

Two alternative data-splitting

Numerous hypothesis tests were performed in this study. To reduce the false positive due to multiple testing, we are not only seeking the results with extremely small p values but also those that can be robustly found in different subsets of our dataset. To do so, we split our dataset into the discovery and validation sets as shown in Table 1. Furthermore, we performed additional two data-splitting on the same dataset. The results shown in the main text are from the original data-splitting. Presented here in Supporting Information were those from the two alternative data-splitting. The results from three data-splitting showed consistent patterns. (Tables S1-S3) Therefore, it is unlikely that our report is simply chance aberration due to particular split.

Analyses of batch effect

During the data preprocessing, normalization with consideration of the batch effect was performed. Even so, the association between the batch and cigarette smoking may still potentially exert undue confounding effects. Because of that, we performed the analyses with adjustment of the batch effect in the genome-wide analyses and focal copy number analyses. The corresponding analyses to Figure 1A and 1B and those to Figure 1C and 1D were shown in Tables S4 and S5, respectively. For the focal copy number-smoking association, 1,000 loci were randomly selected to perform both batch-adjusted and unadjusted analyses. P values from the two analyses were compared to see whether they follow similar distributions. Such comparisons were made for both single-marker analyses (Figure S1A) and 10-marker analyses (Figure S1B). Overall, the analyses adjusted for the batch effect did not change the pattern of the results and our conclusions.

Single-marker analyses of cigarette smoking and focal copy number: the dichotomous version

We also pursued the single-marker analyses in a dichotomous fashion. We tabulated the copy numbers (≥ 2.7 vs. 1.3-2.7 for gains; ≤ 1.3 vs. 1.3-2.7 for losses) and smoking pack-years (>60 vs. ≤ 60) and tested its association by Fisher exact test for 256,554 loci. Logistic regressions were also performed with adjustment of age at diagnosis, gender, two cohorts, clinical stage and histology. The dichotomous-version analyses show that copy number gains in 15 loci and copy number losses in 6 loci are associated with the heavy smoking. (Figure S8A and Table S7) These candidate loci were clustered in 8q24, 12q21, 12q23 and 17q22 (Figures S8B-S8E) for gains and 8p12 and 8p23 for losses (Figures S8F and S8G).

Table S1. P values of comparing % of probes with $CN \geq 2.7$ (or < 1.3) between heavy smokers and non-/light-smokers, with three different data-splitting.

| | Main text | Alternative 1 | Alternative 2 |
|---|-----------|---------------|---------------|
| % of probes with copy number ≥ 2.7 | | | |
| Discovery set | 0.008 | 0.0076 | 0.00467 |
| Validation set | 0.0095 | 0.0108 | 0.0147 |
| Both sets | 0.000246 | | |
| % of probes with copy number < 1.3 | | | |
| Discovery set | 0.44 | 0.88 | 0.64 |
| Validation set | 0.97 | 0.29 | 0.25 |
| Both sets | 0.61 | | |

Table S2. P values of testing whether the mean G/T ratio is different than that at random (40.64%), with three different data-splitting. NS/LS: non-smokers/light-smokers; HS: heavy smokers.

| | | Main text | Alternative 1 | Alternative 2 |
|------------------|-------|-----------------------|-----------------------|-----------------------|
| Copy number gain | | | | |
| Discovery set | NS/LS | 0.8 | 0.11 | 0.095 |
| | HS | 0.59 | 0.16 | 0.87 |
| Validation set | NS/LS | 0.007 | 0.38 | 0.41 |
| | HS | 0.023 | 0.32 | 0.05 |
| Both sets | NS/LS | 0.08 | | |
| | HS | 0.083 | | |
| Copy number loss | | | | |
| Discovery set | NS/LS | 0.011 | 9.42×10^{-4} | 6.45×10^{-5} |
| | HS | 0.78 | 0.562 | 0.6 |
| Validation set | NS/LS | 9.80×10^{-4} | 0.0165 | 0.071 |
| | HS | 0.31 | 0.097 | 0.35 |
| Both sets | NS/LS | 5.15×10^{-5} | | |
| | HS | 0.32 | | |

Table S3. P values of the eleven loci presented in Table S6 from the single-marker CN-smoking analyses with three different data-splitting.

| | Main text | | Alternative 1 | | Alternative 2 | |
|----------------------|-----------------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | Discovery set | Validation set | Discovery set | Validation set | Discovery set | Validation set |
| Copy number >2 | | | | | | |
| rs1086086 | 1.65×10^{-7} | 0.014 | 5.11×10^{-5} | 3.70×10^{-5} | 1.70×10^{-5} | 0.00118 |
| rs2946831 | 5.57×10^{-7} | 0.0061 | 9.55×10^{-5} | 8.57×10^{-5} | 1.18×10^{-5} | 0.00114 |
| rs9308315 | 1.09×10^{-5} | 0.016 | 4.50×10^{-4} | 4.80×10^{-4} | 3.66×10^{-4} | 4.96×10^{-4} |
| rs2072592 | 1.15×10^{-5} | 0.047 | 0.00252 | 2.93×10^{-4} | 1.89×10^{-4} | 2.22×10^{-4} |
| rs1462651 | 2.39×10^{-5} | 0.016 | 2.83×10^{-4} | 3.07×10^{-4} | 0.00285 | 1.17×10^{-4} |
| rs7951476 | 3.30×10^{-6} | 0.0066 | 0.00298 | 0.0113 | 5.44×10^{-4} | 0.0487 |
| Copy number ≤ 2 | | | | | | |
| rs2290033 | 1.15×10^{-4} | 0.047 | 0.0213 | 2.44×10^{-4} | 0.00198 | 1.76×10^{-4} |
| rs4716055 | 4.71×10^{-5} | 0.042 | 0.00376 | 0.00232 | 1.33×10^{-5} | 0.00489 |
| rs9370883 | 3.44×10^{-5} | 0.021 | 0.0113 | 0.044 | 1.43×10^{-4} | 0.0109 |
| rs4577696 | 4.85×10^{-5} | 0.028 | 0.128 | 3.88×10^{-5} | 8.59×10^{-4} | 0.0111 |
| rs312688 | 5.14×10^{-5} | 0.0056 | 0.0859 | 0.0857 | 3.65×10^{-4} | 0.0747 |

Table S4. P values of comparing % of probes with $CN \geq 2.7$ (or < 1.3) between heavy smokers and non-/light-smokers, with and without adjustment of batch effects.

| | Unadjusted | Batch-adjusted |
|---|------------|----------------|
| % of probes with copy number ≥ 2.7 | | |
| Discovery set | 0.0080 | 0.0088 |
| Validation set | 0.0095 | 0.013 |
| Both sets | 0.000246 | 0.000348 |
| % of probes with copy number < 1.3 | | |
| Discovery set | 0.44 | 0.45 |
| Validation set | 0.97 | 0.95 |
| Both sets | 0.61 | 0.58 |

Table S5. P values of testing whether the mean G/T ratio is different than that at random (40.64%), with and without adjustment of batch effects. NS/LS: non-smokers/light-smokers; HS: heavy smokers.

| | | Unadjusted | Batch-adjusted |
|------------------|-------|------------|----------------|
| Copy number gain | | | |
| Discovery set | NS/LS | 0.80 | 0.80 |
| | HS | 0.59 | 0.60 |
| Validation set | NS/LS | 0.0070 | 0.0073 |
| | HS | 0.023 | 0.024 |
| Both sets | NS/LS | 0.080 | 0.080 |
| | HS | 0.083 | 0.086 |
| Copy number loss | | | |
| Discovery set | NS/LS | 0.011 | 0.010 |
| | HS | 0.78 | 0.77 |
| Validation set | NS/LS | 0.00098 | 0.0010 |
| | HS | 0.31 | 0.32 |
| Both sets | NS/LS | 0.000052 | 0.000055 |
| | HS | 0.52 | 0.31 |

Table S6. Summary of eleven loci from the top 100 loci with smallest p value in the discovery set and confirmed in the validation set with $p < 0.05$, under the continuous focal copy number-smoking analyses.

| | | | | | | | | | Linear model with spline of | |
|----------------|------------|-----------------------------|------------------|--|-------------------|-----------------------|--------|-------------------------|-----------------------------|-----------------------|
| | | | | Linear model with up to quadratic term of square root- | | | | square root-transformed | | |
| | | | | transformed smoking pack-years | | | | smoking pack-years | | |
| dbSNP ID | Chromosome | Genomic position (Mb) | Gene | P value unadjusted for covariates | | r ² | | Adjusted p | Unadjusted | Adjusted p |
| | | | | | | | | value * | p value | value * |
| | | | | Discovery set | Validation set | Pooled | Pooled | Pooled | Pooled | Pooled |
| Copy number >2 | | | | | | | | | | |
| rs10860860 | 12q23.2 | 101.283 | - | 1.65×10 ⁻⁷ | 0.014 | 1.79×10 ⁻⁸ | 0.223 | 6.77×10 ⁻⁸ | 1.98×10 ⁻⁷ | 4.96×10 ⁻⁷ |
| rs2946831 | 12q23.2 | 101.289 | - | 5.57×10 ⁻⁷ | 0.0061 | 1.29×10 ⁻⁸ | 0.235 | 2.09×10 ⁻⁸ | 1.36×10 ⁻⁷ | 1.65×10 ⁻⁷ |
| rs9308315 | 12q23.2 | 101.306 | <i>IGF1</i> | 1.09×10 ⁻⁵ | 0.016 | 2.10×10 ⁻⁷ | 0.202 | 8.65×10 ⁻⁷ | 6.36×10 ⁻⁷ | 2.37×10 ⁻⁶ |
| rs2072592 | 12q23.2 | 101.316 | <i>IGF1</i> | 1.16×10 ⁻⁵ | 0.047 | 6.17×10 ⁻⁷ | 0.200 | 5.24×10 ⁻⁶ | 1.61×10 ⁻⁶ | 1.13×10 ⁻⁵ |
| rs1462651 | 3q24 | 149.236 | - | 2.39×10 ⁻⁵ | 0.016 | 2.79×10 ⁻⁷ | 0.164 | 1.21×10 ⁻⁶ | 8.53×10 ⁻⁷ | 4.73×10 ⁻⁶ |
| rs7951476 | 11q24.2 | 123.891 | - | 3.30×10 ⁻⁶ | 0.0066 | 2.15×10 ⁻⁴ | 0.121 | 3.88×10 ⁻⁴ | 0.00113 | 0.0012 |
| Copy number ≤2 | | | | | | | | | | |
| rs2290033 | 8q24.21 | 128.562 | <i>LOC727677</i> | 1.15×10 ⁻⁴ | 0.047 | 1.05×10 ⁻⁵ | 0.416 | 1.36×10 ⁻⁵ | 1.13×10 ⁻⁴ | 0.0028 |
| rs4716055 | 6p24.3 | 9.96190 | - | 4.71×10 ⁻⁵ | 0.042 | 5.67×10 ⁻⁵ | 0.164 | 1.00×10 ⁻⁴ | 1.15×10 ⁻⁵ | 1.34×10 ⁻⁴ |
| rs9370883 | 6p24.3 | 9.96196 | - | 3.44×10 ⁻⁵ | 0.021 | 0.0020 | 0.112 | 0.0058 | 4.19×10 ⁻⁴ | 0.0015 |
| rs4577696 | 5p14.3 | 19.1683 | - | 4.85×10 ⁻⁵ | 0.028 | 6.63×10 ⁻⁶ | 0.394 | 5.31×10 ⁻⁴ | 2.29×10 ⁻⁵ | 0.051 |
| rs312688 | 17q24.3 | 65.8359 | - | 5.14×10 ⁻⁵ | 0.0056 | 0.0051 | 0.133 | 0.012 | 0.022 | 1.27×10 ⁻⁴ |

*P values of smoking pack-years were calculated from linear models with either up to quadratic term or models with spline of square root-transformed smoking pack-years, adjusting for age, gender, clinical stage, two sets and cell type.

Table S7. Summary of twenty one loci from the top 100 loci with smallest p value in the discovery set and confirmed in the validation set with $p < 0.05$, under the dichotomous-version of focal copy number-smoking analyses.

| Affymetrix ID | dbSNP ID | Chromosome | Position (Mb) | Gene | P value | | | OR (95% CI)* | P _{adj} value* |
|--------------------|------------|------------|---------------|-----------|-----------------------|----------------|-----------------------|--------------------|-------------------------|
| | | | | | Discovery set | Validation set | Pooled | | |
| Copy number gains | | | | | | | | | |
| SNP_A-1993896 | rs13274172 | 8q21.11 | 77.343 | - | 3.58×10 ⁻⁵ | 0.0401 | 1.11×10 ⁻⁵ | 4.96 (2.44, 10.11) | 1.03×10 ⁻⁵ |
| SNP_A-1995137 | rs1025524 | 8q24.21 | 130.621 | - | 6.39×10 ⁻⁵ | 0.017 | 5.02×10 ⁻⁶ | 5.25 (2.62, 10.52) | 2.84×10 ⁻⁶ |
| SNP_A-1830605 | rs6470759 | 8q24.21 | 130.771 | - | 6.98×10 ⁻⁵ | 0.00867 | 2.77×10 ⁻⁶ | 6.33 (3.04, 13.2) | 8.36×10 ⁻⁷ |
| SNP_A-2230506 | rs16904143 | 8q24.21 | 130.772 | - | 6.39×10 ⁻⁵ | 0.0209 | 7.42×10 ⁻⁶ | 5.13 (2.52, 10.48) | 6.89×10 ⁻⁶ |
| SNP_A-2013907 | rs4341199 | 8q24.23 | 137.977 | - | 4.98×10 ⁻⁶ | 0.0277 | 2.16×10 ⁻⁶ | 5.89 (2.97, 11.65) | 3.63×10 ⁻⁷ |
| SNP_A-1898911 | rs10745509 | 12q21.33 | 88.5827 | - | 1.30×10 ⁻⁵ | 0.032 | 2.03×10 ⁻⁶ | 16.1 (4.17, 62.11) | 5.50×10 ⁻⁵ |
| SNP_A-1865023 | rs11105391 | 12q21.33 | 88.6307 | - | 6.86×10 ⁻⁵ | 0.0458 | 9.24×10 ⁻⁶ | 19.7 (3.95, 97.9) | 0.000276 |
| SNP_A-4207364 | rs1394376 | 12q23.1 | 97.0512 | - | 1.04×10 ⁻⁵ | 0.00883 | 3.34×10 ⁻⁷ | 17 (5.04, 57.23) | 4.92×10 ⁻⁶ |
| SNP_A-1787215 | rs7488965 | 12q23.1 | 97.0664 | - | 5.25×10 ⁻⁵ | 0.0293 | 5.60×10 ⁻⁶ | 12.3 (3.69, 41.2) | 4.44×10 ⁻⁵ |
| SNP_A-2090552 | rs1842329 | 12q23.1 | 97.0667 | - | 9.37×10 ⁻⁶ | 0.0293 | 1.43×10 ⁻⁶ | 20.4 (4.96, 83.89) | 2.90×10 ⁻⁵ |
| SNP_A-1832480 | rs10861139 | 12q23.3 | 102.794 | LOC253724 | 6.82×10 ⁻⁵ | 0.0377 | 1.49×10 ⁻⁵ | 6.63 (2.84, 15.48) | 1.22×10 ⁻⁵ |
| SNP_A-2120074 | rs979746 | 17q21.32 | 43.6911 | SKAP1 | 6.98×10 ⁻⁵ | 0.0488 | 1.64×10 ⁻⁵ | 6.57 (2.88, 15.03) | 8.02×10 ⁻⁶ |
| SNP_A-2057721 | rs8079095 | 17q22 | 48.5358 | - | 3.81×10 ⁻⁵ | 0.0404 | 6.72×10 ⁻⁶ | 9.31 (3.35, 25.9) | 1.89×10 ⁻⁵ |
| SNP_A-4199171 | rs16952794 | 17q22 | 48.5366 | - | 5.57×10 ⁻⁵ | 0.0229 | 7.57×10 ⁻⁶ | 9.59 (3.33, 27.6) | 2.76×10 ⁻⁵ |
| SNP_A-2005196 | rs998346 | 17q22 | 48.6198 | - | 5.57×10 ⁻⁵ | 0.0361 | 1.13×10 ⁻⁵ | 10.9 (3.56, 33.26) | 2.83×10 ⁻⁵ |
| Copy number losses | | | | | | | | | |
| SNP_A-2081471 | rs1714809 | 8p23.2 | 3.99017 | CSMD1 | 0.00464 | 0.00359 | 3.46×10 ⁻⁵ | 8.25 (2.89, 23.49) | 7.81×10 ⁻⁵ |
| SNP_A-1877228 | rs1714814 | 8p23.2 | 3.99125 | CSMD1 | 0.00464 | 0.0323 | 0.000389 | 6.92 (2.35, 20.38) | 0.000449 |
| SNP_A-4212915 | rs10503973 | 8p12 | 35.0188 | - | 0.00656 | 0.0144 | 0.000215 | 8.81 (2.98, 26.11) | 8.56×10 ⁻⁵ |
| SNP_A-1993023 | rs10503974 | 8p12 | 35.0189 | - | 0.00656 | 0.0144 | 0.000215 | 8.81 (2.98, 26.11) | 8.56×10 ⁻⁵ |
| SNP_A-2256354 | rs10954975 | 8p12 | 35.019 | - | 0.00656 | 0.0144 | 0.000215 | 8.82 (2.98, 26.14) | 8.53×10 ⁻⁵ |
| SNP_A-4212916 | rs9297228 | 8p12 | 35.0194 | - | 0.00116 | 0.0453 | 0.00015 | 11.2 (3.31, 38.16) | 0.000104 |

* Obtained from logistic regressions with adjustment of gender, age at surgery, two sets, clinical stage, and histology.

Figure S1. QQ plots of p values from batch-adjusted and unadjusted associations between focal copy numbers and pack-years of cigarette smoking.

A, 1,000 p values were randomly chosen from 256,554 single-marker analyses. **B**, 1,000 p values were randomly chosen from 25,655 10-marker analyses.

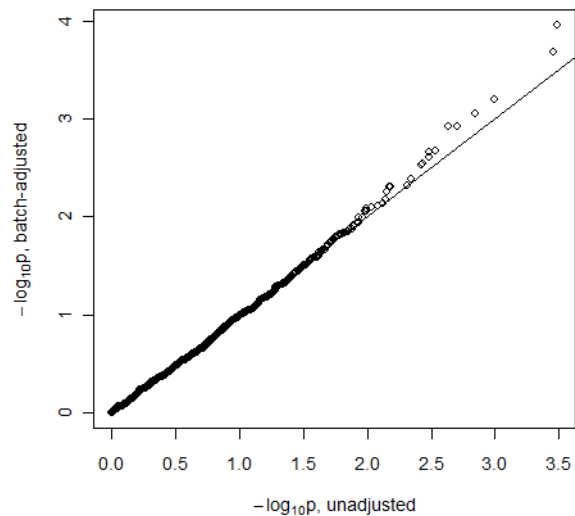
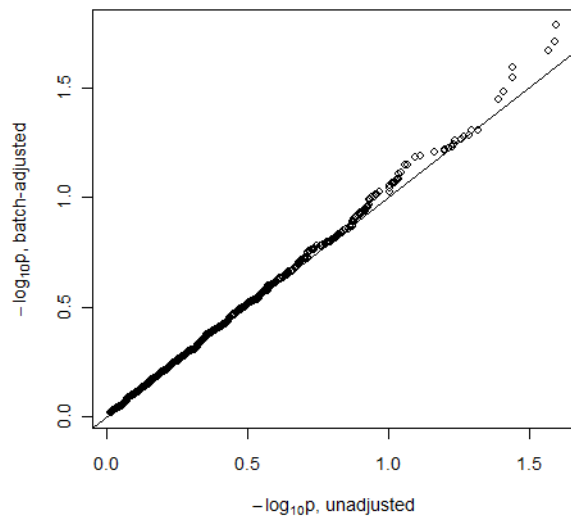
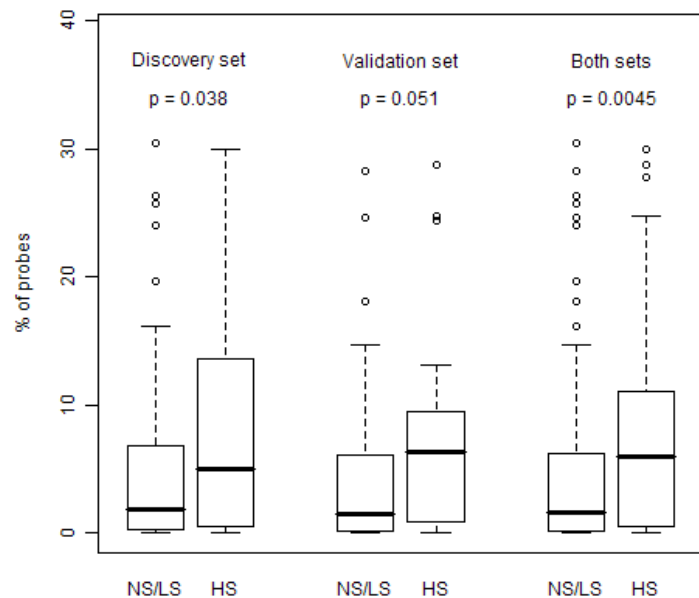
A**Single-marker analyses****B****10-marker analyses**

Figure S2. % of probes in genes with CNAs in heavy (>60 pack-years) and light or non-smokers (≤ 60 pack-years).

Proportion (%) of probes in genes with CNAs (**A**, gains; **B**, losses) events by pack-years of cigarette smoking.

A

% of probes in genes with $CN \geq 2.7$ in early stage samples

**B**

% of probes in genes with $CN \leq 1.3$ in early stage samples

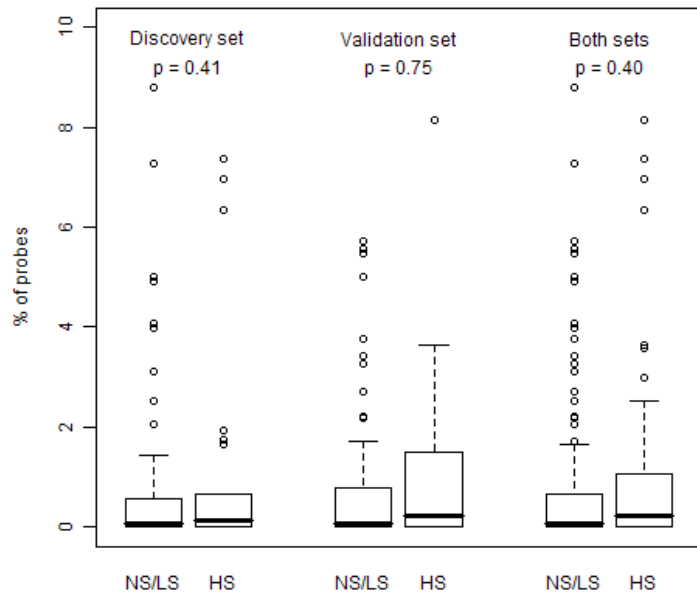


Figure S3. Chromosome-specific copy number pattern between heavy and light/non-smokers.

Average copy numbers in chromosomes 1 (**A**), 3 (**D**), 7 (**C**), 8 (**D**), 10 (**E**), 11 (**F**), 12 (**G**), 16 (**H**) and 17 (**I**) are plotted where copy numbers greater than two are red (heavy smokers) or pink color (light/non-smokers), and those less or equal to two are blue (heavy smokers) or light blue color (light/non-smokers). Vertical dotted lines indicate the centromeres of chromosomes. P value in each plot indicates the statistical significance of testing the averaged copy numbers (across the subjects) in the chromosome between heavy and light/non-smokers.

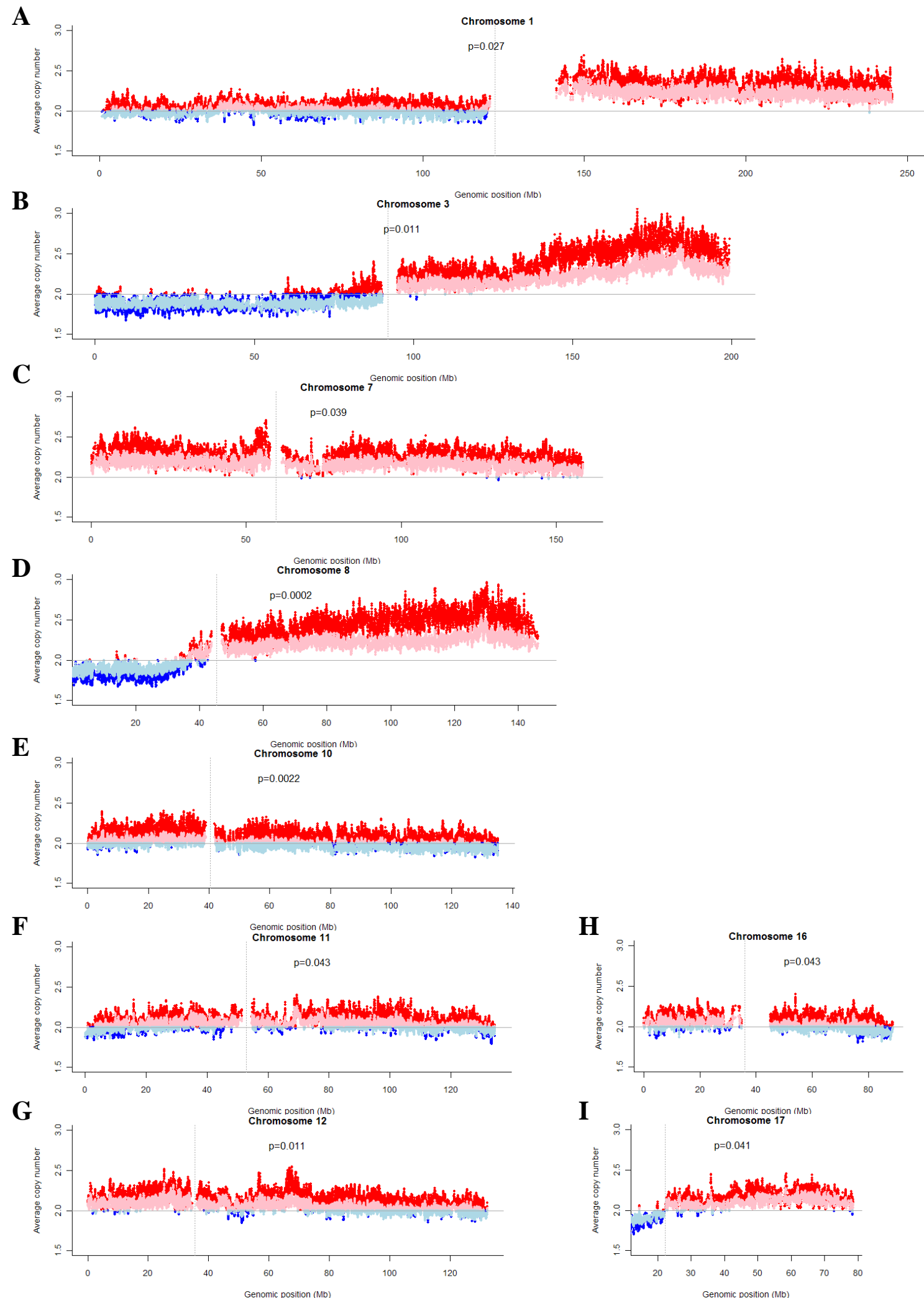


Figure S4. Association of cigarette smoking and chromosome/arm-specific copy numbers. Panel a illustrates the association of CNAs with heavy against light/non-smokers. Genome-wide mean CNs, genome-wide % of CN gains and total CN gains are the three indices, the difference of which between heavy and light/non-smokers is what p values intend to test (see detail in Methods). Panel b illustrates the association of copy numbers with the continuous smoking pack-years. The dashed line indicates $p=0.05$.

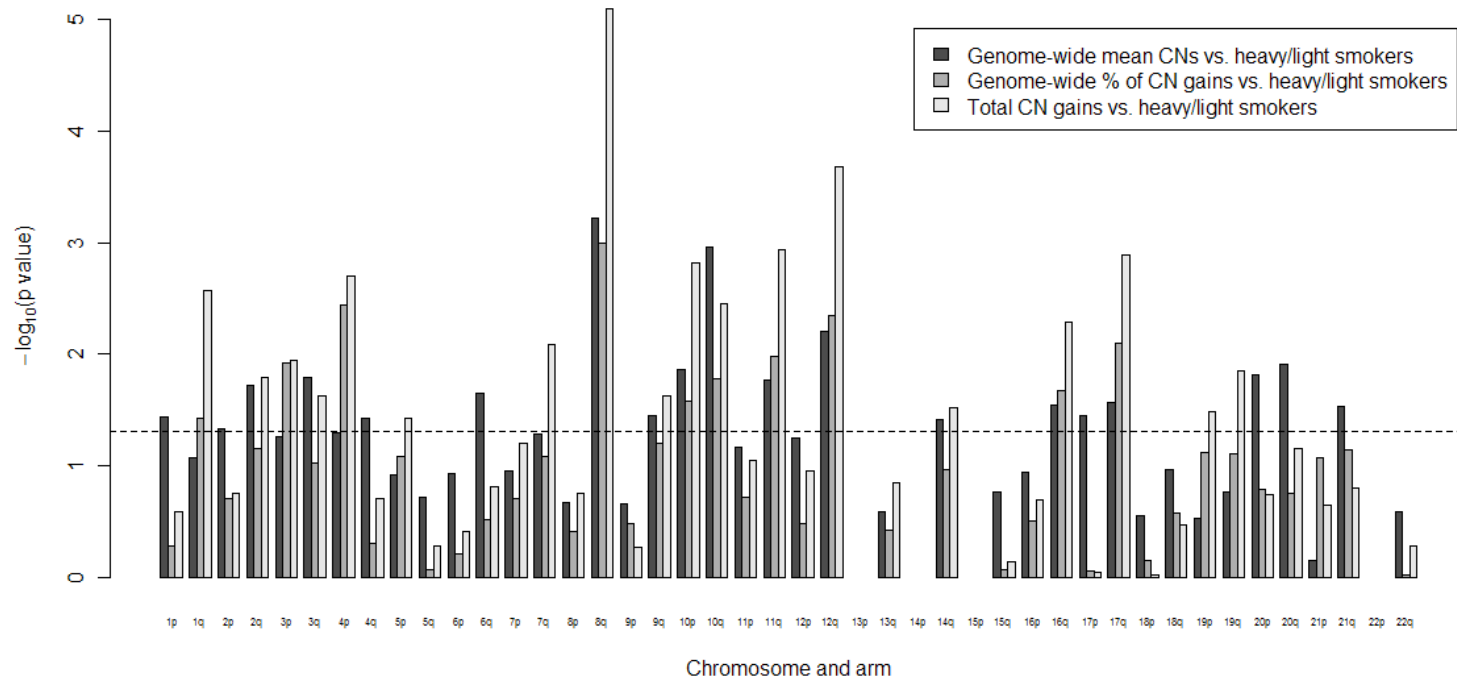
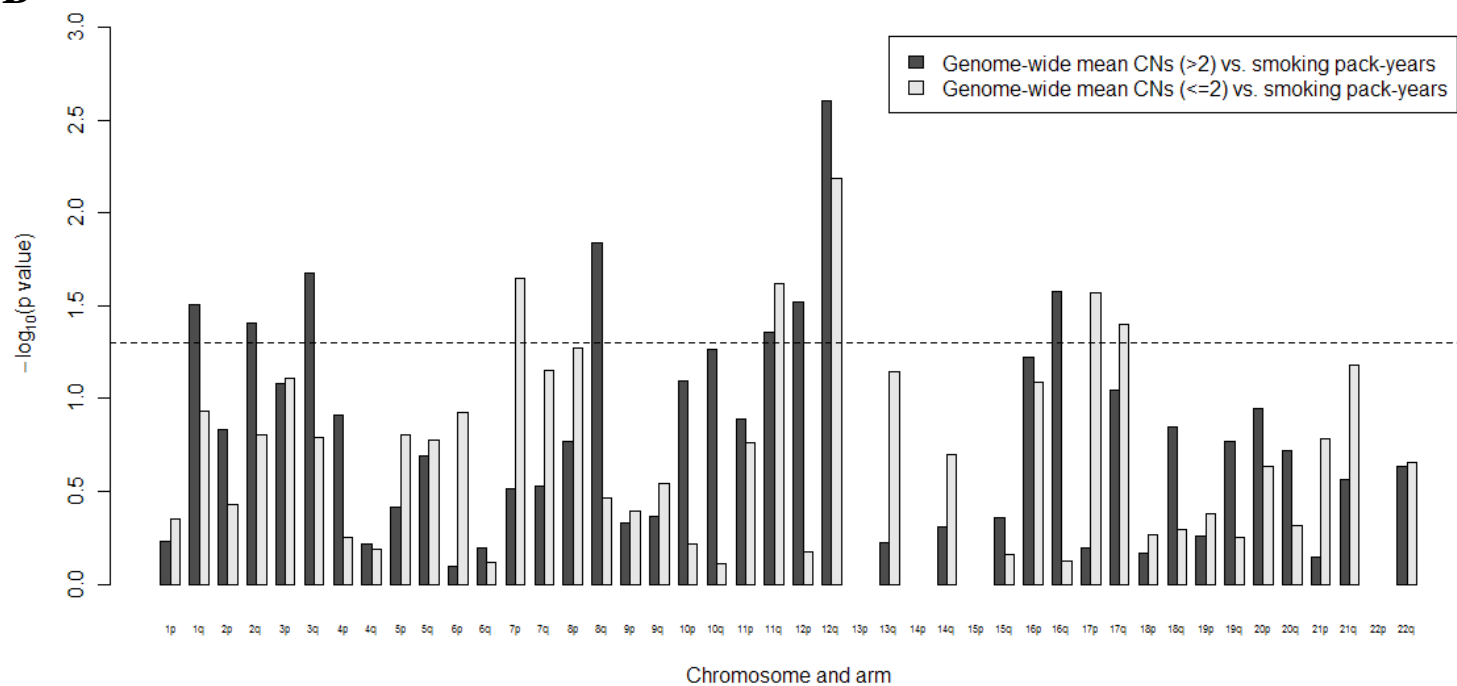
A**B**

Figure S5. Association of cigarette smoking and 256,554 focal copy numbers.

A, A plot of $-\log_{10}P$ of the association between smoking pack-years and copy number, which is analyzed for copy number >2 (upper half) and ≤ 2 (lower half), separately. Dashed lines indicate $p=2.0 \times 10^{-7}$. **B**, P values of probes in the 12q23.2 region (100.6-101.6 Mb), indicated by the black arrow in a. The other two arrows indicate 3q24 and 8q24.21, of which the p values in the focal region are shown in Figure S6. Blue dots in both plots are the top 100 p values in the discovery set and confirmed in the validation set with $p < 0.05$. The dashed lines indicate $p=2.0 \times 10^{-7}$ and 0.01. *IGF1*, insulin-like growth factor 1; *PMCH*, pro-melanin-concentrating hormone; *C12orf48*, chromosome 12 open reading frame 48; *NUP37*, nucleoporin 37kDa; *CCDC53*, coiled-coil domain containing 53. **C-F**, Scatter plots of the four validated loci (a, rs10860860; b, rs2946831; c, rs9308315; d, rs2072592) in the region of 12q23.2 with predicted values from three regression models: quadratic model: ordinary least square linear models with up to quadratic term of square root-transformed smoking pack-years (solid black lines), spline: linear models with spline of pack-years (solid green lines), and LOWESS (solid red lines), and the 95% confidence intervals of predicted mean value from spline models (dashed green lines). The other seven validated loci were plotted in Figure S7.

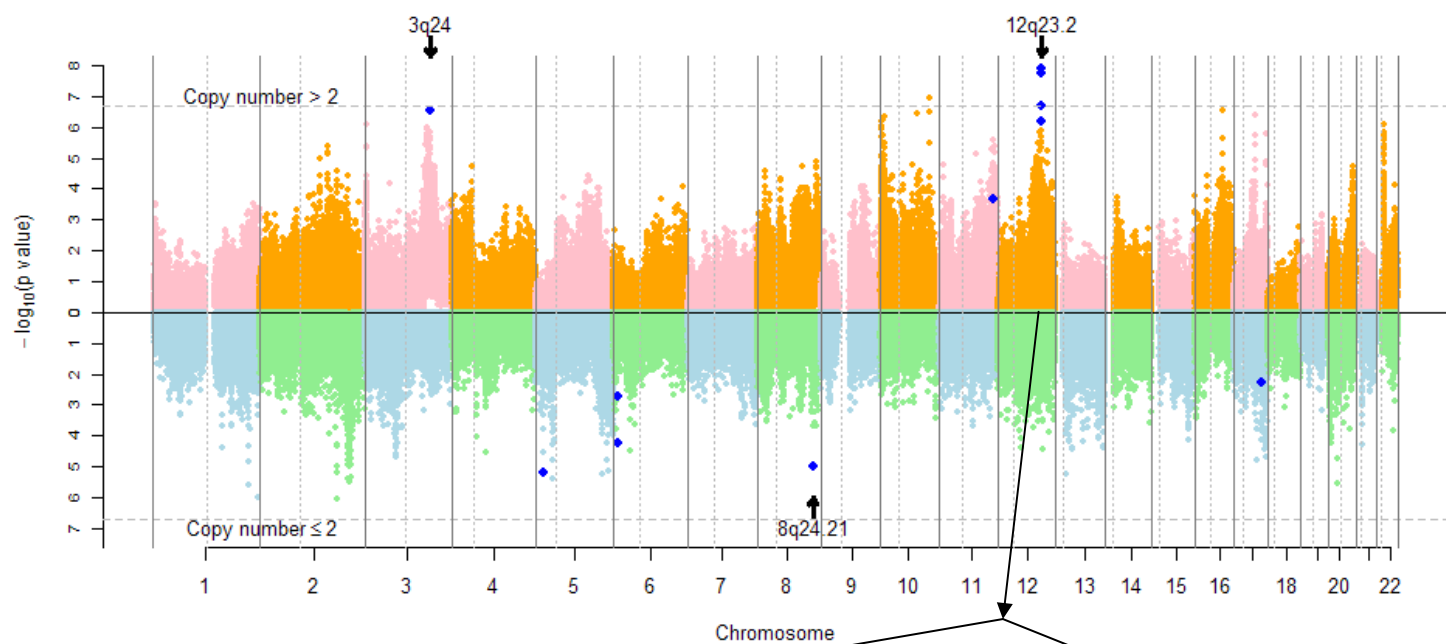
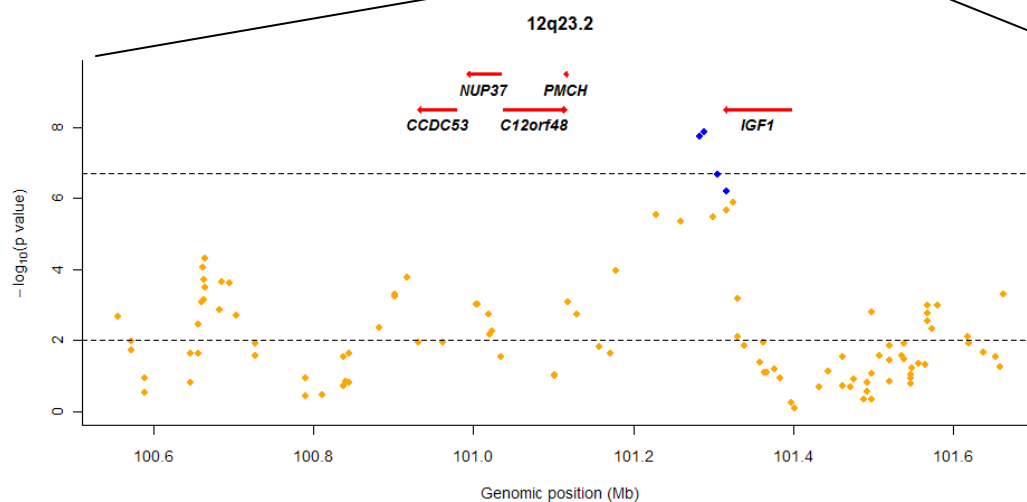
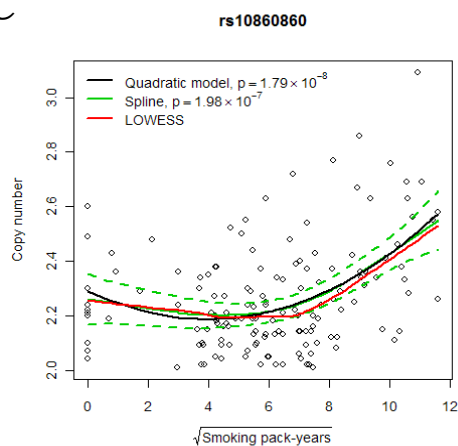
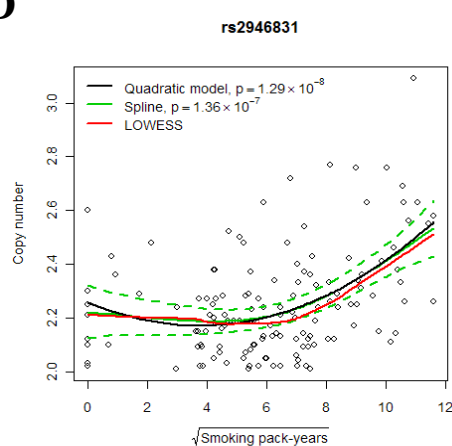
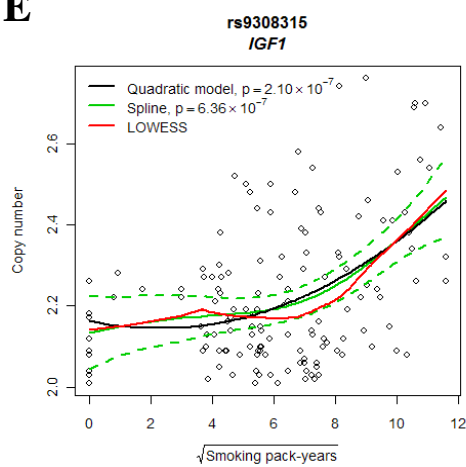
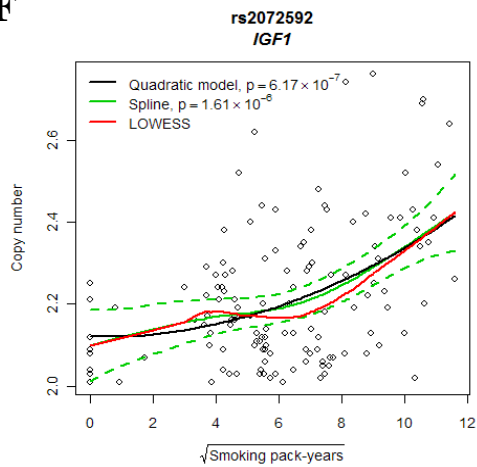
A**B****C****D****E****F**

Figure S6. Two candidate regions (in addition to 12q23.2) identified from focal copy number-smoking analyses.

P values of probes from the association analyses of focal copy numbers and smoking pack-years in the region of 3q24 (**A**) and 8q24.21 (**B**). Blue dots (in **A**, the location of rs1462651; in **B**, the location of rs2290033) are those in the list of top 100 p values in the discovery set and confirmed in the validation set with $p < 0.05$. The dashed lines indicate $p = 2.0 \times 10^{-7}$ and 0.01. *AGTR1*, angiotensin II receptor, type 1; *CPB1*, carboxypeptidase B1 (tissue); *CPA3*, carboxypeptidase A3 (mast cell); *LOC727677*, hypothetical LOC727677; *MYC*, v-myc myelocytomatosis viral oncogene homolog (avian); *PVT1*, Pvt1 oncogene (non-protein coding).

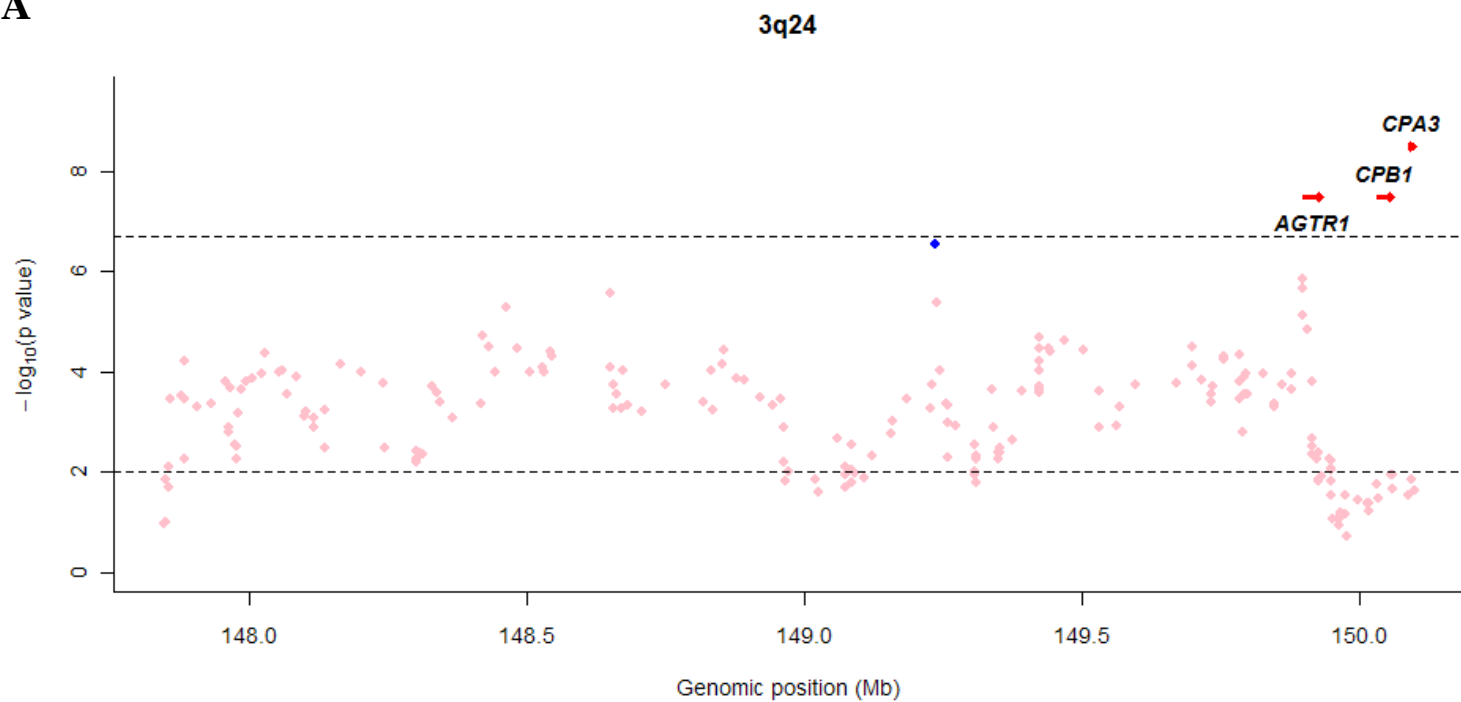
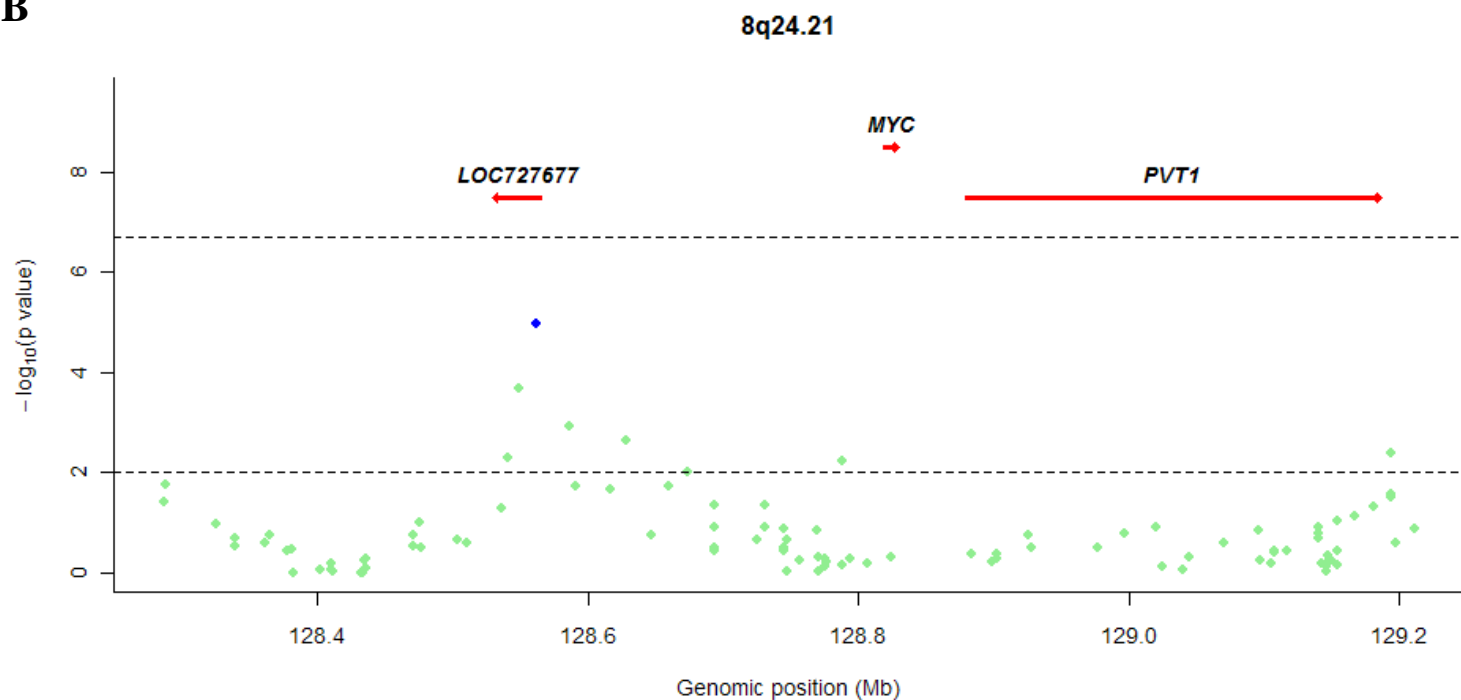
A**B**

Figure S7. Scatter plot and dose-response relationship of smoking pack-years and copy numbers in the seven candidate loci (in addition to the four in 12q23 (Figure S5C-F)).

Scatter plots of the loci (**A**, rs1462651 (at 3q24); **B**, rs2290033 (at 8q24.21); **C**, rs7951476 (11q24.2); **D**, rs4716055 (at 6p24.3); **E**, rs9370883 (at 6p24.3); **F**, rs4577696 (at 5p14.3); **G**, rs312688 (at 17q24.3)) with predicted values from ordinary least square linear models with up to quadratic term of square root-transformed smoking pack-years (solid black lines), models with spline of pack-years (solid green lines), and LOWESS (solid red lines) and the 95% confidence intervals of predicted mean value from spline models (dashed green lines).

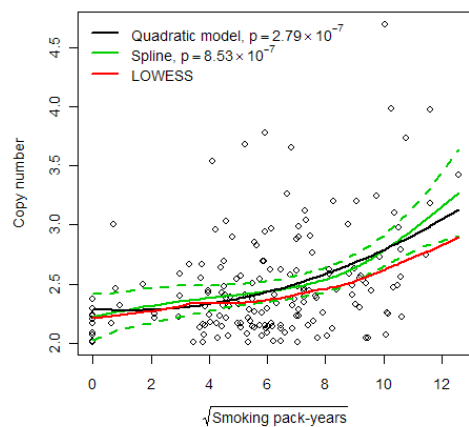
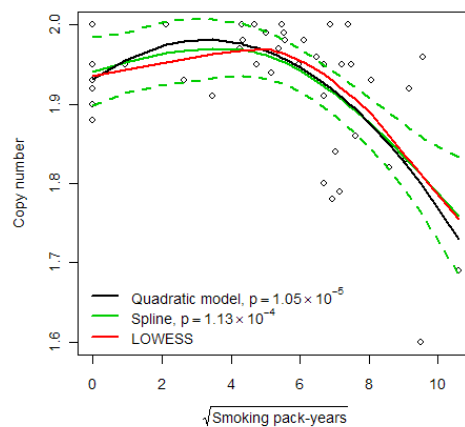
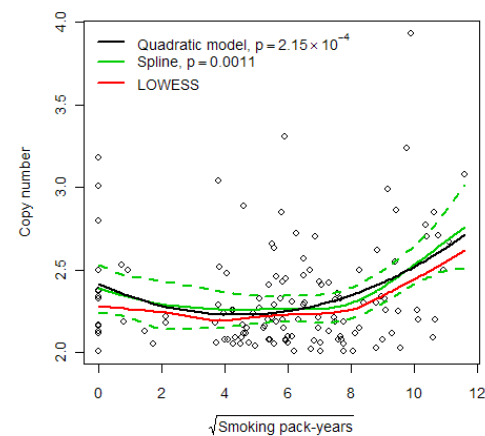
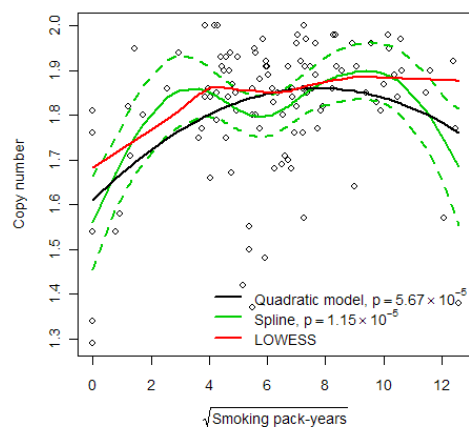
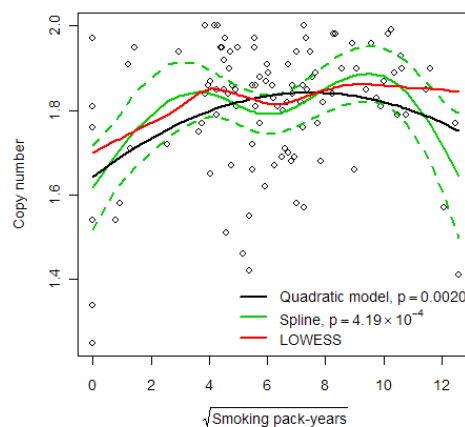
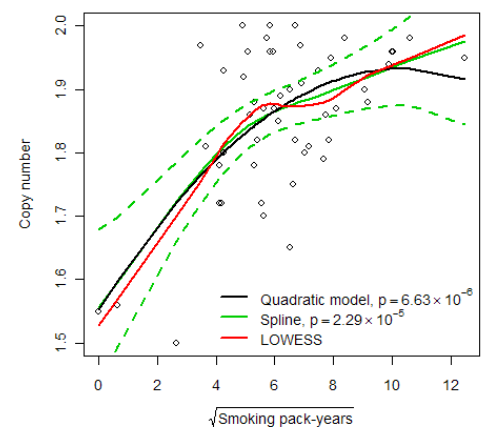
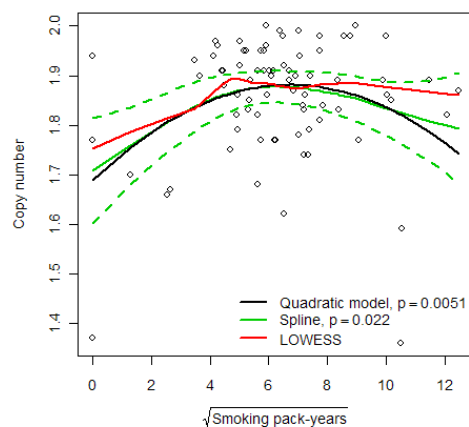
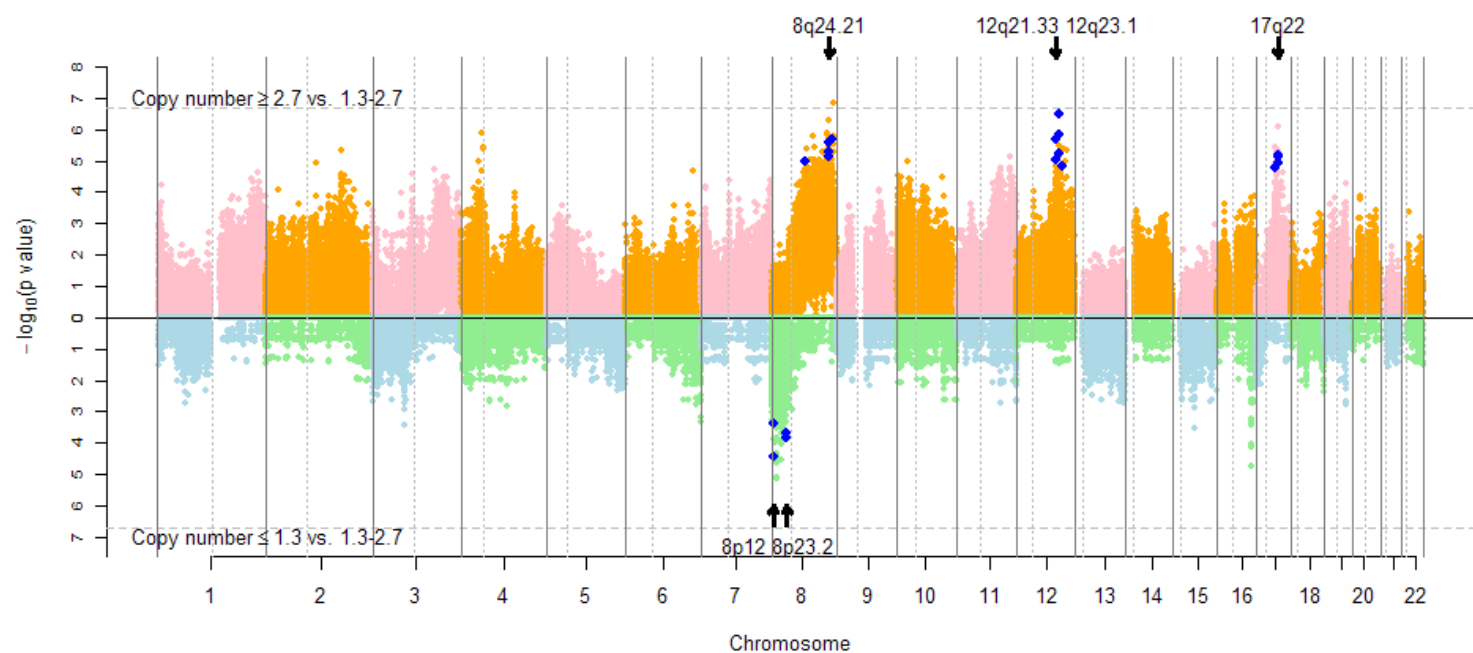
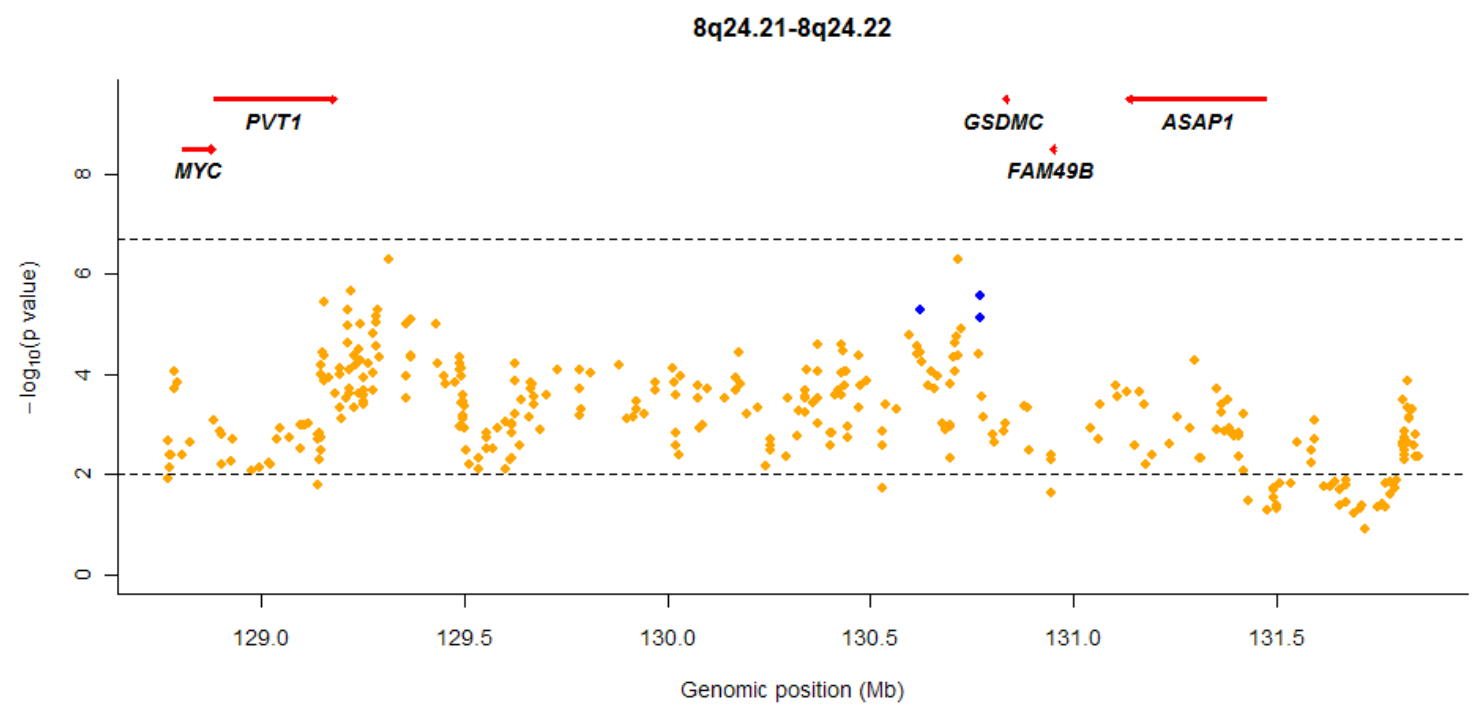
A**rs1462651****B****rs2290033**
LOC727677**C****rs7951476****D****rs4716055****E****rs9370883****F****rs4577696****G****rs312688**

