium research received approval from the Dartmouth Committee for Protection of Human Subjects on 7/30/2014 with id STUDY00023602. All experimental protocols and other methods used comply with institutional, national, or international guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Genetic Epidemiology, University Medical Center, Georg-August-University Göttingen, Göttingen, Germany. [2]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, University of Toronto, Toronto, ON, Canada. [3]Dalla Lana School of Public Health, University of Toronto, Toronto, Canada. [4]Department of Environmental Health, Harvard T.H. Chan School of Public Health and Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA. [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, MD, USA. [6]Medical Oncology and Medical Biophysics, Princess Margaret Cancer Centre, Toronto, ON, Canada. [7]Medicine and Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [8]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark. [9]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [10]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen, Denmark. [11]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA. [12]Department of Thoracic Surgery, Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA. [13]Faculty of Medicine, University of Oviedo, ISPA and CIBERESP, Oviedo, Spain. [14]Clalit National Cancer Control Center at Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel. [15]Department of Molecular and Clinical Cancer Medicine, Roy Castle Lung Cancer Research Programme, The University of Liverpool, Liverpool, UK. [16]Departments of Health Evidence and Urology, Radboud University Medical Center, Nijmegen, The Netherlands. [17]Department of Pharmaceutical Sciences, College of Pharmacy, Washington State University, Spokane, WA, USA. [18]National

Institute of Occupational Health, Oslo, Norway. [19]British Columbia Cancer Agency, Vancouver, BC, Canada. [20]Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [21]Department of Epidemiology, Geisel School of Medicine, Hanover, NH, USA. [22]Unit of Biomarkers and Susceptibility, Oncology Data Analytics Program, Catalan Institute of Oncology (ICO), Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. [23]Markey Cancer Center, University of Kentucky, Lexington, KY, USA. [24]Swedish Medical Group, Seattle, WA, USA. [25]Program in Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [26]Department of Population Health Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA. [27]Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK. [28]University of Salzburg and Cancer Cluster Salzburg, Salzburg, Austria. [29]Member of the German Center for Lung Research (DZL), Translational Lung Research Center (TLRC) Heidelberg, Heidelberg, Germany. [30]Translational Research Unit, Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany. [31]Department of Radiation Sciences, Umeå University, Umeå, Sweden. [32]International Agency for Research on Cancer, World Health Organization, Lyon, France. [33]Department of Biostatistics, Division of Basic Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [34]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA. [35]Institut Für Genetische Epidemiologie, Universitätsmedizin Göttingen, Humboldtallee 32, 37073 Göttingen, Germany.

Received: 7 September 2021 Accepted: 7 January 2022
Published online: 31 January 2022

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66:7–30.
2. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global cancer observatory: cancer today. https://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf. Accessed 1 Jan 2020.
3. Zhu W, Wang H, Zhu D. Wnt/β-catenin signaling pathway in lung cancer. Med Drug Discov. 2021. https://doi.org/10.1016/j.medidd.2021.100113.
4. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet. 2017;49:1126–32.
5. Timofeeva MN, Hung RJ, Rafnar T, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14,900 cases and 29,485 controls. Hum Mol Genet. 2012;21:4980–95.
6. Rosenberger A, Sohns M, Friedrichs S, et al. Gene-set meta-analysis of lung cancer identifies pathway related to systemic lupus erythematosus. PLOS ONE. 2017;12:e0173339.
7. Brenner DR, Brennan P, Boffetta P, et al. Hierarchical modeling identifies novel lung cancer susceptibility variants in inflammation pathways among 10,140 cases and 11,012 controls. Hum Genet. 2013;132:579–89.
8. Ji X, Bossé Y, Landi MT, et al. Identification of susceptibility pathways for the role of chromosome 15q25.1 in modifying lung cancer risk. Nat Commun. 2018;9:3221.
9. Truong T, Sauter W, McKay JD, et al. International Lung Cancer Consortium: coordinated association study of 10 potential lung cancer susceptibility variants. Carcinogenesis. 2010;31:625–33.
