1	AN	ASSESSMENT OF THE VALUE OF DEEP NEURAL NETWORKS
2	IN	GENETIC RISK PREDICTION FOR SURGICALLY RELEVANT
3		OUTCOMES
4	Mathia	as A Christensen, MD ^{1,2} , Arnór Sigurdsson, MSc ^{3,4} , Alexander Bonde, MD ^{1,2} , Simon
5	Rasmu	ussen PhD ^{3,4} , Sisse R Ostrowski ^{5,6} , MD, PhD, DMSc, Mads Nielsen ⁷ PhD and Martin
6	Sillese	n* MD, PhD ^{1,2,5}
7	1)	Department of Organ Surgery and Transplantation.
8		Copenhagen University Hospital, Rigshospitalet, Denmark
9	2)	Center for Surgical Translational and Artificial Intelligence Research (CSTAR)
10		Copenhagen University Hospital, Rigshospitalet, Denmark.
11 12 13	3)	Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
13 14 15 16	4)	The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
10	5)	Department of Clinical Medicine, Faculty of Health and Medical Sciences, University
18		of Copenhagen Medical School, Copenhagen, Denmark.
19	6)	Department of Clinical Immunology, Rigshospitalet, Copenhagen University
20		Hospital, Copenhagen, Denmark.
21	7)	Department of Computer Science, University of Copenhagen, Copenhagen, Denmark.
22	Corres	ponding Author:
23 24 25	Martin Associ	a Sillesen MD PhD Jate Professor of Surgery
25 26 27	Depart Blegda	magen University Hospital, Rigsnospitalet tment of Surgery and Transplantation amsvej 9
28	2100 0	Copenhagen O
29	Denma	ark
30	Mail: I	Martin.Sillesen@regionh.dk

31 Phone: +45 3545 2122 NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

32

Abstract

33 34 Introduction

35

36 Postoperative complications affect up to 15% of surgical patients constituting a major part of

the overall disease burden in a modern healthcare system. While several surgical risk 37

- 38 calculators have been developed, none have so far been shown to decrease the associated
- 39 mortality and morbidity. Combining deep neural networks and genomics with the already
- 40 established clinical predictors may hold promise for improvement.
- 41

42 **Methods**

43

The UK Biobank was utilized to build linear and deep learning models for the prediction of 44

- 45 surgery relevant outcomes. An initial GWAS for the relevant outcomes was initially
- 46 conducted to select the Single Nucleotide Polymorphisms for inclusion in the models. Model
- 47 performance was assessed with Receiver Operator Characteristics of the Area Under the
- 48 Curve and optimum precision and recall. Feature importance was assessed with SHapley
- Additive exPlanations. 49
- 50

51 **Results**

52

Models were generated for atrial fibrillation, venous thromboembolism and pneumonia as 53

- 54 genetics only, clinical features only and a combined model. For venous thromboembolism,
- 55 the ROC-AUCs were 59.6% [59.0%-59.7%], 63.4% [63.2%-63.4%] and 66.1% [65.7%-
- 66.1%] for the linear models and 60.0% [57.8%-61.8%], 63.2% [61.2%-65.0%] and 65.4% 56
- 57 [63.6%-67.2%] for the deep learning SNP, clinical and combined models, respectively. For 58 atrial fibrillation, the ROC-AUCs were 60.9% [60.6%-61.0%], 78.7% [78.7%-78.7%] and
- 59 80.1% [80.0%-80.1%] for the linear models and 59.9% [.6%-61.3%], 78.8% [77.8%-79.8%]
- and 79.4% [78.8%-80.5%] for the deep learning SNP, clinical and combined models, 60
- respectively. For pneumonia, the ROC-AUCs were 57.3% [56.5%-57.4%], 69.2% [69.1%-61
- 69.2%] and 70.5% [70.2%-70.6%] for the linear models and 55.5% [54.1%-56.9%], 69.7% 62
- [.5%-70.8%] and 69.9% [68.7%-71.0%] for the deep learning SNP, clinical and combined 63
- models, respectively. 64
- 65

Conclusion 66

67

68 In this report we presented linear and deep learning predictive models for surgery relevant 69 outcomes. Overall, predictability was similar between linear and deep learning models and

- 70 inclusion of genetics seemed to improve accuracy.
- 71
- 72
- 73
- 74
- 75
- 76

77

INTRODUCTION

 result of surgery, v an uneventful reco result of surgery, v hospital length-of- While treatment ac Recovery after Sur have remained rem still experiences Pe deviation from the move towards a pr To achieve this go 	the global burden of disease.(1, 2) While most surgical patients proceed to overy, current estimates indicate that roughly 4% die as a direct or indirect
 an uneventful reco result of surgery, v hospital length-of- While treatment at Recovery after Sur have remained rem still experiences P deviation from the move towards a pr To achieve this go 	overy, current estimates indicate that roughly 4% die as a direct or indirect
 81 result of surgery, v 82 hospital length-of- 83 While treatment ac 84 Recovery after Sur 85 have remained ren 86 still experiences Pe 87 deviation from the 88 move towards a pr 89 To achieve this go 	
 kospital length-of- While treatment at Recovery after Sur have remained rem still experiences Per deviation from the move towards a pr To achieve this go 	while up to 15% experience a postoperative complication (PC), prolonging
 83 While treatment at 84 Recovery after Sur 85 have remained rem 86 still experiences Point 87 deviation from the 88 move towards a pr 89 To achieve this go 	-stay with consequential morbidity.(2)
 Recovery after Sur have remained rem still experiences Per deviation from the move towards a pr To achieve this go 	dvances following the implementation of approaches such as Enhanced
 have remained ren still experiences P deviation from the move towards a pr To achieve this go 	rgery (ERAS) protocols have been well documented, the incidences of PCs
 86 still experiences P 87 deviation from the 88 move towards a pr 89 To achieve this go 	narkably stable over the last decade.(3) As such, a stable subset of patients
87 deviation from the88 move towards a pr89 To achieve this go	Cs, suggesting that this patient group could potentially benefit from a
88 move towards a pr89 To achieve this go	e current one-size-fits all approach deployed by most ERAS protocols and a
89 To achieve this go	recision medicine approach in the surgical setting.
	al, risk predictions models are, however, needed for the identification of
90 patients who will f	fail standard ERAS protocols.
91 To this end, many	risk assessment tools have been fielded to identify at-risk patients
92 including the regre	ession-based American College of Surgeons National Surgical Quality
93 Improvement Prog	gram (ACS-NSQIP) risk calculator as well as newer machine learning
94 approaches investi	igating the value of random forests or deep neural networks (DNNs).(4, 5)
95 These models are,	however, limited by the fact that they only perform predictions on
96 available clinical d	data, which may provide insights into parts of the driving factors of the
97 patients risks only	•
98 As such, recent da	ta has suggested that genetic susceptibility could, in part, be a modifier of
99 individual PCs risk	

100 prediction models in order to improve model performance.(6, 7)

101 Genetic variations are increasingly being recognized as an important modality for various 102 surgical adverse events including venous thromboembolisms, renal complications and cardiac arrythmias.(6, 8, 9) However, it is currently not clear to what degree genetic susceptibility 103 104 contribute to the overall risk compared with other well-known clinical risk factors. 105 Furthermore, as genetic susceptibility may include complex non-linear effects such as 106 previously non-identified complex interactions between genes that possibly lie far from each 107 other in the human genome, optimal modelling strategies remain unknown. As such, whether legacy risk prediction approaches such as the linear Polygenic Risk Scores (PGS), 108 109 traditionally utilized to assess an overall genetic risk composition and weighted sum for the phenotype in question, could be inferior to a DNN approach, is currently unknown.(10) 110 111 Using the clinical question of assessing whether DNNs can outperform a classic PGS 112 approach for assessing genotype-associated risk of PCs, we target three high impact PCs with 113 proven genetic susceptibility.(11) These include postoperative pneumonias, postoperative 114 venous thromboembolisms (pVTEs) and postoperative atrial fibrillation (pAFLI). 115 Furthermore, we investigate whether single nucleotide polymorphisms (SNPs) highlighted as 116 driving the phenotype, differ between DNN and PGS approaches, thus potentially indicating 117 that non-linear genotype-phenotype associations can be identified by the DNN approach. We hypothesize that DNNs will achieve superior predictive performance in predicting the 118 119 genotype-associated risk of these PCs compared with a linear PGS, and that the DNN models 120 will highlight a different subset of important SNPs compared with a linear PGS.

1	2	1	

METHODS

122	This study utilized genotype data from the United Kingdom biobank (UKB) consortium.(12)
123	Access to the UKB data was approved by the consortium (Study ID #60861). Under Danish
124	law, the study was exempt from ethical board approval due to the anonymized nature of the
125	dataset.
126	We conducted a comparative study of different methodologies for genotyping risk prediction
127	and Single Nucleotide Polymorphism (SNP)-identification in a general as well as a surgical,
128	national cohort.
129	For the initial approach, we conducted standard GWAS-analyses without covariates on the
130	chosen phenotypes with a high prevalence following surgery. Details for the GWAS are
131	described below. These phenotypes included venous thromboembolisms (VTE), atrial
132	fibrillation (AF) and bacterial pneumonia.
133	UKB has more than 500,000 individuals enrolled and consented across the United Kingdom
134	of the age from 40 to 69. Patients were invited for participation through National Health
135	service (NHS) registries and asked to fill surveys on basic demographic data, general lifestyle
136	measures as well as medical history. Inclusion of all participants took place from 2006 to
137	2010.
138	
139	Identification of cohort

All patients with available genomic data in the UKB were initially included foranalysis. Cases were identified depending on the phenotype in question. For AF, VTE and

142 pneumonia, cases were defined using relevant International Statistical Classification of

143 *Disease*, 9th revision (ICD-9) and ICD-10 codes.

144 The phenotypes in question were identified with the ICD-9 and ICD-10 codes listed in

supplementary table 1. The cohorts were split into training/validation and test sets. The

146 training/validation set consisted of all non-surgical patients and a random sample of 80% of 147 the surgical cohort. The test set consisted of the remaining 20% of the surgical cohort. Surgery was defined with the OPCS-4 codes listed in supplementary table 2. The post-148 149 surgical phenotypes were defined with the same ICD-codes as above registered up to 30 days 150 after the given procedure. For AF, only first-time diagnoses were counted as post-surgery 151 cases. For VTE and pneumonia, any diagnoses within 30 days were counted as cases, 152 regardless of previous history. For each outcome of interest (pAFLI, pVTE and pneumonia, both deep learning and linear 153 154 models were created using three distinct input strategies (see below for model descriptions): 155 1. A genotype only model: using only the identified SNPs (see below) as input (SNP 156 model) 157 2. A clinical data only model: using only clinical data as input (Clinical model) 3. A combined model: using both SNPs and clinical data as input (Combined model) 158 159 Input SNPs were the top 100 SNPs from the discovery GWAS for each phenotype of interest, 160 with clinical data including demographics and comorbidities (supplementary table 3) and combined models including both genetic and clinical data. 161 162 *Quality control* The first 50,000 individuals included in UKB were genotyped using the Applied 163 164 Biosystems UK BiLEVE Axiom Array. The remaining were genotyped using the Applied 165 Biosystems UK Biobank Axiom Array. The two array types are equal, and the differences are not of significance. The arrays interrogated 850,000 SNPs in total. To account for potential 166 biases, patients with outlying heterozygosity rates, cryptic relatedness (PIHAT cut-off 0.2) 167

168 and sex discrepancies in data were excluded. To ensure that only participants with high-

169 quality genomic information were included for analysis, everyone with a genotyping rate of

170 98% or less were excluded. To ensure that only high-quality genetic variants were left for

171	analyses, a missingness rate of 2% were used as a cut-off point. Lastly, a Minor Allele
172	Frequency (MAF) of $> 5\%$ was used, and variants found not to be in Hardy-Weinberg
173	equilibrium were excluded (threshold: 1×10^{-6} for both cases and controls).
174	
175	Linear Polygenic risk score (PGS) modelling approach
176	The initial GWAS-analyses were analyzed using a mixed linear model (MLM)
177	approach. GCTA version 1.93 beta for Windows was used to conduct the analyses. The
178	MLM-model was created using fastGWA with a sparse genetic relationship matrix (GRM)
179	with non-imputed data from the UKB. For all the phenotypes analyzed in the respective
180	GWAS, the 100 most significant SNPs were included in the genetic and mixed models. The
181	choice to utilize only the top 100 SNPs was made to optimize the balance between predictive
182	power and keeping the model computational pragmatic. SNPs are referenced using the
183	dbSNP (rs) reference number. The cohorts were split into training/validation and test sets.
184	Relevant GWAS plots, including Manhattan and Quantile-Quantile (QQ) plots were
185	generated using qqman (R version 4.0.2).(13) Performance plots including ROC-AUCs and
186	heatmaps were created using Scikit-learn 1.2.1 (Python 3).(14)
187	A linear PGS was generated using the logistic regression module as implemented in scikit-
188	learn 1.2.1 for Python 3. Models were created with both L1 (lasso) and L2 (ridge)
189	regularization. Feature importance was determined by coefficients of the SNPs.
190	
191	Deep neural network (DNN) modelling approach
192	All DNN models were implemented using EIR (version 0.1.25-alpha).(15) EIR is a
193	framework that incorporates genetic, clinical, image, sequencing, and binary data for
194	supervised training of deep learning models. A held-out test set was used for all models to

