Supplement

Content

[Material and Methods 2](#_Toc514150379)

[Scaling residential and occupational radon exposure 2](#_Toc514150380)

[Study population 3](#_Toc514150381)

[Study population 4](#_Toc514150382)

[Genotyping and QC 5](#_Toc514150383)

[Merging samples 5](#_Toc514150384)

[Definition of LD blocks 5](#_Toc514150385)

[Statistical analysis 6](#_Toc514150386)

[Single-marker interaction analysis 6](#_Toc514150387)

[Multi-marker interaction analysis 8](#_Toc514150388)

[Gene-set analysis 8](#_Toc514150389)

[Selected gene sets 9](#_Toc514150390)

[Genes and references for ii) due to published genetic radon interaction with respect to lung cancer 9](#_Toc514150391)

[Genes and references for iii) pathways associated with radiation are associated with the progression of lung caner 9](#_Toc514150392)

[Results 13](#_Toc514150393)

[Single marker interaction analysis 13](#_Toc514150394)

[Multi-marker interaction analysis 14](#_Toc514150395)

[Discussion 15](#_Toc514150396)

[Discussion of the GxE interaction at 10p13 15](#_Toc514150397)

[Discussion of the GxE interaction at 12p12.1 15](#_Toc514150398)

[References 17](#_Toc514150399)

[Supplementary Table I Characteristics of the source samples 4](#_Toc514149679)

[Supplementary Table II Occupational radon exposure in working level months 5](#_Toc514149680)

[Supplementary Table III regions of long-distance LD, SNPS correlated to PCs or known association with lung cancer 7](#_Toc514149681)

[Supplementary Table IV investigated gene sets 11](#_Toc514149682)

Supplementary Figure 1 Manhattan-Plot displaying significance of GxE interaction for each marker 13

Supplementary Figure 2: Significance of selected markers corrected with the H2-method across several ρ-values 13

# Material and Methods

The participating studies of TRICL/ILCCO are individually described in the supplement of McKay et al. (2017), Table 1 (main document) and Supplementary Table I. The LC cases of the BfS sample collection were recruited for a study investigating indoor-radon exposure between 1990 und 1997 (Brüske-Hohlfeld et al. 2006). The cancer-free BfS controls are former uranium miners recruited from 2009-2012, who continuously participated in health surveillance program of the German Social Accident Insurance and are long term survivors (Pesch et al. 2015). This control samples stored in the BfS Bio- and Databank (GUMB) were drawn from these miners, which were either very high (>750 Working Level Months, WLM) or low (≤50 WLM) radiation exposed in a targeted and no-representative ratio of 2:1 (Pesch et al. 2015). The method how radon exposures was measured is given elsewhere (Kreuzer et al. 2010b) (see Scaling residential and occupational radon exposure).

## Scaling residential and occupational radon exposure

To measure residential and occupational radon exposure usually different scales are used. The scale working level moth (WLM) was introduced in the 1950s for risk assessment of occupational radon exposure instead of a calculated dose in Millisievert (mSv). To calculate the cumulative exposure to radon in WLM, the measured alpha energy concentration (unit: Working Level (WL)) in one litre air is multiplied by the time the miner has worked in this surrounding. 1 WLM equals an exposure of 1 WL (1.3 \* 105 Megaelectron-volt (MeV) potential alpha energy per litre air) over 170 working hours (monthly working time), or a half WL over 2 months (340 working hours), respectively. The following conversion can be used:(Hauptmann et al. 2003)

Assuming a constant exposure and an exposure period of 30 years covering the biologically relevant time:

Mean indoor radon concentration can be categorised in 0-50, 50-80, 80-140, 140-250, >250 , sometimes also >400 .(Darby et al. 2005; Wichmann et al. 1998) The average radon exposure in Germany of over an assumed period of 30 years between cancer initiation and diagnosis can be equated to . A very high mean exposure of over a period of 30 years can be equated to . In contrast, the lowest observed exposure in “exposed miners” was 200 WLM and equivalent to a mean indoor radon concentration of over 30 years. The mean occupational exposure of can be equated to . Both values are far above observed indoor exposure levels. Hence, a misclassification of any case or control from ILCCO/TRICL-studies in the general population by assuming WLM<50 is pretty unlikely.

## Study population

The analysed sample consisted of 28 599 study participants with European ancestry and valid information on age at diagnosis/interview, sex and smoking status (15 077 cases : 13 522 controls); 463 thereof are former uranium miners of the Wismut mining company (61 cases : 402 controls), 949 are from the German Lung Cancer Study (471 cases : 478 controls), the remaining are from 25 studies of TRICL/ILCCO (14 545 cases : 12 642 controls) (see Table 1 (main document) and Supplementary Table I). 49 of 15 077 (0.3%) LC cases and 259 of 13 522 cancer-free controls (1.9%) had been occupationally exposed by a high cumulative dose exposure to radon and its progeny, external gamma radiation and long-lived radionuclides (LRN) (WLM>50) (see Supplementary Table II). The exposure to occupational radiation of the uranium miners was estimated retrospectively using a comprehensive job-exposure matrix (JEM). For each work place and each type of job the JEM provided annual values of the exposures to radon and its progeny (WLM) (Kreuzer et al. 2011). A working level (WL) is defined as 1.3x105 MeV (million electron volts) of alpha energy/l air which will be emitted by short lived radon progeny. Thus WLM quantifies only the exposure to radon, but neither gamma radiation, nor LRN or dust. One WLM equals exposure to 1 WL for 170 hours. It is unlikely that a cumulative lifetime exposure solely due to an exposure by indoor or other environmental radon sums up to more than 50 WLM. Using the conversion factors from Hunter et al. (2013), a residential cumulative radon exposure of 425 Bq/m3 corresponds to 50 WLM. Thus, we categorised occupational radon exposure into ≤50 (“unexposed”) and >50 WLM (“exposed”), a threshold for significant elevated relative LC-risk (Kreuzer et al. 2010a). All TRICL/ILCCO participants were assigned to the exposure categories ≤50 WLM. Misclassification would be conservative. The general population represents a suitable reference group of less exposed persons.

## Study population

Supplementary Table I Characteristics of the source samples

|  |  | **Lung cancer** | | **Age\*** | **Sex** | | **Smoking** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | n | Controls | Cases | Median | Males | Females | Never smokers | Former smokers | Current smokers | Ever smokers |
| *Total* | 28,599 | 13,522 | 15,077 | 28,599 | 18,059 | 10,540 | 5,676 | 9,518 | 12,039 | 1,366 |
|  |  | 47% | 53% | 63 | 63% | 37% | 20% | 33% | 42% | 5% |
| *Source study* |  |  |  |  |  |  |  |  |  |  |
| *Indoor-Radon* | 58 |  | 58 | 67 | 58 |  | 1 |  |  | 57 |
| *WISMUT* | 405 | 402 | 3 | 77 | 405 |  | 133 | 22 | 246 | 4 |
| *GLC-550K* | 949 | 478 | 471 | 46 | 536 | 413 | 256 | 217 | 476 |  |
| *OncoArray-C$* |  |  |  |  |  |  |  |  |  |  |
| *ATBC* | 1,683 | 666 | 1017 | 59 | 1,683 |  |  |  | 1,683 |  |
| *CANADA* | 656 | 442 | 214 | 65 | 283 | 373 | 3 | 279 | 374 |  |
| *CAPUA* | 1,399 | 684 | 715 | 68 | 1,227 | 172 | 240 | 584 | 571 | 4 |
| *COPENHAGEN* | 1,823 | 1,341 | 482 | 64 | 804 | 1019 | 488 | 114 | 195 |  |
| *EAGLE* | 3,494 | 1,702 | 1792 | 67 | 2,744 | 750 | 659 | 1,326 | 1,509 |  |
| *CARET* | 1,065 | 519 | 546 | 60 | 712 | 353 |  | 209 | 856 |  |
| *LLP-2008* | 200 | 101 | 99 | 69 | 118 | 82 | 35 | 106 | 59 |  |
| *LLP-2013* | 675 | 355 | 320 | 67 | 376 | 299 | 251 | 316 | 107 | 1 |
| *GLC* | 1,014 | 221 | 793 | 47 | 557 | 457 | 129 | 170 | 687 | 28 |
| *HLCS* | 1,605 | 512 | 1093 | 64 | 763 | 842 | 393 | 817 | 395 |  |
| *NICCC-LCA* | 1,149 | 508 | 641 | 68 | 725 | 424 | 380 | 391 | 378 |  |
| *LCRI-DOD* | 220 | 128 | 92 | 63 | 105 | 115 | 63 | 71 | 85 | 1 |
| *MDCS* | 325 | 167 | 158 | 62 | 143 | 182 | 83 | 101 | 141 |  |
| *MEC* | 430 | 217 | 213 | 73 | 229 | 201 | 123 | 190 | 117 |  |
| *NELCS* | 329 | 169 | 160 | 62 | 145 | 184 | 83 | 142 | 104 |  |
| *NIJMEGEN* | 816 | 442 | 374 | 61 | 501 | 315 | 118 | 366 | 332 | 40 |
| *NORWAY* | 725 | 416 | 309 | 62 | 502 | 223 | 19 | 96 | 194 |  |
| *NSHDC* | 473 | 236 | 237 | 60 | 238 | 235 | 55 | 132 | 286 |  |
| *PLCO* | 2,231 | 885 | 1346 | 68 | 1,363 | 868 | 201 | 974 | 1,056 | 416 |
| *RESOLUCENT* | 750 | 258 | 492 | 56 | 357 | 393 | 132 | 201 | 409 |  |
| *L2* | 2,009 | 1,025 | 984 | 61 | 1352 | 657 | 604 | 415 | 990 |  |
| *TAMPA* | 242 | 144 | 98 | 65 | 163 | 79 | 54 |  |  | 188 |
| *TLC* | 419 |  | 419 | 66 | 197 | 222 | 28 | 252 | 139 |  |
| *MSH-PMH* | 2,295 | 946 | 1,349 | 64 | 1,152 | 1,143 | 604 | 1,001 | 650 | 40 |
| *VANDERBILT* | 1,160 | 558 | 602 | 66 | 621 | 539 | 541 |  |  | 619 |

\* age at diagnosis/interview; ***$*** OncoArray-consortium

CARET: The Carotene and Retinol Efficacy Trial; BioVU: The Vanderbilt Lung Cancer Study; HLCS: Harvard Lung Cancer Study; ATBC: The Alpha-Tocopherol, Beta-Carotene Cancer Prevention; PLCO: The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; MSH-PMH: Mount Sinai Hospital-Princess Margaret Hospital Study; HLCS: Harvard Lung Cancer Study; LCRI-DOD: Study of Lung Cancer in Appalachian Kentucky; Tampa: Tampa Lung Cancer Study; NELCS: New England Lung Cancer Study; TLC: Total Lung Cancer: Molecular Epidemiology of Lung Cancer Survival; MEC: Multiethnic Cohort Study; Canada: Pan-Canadian screening study; EAGLE: Environment and Genetics in Lung Cancer Study Etiology; Copenhagen: Copenhagen lung cancer study; CAPUA: Cancer de Pulmon en Asturias; GLC: German lung cancer study; GLC-500K: German lung cancer study; Nijmegen: The Nijmegen Lung Cancer Study; ReSoLucent: Resource for the Study of Lung Cancer Epidemiology in North Trent; Norway: Norway Lung Cancer Study; LLP-2008,; LLP-2013: Liverpool Lung Cancer Project; NSHDC: Northern Sweden Health and Disease Cohort; Wismut: Former uranium miners recruited from the medical follow-up care; MDCS: The Malmö Diet and Cancer Study; Indoor-Radon: Case-control study on lung cancer among Wismut miners in Germany; NICCC-LCA: Clalit National Israeli Cancer Control Center- lung cancer study; L2: the IARC L2 Study

Supplementary Table II Occupational radon exposure in working level months

|  | Working Level Months | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| n | Min | Max | Median | Mean | s |
| *Total* | 28,599 | 0 | 2,479 | 0 | 11 | 110.62 |
| *Exposed (WLM≥50)* | 308 | 51 | 2,479 | 966 | 986 | 419.39 |
| *Not exposed (WLM<50)* | 28,291 | 0 | 46 | 0 | 0 | 1.49 |
| *thereof WISMUT* | 155 | 0 | 46 | 14 | 16 | 12.44 |

S standard deviation

## Genotyping and QC

Genotyping with the OncoArray was completed at the Centre for Inherited Disease Research in Baltimore, Copenhagen University Hospital, and the University of Cambridge and the Helmholtz Center Munich. Quality control (QC) steps followed the approach previously described for the OncoArray (Amos et al. 2017). QC comprised checks for missing genotypes, Hardy-Weinberg equilibrium in controls, genomic sex, inbreeding and heterozygosity, genomic relationship, batch effects and population stratification. To validate the European ancestry of the participants we applied the program ADMIXTURE (Alexander et al. 2009) to estimate the probability of being Caucasian based on a set of 159 ancestry-informative markers and PCA-informative markers covering European fine-structures, in particular tagSNPs thereof in close linkage disequilibrium (LD) (Huckins et al. 2014; Kosoy et al. 2009; Setsirichok et al. 2012).

456 699 markers passed QC in the TRICL/ILCCO and BfS samples. 792 markers were monomorph. 48 790 markers had minor allele frequencies (MAF) <1%. After excluding those, 407 117 markers entered the analysis. These could be clustered into 103 983 blocks (67 161 LD blocks and 36 822 hot spots; for definition see Online Resource 1).

## Merging samples

The ratio of cases to controls within the 308 “exposed” participants of the BfS sample collection was 49 : 259, and within the 155 “unexposed” participants 12 : 143. Hence, the crude odds ratio (OR) for the occupational radon exposure would be OR=2.25. After naïvely adding the TRICL/ILCCO participants, with a case : control ratio of 15 016 : 23 121, this estimate would be biased to OR=0.17. This shift in the ratios of exposed : unexposed is due to adding solely unexposed cases and controls. We down-weighted the cases of TRICL/ILCCO by the factor, to avoid this unjustified inversion, and still use all available information for analysis. Thus, we have fixed the marginal risk of a radon exposure at the point estimate from the BfS sample collection.

## Definition of LD blocks

LD-blocks are defined based either on the estimation of haplotypes or on the estimation of the LD between markers.(Barrett et al. 2005; Wall and Pritchard 2003)

Haplotype-blocks were determined according to the routine implemented in PLINK (Purcell et al. 2007) (identical to the presetting of Haploview (Barrett et al. 2005)): the distance between any two markers of the same block need to be less than 500 kb; markers need at least 50% available genotypes and a minimal minor allele frequency MAF of 5%. A haplotype-block contains at least 95% pairs of markers with "strong LD" according to the 95% confidence intervals for D '. (Gabriel et al. 2002)

LD-blocks are further defined from all adjacent pairs of markers with r²≥0.025.

The blocks used for analysis are a result from the combination of both definitions (haplotype-blocks and LD-blocks).

## Statistical analysis

We fitted two models to individual data and also carried out a gene-set analysis (GSA) to search for accumulated significance in pre-defined groups of genes for pathways and gene families of interest. All calculations, data handling and image acquire were performed using PLINK 1.9 (Purcell et al. 2007) and SAS 9.4 of the SAS Institute Inc., Cary. NC. USA.

### Single-marker interaction analysis

We first performed single-marker interaction analysis fitting the log-additive model:

[1]

where D is the disease status (D=1: LC patient; D=0: control); G is minor allele count at marker m; E is the exposure category (0: ≤50 WLM, 1: WLM>50); PS is a propensity score comprising the probability being a case explained by age, sex and smoking. To adjust for genomic population stratification we calculated the principal components (PC) of genotypes of a set of 26 600 uncorrelated, randomly selected markers applying SMARTPCA und EIGENSTRAT (Price et al. 2006). Markers in „long-range LD“-regions (Price et al. 2008), in known „susceptibility genes“ for LC or novel identified susceptibility loci were excluded from being selected (see Supplementary Table III) (McKay et al. 2017; Truong et al. 2010b). Only the first four PCs were included in the statistical modelling, because the fifth PC was significantly correlated with the disease status. The remaining inflation factor (median of the χ²-distribution for unadjusted association) was λ~1.1, which is acceptable close to 1.0 (Yang et al. 2011).

The propensity-score (PS) method was applied to adjust for sex, age and smoking in a robust way, to adjust for background case probability in a single quantity, and to prune cases or controls with no comparable counterpart (Arbogast and Ray 2011).

The data at hand are not a representative data set of a well-defined source population. Thus the effect estimate of interaction, expressed as odds ratio, is potentially proportionally biased. Therefore, the tilde is added to indicate that a weighted sample was used for estimation (see Merging samples). However, estimating is not our main interest, rather than testing the null hypothesis , which is still valid (Mukherjee et al. 2008; Stenzel et al. 2015).

With as global level of significance, we use as Bonferroni-corrected, genome-wide level of significance, adjusted for the number of independent LD-blocks tested. A suggestive level of significance was set to 1. This corresponds to a p-value corrected for multiple testing of . Significance was determined according the Hybrid 2-step (H2) method of Murcray et al. (2011). All markers were first inspected for marginal disease-gene (DxG) or environmental-gene (ExG) effects (i.e. association between outcome G with D or E, respectively, as the explained variable), and grouped into four classes: (a) DxG effect only, (b) ExG effect only, (c) both or (d) none. The significance threshold for group (a) is , with the number of markers in group (a). The significance threshold for group (b) is , with the number of markers in group (b). The significance threshold for group (c) is . The significance of markers in group (d) is not determined. is a tuning parameter and can take any value between 0 and 1. Equal weights for the DxG and ExG screening are given when .

We choose markers for group (a) according the TRICL/ILCCO analysis of McKay et al. (2017), and markers for group (b) by fitting ExG models with the data at hand. Values for between 0.5 to 1-10-20 were applied.

Supplementary Table III regions of long-distance LD, SNPS correlated to PCs or known association with lung cancer

| Chromosome | Position from | to | LD or correlation | Association with lung cancer |
| --- | --- | --- | --- | --- |
| 1 | 78.300 | 78.700 |  | 1p31 |
| 1 | 8.500.000 | 9.000.000 | PC-SNP correlation |  |
| 1 | 42.000.000 | 52.000.000 | long-distance LD |  |
| 1 | 110.000.000 | 120.000.000 | long-distance LD |  |
| 1 | 182.000.000 | 195.000.000 | long-distance LD |  |
| 1 | 200.000.000 | 240.000.000 | PC-SNP correlation |  |
| 2 | 8.000.000 | 8.500.000 | PC-SNP correlation |  |
| 2 | 69.000.000 | 69.500.000 | PC-SNP correlation |  |
| 2 | 86.000.000 | 100.500.000 | long-distance LD |  |
| 2 | 111.500.000 | 143.000.000 | long-distance LD |  |
| 2 | 160.000.000 | 230.000.000 | long-distance LD |  |
| 3 | 4.000.000 | 4.500.000 | PC-SNP correlation |  |
| 3 | 21.000.000 | 25.500.000 | long-distance LD |  |
| 3 | 47.500.000 | 50.000.000 | long-distance LD |  |
| 3 | 58.000.000 | 68.000.000 | long-distance LD |  |
| 3 | 75.500.000 | 76.550.000 | long-distance LD |  |
| 3 | 83.500.000 | 87.000.000 | long-distance LD |  |
| 3 | 89.000.000 | 97.500.000 | long-distance LD |  |
| 3 | 108.000.000 | 140.000.000 | PC-SNP correlation |  |
| 3 | 189.200.000 | 189.400.000 |  | 3q28 |
| 4 | 9.600.000 | 9.800.000 |  | 4p16 |
| 4 | 20.000.000 | 26.000.000 | PC-SNP correlation |  |
| 4 | 75.000.000 | 123.000.000 | PC-SNP correlation |  |
| 5 | 1.200.000 | 6.000.000 |  | 5p15\_TERT |
| 5 | 1.200.000 | 6.000.000 |  | 5p15 |
| 5 | 41.000.000 | 52.500.000 | long-distance LD |  |
| 5 | 71.000.000 | 100.500.000 | long-distance LD |  |
| 5 | 129.000.000 | 132.000.000 | long-distance LD |  |
| 5 | 135.500.000 | 138.500.000 | long-distance LD |  |
| 6 | 14.000.000 | 20.000.000 | PC-SNP correlation |  |
| 6 | 25.500.000 | 33.500.000 | long-distance LD |  |
| 6 | 31.200.000 | 31.600.000 |  | 6p21 |
| 6 | 31.600.000 | 46.000.000 |  | 6p21\_BAG6 |
| 6 | 57.000.000 | 86.000.000 | long-distance LD |  |
| 6 | 106.000.000 | 118.000.000 | long-distance LD |  |
| 6 | 138.000.000 | 139.000.000 | long-distance LD |  |
| 6 | 139.000.000 | 142.500.000 | long-distance LD |  |
| 6 | 167.200.000 | 167.600.000 |  | 6q27 |
| 7 | 50.000.000 | 72.000.000 | long-distance LD |  |
| 7 | 111.000.000 | 140.000.000 | PC-SNP correlation |  |
| 8 | 8.000.000 | 12.000.000 | long-distance LD |  |
| 8 | 27.000.000 | 30.000.000 | long-distance LD |  |
| 8 | 27.200.000 | 27.600.000 |  | 8p21 |
| 8 | 32.200.000 | 32.600.000 |  | 8p12 |
| 8 | 43.000.000 | 60.000.000 | long-distance LD |  |
| 8 | 112.000.000 | 115.000.000 | long-distance LD |  |
| 8 | 94.000.000 | 95.000.000 | PC-SNP correlation |  |
| 9 | 20.000.000 | 22.200.000 |  | 9p21 |
| 9 | 77.500.000 | 125.000.000 |  | 9q31 |
| 10 | 2.000.000 | 9.000.000 | long-distance LD |  |
| 10 | 37.000.000 | 43.000.000 | long-distance LD |  |
| 10 | 90.000.000 | 107.000.000 |  | 10q24 |
| 11 | 7.000.000 | 59.000.000 | long-distance LD |  |
| 11 | 57.200.000 | 57.400.000 |  | 11q12 |
| 11 | 87.500.000 | 90.500.000 | long-distance LD |  |
| 11 | 118.000.000 | 118.200.000 |  | 11q23 |
| 12 | 800.000 | 1.200.000 |  | 12p13 |
| 12 | 23.000.000 | 58.000.000 | long-distance LD |  |
| 12 | 109.500.000 | 128.000.000 | long-distance LD |  |
| 13 | 32.800.000 | 50.000.000 |  | 13q13 |
| 14 | 28.000.000 | 70.000.000 | PC-SNP correlation |  |
| 15 | 47.400.000 | 47.600.000 |  | 15q21 |
| 15 | 49.200.000 | 51.000.000 |  | 15q21 |
| 15 | 78.600.000 | 79.000.000 |  | 15q25\_CHRNA3 |
| 16 | 12.000.000 | 23.000.000 | PC-SNP correlation |  |
| 17 | 46.000.000 | 54.000.000 | PC-SNP correlation |  |
| 18 | 4.000.000 | 5.000.000 | PC-SNP correlation |  |
| 19 | 41.200.000 | 41.400.000 |  | 19q13 |
| 20 | 18.000.000 | 34.500.000 | long-distance LD |  |
| 20 | 59.000.000 | 62.400.000 |  | 20q13 |
| 21 | 19.000.000 | 27.000.000 | PC-SNP correlation |  |
| 22 | 29.000.000 | 29.200.000 |  | 22q12 |

Region with known association with LD are defined according to Amos et al. (2008), Brennan et al. (2011), Fehringer et al. (2012), Hung et al. (2008), Timofeeva et al. (2012), Truong et al. (2010a), Wang et al. (2008) and Wang et al. (2014).

### Multi-marker interaction analysis

We also searched for the best fitting model of each LD block, allowing all markers of a block to enter the model (denoted as complete model). We then applied a backward selection with the best model chosen according to Akaike’s information criteria (AIC), requiring at least one interaction with a marker (denoted as AIC-best model).

### Gene-set analysis

For GSA we assigned markers to genes according to ENSEMBL (Cunningham et al. 2014), and genes to gene sets according to Gene Ontology (GO) and the Human Genome Nomenclature Committee (HGNC) (Ashburner et al. 2000; Gray et al. 2015). In addition, the gene set of homeobox (HOX) genes in regulatory networks with respect to LC was defined based on literature (Bhatlekar et al. 2014). In total 119 gene sets were considered for analysis. These were selected either i) due to findings of the previous approaches, ii) due to published genetic radon interaction with respect to LC or iii) because genes of pathways associated with radiation are associated with the progression of LC. More detail are given in chapter Selected gene sets. The gene sets were thinned out to assigned and genotyped markers. Gene sets with genotyped markers in less than 5 genes were excluded. Due to the subjective and in parts data driven selection of gene sets, the GSA was performed as explorative data analysis. The global level of significance of α=0.05 was used. A list of all investigated gene sets, along with literature references, is given in Supplementary Table IV.

We applied Gene-Set Enrichment Analyse (GSEA), based on the p-values obtained from the multi-marker interaction analysis (Subramanian et al. 2005). This method highlights gene sets with an accumulation of low p-values (per LD block) in comparison to all genes not included in the gene set of interest. Genes most responsible for such an accumulation are characterized as “significance driving genes”.

## Selected gene sets

In total 119 gene sets were considered for analysis. These were selected either

1. due to findings of the previous approaches,
2. due to published genetic radon interaction with respect to lung cancer   
   or
3. because genes of pathways associated with radiation are associated with the progression of lung cancer.

### Genes and references for ii) due to published genetic radon interaction with respect to lung cancer

SIRT1 Leng et al. (2013)

EPHX1, GSTM1, GSTT1 Ruano-Ravina et al. (2014); Bonner et al. (2006)

p53 (TP53) Vahakangas et al. (1992); Taylor et al. (1994); Yngveson et al. (1999)

CDKN2A, MGMT Su et al. (2006)

IL6 Leng et al. (2016)

### Genes and references for iii) pathways associated with radiation are associated with the progression of lung caner

DNA repair Brambilla (2009); Hornhardt et al. (2014)

* [GO:0036473 cell death in response to oxidative stress](http://amigo.geneontology.org/amigo/term/GO:0036473#display-lineage-tab)
* [GO:0070265 necrotic cell death](http://amigo.geneontology.org/amigo/term/GO:0070265#display-lineage-tab)
* [GO:0006915 apoptotic process](http://amigo.geneontology.org/amigo/term/GO:0006915#display-lineage-tab)
* [GO:0097468 programmed cell death in response to reactive oxygen species](http://amigo.geneontology.org/amigo/term/GO:0097468#display-lineage-tab)
* [GO:0097300 programmed necrotic cell death](http://amigo.geneontology.org/amigo/term/GO:0097300#display-lineage-tab)
* GO:0006281 DNA repair
* GO:0007165 signal transduction

*Epigenetic modifications with respect to let-7* Brambilla and Gazdar (2009)  
 Takamizawa et al. (2004)

* HGNC gene-family „*MicroRNAs (MIR)*“

Abnormalities in „growth-stimulatory signalling“ pathways Brambilla and Gazdar (2009)

„Epidermal growth factor receptor signalling“

* GO:0000165 MAPK cascade Ding et al. (2008)
* GO:0038127 ERBB signalling pathway
* [GO:0007173 epidermal growth factor receptor signalling pathway](http://amigo.geneontology.org/amigo/term/GO:0007173#display-lineage-tab)
* [GO:0038128 ERBB2 signalling pathway](http://amigo.geneontology.org/amigo/term/GO:0038128#display-lineage-tab)
* [GO:0038129 ERBB3 signalling pathway](http://amigo.geneontology.org/amigo/term/GO:0038129#display-lineage-tab)
* [GO:0038130 ERBB4 signalling pathway](http://amigo.geneontology.org/amigo/term/GO:0038130#display-lineage-tab)
* [GO:1901185 negative regulation of ERBB signalling pathway](http://amigo.geneontology.org/amigo/term/GO:1901185#display-lineage-tab)
* [GO:1901186 positive regulation of ERBB signalling pathway](http://amigo.geneontology.org/amigo/term/GO:1901186#display-lineage-tab)
* [GO:1901184 regulation of ERBB signalling pathway](http://amigo.geneontology.org/amigo/term/GO:1901184#display-lineage-tab)

„Ras/mitogen-activated protein kinase” and PI3K/Akt pathways

* GO:0038201 TOR complex

„Anaplastic lymphoma kinase fusion“-proteins

* GO:0007169 transmembrane receptor protein tyrosine kinase signalling pathway

Thyroid transcription factor 1 (NKX2-1 ; alternative name TITF1) Bhatlekar et al. (2014)

* HGNC gene-families 518 to 530, but not 520

Abnormalities in tumour suppressor gene pathways Brambilla and Gazdar (2009)

p16INK4/cyclin D1/Rb- pathway

* GO:0000083 regulation of transcription involved in G1/S transition of mitotic   
  cell cycle

Evasion of apoptosis Brambilla and Gazdar (2009)

Mitochondrial apoptosis (Bax/Bcl-2)

* GO:0097345 mitochondrial outer membrane permeabilization

*Death Receptor*“-deregulation

* GO:0036337 Fas signalling pathway
* GO:0097190 apoptotic signalling pathway

Cell immortalisation and telomerase activation Brambilla and Gazdar (2009)  
 Jafri et al. (2016)

* GO: 0003720 telomerase activity
* Wnt/β-catenin pathway Ding et al. (2008); Hubaux et al. (2012)
* GO:0060070 canonical Wnt signalling pathway
* [GO:1904886 beta-catenin destruction complex disassembly](http://amigo.geneontology.org/amigo/term/GO:1904886#display-lineage-tab) (part of GO:0060070)
* GO:0060071 Wnt signalling pathway, planar cell polarity pathway
* GO:0007223 Wnt signalling pathway, calcium modulating pathway

Supplementary Table IV investigated gene sets

| Gene set | Exclusion | Description | p-value |
| --- | --- | --- | --- |
| GO:0000012 |  | single strand break repair | 0.9204 |
| GO:0000083 |  | regulation of transcription involved in G1/S transition of mitotic cell cycle | 0.9811 |
| GO:0000165 |  | MAPK cascade | 1.0000 |
| GO:0000209 |  | protein polyubiquitination | 0.5889 |
| GO:0000725 | n≤5 genes | recombinational repair |  |
| GO:0000726 |  | non-recombinational repair | 0.6126 |
| GO:0000731 |  | DNA synthesis involved in DNA repair | 0.4323 |
| GO:0000790 |  | nuclear chromatin | 1.0000 |
| GO:0001894 |  | tissue homeostasis | 0.2649 |
| GO:0003677 |  | DNA binding | 0.7061 |
| GO:0003700 |  | transcription factor activity, sequence-spec | 0.5544 |
| GO:0003720 |  | telomerase activity | 0.9741 |
| GO:0003996 |  | acyl-CoA ligase activity | 1.0000 |
| GO:0004321 |  | fatty-acyl-CoA synthase activity | 1.0000 |
| GO:0004672 |  | protein kinase activity | 1.0000 |
| GO:0004674 |  | protein serine/threonine | 0.6434 |
| GO:0004872 |  | receptor activity | 0.6898 |
| GO:0005044 |  | scavenger receptor activity | 0.1295 |
| GO:0005215 |  | transporter activity | 0.9522 |
| GO:0005509 |  | calcium ion binding | 0.9831 |
| GO:0005524 |  | ATP binding | 0.2849 |
| GO:0005576 |  | extracellular region | 0.5145 |
| GO:0005737 |  | cytoplasm | 0.4482 |
| GO:0005759 |  | mitochondrial matrix | 0.6120 |
| GO:0005765 |  | lysosomal membrane | 0.9082 |
| GO:0005783 |  | endoplasmic reticulum | 0.5100 |
| GO:0005794 |  | Golgi apparatus | 0.6474 |
| GO:0005905 |  | clathrin-coated pit | 0.0777 |
| GO:0006281 |  | DNA repair | 1.0000 |
| GO:0006282 |  | regulation of DNA repair | 1.0000 |
| GO:0006284 |  | base-excision repair | 0.9087 |
| GO:0006289 |  | nucleotide-excision repair | 0.7461 |
| GO:0006290 | n≤5 genes | pyrimidine dimer repair |  |
| GO:0006298 |  | mismatch repair | 0.9314 |
| GO:0006301 |  | Post-replication repair | 0.6675 |
| GO:0006302 |  | double-strand break repair | 0.8340 |
| [GO:0006303](http://amigo.geneontology.org/amigo/term/GO:0006303#display-lineage-tab) |  | double-strand break repair via non-homologous end joining | 0.7170 |
| GO:0006307 |  | DNA dealkylation involved in DNA repair | 0.0139 |
| GO:0006355 |  | regulation of transcription, DNA-templated | 0.9044 |
| GO:0006366 |  | transcription from RNA polymerase II promote | 0.5979 |
| GO:0006464 |  | cellular protein modification process | 0.7365 |
| GO:0006633 |  | fatty acid biosynthetic process | 0.8846 |
| GO:0006637 |  | acyl-CoA metabolic process | 0.0538 |
| GO:0006897 |  | endocytosis | 0.6534 |
| GO:0006898 |  | receptor-mediated endocytosis | 1.0000 |
| GO:0006915 |  | apoptotic process | 0.9821 |
| GO:0007165 |  | signal transduction | 0.4701 |
| GO:0007169 |  | transmembrane receptor protein tyrosine kinase signalling pathway | 0.8028 |
| GO:0007173 |  | epidermal growth factor receptor signalling pathway | 0.2550 |
| GO:0007223 |  | Wnt signalling pathway, calcium modulating pathway | 0.8738 |
| GO:0008203 |  | cholesterol metabolic pro | 0.6295 |
| GO:0008360 |  | regulation of cell shape | 0.5113 |
| GO:0009235 |  | cobalamin metabolic process | 0.6574 |
| GO:0009380 |  | excinuclease repair complex | 0.3406 |
| GO:0010008 |  | endosome membrane | 0.7332 |
| GO:0010213 | =GO: 0009380 | non-photoreactive DNA repair |  |
| GO:0015031 |  | protein transport | 0.6618 |
| GO:0015645 |  | fatty acid ligase activity | 1.0000 |
| GO:0015889 |  | cobalamin transport | 0.8486 |
| GO:0016020 |  | membrane | 0.0558 |
| GO:0016055 |  | Wnt signalling pathway | 1.0000 |
| GO:0016324 |  | apical plasma membrane | 1.0000 |
| GO:0016574 |  | histone ubiquitination | 0.1434 |
| GO:0018105 |  | peptidyl-serine phosphorylation | 0.8765 |
| GO:0030139 |  | endocytic vesicle | 0.7497 |
| GO:0031232 |  | extrinsic component of ex | 0.0817 |
| GO:0031419 |  | cobalamin binding | 0.6474 |
| GO:0031526 |  | brush border membrane | 0.5951 |
| GO:0031625 |  | ubiquitin protein ligase binding | 0.9470 |
| GO:0032332 |  | positive regulation of chondrocyte different | 0.5697 |
| GO:0033503 | n≤5 genes | HULC complex |  |
| GO:0036297 |  | interstrand cross-link repair | 0.8568 |
| [GO:0036299](http://amigo.geneontology.org/amigo/term/GO:0036299#display-lineage-tab) | =GO: 0009380 | non-recombinational interstrand cross-link repair |  |
| GO:0036337 | n≤5 genes | Fas signalling pathway |  |
| GO:0036473 | n≤5 genes | cell death in response to oxidative stress |  |
| GO:0038127 | n≤5 genes | ERBB signalling pathway |  |
| GO:0038128 |  | ERBB2 signalling pathway | 0.9582 |
| GO:0038129 |  | ERBB3 signalling pathway | 0.4044 |
| GO:0038130 | =GO: 0009380 | ERBB4 signalling pathway |  |
| GO:0038201 |  | TOR complex | 0.9064 |
| GO:0042157 |  | lipoprotein metabolic pro | 0.6733 |
| GO:0042359 |  | vitamin D metabolic process | 0.7351 |
| GO:0042803 |  | protein homodimerization | 0.8167 |
| GO:0042953 |  | lipoprotein transport | 0.8267 |
| GO:0043161 |  | proteasome-mediated ubiquitin-dependent protein catabolic process | 0.8163 |
| GO:0043202 |  | lysosomal lumen | 0.9265 |
| GO:0043504 | n≤5 genes | mitochondrial DNA repair |  |
| [GO:0045002](http://amigo.geneontology.org/amigo/term/GO:0045002#display-lineage-tab) |  | double-strand break repair via single-strand annealing | 0.1574 |
| GO:0045004 | n≤5 genes | DNA replication proofreading |  |
| GO:0045738 | n≤5 genes | negative regulation of DNA repair |  |
| GO:0045739 |  | positive regulation of DNA repair | 0.9980 |
| GO:0046787 | =GO: 0009380 | viral DNA repair |  |
| GO:0046872 |  | metal ion binding | 0.9658 |
| GO:0047760 |  | butyrate-CoA ligase activity | 1.0000 |
| GO:0051103 |  | DNA ligation involved in DNA repair | 0.9975 |
| GO:0055059 | n≤5 genes | asymmetric neuroblast division |  |
| GO:0060070 |  | canonical Wnt signalling pathway | 0.9689 |
| GO:0060071 |  | Wnt signalling pathway, planar cell polarity pathway | 0.9841 |
| GO:0061036 |  | positive regulation of cartilage development | 0.5718 |
| GO:0061630 |  | ubiquitin protein ligase activity | 0.8098 |
| GO:0070062 |  | extracellular exosome | 0.7867 |
| GO:0070265 |  | necrotic cell death | 0.5159 |
| GO:0070914 |  | UV-damage excision repair | 0.9709 |
| GO:0071560 |  | cellular response to transforming growth factor | 0.7550 |
| GO:0072331 | n≤5 genes | signal transduction by p53 class mediator |  |
| GO:0097190 |  | apoptotic signalling pathway | 0.6651 |
| GO:0097196 | n≤5 genes | Shu complex |  |
| GO:0097300 | n≤5 genes | programmed necrotic cell death |  |
| GO:0097345 |  | mitochondrial outer membrane permeabilization | 0.5416 |
| GO:0097468 |  | programmed cell death in response to reactive oxygen species | 0.4124 |
| GO:0098504 | n≤5 genes | DNA 3' dephosphorylation involved in DNA repair |  |
| GO:0100026 | =GO: 0009380 | positive regulation of DNA repair by transcription from RNA polymerase II promoter |  |
| GO:1901184 | n≤5 genes | regulation of ERBB signalling pathway |  |
| GO:1901185 |  | negative regulation of ERBB signalling pathway | 0.5527 |
| GO:1901186 | n≤5 genes | positive regulation of ERBB signalling pathway |  |
| GO:1902113 | =GO: 0009380 | nucleotide phosphorylation involved in DNA repair |  |
| GO:1904886 |  | beta-catenin destruction complex disassembly | 0.9398 |
| GO:1990391 | n≤5 genes | DNA repair complex |  |
| GO:2000741 | n≤5 genes | positive regulation of mesenchymal stem cell |  |
| HGNC:102 |  | Ubiquitin conjugating enzymes E2 gene family | 0.6487 |
| HGNC:1022 |  | ATG gene family | 0.2550 |
| HGNC:1253 |  | Scavenger receptors gene family | 0.3745 |
| HGNC:1256 | n≤5 genes | FOS gene family |  |
| HGNC:1257 | n≤5 genes | JUN gene family |  |
| HGNC:1264 |  | IL6 gene family | 1.0000 |
| HGNC:40 |  | Acyl-CoA synthetase family (ACS) | 0.6581 |
| HGNC:476 |  | microRNAs | 0.0159 |
| HGNC:476b | no LD-blocks assigned | miRNA gene family (restricted to LET7-genes) | 1.0000 |
| HGNC:496 |  | CDK gene family | 0.7716 |
| HGNC:508 |  | FOXO gene family | 0.1633 |
| HGNC:518 |  | ANTP/HOXL subclass homeoboxes gene family | 0.3765 |
| HGNC:519 |  | ANTP/NKL subclass homeoboxes and pseudogenes gene family | 0.7112 |
| HGNC:521 |  | PRD//PAX+PAXL subclass homeoboxes gene family | 0.7450 |
| HGNC:522 |  | LIM subclass homeoboxes gene family | 0.1992 |
| HGNC:523 |  | POU subclass homeoboxes gene family | 0.7676 |
| HGNC:524 | n≤5 genes | HNF subclass homeoboxes gene family |  |
| HGNC:525 |  | SINE subclass homeoboxes gene family | 1.0000 |
| HGNC:526 |  | TALE subclass homeoboxes gene family | 0.6920 |
| HGNC:527 |  | CUT subclass homeoboxes gene family | 0.6495 |
| HGNC:528 | n≤5 genes | PROS/PROX subclass homeoboxes gene family |  |
| HGNC:529 |  | ZF subclass homeoboxes gene family | 0.9566 |
| HGNC:530 |  | CERS subclass homeoboxes gene family | 0.9885 |
| HGNC:567 |  | Glutathione S-transferases | 0.7652 |
| HGNC:598 |  | Interferons IFN gene family | . |
| HGNC:750 |  | SMAD gene family | 0.7218 |
| HGNC:757 |  | SRY-boxes | 0.9263 |
| HGNC:938 |  | SIRT gene family | 0.8095 |
| literature based |  | Homeoboxes-Gene in regulatory networks related to lung cancers | 0.4402 |

in total 148 gene sets; thereof 119 GO terms, 28 HGNC gene-families and 1 literature based gene set.

# Results

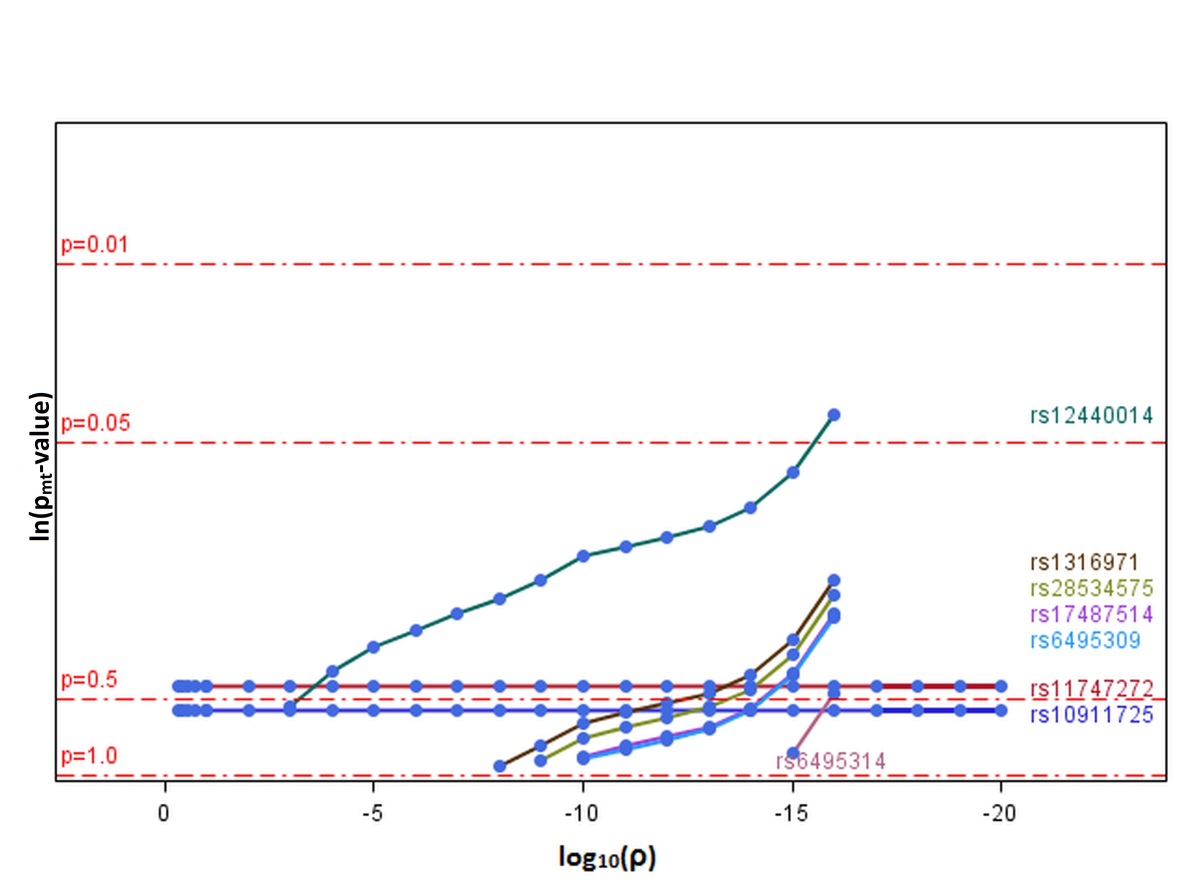
## Single marker interaction analysis

Supplementary Figure 1 Manhattan-Plot displaying significance of GxE interaction for each marker



Each point represents the significance of a GxE interaction for a single marker

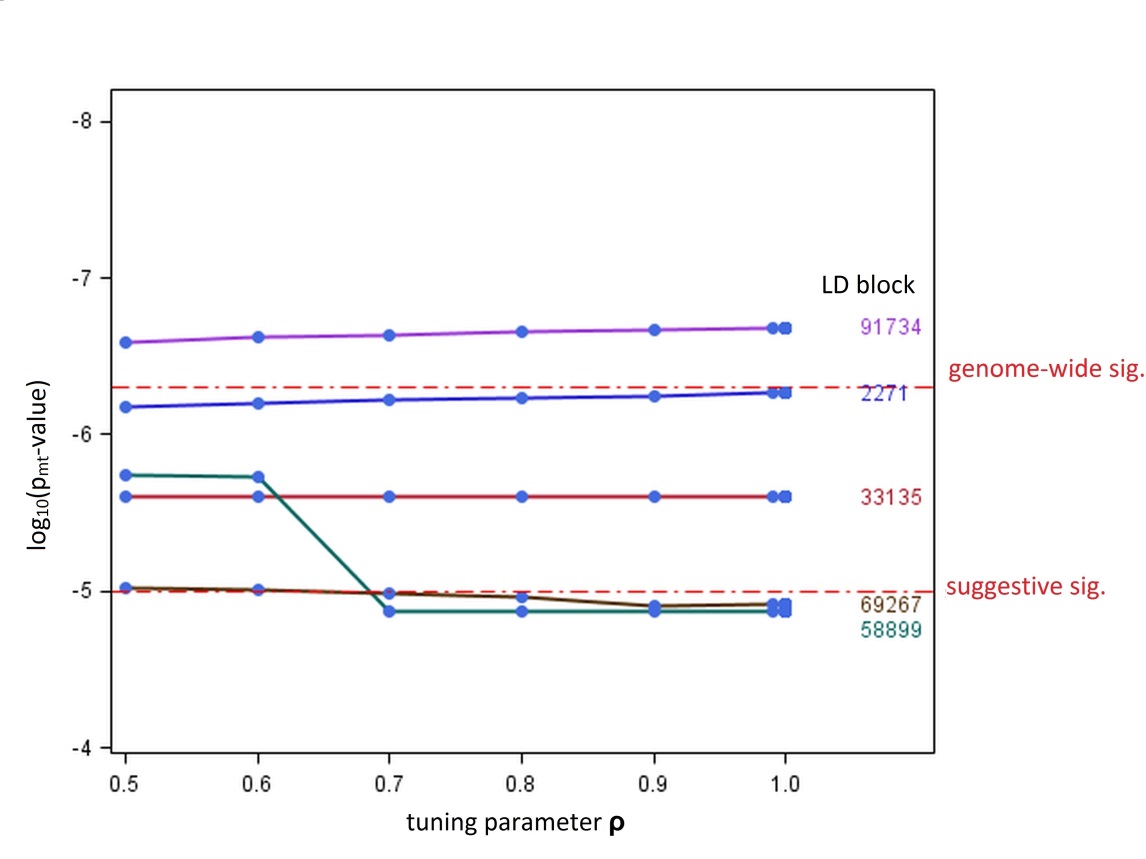
Supplementary Figure 2: Significance of selected markers corrected with the H2-method across several ρ-values



P-values are adjusted for multiple testing according to Hybrid 2-step (H2)-method

## Multi-marker interaction analysis

Supplementary Figure 3: Significance of selected LD blocks corrected with the Hybrid 2-step (H2)-method across several ρ-values



Significance of LD block according to the AIC-best model; p-values are adjusted for multiple testing according to Hybrid 2-step (H2)-method

# Discussion

## Discussion of the GxE interaction at 10p13

Setting ρ=0.5 of the Hybrid 2-step (H2)-method, the block **no. 58899** on chromosome **10p13** (*pmt*=0.1878) advanced to suggestive significance (see Figure 1 and Supplementary Figure 3).

The gene set GO:0016020, which was found borderline significant by GSA, hosts 90 “driving” genes including the LD block no. 58899 next to the gene CUBN.

The block **no. 58899** on chromosome **10p13** contains in total 10 markers, 7 of these remained in the AIC-best model; six thereof with a local significant GxE interaction, while no marker carried a G-main effect. However, fitting the model results in an inverse association of the E-main effect (OR=0.1 instead of OR~2.25). Thus it can be assumed that the E-main effect is absorbed by the genomic markers potentially due to the low number of cases, which are informative for the radon exposure. Hence, the estimated ORs are untrustworthy and no marker can be highlighted. The block is located within the gene **CUBN** that encodes the protein cubilin, a receptor for intrinsic factor-vitamin B12 complexes.

The gene **CUBN** (10p13) was highlighted by the multi-marker analysis (*pmt*=0.1878) and is the most important gene driving the borderline significant gene set “membrane” (GO:0016020, p=0.0558). CUBN encodes the protein cubilin, which was shown to be involved in the endocytosis and transcellular transport of numerous ligands, including vitamin D.(Kaseda et al. 2011) An increased risk for lung cancer was associated with a low vitamin D status.(Zhang et al. 2015) It is well understood that the human body depend on sunshine for its vitamin D requirement.(Holick 2008) Most recently a comprehensive meta-analysis demonstrated a significantly high risk to develop vitamin D deficiency for shift workers and in general for indoor workers.(Sowah et al. 2017) Surprisingly, the vitamin D status of underground miners was not significantly different from surface miners. However the sample size of miners was small, exposure to sunlight in-between working shifts could not been excluded and the vitamin D level can be attributed to others lifestyle factors, too. All this taken together gives reasons for suspecting the observed GxE interaction being spurious due to confounding by unmeasured vitamin D status. On the other hand, it cannot be ruled out that radiation-induced oxidative stress or DNA injuries in skin cells interfere with metabolising towards vitamin D. Hence the risk for lung cancer attributable to radon would truly be stratified across CUBN genotypes.

## Discussion of the GxE interaction at 12p12.1

Setting ρ=0.5 of the Hybrid 2-step (H2)-method, the block **no. 69267** on chromosome **12p12.1** (*pmt*=0.9875) advanced to suggestive significance. (see Figure 1 and Supplementary Figure 3)

The gene **SOX5** (12p12.1) was detected by the multi-marker analysis with suggestive significance. It encodes a member of the SOX (SRY-related HMG-box) family of transcription factors that is involved in the development of the lung. A number of studies have shown strong expression of family members (e.g. SOX2, SOX4 or SOX11) in most SCLCs, some also in NSCLC.(Zhu et al. 2012) Notably, the marker rs11046966 located within SOX5 was found to be associated with COPD in two studies (National Emphysema Treatment Trial: OR=1.48, p=6.0x10-4; Boston Early-Onset COPD Study: p=1.5x10-5), but this could not be replicated in a family-based study of then International COPD Genetics Network (p=0.16).(Hersh et al. 2011) Because COPD was demonstrated as risk factor for lung cancer (Brenner et al. 2012), there is a not negligible chance that the observed interaction is confounded.

# References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome research 19(9):1655-64 doi:10.1101/gr.094052.109

Amos CI, et al. (2017) The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 26(1):126-135 doi:10.1158/1055-9965.EPI-16-0106

Amos CI, et al. (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 40(5):616-622

Arbogast PG, Ray WA (2011) Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. Am J Epidemiol 174(5):613-20 doi:10.1093/aje/kwr143

Ashburner M, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1):25-29

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263-265

Bhatlekar S, Fields JZ, Boman BM (2014) HOX genes and their role in the development of human cancers. Journal of Molecular Medicine 92(8):811-823 doi:10.1007/s00109-014-1181-y

Bonner MR, et al. (2006) Radon, secondhand smoke, glutathione-S-transferase M1 and lung cancer among women. International journal of cancer Journal international du cancer 119(6):1462-7 doi:10.1002/ijc.22002

Brambilla E (2009) [Epigenetic modifications in lung cancer]. Ann Pathol 29 Spec No 1:S31-3 doi:10.1016/j.annpat.2009.07.010

Brambilla E, Gazdar A (2009) Pathogenesis of lung cancer signalling pathways: roadmap for therapies. The European respiratory journal 33(6):1485-97 doi:10.1183/09031936.00014009

Brennan P, Hainaut P, Boffetta P (2011) Genetics of lung-cancer susceptibility. The lancet oncology 12(4):399-408 doi:10.1016/s1470-2045(10)70126-1

Brenner DR, et al. (2012) Previous lung diseases and lung cancer risk: a pooled analysis from the International Lung Cancer Consortium. Am J Epidemiol 176(7):573-85 doi:10.1093/aje/kws151

Brüske-Hohlfeld I, et al. (2006) Lung cancer risk among former uranium miners of the WISMUT Company in Germany. Health physics 90(3):208-16 doi:10.1097/01.HP.0000175443.08832.84

Cunningham F, et al. (2014) Ensembl 2015. Nucleic Acids Res doi:10.1093/nar/gku1010

Darby S, et al. (2005) Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ 330(7485):223

Ding L, et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455(7216):1069-75 doi:10.1038/nature07423

Fehringer G, et al. (2012) Association of the 15q25 and 5p15 lung cancer susceptibility regions with gene expression in lung tumor tissue. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 21(7):1097-104 doi:10.1158/1055-9965.epi-11-1123-t

Gabriel SB, et al. (2002) The structure of haplotype blocks in the human genome. Science 296(5576):2225-9 doi:10.1126/science.1069424

Gray KA, Yates B, Seal RL, Wright MW, Bruford EA (2015) Genenames.org: the HGNC resources in 2015. Nucleic Acids Res 43(Database issue):D1079-85 doi:10.1093/nar/gku1071

Hauptmann M, et al. (2003) SNPs in genes involved in DNA repair and carcinogen metabolism in a German case-control study on lung cancer In: (NCI) NCI, (GSF) NRCfEH (eds).

Hersh CP, et al. (2011) SOX5 is a candidate gene for chronic obstructive pulmonary disease susceptibility and is necessary for lung development. Am J Respir Crit Care Med 183(11):1482-9 doi:10.1164/rccm.201010-1751OC

Holick MF (2008) Sunlight, UV-radiation, vitamin D and skin cancer: how much sunlight do we need? Advances in experimental medicine and biology 624:1-15 doi:10.1007/978-0-387-77574-6\_1

Hornhardt S, et al. (2014) Genetic factors in individual radiation sensitivity. DNA Repair (Amst) 16:54-65 doi:10.1016/j.dnarep.2014.02.001

Hubaux R, Becker-Santos DD, Enfield KS, Lam S, Lam WL, Martinez VD (2012) Arsenic, asbestos and radon: emerging players in lung tumorigenesis. Environ Health 11:89 doi:10.1186/1476-069X-11-89

Huckins LM, et al. (2014) Using ancestry-informative markers to identify fine structure across 15 populations of European origin. European journal of human genetics : EJHG 22(10):1190-200 doi:10.1038/ejhg.2014.1

Hung RJ, et al. (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452(7187):633-637

Hunter N, et al. (2013) Joint analysis of three European nested case-control studies of lung cancer among radon exposed miners: exposure restricted to below 300 WLM. Health physics 104(3):282-92 doi:10.1097/HP.0b013e3182765857

Jafri MA, Ansari SA, Alqahtani MH, Shay JW (2016) Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. Genome Medicine 8:69 doi:10.1186/s13073-016-0324-x

Kaseda R, Hosojima M, Sato H, Saito A (2011) Role of megalin and cubilin in the metabolism of vitamin D(3). Therapeutic apheresis and dialysis : official peer-reviewed journal of the International Society for Apheresis, the Japanese Society for Apheresis, the Japanese Society for Dialysis Therapy 15 Suppl 1:14-7 doi:10.1111/j.1744-9987.2011.00920.x

Kosoy R, et al. (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Human mutation 30(1):69-78 doi:10.1002/humu.20822

Kreuzer M, Grosche B, Dufey F, Schnelzer M, Tschense A, Walsh L (2011) The German Uranium Miners Cohort Study (Wismut cohort), 1946–2003 Bundesamt für Strahlenschutz (German Federal Office of Radiation Prodection), Oberschleissheim, February 2011

Kreuzer M, Grosche B, Schnelzer M, Tschense A, Dufey F, Walsh L (2010a) Radon and risk of death from cancer and cardiovascular diseases in the German uranium miners cohort study: follow-up 1946-2003. Radiation and environmental biophysics 49(2):177-85 doi:10.1007/s00411-009-0249-5

Kreuzer M, Schnelzer M, Tschense A, Walsh L, Grosche B (2010b) Cohort profile: the German uranium miners cohort study (WISMUT cohort), 1946-2003. Int J Epidemiol 39(4):980-7 doi:10.1093/ije/dyp216

Leng S, et al. (2013) Genetic variation in SIRT1 affects susceptibility of lung squamous cell carcinomas in former uranium miners from the Colorado plateau. Carcinogenesis 34(5):1044-50 doi:10.1093/carcin/bgt024

Leng S, et al. (2016) Radon Exposure, IL-6 Promoter Variants, and Lung Squamous Cell Carcinoma in Former Uranium Miners. Environmental health perspectives 124(4):445-51 doi:10.1289/ehp.1409437

McKay JD, et al. (2017) Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet 49(7):1126-1132 doi:10.1038/ng.3892

Mukherjee B, Ahn J, Gruber SB, Rennert G, Moreno V, Chatterjee N (2008) Tests for gene-environment interaction from case-control data: a novel study of type I error, power and designs. Genetic epidemiology 32(7):615-26 doi:10.1002/gepi.20337

Murcray CE, Lewinger JP, Conti DV, Thomas DC, Gauderman WJ (2011) Sample size requirements to detect gene-environment interactions in genome-wide association studies. Genetic epidemiology 35(3):201-10 doi:10.1002/gepi.20569

Pesch B, Johnen G, Lehnert M (2015) Aufbau einer Bioproben-Bank von ehemaligen Beschäftigten der SAG / SDAG Wismut – Pilotstudie Ressortforschungsberichte zur kerntechnischen Sicherheit und zum Strahlenschutz. BfS - Bundesamt für Strahlenschutz

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8):904-9 doi:10.1038/ng1847

Price AL, et al. (2008) Long-range LD can confound genome scans in admixed populations. Am J Hum Genet 83(1):132-5; author reply 135-9 doi:10.1016/j.ajhg.2008.06.005

Purcell S, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559-75 doi:10.1086/519795

Ruano-Ravina A, Pereyra MF, Castro MT, Perez-Rios M, Abal-Arca J, Barros-Dios JM (2014) Genetic susceptibility, residential radon, and lung cancer in a radon prone area. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer 9(8):1073-80 doi:10.1097/JTO.0000000000000205

Setsirichok D, et al. (2012) Small Ancestry Informative Marker panels for complete classification between the original four HapMap populations. Int J Data Min Bioinform 6(6):651-74

Sowah D, Fan X, Dennett L, Hagtvedt R, Straube S (2017) Vitamin D levels and deficiency with different occupations: a systematic review. BMC public health 17(1):519 doi:10.1186/s12889-017-4436-z

Stenzel SL, Ahn J, Boonstra PS, Gruber SB, Mukherjee B (2015) The impact of exposure-biased sampling designs on detection of gene-environment interactions in case-control studies with potential exposure misclassification. European journal of epidemiology 30(5):413-23 doi:10.1007/s10654-014-9908-1

Su S, et al. (2006) Aberrant promoter methylation of p16(INK4a) and O(6)-methylguanine-DNA methyltransferase genes in workers at a Chinese uranium mine. J Occup Health 48(4):261-6

Subramanian A, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. ProcNatlAcadSciUSA 102(43):15545-15550

Takamizawa J, et al. (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer research 64(11):3753-6 doi:10.1158/0008-5472.CAN-04-0637

Taylor JA, Watson MA, Devereux TR, Michels RY, Saccomanno G, Anderson M (1994) p53 mutation hotspot in radon-associated lung cancer. Lancet 343(8889):86-7

Timofeeva MN, et al. (2012) Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. Human molecular genetics 21(22):4980-95 doi:10.1093/hmg/dds334

Truong T, et al. (2010a) Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. J NatlCancer Inst 102(13):959-971

Truong T, et al. (2010b) International Lung Cancer Consortium: coordinated association study of 10 potential lung cancer susceptibility variants. Carcinogenesis 31(4):625-633

Vahakangas KH, et al. (1992) Mutations of p53 and ras genes in radon-associated lung cancer from uranium miners. Lancet 339(8793):576-80

Wall JD, Pritchard JK (2003) Haplotype blocks and linkage disequilibrium in the human genome. Nature reviews Genetics 4(8):587-97 doi:10.1038/nrg1123

Wang Y, et al. (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. NatGenet 40(12):1407-1409

Wang Y, et al. (2014) Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 46(7):736-41 doi:10.1038/ng.3002

Wichmann EH, et al. (1998) Lungenkrebsrisiko durch Radon in der Bundesrepublik Deutschland (West). Fortschritte in der Umweltmedizin. ecomed, Landsberg

Yang J, et al. (2011) Genomic inflation factors under polygenic inheritance. European journal of human genetics : EJHG 19(7):807-12 doi:10.1038/ejhg.2011.39

Yngveson A, Williams C, Hjerpe A, Lundeberg J, Soderkvist P, Pershagen G (1999) p53 Mutations in lung cancer associated with residential radon exposure. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 8(5):433-8

Zhang L, Wang S, Che X, Li X (2015) Vitamin D and lung cancer risk: a comprehensive review and meta-analysis. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology 36(1):299-305 doi:10.1159/000374072

Zhu Y, Li Y, Jun Wei JW, Liu X (2012) The role of Sox genes in lung morphogenesis and cancer. International journal of molecular sciences 13(12):15767-83 doi:10.3390/ijms131215767