10. Wang Y, Wei Y, Gaborieau V, et al. Deciphering associations for lung cancer risk through imputation and analysis of 12,316 cases and 16,831 controls. Eur J Hum Genet. 2015;23:1723–8.
11. Feng Y, Wang Y, Liu H, et al. Novel genetic variants in the P38MAPK pathway gene ZAK and susceptibility to lung cancer. Mol Carcinog. 2018;57:216–24.
12. Landi MT, Chatterjee N, Yu K, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009;85:679–91.
13. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. Oncogene. 2016;36:1461–73. https://doi.org/10.1038/onc.2016.304.
14. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47:D1005–12.
15. GWAS Catalog. In: The NHGRI-EBI catalog of human genome-wide association studies. https://www.ebi.ac.uk/gwas/. Accessed 22 Dec 2020. 2020
16. Kerdidani D, Chouvardas P, Arjo AR, et al. Wnt1 silences chemokine genes in dendritic cells and induces adaptive immune resistance in lung adenocarcinoma. Nat Commun. 2019. https://doi.org/10.1038/s41467-019-09370-z.
17. Fang L, Cai J, Chen B, et al. Aberrantly expressed miR-582-3p maintains lung cancer stem cell-like traits by activating Wnt/beta-catenin signalling. Nat Commun. 2015;6:8640.
18. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet. 2008;40:616–22.
19. Yuan Y, Lu C, Xue L, Ge D. Association between TERT rs2736100 polymorphism and lung cancer susceptibility: evidence from 22 case–control studies. Tumor Biol. 2014;35:4435–42.
20. Akiyama T. Wnt/beta-catenin signaling. Cytokine Growth Factor Rev. 2000;11:273–82.
21. Schneider AJ, Branam AM, Peterson RE. Intersection of AHR and Wnt signaling in development, health, and disease. Int J Mol Sci. 2014;15:17852–85.
22. Chang JT, Chang H, Chen PH, Lin SL, Lin P. Requirement of aryl hydrocarbon receptor overexpression for CYP1B1 up-regulation and cell growth in human lung adenocarcinomas. Clin Cancer Res. 2007;13:38–45.
23. Lin P, Chang H, Tsai WT, Wu MH, Liao YS, Chen JT, Su JM. Overexpression of aryl hydrocarbon receptor in human lung carcinomas. Toxicol Pathol. 2003;31:22–30.
24. Wang CK, Chang H, Chen PH, Chang JT, Kuo YC, Ko JL, Lin P. Aryl hydrocarbon receptor activation and overexpression upregulated fibroblast growth factor-9 in human lung adenocarcinomas. Int J Cancer. 2009;125:807–15.
25. Bahl C, Singh N, Behera D, Sharma S. High-order gene interactions between the genetic polymorphisms in Wnt and AhR pathway in modulating lung cancer susceptibility. Pers Med. 2017. https://doi.org/10.2217/pme-2017-0018.
26. Bahl C, Singh N, Behera D, Sharma S. Association of polymorphisms in Dickopff (DKK) gene towards modulating risk for lung cancer in north Indians. Future Oncol. 2017;13:213–32.
27. Yilmaz M, Donmez G, Kacan T, Sari I, Akgül Babacan N, Sari M, Kilickap S. Significant association between polymorphisms of Wnt antagonist genes and lung cancer. J Investig Med. 2015. https://doi.org/10.1097/JIM.0000000000000241.
28. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. Cancer Epidemiol Biomarkers Prev. 2017;26:126–35.
29. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31:3555–7.
30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:1–16.
31. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
32. Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. J Roy Stat Soc. 1977;39:44–7.
33. Fang Y. Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. JDS. 2011;9:15–21.
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837–45.
35. Andri S. DescTools: tools for descriptive statistics. Tukey: CRAN; 2021.
36. Therneau T, Atkinson B, Ripley B. rpart: recursive partitioning for classification, regression and survival trees. Tukey: CRAN; 2019.
37. Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. Science. 2015. https://doi.org/10.1126/science.1260419.
38. The Human Protein Atlas. https://www.proteinatlas.org/. Accessed 5 Feb 2021.
39. LungGENS. https://research.cchmc.org/pbge/lunggens/. Accessed 5 Feb 2021.

40. Du Y, Guo M, Whitsett JA, Xu Y. "LungGENS": a web-based tool for mapping single-cell gene expression in the developing lung. Thorax. 2015;70:1092–4.
41. Xue P, Fu J, Zhou Y. The aryl hydrocarbon receptor and tumor immunity. Front Immunol. 2018. https://doi.org/10.3389/fimmu.2018.00286.
42. Xu H, Wu J, Chen B, et al. Serum Dickkopf-1 (DKK1) is significantly lower in patients with lung cancer but is rapidly normalized after treatment. Am J Transl Res. 2014;6:850–6.
43. Shen T, Chen Z, Qiao J, Sun X, Xiao Q. Neutralizing monoclonal antibody against Dickkopf2 impairs lung cancer progression via activating NK cells. Cell Death Discov. 2019;5:1–12.
44. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. BMC Cancer. 2012;12:385.
45. Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs J, Jemal A, Berg CD, Chaturvedi AK. Implications of nine risk prediction models for selecting ever-smokers for computed tomography lung cancer screening. Ann Intern Med. 2018. https://doi.org/10.7326/M17-2701.
46. Jia G, Lu Y, Wen W, Long J, Liu Y, Tao R, Li B, Denny JC, Shu X-O, Zheng W. Evaluating the utility of polygenic risk scores in identifying high-risk individuals for eight common cancers. JNCI Cancer Spectr. 2020. https://doi.org/10.1093/jncics/pkaa021.
47. Jia G, Wen W, Massion PP, Shu X-O, Zheng W. Incorporating both genetic and tobacco smoking data to identify high-risk smokers for lung cancer screening. Carcinogenesis. 2021. https://doi.org/10.1093/carcin/bgab018.
48. Hung RJ, Warkentin MT, Brhane Y, et al. Assessing lung cancer absolute risk trajectory based on a polygenic risk model. Cancer Res. 2021;81:1607–15.
49. Li Y, Xiao X, Bossé Y, et al. Genetic interaction analysis among oncogenesis-related genes revealed novel genes and networks in lung cancer development. Oncotarget. 2019;10:1760–74.
50. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. Nat Rev Cancer. 2007;7:778–90.
51. Saccone NL, Culverhouse RC, Schwantes-An T-H, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. PLOS Genetics. 2010;6:e1001053.
52. Liu CY, Stucker I, Chen C, Goodman G, McHugh MK, D'Amelio AM Jr, Etzel CJ, Li S, Lin X, Christiani DC. Genome-wide gene-asbestos exposure interaction association study identifies a common susceptibility variant on 22q13.31 associated with lung cancer risk. Cancer Epidemiol Biomarkers Prev. 2015;24:1564–73.
53. Wei S, Wang L-E, McHugh MK, Han Y, Xiong M, Amos CI, Spitz MR, Wei QW. Genome-wide gene-environment interaction analysis for asbestos exposure in lung cancer susceptibility. Carcinogenesis. 2012;33:1531–7.
54. Rosenberger A, Hung RJ, Christiani DC, et al. Genetic modifiers of radon-induced lung cancer risk: a genome-wide interaction study in former uranium miners. Int Arch Occup Environ Health. 2018. https://doi.org/10.1007/s00420-018-1334-3.
55. Lorenzo-González M, Ruano-Ravina A, Torres-Durán M, et al. Residential radon, genetic polymorphisms in DNA damage and repair-related. Lung Cancer. 2019;135:10–5.
56. Hung RJ, Spitz MR, Houlston RS, et al. Lung cancer risk in never-smokers of european descent is associated with genetic variation in the 5p15.33 TERT-CLPTM1Ll region. J Thorac Oncol. 2019;14:1360–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC