Supplementary information for

Germline Functional Variants Contribute to Somatic Mutation and Outcomes in Neuroblastoma

Eun Seop Seo[#1,2], Ji Won Lee[#1], Jinyeong Lim[3], Sunghwan Shin[4], Hee Won Cho[1], Hee Young Ju[1], Keon Hee Yoo[1], Ki Woong Sung[*1] and Woong-Yang Park[*2,3,5]

[1]Department of Pediatrics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

[2]Department of Digital Health, Samsung Advanced Institute for Health Science and Technology, Sungkyunkwan University, Seoul, Korea

[3]Samsung Genome Institute, Samsung Medical Center, Seoul, Korea

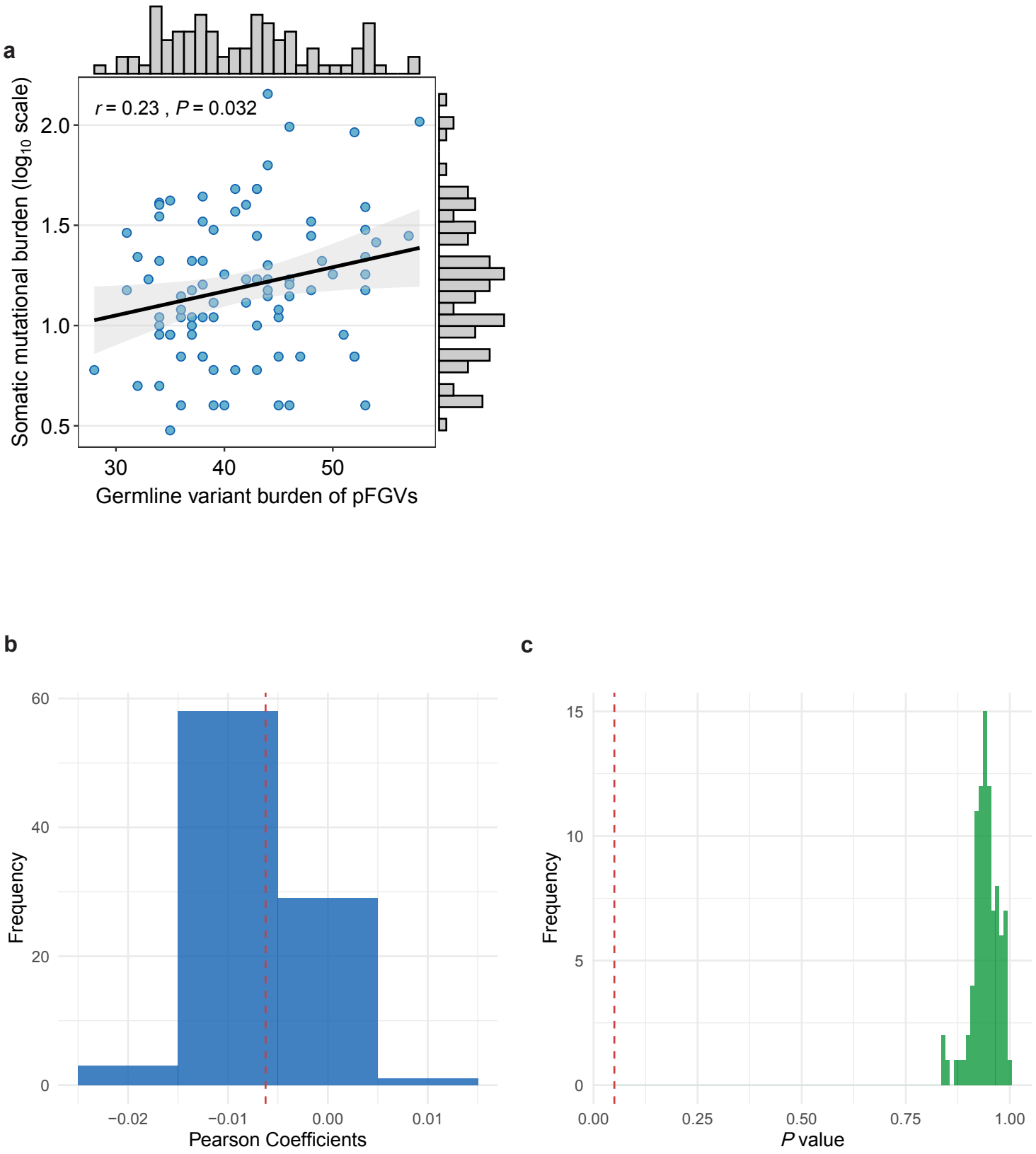
[4]Department of Laboratory Medicine, Inje University Ilsan Paik Hospital, Goyang, Korea [5]Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon, Korea

[#]Eun Seop Seo and Ji Won Lee contributed equally to this work.

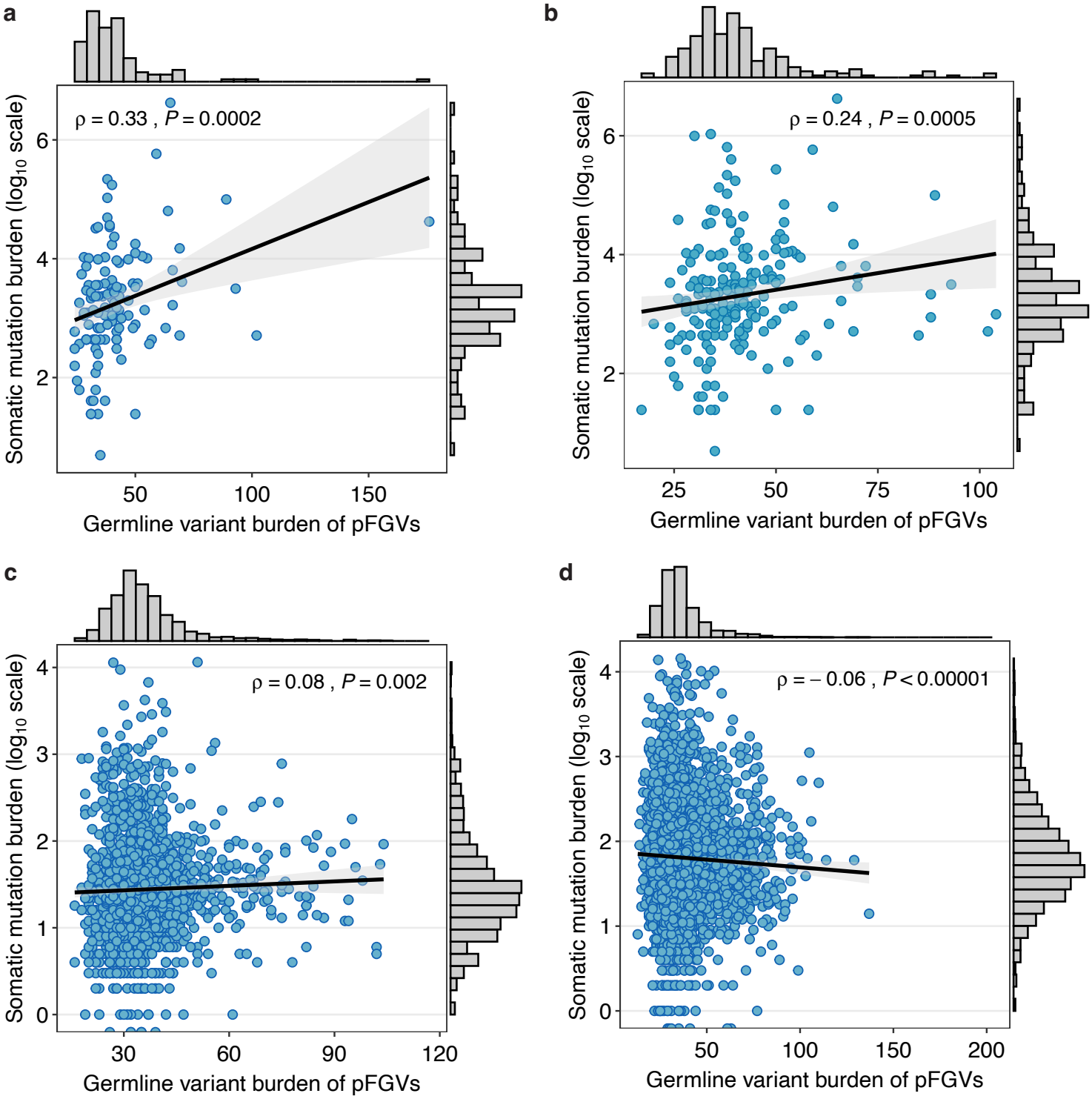
Correspondence Authors: Woong-Yang Park, Samsung Genome Institute, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea, E-mail: woongyang.park@samsung.com and Ki Woong Sung, Department of Pediatrics, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea, E-mail: kwsped@skku.edu

This file includes

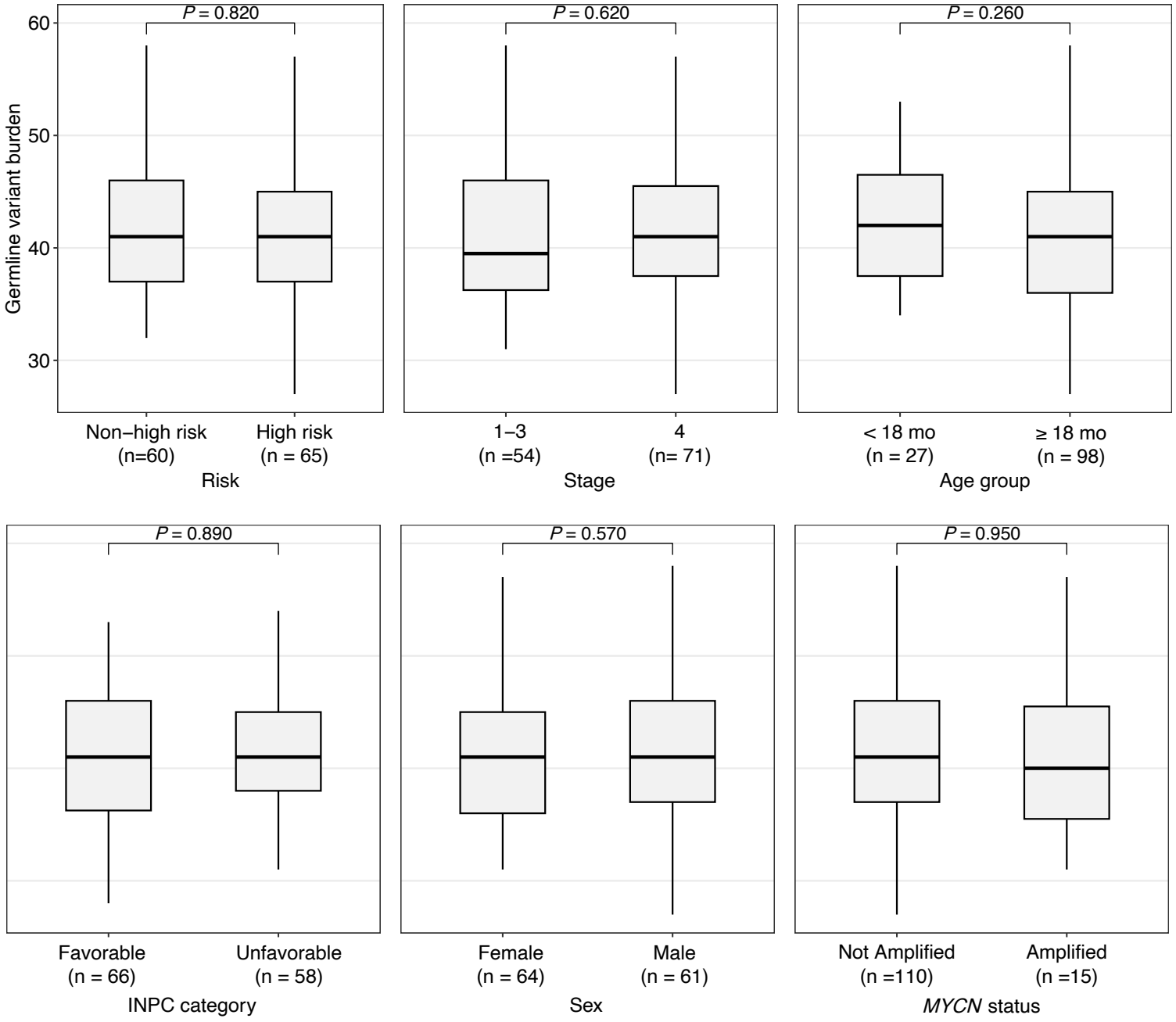
- Supplementary Figures 1 to 8
- Supplementary Tables 1 to 2



Supplementary Fig. 1 Correlation between germline variant burden and somatic mutational burden.
a, Correlation in patients without pFGVs in DDR genes in SMC cohort. The *r* represents Pearson's correlation coefficient and the black line represents the fitted values from linear regressions, with 95% confidence intervals in grey. **b**, Histogram of the two-sided Pearson coefficients from down-sampling analysis in synonymous germline variants. The red dashed line indicates mean of coefficient. **c**. Histogram of *P* value from down-sampling analysis in synonymous germline variants. The red dashed line indicates a *P* value of 0.05. Source data are provided as a Source Data file.

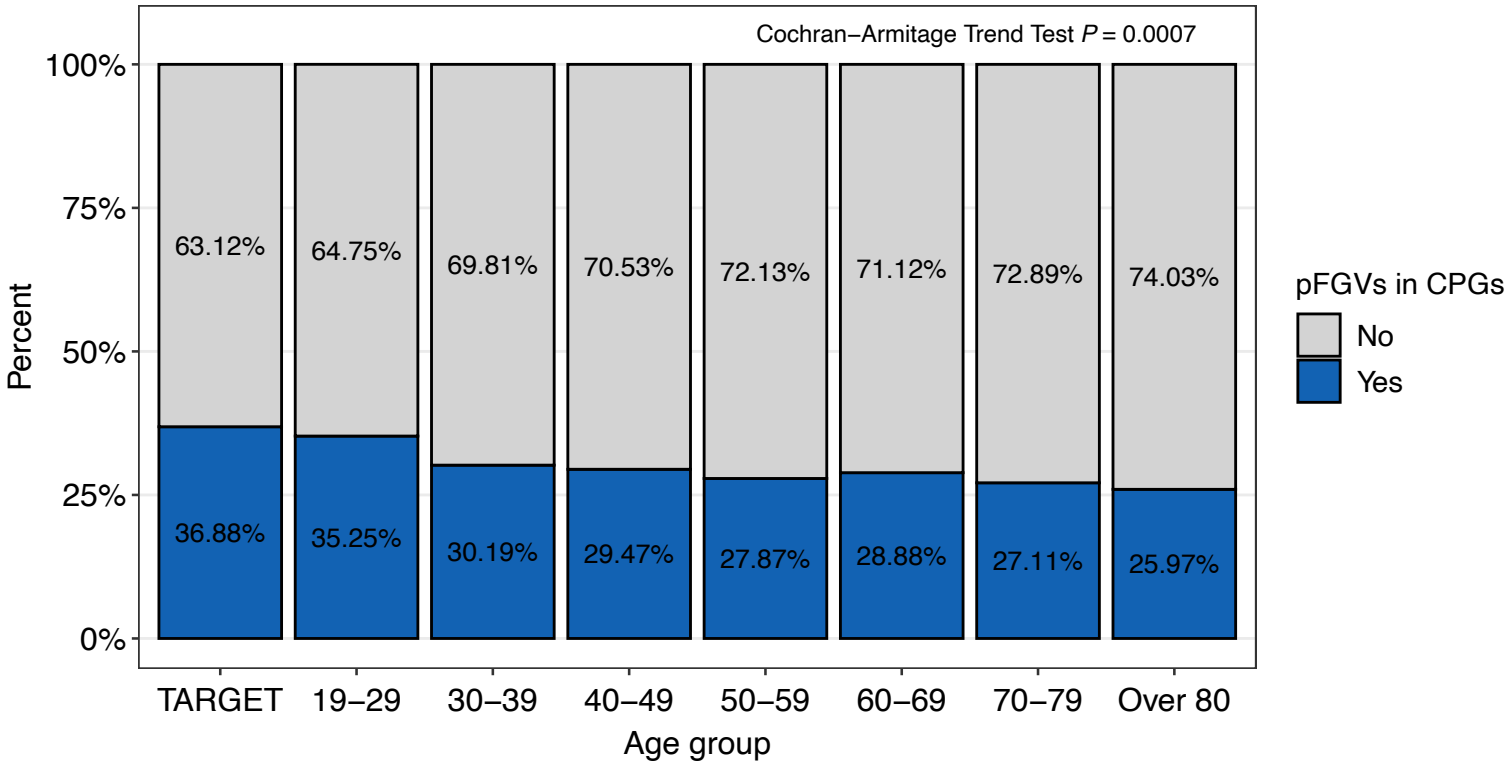


Supplementary Fig. 2 Correlation between germline variant burden and somatic mutational burden in TARGET and TCGA cohort (subgroup analysis).
a, Correlation in patients without pFGVs in DDR genes (TARGET). **b**, Correlation in patients after excluding outliers determined by a Z-score threshold of 3 (TARGET). **c**, Correlation in early-onset cancer (TCGA). **d**, Correlation in late-onset cancer (TCGA). For all the scatter plots, the ρ represents Spearman's correlation coefficient and the black line represents the fitted values from linear regressions, with 95% confidence intervals in grey. Statistical analysis was performed using two-sided tests without correction for multiple comparisions. Source data are provided as a Source Data file.

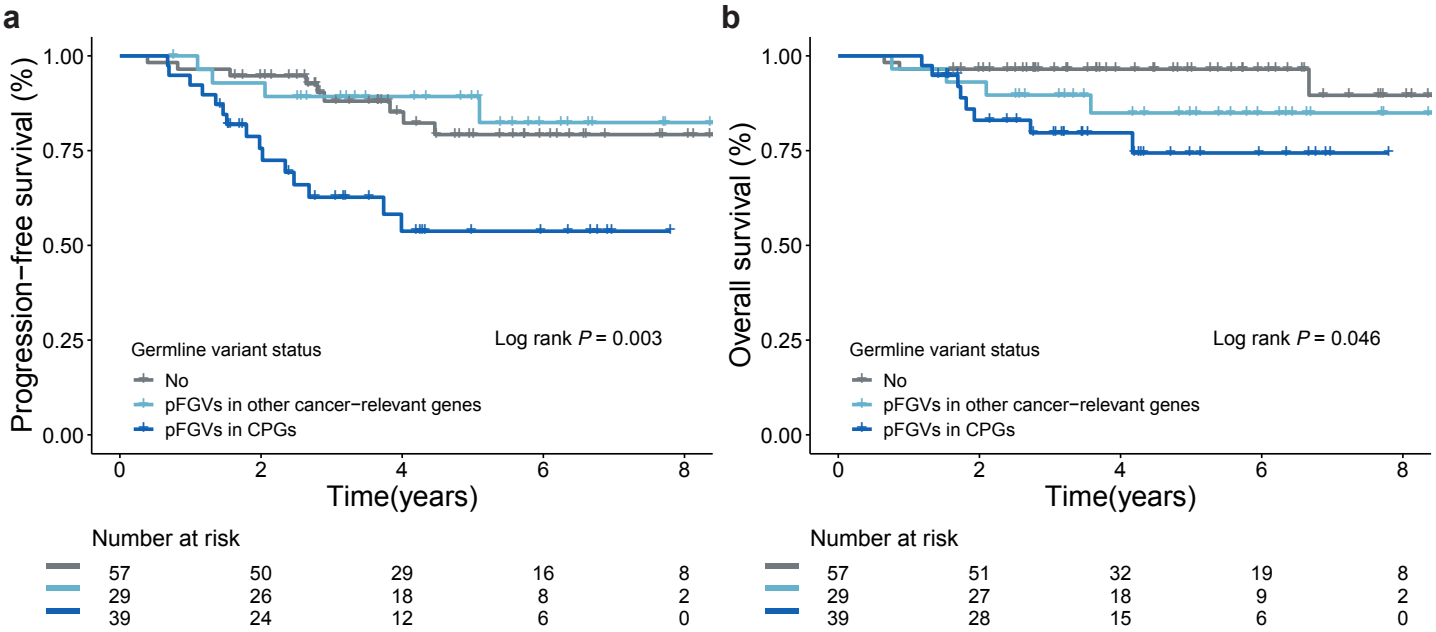


Supplementary Fig. 3 Box plot comparison of germline variant burden based on clinical risk factors.

Each box plot displays the median value as the center line, the upper and lower box boundaries at the first and third quartiles (25th and 75th percentiles). All *P* values are derived from two-sided test. Source data are provided as a Source Data file.

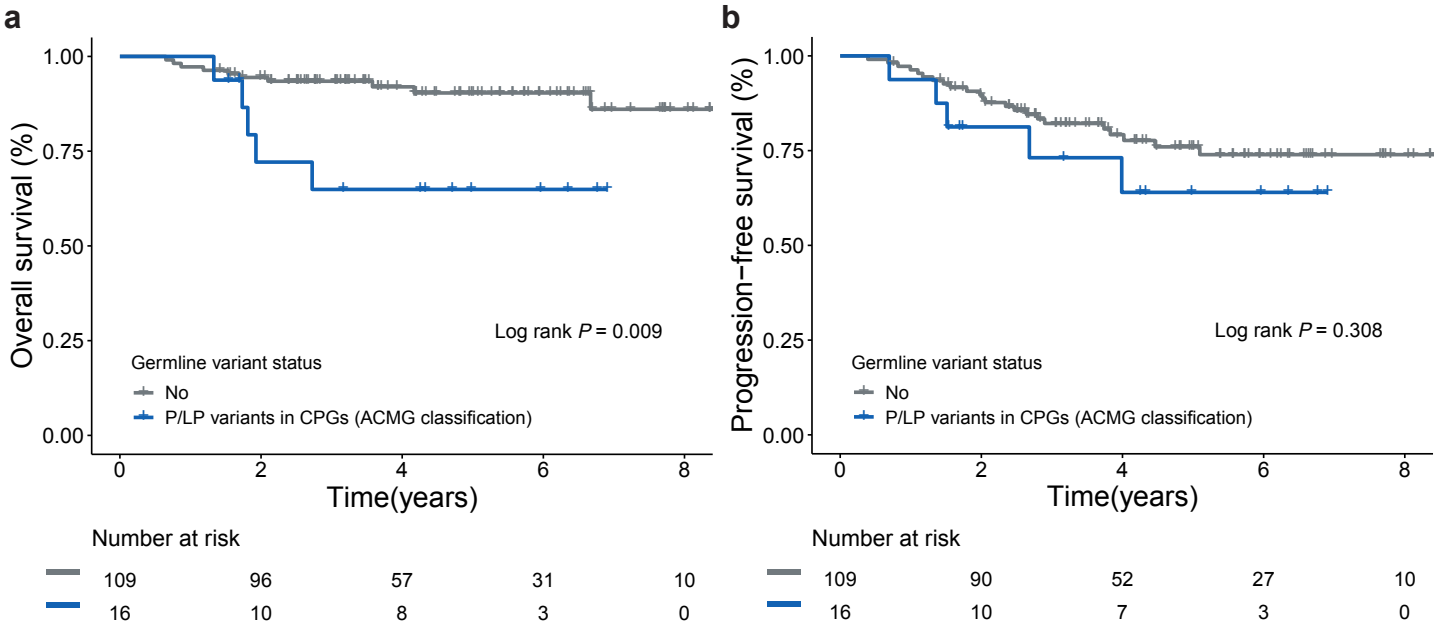


Supplementary Fig. 4 Prevalence trends of pFGVs in CPGs across age groups in the TARGET and TCGA cohort.
Source data are provided as a Source Data file.



Supplementary Fig. 5 Kaplan-Meier survival curves of pFGVs in CPGs and other cancer-relevant genes in the SMC cohort.

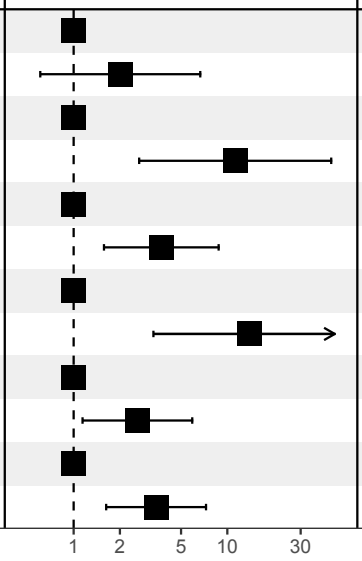
a, PFS, **b**, OS. All *P* values are two-sided without correction for multiple comparisons. Source data are provided as a Source Data file.



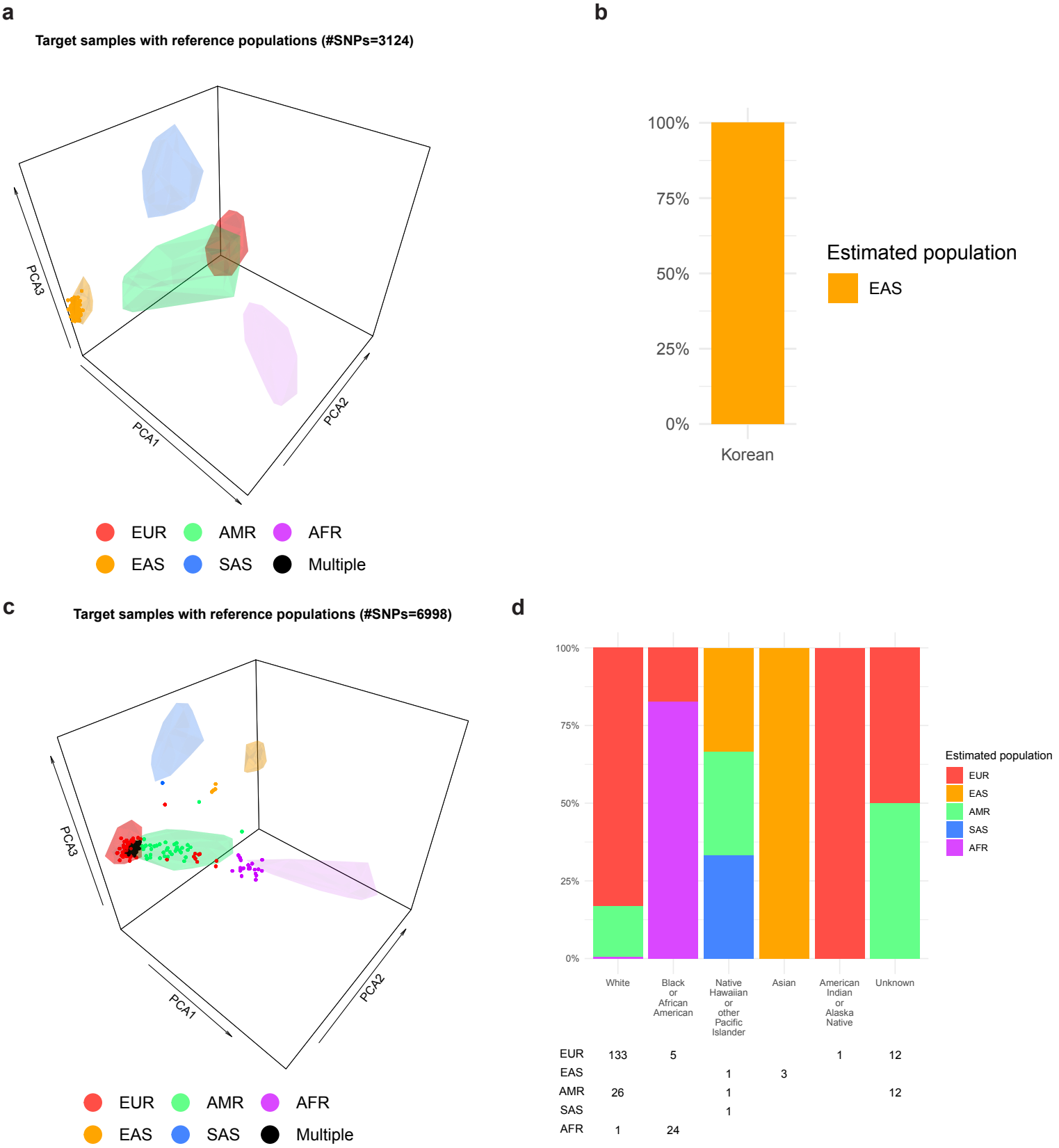
Supplementary Fig. 6 Kaplan-Meier survival curves of P/LP variants presence in CPGs in the SMC cohort.

a, OS. **b**, PFS. All *P* values are two-sided without correction for multiple comparisons. Source data are provided as a Source Data file.

Variable		N	Hazard ratio	*P* value
Age	< 18 mo	27	Reference	
	≥ 18 mo	98	2.01 (0.61, 6.67)	0.254
Stage	1¹3	54	Reference	
	4	71	11.25 (2.67, 47.43)	< 0.001
***MYCN* status**	Not amplified	110	Reference	
	Amplified	15	3.72 (1.58, 8.78)	0.003
Risk	Non¹high risk	60	Reference	
	High risk	65	13.93 (3.30, 58.71)	< 0.001
Germline variant burden	Low	59	Reference	
	High	66	2.60 (1.14, 5.91)	0.023
pFGVs in CPGs	No	86	Reference	
	Yes	39	3.44 (1.63, 7.27)	0.001



Supplementary Fig. 7 Univariable Cox regression analysis for PFS in the SMC cohort.
All *P* values are two-sided without correction for multiple comparisons. Source data are provided as a Source Data file.



Supplementary Fig. 8 Ancestry analysis using genotype principal components analysis (PCA)-based approach.

a, Three-dimensional representation of the first three PCs for the SMC cohort. **b**, Bar plot of estimated population proportions by ancestry analysis according to reported ethnicity in the SMC cohort. **c**, Three-dimensional representation of the first three PCs. **d**, Bar plot of estimated population proportions by ancestry analysis according to reported ethnicity in the TARGET cohort. Individuals who were annotated with multiple ancestry was categorized by the most contributory ethnicity. Source data are provided as a Source Data file.

Supplementary Table 1.
Patient and Disease Characteristics

Characteristic	N = 125
Age at diagnosis	
< 18 months	27 (22%)
≥ 18 months	98 (78%)
Sex	
Female	64 (51%)
Male	61 (49%)
INSS stage	
1	13 (10%)
2	30 (24%)
3	11 (8.8%)
4	71 (57%)
Histology	
Favorable	66 (53%)
Unfavorable	58 (47%)
Family history of cancer[a]	
No	28 (38%)
Yes	45 (62%)
Primary site	
Abdomen	88 (70%)
Mediastinum	37 (30%)
***MYCN* status**	
Not amplified	110 (88%)
Amplified	15 (12%)
Risk[b]	
Low	40 (32%)
Intermediate	20 (16%)
High	65 (52%)

[a]Family history of cancer were defined as presence of any cancer in at least one first or second-degree relative. [b]Stage 1, 2, and 4S tumors according to the INSS were stratified into the low-risk group if *MYCN* was not amplified, and stage 4 tumors in patients ≥ 18 months of age or any tumors with amplified *MYCN* were classified as the high-risk group.
Abbreviations: INSS, International Neuroblastoma Staging System

Supplementary Table 2.
Baseline Characteristics According to the Presence of pFGVs in CPGs

Characteristics	N	No, N = 86[1]	Yes, N = 39[1]	*P* value
Age	125			0.842
< 18 months		19 / 86 (22%)	8 / 39 (21%)	
³ 18 months		67 / 86 (78%)	31 / 39 (79%)	
Sex	125			0.252
Female		47 / 86 (55%)	17 / 39 (44%)	
Male		39 / 86 (45%)	22 / 39 (56%)	
Family history of cancer[a]	73	25 / 47 (53%)	20 / 26 (77%)	0.046
INSS Stage	125			0.467
1		11 / 86 (13%)	2 / 39 (5.1%)	
2		21 / 86 (24%)	9 / 39 (23%)	
3		6 / 86 (7.0%)	5 / 39 (13%)	
4		48 / 86 (56%)	23 / 39 (59%)	
Histology	124			0.762
Favorable		45 / 86 (52%)	21 / 38 (55%)	
Unfavorable		41 / 86 (48%)	17 / 38 (45%)	
Primary site	125			0.282
Abdomen		58 / 86 (67%)	30 / 39 (77%)	
Mediastinum		28 / 86 (33%)	9 / 39 (23%)	
***MYCN* Status**	125			>0.999
Not amplified		76 / 86 (88%)	34 / 39 (87%)	
Amplified		10 / 86 (12%)	5 / 39 (13%)	
Risk	125			0.478
Low		30 / 86 (35%)	10 / 39 (26%)	
Intermediate		12 / 86 (14%)	8 / 39 (21%)	
High		44 / 86 (51%)	21 / 39 (54%)	
Progression	125			0.004
No		73 / 86 (85%)	24 / 39 (62%)	
Yes		13 / 86 (15%)	15 / 39 (38%)	
Death	125			0.072
No		79 / 86 (92%)	31 / 39 (79%)	
Yes		7 / 86 (8.1%)	8 / 39 (21%)	

Somatic mutational burden	125	25.01	20.92	0.654
Germline variant burden	125	41.14	42.56	0.137

[1]n / N (%)

[a]Family history of cancer were defined as presence of any cancer in at least one first or second-degree relative.

All *P* values are derived from two-sided test without correction for multiple comparisons.

Abbreviations: pFGV, putatively functional germline variant; CPG, cancer predisposition gene; INSS, International Neuroblastoma Staging System