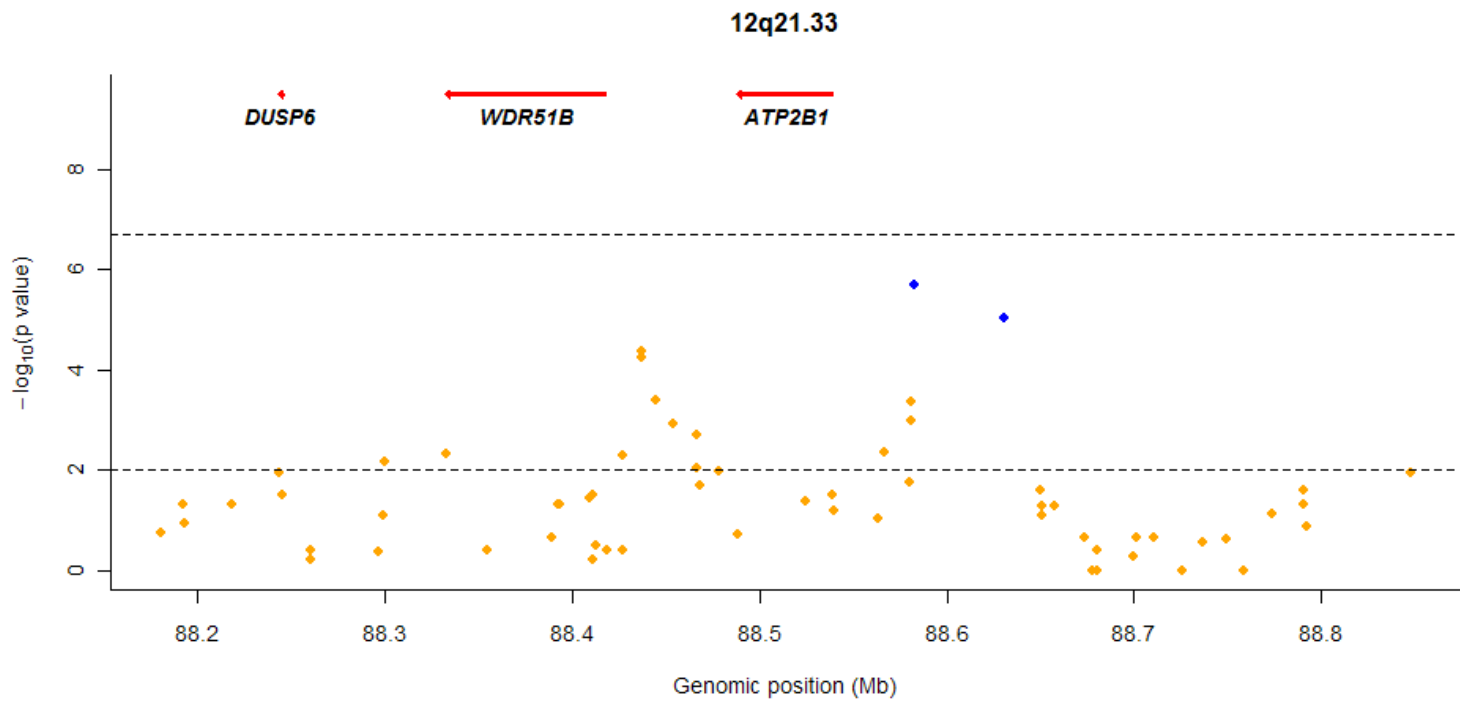
Figure S8. Association of dichotomous smoking pack-years (>60 vs. <60) and 256,554 focal CNAs (≥ 2.7 or ≤ 1.3 vs. 1.3-2.7).

A, A plot of $-\log_{10}P$ of the association between smoking pack-years and copy number by Fisher exact test, which is analyzed for copy number gains (upper half) and losses (lower half), separately. Dashed lines indicate $p=2.0 \times 10^{-7}$. **B-G**, P values of probes in the 8q24.21-8q24.22 (**B**), 12q21.33 (**C**), 12q.23.1 (**D**), 17q21.33-17q22 (**E**), 8p23.2 (**F**) and 8p12 (**G**), indicated by the black arrow in **A**. Blue dots in these plots are the top 100 p values in the discovery set and confirmed in the validation set with $p < 0.05$. The dashed lines indicate $p=2.0 \times 10^{-7}$ and 0.01.

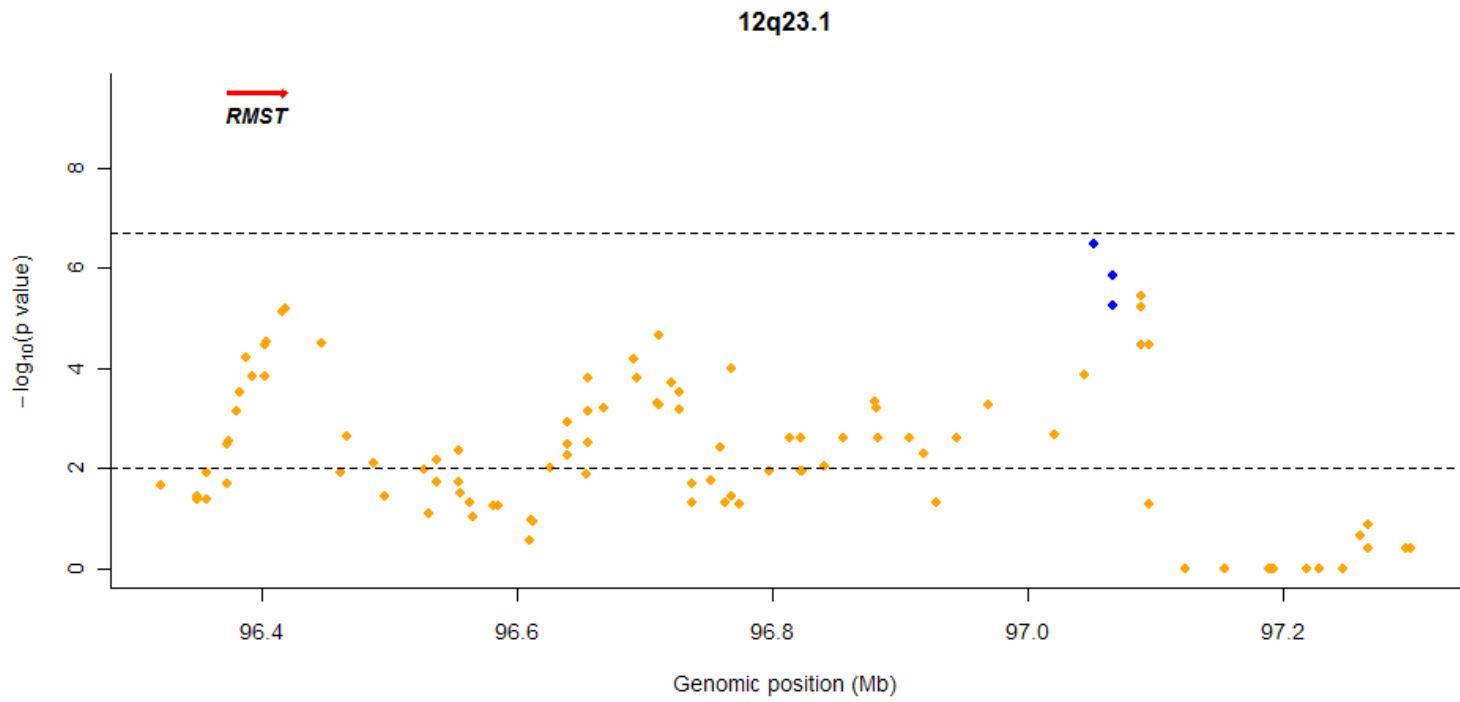
MYC, v-myc myelocytomatosis viral oncogene homolog (avian); *PVT1*, Pvt1 oncogene; *GSDMC*, gasdermin C; *FAM49B*, family with sequence similarity 49, member B; *ASAP1*, ArfGAP with SH3 domain, ankyrin repeat and PH domain 1; *DUSP6*, dual specificity phosphatase 6; *WDR51B* (*POC1B*), POC1 centriolar protein homolog B; *ATP2B1*, ATPase, Ca^{++} transporting, plasma membrane 1; *RMST*, rhabdomyosarcoma 2 associated transcript; *CA10*, carbonic anhydrase X; *KIF2B*, kinesin family member 2B; *CSMD1*, CUB and Sushi multiple domains 1; *UNC5D*, unc-5 homolog D (*C. elegans*).