195 obtain a final performance after training and validation. The Cross Entropy loss was

employed during training for the classification tasks. All models were trained with a batch 196 size of 64. During training, plateau learning rate scheduling was used to reduce the learning 197 rate by a factor of 0.2 if the validation performance had not improved for 10 steps, with a 198 199 validation interval of 500 steps. Early stopping was used to terminate training when performance had not improved with a patience of 16 steps. The early stopping criterion was 200 201 activated after a buffer of 2,000 iterations. All models were trained with the Adam optimizer with a weight decay of 1×10^{-4} and a base learning rate of 1×10^{-3} .(16) For the neural network 202 models, we augmented the genotype input by randomly setting 40% of the SNPs as missing 203 204 in the one-hot encoded array. All DNN models utilize the genome-localnet (GLN) architecture for the genotype feature extraction.(15) The same cohort splits were used as in 205 206 the linear PGS-approach. Importance of features were determined using SHapley Additive 207 exPlanations (SHAP) values.(17)

209	RESULTS
210	Cohort
211	We identified 488,377 patients in the UKB with available genetic and relevant phenotypic
212	data, with 446,180 patients available for analyses after genetic quality measures were applied
213	and were used for both the linear and deep learning modelling approaches.
214	For the outcomes of interest, 19,704 had a diagnosis of AF, 9,101 had a diagnosis of VTE
215	and 13,757 had a diagnosis of pneumonia overall in the UKB.
216	
217	Linear models
218	Atrial fibrillation
219	Baseline characteristics are listed in table 1. The SNP model reached a ROC-AUC of 60.9%
220	[95% CI, 60.6%-61.0%]. All individuals were classified as not having AF. The clinical model
221	reached a ROC-AUC of 78.7% [95% CI, 78.7%-78.7%] with a recall of 9% and a precision
222	of 53%. The combined model reached a ROC-AUC of 80.1% [95% CI, 80.0%-80.1%] with a
223	recall of 9% and a precision of 57%. All performances are depicted in figure 3A. The SNPs
224	and the associated genes with the highest feature importance are listed in table 4B.
225	
226	Venous thromboembolism
227	Baseline characteristics for VTE are listed in table 2. The SNP model reached a ROC-AUC
228	of 59.6% [95% CI, 59.0%-59.7%]. All individuals were classified as not having VTE. The
229	clinical model reached a ROC-AUC of 63.4% [95% CI, 63.2%-63.4%]. All individuals were
230	classified as not having VTE. The combined model reached a ROC-AUC of 66.1% [95% CI,
231	65.7%-66.1%]. All individuals were classified as not having VTE. All performances are
232	depicted in figure 3B. The SNPs and the associated genes with the highest feature importance
233	are listed in table 4D.

2	2	
,	-	/
2	J	-

235 Pneumonia

236	Baseline characteristics are listed in table 3. The SNP model reached a ROC-AUC of 57.3%
237	[95% CI, 56.5%-57.4%]. All individuals were classified as not having pneumonia. The
238	clinical model reached a ROC-AUC of 69.2% [95% CI, 69.1%-69.2%]. All individuals were
239	classified as not having pneumonia. The combined model reached a ROC-AUC of 70.5%
240	[95% CI, 70.2%-70.6%] with a recall of 0.01% and a precision of 0.4%. The SNPs and the
241	associated genes with the highest feature importance are listed in table 4F. All performances
242	are depicted in figure 3C.
243	
244	Deep learning models
245	Atrial fibrillation
246	The SNP model reached a ROC-AUC of 59.9% [95% CI, 58.6%-61.3%] in the test set.
247	Recall was 36.9% and precision was 9.3%. The clinical model reached a ROC-AUC of
248	78.8% [95% CI, 77.8%-79.8%] with recall and precision of 72.0% and 13.5%, respectively.
249	The combined model reached a ROC-AUC of 79.4% [95% CI, 78.8%-80.5%] with a recall
250	and precision of 74.8% and 13.5%, respectively. The SNPs and the associated genes with the
251	highest feature importance are listed in table 4A. All performances are depicted in figure 3A.
252	Hard predictions are depicted in figure 1A. ROC-AUC development is depicted in figure 2A.

- 253
- 254 *VTE*

The SNP model reached a ROC-AUC of 60.0% [95% CI, 57.8%-61.8%] with a recall of
50.8% and precision of 4%. The clinical model reached a ROC-AUC of 63.2% [95% CI,
61.2%-65.0%] with a recall and precision of 67.5% and 4.0%, respectively. The combined
model reached a ROC-AUC of 65.4% [95% CI, 63.6%-67.2%] with a recall and precision

259	68.8% and 4.0%, respectively. The SNPs and the associated genes with the highest feature
260	importance are listed in table 4C. All performances are depicted in figure 3B. Hard
261	predictions are depicted in figure 1B. ROC-AUC development is depicted in figure 2B.
262	
263	Pneumonia
264	The SNP model reached a ROC-AUC of 55.5% [95% CI, 54.1%-56.9%] with a recall of
265	55.0% and precision of 5%. The clinical model reached a ROC-AUC of 69.7% [95% CI,
266	68.5%-70.8%] with a recall and precision of 67.7% and 7.4%, respectively. The combined
267	model reached a ROC-AUC of 69.9% [95% CI, 68.7%-71.0%] with a recall and precision of
268	70.1% and 7.3%, respectively. The SNPs and the associated genes with the highest feature
269	importance are listed in table 4E. All performances are depicted in figure 3C. Hard
270	predictions are depicted in figure 1C. ROC-AUC development is depicted in figure 2C.
271	

2	7	2
2	1	Z

DISCUSSION

In this study, we assessed the performance of linear and deep learning models including genotypic information on specific phenotypes relevant for pAFLI, pVTE and postoperative pneumonia. Overall, we found that adding SNP data to clinical risk prediction models enhanced the predictive power, and that the GLN approach seemed superior to a legacy linear risk prediction approach.

278

279 *Modelling approaches*

280 All three SNP linear models failed to make any meaningful hard predictions, as they classified all individuals in the cohorts as not having the disease in question. However, the 281 separations were roughly similar to the GLN-models, as demonstrated by similarity in ROC-282 283 AUC performance, and the lack of positive predictions may be due to imbalanced data and 284 skewed threshold for hard predictions. The GLN-models performed better on recall and 285 precision, and it was able to classify positives correctly with just genomic information. Given 286 that the linear and GLN models utilize distinct tuning parameters for hard predictions, a 287 direct comparison of recall and precision may not be critically significant. However, as hard 288 predictions are necessary in a clinical setting, a discussion is still warranted. Precision-recall curves are depicted in supplementary materials. 289

290 It is exemplified for AF, where the genetic linear model had a ROC-AUC of 60.9% [95% CI

291 69.6%-61.0%] while the GLN had a ROC-AUC of 59.9% [95% CI 5.8.6%-61.3%]. The

recall, however, was 0% and 36.9%, respectively. As the linear model performed very poorly

in terms of recall, the result is likely to be the result of the imbalanced data and failure to

294 capture feature relevance and possibly non-linearity and would need optimization before any

form of utilization for positive prediction in a clinical setting. On the contrary, the GLN-

model had a recall of 36.9% and therefore identifies around one third of cases correctly,

297 which heightens the likelihood of clinical meaningful utilization considering only SNPs were 298 included in the model. However, the precision was calculated low at 4%, which would lead to overdiagnosis and possibly overtreatment. If used in clinical practice, it is therefore of 299 300 paramount importance that any possible intervention would carry little to no risk of harm. It, 301 however, cannot be ruled out that the differences between the models are not due to an inherent predictive advantage in the GLN-model, but simply due to different hyperparameter 302 303 tuning. When combining SNP and clinical data for all phenotypes in question, we observed a trend 304 305 towards better performance compared with SNP or clinical data only models, although most 306 confidence intervals were overlapping with clinical model performances. This thus indicates 307 that limited performance gains could obtained by combining genetic and clinical data and 308 may suggest that genotype effects may already in part be captured by diagnoses codes. 309 Alternatively, the lack of performance improvement could be affected by limited study power 310 due to factors such as lack of correct PC diagnoses codes, a problem often encountered when 311 administrative codes are used for PC curation.(18) 312 GLN based models did, however, outperform linear approaches in terms of recall performance, might indicate that the ability to capture the effects of non-linear genetic traits 313 on the overall phenotype, may be possible through this modelling approach. 314 315 316 Identified Single Nucleotide Polymorphisms In the GLN-model, rs3807989 was the most activated SNP in regards of classifying 317

318 individuals with AF. It is an intron variant in *CAV1* which codes for a main component in

319 caveolae plasma membrane and further acts as a tumor suppressor.(19, 20) It has been

associated with a large variety of diseases including AF in numerous populations.(21, 22)

321 Interestingly, the prevalence of the reference and risk allele is roughly equal which suggests

322 the possibility of a relatively new mutation or that the risk variant has a different functional advantage which balances the selection. The most highly activated SNP in the GLN-model 323 with importance for classifying patients for not having AF was rs17042081, a variant near 324 325 4q25, which has been extensively associated with AF in a variety of populations and in close 326 proximity to PITX2.(23, 24), The variants most highly associated with AF in the linear model 327 was rs17042171, also a variant near 4q25. The alternative allele has worldwide prevalence of 328 up to 16% and 13% in the European population, which makes the risk variant very common, although not equal to the reference allele suggesting a negative selection pressure of the risk 329 330 allele.(25) As the testing set consists of purely surgical patients, it is not unexpected that 331 variants near 4q25 are important for the models, as the same region was the only one associated with postoperative AF in a recent GWAS-analysis from our group.(7) 332 333 The difference in which variants show importance for the GLN and linear model, 334 respectively, and that the GLN-models in general performed significantly better in recall 335 compared with the matching linear, shows that non-linear interactions between genes which 336 are potentially of great importance in the risk of a particular trait. Other explanations include non-linear effects in non-genetic features, such as age and sex, or dominant/recessive effects 337 338 of the SNPs in question. When exploring pathways and interactions for the most highly activated genes in online 339

340 repositories such as the Reactome Pathway Database and BioGRID, it appears that none of

341 the genes have previously been described to be in a direct pathway or in any kind of

interaction. Interestingly, Gao et al. showed that the level of caveolin-1 determines the level

343 of product of *KCNN1* which previously has been highly associated with AF in several

344 GWAS-studies.(21, 26-28)

345 The SNPs with highest importance for classifying VTE in the GLN-model was rs505922, an

intron variant in ABO.(29) The variant has previously been associated with VTE.(6) The

ABO blood group antigen genes are amongst the most heavily associated with VTE, and it
has a biological plausible explanation, thus it is expected that specific variants within these
genes would play a significant role in predictive models for VTE risk.

The SNP that had the highest feature importance in the GLN-model for classifying bacterial pneumonia was rs11080143. However, as the model performed poorly in terms of the overall accuracy, there is a high likelihood that the highest activated variants are not due to genuine, biological phenomena and interactions, but rather due to chance alone. This sentiment is further supported by the fact that rs11080143 has no reported clinical significance in the

literature and does not lie close to any biological meaningful genes. One downstream gene,

356 KSR1, has been associated with different malignancies including breast adenocarcinoma and

thyroid cancer.(30, 31) To our knowledge, *KSR1* has not been associated directly with lung

358 cancer, but considering the protein product has a positive downstream signaling function of

the RAS/MAPK pathway, and the association with other cancers, a connection seems likely.

360 Although a history of cancer was included as a covariate in the initial GWAS and in both

361 models, it cannot be ruled out that the phenotype and models are confounded by occult lung

362 malignancy.

363 In the linear model, rs10519203 was the most associated with classifying pneumonia. It is an

intron variant in *HYKK* and has previously been associated with lung cancer and smoking

365 behavior, which may indicate the basis for its ability to classify pneumonia, and not an

inherent increased risk to infection.(32, 33)

We again observed a discrepancy between the variants with highest feature importance between the models, which suggests that complex non-linear effects may exist between the genes in question activated by the GLN-model. It should be noted that while all the phenotypes of interest in this study are complex diseases, we find it likely that the susceptibility to bacterial pneumonia is less driven by genetics compared with AF and VTE.

Although certain genetic variants have been associated pneumonia susceptibility, the genetic
landscape has not been explored to the same extent as with AF and especially VTE.
Consequently, the ratio of importance of the input variants compared with the clinical factors
included is likely lower compared with AF and VTE, and we anticipate that future research
will highlight the importance of genetics compared with clinical factors in predictive
modelling.

378

379 *Potentials for clinical use*

380 As these analyses were specifically made on phenotypes relevant for the postoperative course of surgical patients, it is of utmost importance that the models in question can be validated 381 and potentially optimized in a specific surgical cohort. Above all, this will ascertain the 382 383 utilization on this specific population, and it will further establish a foundation for the 384 investigation into if the models are able to improve the outcomes for the phenotypes in 385 question or be of prophylactic benefit. Further, as multiple surgical risk predictors built on 386 clinical data already exist, investigate whether the addition of genetic data will enhance the predictability of such models could offer a promising pathway for increasing the model 387 388 performances further. It is key to establish methods to improve prediction, as the current standard of models fail to demonstrate any clear clinical benefit compared with standard 389 390 practice.(34) Although multiple factors account for the current limited applicability, including 391 lacking external validity and variance in the retro- and prospective data, a lack of important factors such as genetics may also be of significance. Further, as we present a model where a 392 393 deep learning framework specifically made to incorporate genetics with clinical variables that 394 performs better compared with a linear PGS in terms of recall and precision, it is important to 395 consider the quality of the used software as well as the pragmatic applicability of the models

in question in a real-life scenario. At this time, neither model have applicability ifincorporating only the top 100 SNPs, as determined by the poor accuracy performance.

398

399

400 *Limitations*

401 This study has limitations. First, all phenotypes were established using only ICD-codes which 402 may have a low accuracy for the phenotypes in question. We suspect especially bacterial pneumonia to have an overall low accuracy due to the high hospital incidence and difference 403 404 in presentation as well differences in the microbiological organism and treatments. This 405 generates a heterogenous group which lowers the predictability and clinical utility. 406 Further, we assumed that the one hundred most significant SNPs from an initial GWAS for 407 the phenotype in question would be of interest, although this was an arbitrary choosing due to 408 the need to find an optimum between predictive power and computational efficacy. Other 409 SNPs may also be of importance, and using a different set or potentially the entire genome 410 has the potential to achieve similar or even better genetic predictability, although the latter 411 would be too computational costly and of less clinical utility. A significant challenge in our 412 study is the imbalanced data, which especially proved problematic in the linear models which all had a recall of 0. Larger cohorts, or a more balanced dataset may improve this. 413

414

415 *Conclusion*

In conclusion, we present predictive models on surgery relevant phenotypes incorporating a small sample of genetic variants. Overall, GLN-based models performed equally when compared with linear models based on the AUC metric. However, recall and precision were better in the GLN-based model, making them more useful in a potential clinical setting. Further, different SNPs were important for the same phenotypes between

421	models suggesting importance off non-linear interactions. Lastly, in a comparison between
422	clinical models with and without inclusion of SNPs, the inclusion of genetic data seemed to
423	increase the accuracy, albeit with overlapping confidence intervals. This is a preliminary
424	report assessing the utility of using a small sample of SNPs for clinical risk prediction. Future
425	research needs to validate models in surgical cohorts and assess the utility of incorporating
426	genetics and clinical variables in predictive models to improve surgical outcomes.
427	
428	
429	
430	
431	
432	
433	
434	
435	
436	
437	
438	
439	
440	
441	
442	
443	
444	
445	

446

FIGURES







472 2 = cases.







496 Figure 2B: ROC-AUC development of the genetic GLN-model of venous thromboembolism



Figure 3A: Bar plot of ROC-AUCs of all atrial fibrillation models



Figure 3C: Bar plot of ROC-AUCs of all pneumonia models 546

Table 1: Baseline characteristics for atrial fibrillation cohort

Characteristic	AF	No AF
	N = 19,900	N = 426,280
Age, mean, SD	62.2 ± 5.9	56.4 ± 8.0
Female, N, %	6,623 (33.2)	233,351 (54.7)
BMI, mean, SD	27.4 ± 5.4	29.1 ± 1
Previous or current smoker,	9,633 (48.4)	159,702 (37.5)
N, %		
Previous or current cancer,	2,307 (11.6)	35,338 (8.3)
N, %		
Heart failure, N, %	2,922 (14.7)	3,252 (0.7)
Hypertension, N, %	12,385 (62.2)	86,999 (20.4)

Table 2: Baseline characteristics for venous thromboembolism cohort

Characteristic	VTE	No VTE
	N = 9,193	N = 436,987
Age, mean, SD	59.7 ± 7.1	56.6 ± 8.0
Female, N, %	4,156 (45.2)	235,818 (54.0)
BMI, mean, SD	29.4 ± 5.5	27.4 ± 4.8
Previous or current smoker,	3,842 (41.8)	165,493 (37.9)
N, %		

Previous or current cancer,	1,403 (15.3)	36,242 (8.3)
N, %		
Heart failure, N, %	482 (5.2)	5.692 (1.3)
Hypertension, N, %	3,906 (42.5)	95,478 (21.8)

578 Table 3: Baseline characteristics for pneumonia

Characteristic	Pneumonia	No pneumonia
		1
	N = 14,101	N = 432,079
Age, mean, SD	60.3 ± 7.2	56.5 ± 8.0
Female, N, %	6,101 (43.3)	233,873 (54.1)
BMI, mean, SD	28.4 ± 5.6	27.5 ± 4.8
Previous or current smoker,	6,464 (45.8)	162,871 (37.7)
N, %		
Previous or current cancer,	2,144 (15.2)	35,501(8.2)
N, %		
Heart failure, N, %	1,506 (10.7)	4,668 (1.11)
Hypertension, N, %	7,120 (50.5)	92,264 (21.4)

- **Table 4A:** Table of the SNPs with the highest feature importance for the single nucleotide
- 586 polymorphism (SNP), and combined atrial fibrillation GLN-models

587

Top SNP in SNP model	Gene	Top SNP in combined	Gene
		model	
rs3807989	CAVI	rs3807989	CAVI
rs56250774	PRRXI	rs13124249	Intergenic
rs11047543	LOC105369698	rs13125644	Intergenic
rs1570220	SH3PXD2A	rs4835669	NME5
rs3825214	TBX5	rs6658392	KCNN3

588

Table 4B: Table of the SNPs with the highest feature importance for the genetic, and mixed

590 atrial fibrillation linear models

591

Genetic, SNP	Gene	Mixed, SNP	Gene
rs17042171	Intergenic	rs17042171	Intergenic
rs10033464	Intergenic	rs10033464	Intergenic
rs3807988	CAVI	rs3731748	TTN/TTN-AS1
rs3731748	TTN/TTN-AS1	rs3807988	CAVI
rs2723065	LINC02576	rs3829747	TTN/TTN-AS1

592

593

594

595

- **Table 4C:** Table of the SNPs with the highest feature importance for the single nucleotide
- 598 polymorphism (SNP), and combined venous thromboembolism GLN-models

Top SNP in SNP model	Gene	Top SNP in combined	Gene
		model	
rs505922	ABO	Rs6993770	ZFPM2
rs657152	ABO	Rs687621	ABO
rs8176740	ABO	Rs75112989	ATP1B1
rs7868232	Intergenic	Rs3746438	МҮН7В
rs581107	ABO	Rs56103207	TSPAN15

Table 4D: Table of the SNPs with the highest feature importance for the genetic, and mixed

602 Venous thromboembolism linear models

Genetic, SNP	Gene	Mixed, SNP	Gene
rs4524	<i>F5</i>	rs4524	<i>F5</i>
rs6030	<i>F</i> 5	rs6120849	EDEM2
rs6120849	EDEM2	rs6030	<i>F5</i>
rs75112989	ATP1B1	rs75112989	ATP1B1
rs6050	FGA	rs507666	ABO

- **Table 4E:** Table of the SNPs with the highest feature importance for the single nucleotide
- 610 polymorphism (SNP), and combined pneumonia GLN-models

Top SNP in SNP model	Gene	Top SNP in combined	Gene
		model	
rs11080143	Intergenic	rs17143419	GALNT17
rs72976957	PIAS4	rs2476601	PTPN22/AP4B1- AS1
rs2381116	FAM219A	rs72793809	LOC124903672
rs76002435	MED27	rs361594	PEX26
rs75766461	LOC124902060	rs219258	Intergenic

614 Table 4F: Table of the SNPs with the highest feature importance for the genetic, and mixed

615 pneumonia linear models

Genetic, SNP	Gene	Mixed, SNP	Gene
rs10519203	HYKK	rs10519203	НҮКК
rs7498665	SH2B1	rs61921073	Intergenic
rs8062405	ATXN2l	rs8062405	ATXN2l
rs77139199	GASK1B-AS1	rs77139199	GASK1B-AS1
rs61921073	Intergenic	rs62531875	Intergenic

620 Figure 4: Selection and quality control steps of individuals and SNPs in the UKB



631 Supplementary table 1: List of ICD-9 and ICD-10 codes used for the phenotypes in

632 question.

ICD-9 codes for atrial fibrillation	4273
ICD-10 codes for atrial fibrillation	I48, I480, I481, I483, I484, I489
ICD-9 codes for venous thromboembolism	4151, 4511, 4512, 4519, 4531, 4532, 4534,
	4538, 4539, 4534, 4531, 4532, 4539
ICD-10 codes for venous thromboembolism	1260, 1269, 1801, 1802, 1803, 1808, 1809,
	1820, 1821, 1822, 1823, 1828, 1829, 0082,
	0223, 0871, 0882, 181
ICD-9 codes for pneumonia	4810, 4820, 4821, 4823, 4824, 4828, 4829,
	4830, 4831, 4838, 4840, 4841, 4843, 4835,
	4836, 4847, 4848, 4850, 4860
ICD-10 codes for pneumonia	J13, J14, J150, J151, J152, J153, J154, J155,
	J156, J157, J158, J159, J16, J160, J168,
	J170, J172, J173, J178, J180, J181, J182,
	J188, J189, J851

633

634

635

636

637 **Supplementary table 2:** List of OPCS-4 codes used to define surgery.

['W401', 'W371', 'W381', 'W822', 'Q074', 'M021', 'K453', 'J183', 'M611', 'G693', 'W201', 'W205', 'M025', 'T202', 'T413', 'W852', 'W192', 'W411', 'H335', 'J692', 'W164', 'W191', 'T252', 'K262', 'W421', 'G753', 'W283', 'Q221', 'A054', 'S065', 'W373', 'W391', 'B282', 'W593', 'M022', 'W153', 'A401', 'H114', 'H333', 'H114', 'J021', 'V255', 'T309', 'H053', 'W199', 'W581', 'L841', 'L931', 'L948', 'W242', 'A411', 'S472', 'A651', 'W461 ', 'L851', 'S069', 'C751', 'W192', 'W941', 'W403', 'L622', 'W241', 'E543', 'T676', 'S571', 'B274', 'Q075', 'H071', 'H336', 'R182', 'G743', 'A025', 'W879', 'T962', 'G011', 'H012', 'W742' , 'G331', 'W858', 'J011', 'T791', 'H041', 'H511', 'W833', 'M341', 'L185', 'W931', 'A021', 'H334', 'T439', 'B285', 'A022', 'W083', 'W246', 'H338', 'W791', 'B279', 'G031'

'K403', 'O291', 'P232', 'H331', 'A671', 'Q013', 'K263', 'W042', 'H073', 'T391', 'L843', 'H011', 'W198', 'H095', 'L845', 'L192', 'E528', 'L516', 'G218', 'A023', 'T678', 'M012', 'M612', 'K402', 'L871', 'W571', 'V339', 'H074', 'A384', 'A388', 'W208', 'M538', 'W871', 'M278', 'W424', 'M421', 'V337', 'H072', 'G698', 'Q231', 'W152', 'V468', 'H042', 'T273', 'G243', 'B293', 'A295', 'G784', 'J576', 'B311', 'W222', 'T423', 'G699', 'L842', 'W802', 'R172', 'V254', 'W082', 'T272', 'W451', 'T521', 'L593', 'V253', 'V256', 'V209', 'P242', 'H103', 'W431', 'W382', 'S551', 'T723', 'V411', 'W471', 'W423', 'A389', 'B083', 'T748', 'M013', 'J023', 'T242', 'T679', 'V291', 'W301', 'H062', 'H051', 'J561', 'V221', 'S068', 'A383', 'O089', 'W803', 'G648', 'G352', 'E032', 'E248', 'V031', 'G521', 'T331', 'J091', 'W122', 'G532', 'H092', 'W821', 'W402', 'F349', 'M039', 'F442', 'W961', 'H013', 'L591', 'J185', 'T162', 'G748', 'T531', 'B181', 'V462', 'W399', 'M231', 'B312', 'W384', 'V433', 'G694', 'L181', 'Q092', 'W951', 'A108', 'H105', 'W194', 'L258', 'V143', 'H079', 'E541', 'T316', 'H091', 'L601', 'J569', 'W712', 'G289', 'L541', 'W282', 'G303', 'E542', 'W211', 'O079', 'M051', 'J552', 'K255', 'H332', 'W202', 'H151', 'S022', 'W752', 'A511', 'G028', 'L131', 'L233', 'S172', 'W158', 'R181', 'W743', 'A473', 'M014', 'V336', 'W288', 'H029', 'W212', 'T278', 'W891', 'V403', 'F222', 'H412', 'K334', 'A445', 'T551', 'G234', 'W859', 'V037', 'J582', 'J042', 'J571', 'K301', 'J024', 'J571', 'K301', 'J024', 'W195', 'L581', 'H131', 'G633', 'M373', 'N288', 'W834', 'V294', 'V039', 'V152', 'H052', 'M191', 'W393', 'L198', 'S251', 'K412', 'A412', 'T792', 'S552', 'H142', 'H064', 'V241', 'J012', 'G283', 'L343', 'F231', 'F451', 'W193', 'L844', 'H085', 'W308', 'W332', 'T972', 'L582', 'O182', 'T229', 'G492', 'T458', 'E294', 'K335', 'T262', 'Q518', 'J578', 'V392', 'E148', 'H113', 'L583', 'G242', 'L682', 'L621', 'T412', 'T342', 'L798', 'V104', 'J025', 'A652', 'T622', 'W214', 'W629', 'H104', 'G281', 'K454', 'T671', 'W395', 'T243', 'H101', 'E036', 'A028', 'H112', 'L201', 'T525', 'L624', 'W422', 'H118', 'V383', 'B223', 'L628', 'B012', 'W232', 'T251', 'J028', 'K451', 'G312', 'L124', 'H061', 'L818', 'T203', 'O181', 'K023', 'L602', 'G278', 'W144', 'V313', 'P251', 'N113', 'E192', 'V242', 'W033', 'C134', 'L193', 'G734', 'M348', 'F443', 'E042',

'B284', 'G601', 'W523', 'G789', 'G013', 'W832', 'J182', 'J219', 'L592', 'W196', 'T212', 'W068', 'M619', 'W433', 'T675', 'G362', 'W238', 'V402', 'G584', 'W131', 'T798', 'V259', 'T724', 'K253', 'H152', 'K021', 'W038', 'W231', 'V223', 'M359', 'T259', 'E532', 'T318', 'H019', 'T222', 'G232', 'W091', 'A386', 'H358', 'L194', 'J189', 'K272', 'W758', 'G251', 'T209', 'W372', 'L598', 'L587']

638

639 Supplementary table 3: Demographics and comorbidities with ICD-codes

Demographics	Age (UKB code field: 21022-0.0)	
	Sex (UKB code field: 31-0.0)	
Comorbidities	Previous or current smoker (UKB code	
	field: 1249-0.0, 1249-1.0, 1249-2.0, 1249-	
	3.0)	
	Hypertension (ICD-10: I10, I100, I101,	
	1102, 1103, 1104, 1105, 1106, 1107, 1108,	
	1109, 113, 1130, 1131, 1132, 1133, 1134,	
	1135, 1136, 1137, 1138, 1139, 115, 1150,	
	1151, 1152, 1153, 1154, 1155, 1156, 1157,	
	I158, I159)	
	Cancer (UKB code field: 2453-0.0, 2453-	
	1.0, 2453-2.0, 2453-3.0), heart failure	
	(ICD-10: I50, I501, I502, I5020, I5021,	
	15022, 15023, 1503, 15030, 15031, 15032,	
	15033, 1504, 15041, 15042, 15043, 1508,	
	15081, 150810, 150810, 150811, 150812,	

	150813, 150814, 15082, 15083, 15084, 15089,
	1509)
	BMI (UKB code field: 21001-0.0, 21001-
	1.0, 21001-2.0, 21001-3.0.
640	
641	
642	
643	
644	
645	
646	
647	
648	
649	
650	
651	
652	
653	
654	
655	
656	
657	
658	
659	
660	

661

REFERENCES

Ozgediz D, Jamison D, Cherian M, McQueen K. The burden of surgical conditions 662 1. 663 and access to surgical care in low- and middle-income countries. Bull World Health Organ. 2008;86(8):646-7. 664

Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-Leitz T, et al. 665 2. 666 Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. Lancet. 2015;385 Suppl 2:S11. 667

668 Dencker EE, Bonde A, Troelsen A, Varadarajan KM, Sillesen M. Postoperative 3. 669 complications: an observational study of trends in the United States from 2012 to 2018. BMC Surg. 2021;21(1):393. 670

Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical Risk Is Not Linear: 671 4. 672 Derivation and Validation of a Novel, User-friendly, and Machine-learning-based Predictive OpTimal Trees in Emergency Surgery Risk (POTTER) Calculator. Ann Surg. 673

674 2018;268(4):574-83.

Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development 675 5. 676 and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and

677 informed consent tool for patients and surgeons. J Am Coll Surg. 2013;217(5):833-42 e1-3. M AC, Bonde A, Sillesen M. An assessment of the effect of the genotype on

678 6. 679 postoperative venous thromboembolism risk in 140,831 surgical patients. Ann Med Surg 680 (Lond). 2021;71:102938.

Christensen MA, Bonde A, Sillesen M. Genetic risk factors for postoperative atrial 681 7. 682 fibrillation-a nationwide genome-wide association study (GWAS). Frontiers in 683 Cardiovascular Medicine. 2023;10.

Gaudino M, Di Castelnuovo A, Zamparelli R, Andreotti F, Burzotta F, Iacoviello L, et 684 8. 685 al. Genetic control of postoperative systemic inflammatory reaction and pulmonary and renal complications after coronary artery surgery. J Thorac Cardiovasc Surg. 2003;126(4):1107-12. 686

Kolek MJ, Muehlschlegel JD, Bush WS, Parvez B, Murray KT, Stein CM, et al. 687 9.