A**B**

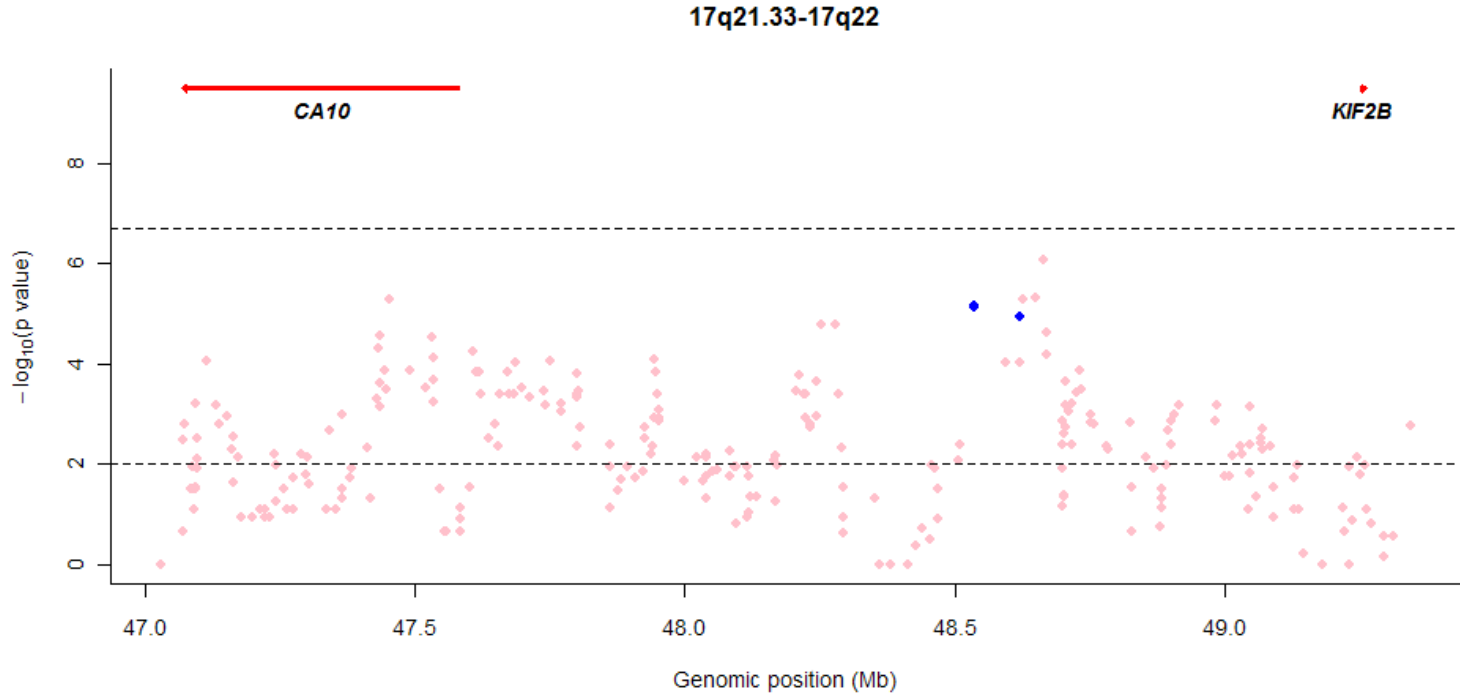
C



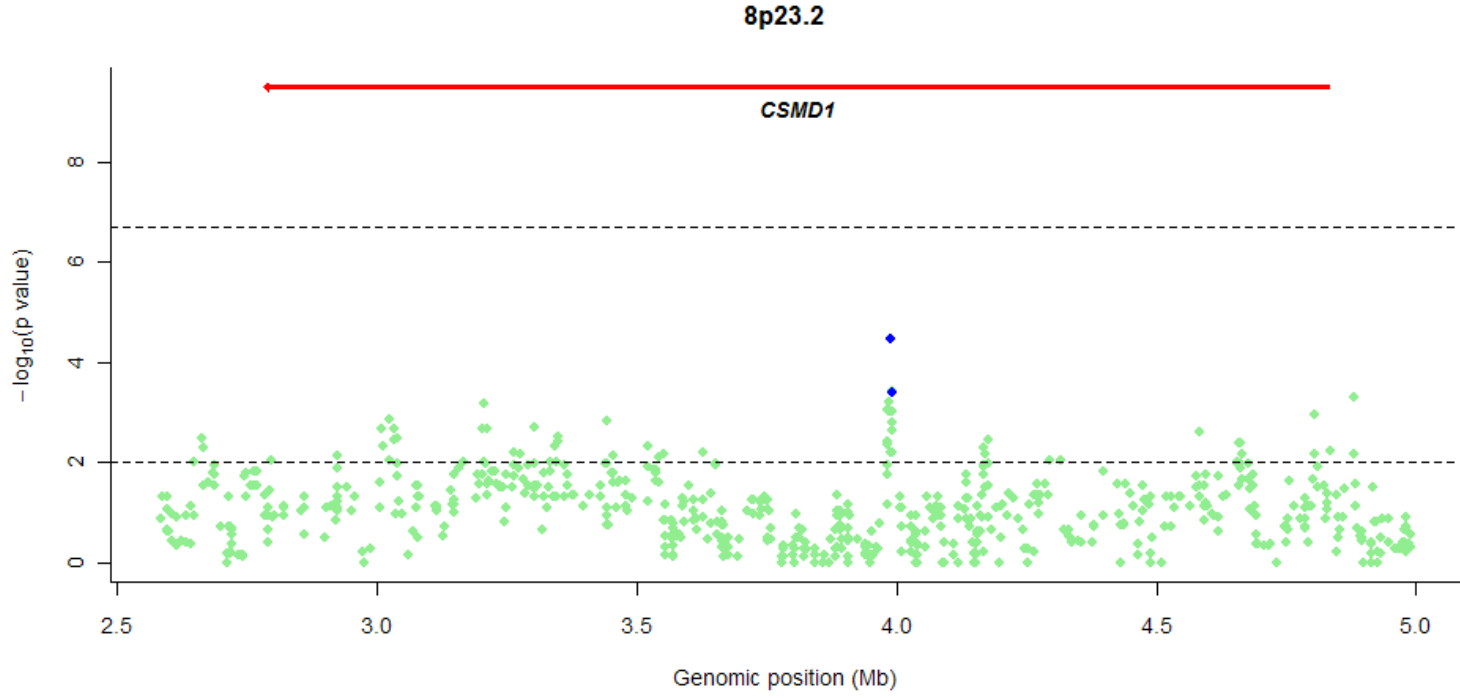
D



E



F



G

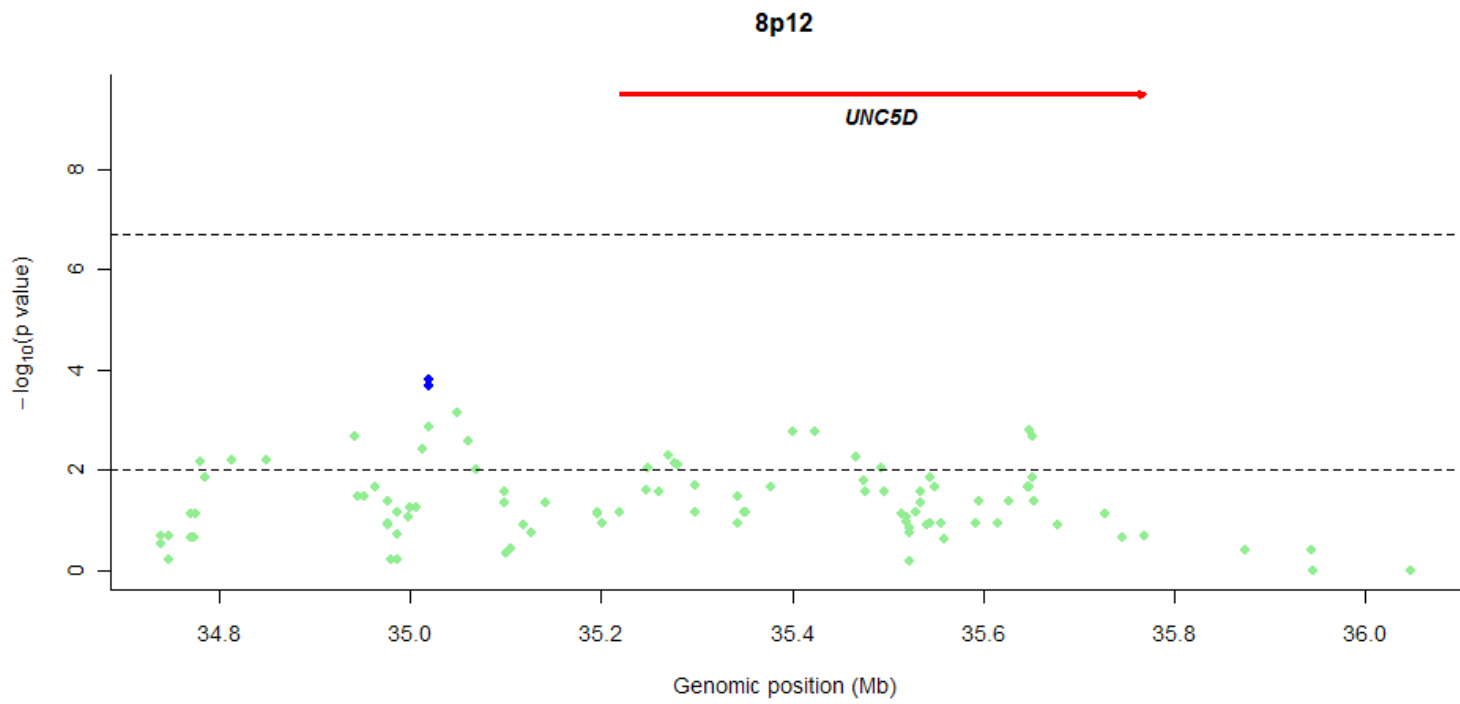
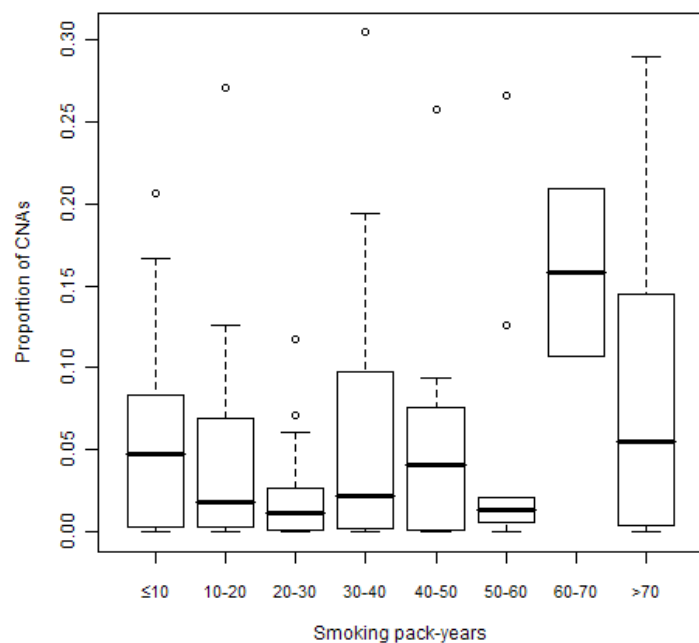
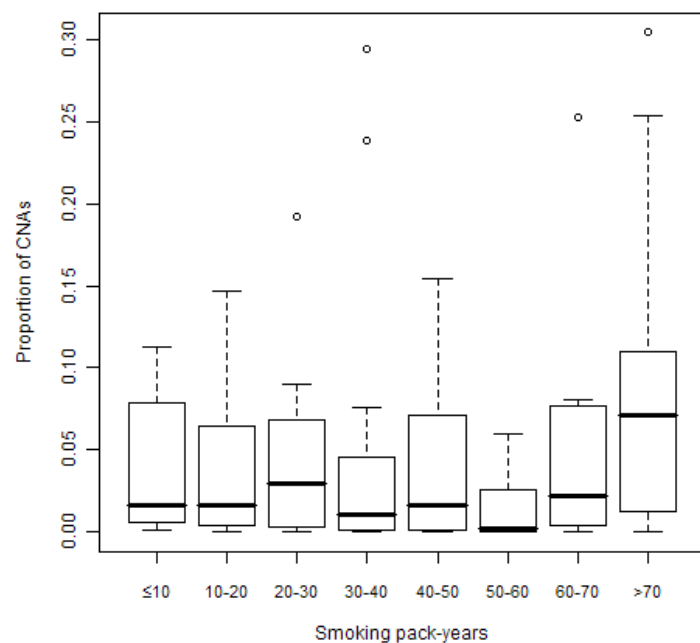


Figure S9. Whole-genome copy number pattern of every 10 smoking pack-years.

Total probes with events of copy number gains by pack-years of cigarette smoking in the discovery set (**A**), the validation set (**B**) and both sets (**C**).

A**Discovery set, copy number gain****B****Validation set, copy number gain****C****Both sets, copy number gain**