688 Genetic and clinical risk prediction model for postoperative atrial fibrillation. Circ Arrhythm 689 Electrophysiol. 2015;8(1):25-31.

690 Bonde A, Varadarajan KM, Bonde N, Troelsen A, Muratoglu OK, Malchau H, et al. 10. 691 Assessing the utility of deep neural networks in predicting postoperative surgical

692 complications: a retrospective study. Lancet Digit Health. 2021;3(8):e471-e85.

Dela Cruz CS, Wunderink RG, Christiani DC, Cormier SA, Crothers K, Doerschuk 693 11.

694 CM, et al. Future Research Directions in Pneumonia. NHLBI Working Group Report. Am J 695 Respir Crit Care Med. 2018;198(2):256-63.

696 Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK 12.

697 Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

698 13. R Core Team (2020). R: A language and environment for statistical

699 computing. R Foundation for Statistical Computing, Vienna, Austria.

URL https://www.R-project.org/. 700

- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-701 14. 702 learn: Machine Learning in Python. J Mach Learn Res. 2011;12(null):2825–30.
- 703 Sigurdsson AI, Louloudis I, Banasik K, Westergaard D, Winther O, Lund O, et al. 15.
- 704 Deep integrative models for large-scale human genomics. Nucleic Acids Research. 2023.

Kingma D, Ba J. Adam: A Method for Stochastic Optimization. International 705 16.

706 Conference on Learning Representations. 2014.

707 17. Lundberg SM LS. A unified approach to interpreting model predictions. Advances in 708 Neural Information Processing Systems 30 (NIPS). 2017.

medRxiv preprint doi: https://doi.org/10.1101/2024.01.09.23297913; this version posted January 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in

perpetuity. It is made available under a CC-BY 4.0 International license .

- 709 18. Lawson EH, Louie R, Zingmond DS, Brook RH, Hall BL, Han L, et al. A comparison 710 of clinical registry versus administrative claims data for reporting of 30-day surgical
- 711 complications. Ann Surg. 2012;256(6):973-81.
- 712 19. Zhang T, Shang F, Ma Y, Xu Y, Sun W, Song H. Caveolin-1 Promotes the Imbalance
- 713 of Th17/Treg in Chronic Obstructive Pulmonary Disease by Regulating Hsp70 Expression.
- 714 Int J Chron Obstruct Pulmon Dis. 2023;18:565-74.
- 715 20. Williams TM, Lisanti MP. The Caveolin genes: from cell biology to medicine. Ann 716 Med. 2004;36(8):584-95.
- 717 Jia W, Qi X, Li Q. Association Between Rs3807989 Polymorphism in Caveolin-1 21.
- 718 (CAV1) Gene and Atrial Fibrillation: A Meta-Analysis. Med Sci Monit. 2016;22:3961-6.
- 719 Liu Y, Ni B, Lin Y, Chen XG, Chen M, Hu Z, et al. The rs3807989 G/A 22.
- 720 polymorphism in CAV1 is associated with the risk of atrial fibrillation in Chinese Han 721 populations. Pacing Clin Electrophysiol. 2015;38(2):164-70.
- 722 Zhao L, Chen XG, Liu Y, Fang Z, Zhang F. Association of rs17042171 with 23.
- chromosome 4q25 with atrial fibrillation in Chinese Han populations. Anatol J Cardiol. 723 724 2016;16(3):165-9.
- 725 24. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A,
- 726 et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature.
- 727 2007;448(7151):353-7.
- 728 25. L. Phan YJ, H. Zhang, W. Qiang, E. Shekhtman, D. Shao, D. Revoe, R. Villamarin, E.
- 729 Ivanchenko, M. Kimura, Z. Y. Wang, L. Hao, N. Sharopova, M. Bihan, A. Sturcke, M. Lee,
- N. Popova, W. Wu, C. Bastiani, M. Ward, J. B. Holmes, V. Lyoshin, K. Kaur, E. Moyer, M. 730
- 731 Feolo, and B. L. Kattman. ALFA: Allele Frequency Aggregator." National Center for
- 732 Biotechnology Information, U.S. National Library of Medicine. 2020.
- 733 26. Gao Y, Bertuccio CA, Balut CM, Watkins SC, Devor DC. Dynamin- and Rab5-
- 734 dependent endocytosis of a Ca2+ -activated K+ channel, KCa2.3. PLoS One.
- 2012;7(8):e44150. 735
- 736 Yi SL, Liu XJ, Zhong JQ, Zhang Y. Role of caveolin-1 in atrial fibrillation as an anti-27. 737 fibrotic signaling molecule in human atrial fibroblasts. PLoS One. 2014;9(1):e85144.
- 738 Ellinor PT, Lunetta KL, Glazer NL, Pfeufer A, Alonso A, Chung MK, et al. Common 28. 739 variants in KCNN3 are associated with lone atrial fibrillation. Nat Genet. 2010;42(3):240-4.
- 740 Dentali F, Sironi AP, Ageno W, Turato S, Bonfanti C, Frattini F, et al. Non-O blood 29.
- 741 type is the commonest genetic risk factor for VTE: results from a meta-analysis of the 742 literature. Semin Thromb Hemost. 2012;38(5):535-48.
- 743 Stebbing J, Zhang H, Xu Y, Lit LC, Green AR, Grothey A, et al. KSR1 regulates 30.
- 744 BRCA1 degradation and inhibits breast cancer growth. Oncogene. 2015;34(16):2103-14.
- 745 31. Lee J, Seol MY, Jeong S, Kwon HJ, Lee CR, Ku CR, et al. KSR1 is coordinately
- 746 regulated with Notch signaling and oxidative phosphorylation in thyroid cancer. J Mol 747 Endocrinol. 2015;54(2):115-24.
- 748 Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A 32. 749 susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 750 15q25. Nature. 2008;452(7187):633-7.
- 751 Schwartz AG, Cote ML, Wenzlaff AS, Land S, Amos CI. Racial differences in the 33. 752 association between SNPs on 15q25.1, smoking behavior, and risk of non-small cell lung 753 cancer. J Thorac Oncol. 2009;4(10):1195-201.
- 754 34. Marwaha JS, Chen HW, Habashy K, Choi J, Spain DA, Brat GA. Appraising the
- Quality of Development and Reporting in Surgical Prediction Models. JAMA Surg. 755 756 2023;158(2):214-6.
- 757