



Софийски Университет „Св. Климент Охридски“

Биологически Факултет

КАТЕДРА Генетика

ГЕОРГИ ДИМИТРОВ БЛАЖЕВ

**„Мулти-омикс подход за анализ на
биологичната и клиничната хетерогенност на
някои редки малигнени заболявания“**

АВТОРЕФЕРАТ

на дисертационен труд за придобиване на образователна и научна степен

„Доктор“ в професионално направление

4.3 Биологически науки (Генетика – Генетика на рака)

Научен ръководител: доц. д-р Велизар Шиваров, д.м.

София

2024

Докторантът Георги Блажев е разработил дисертационния труд като редовен докторант в направление „Биологически науки (Генетика – Генетика на рака) към Биологически Факултет на Софийски Университет „Св. Климент Охридски“.

Дисертационният труд е обсъден на разширено заседание на катедра „Генетика“ към Биологически факултет на Софийски Университет „Св. Климент Охридски“, проведено на .04.2024 г., въз основа на ректорска заповед № РД.....

Дисертационният труд е насрочен за защита пред научно жури, потвърдено със Заповед № РД..... на Ректора на Софийски Университет „Св. Климент Охридски“ проф. д-р Георги Вълчев.

Научно жури:

Вътрешни членове на научно жури:

1.
2.

Външни членове на научно жури:

3.
4.
5.

Защитата на дисертационния труд ще се проведе на Г. В Ч. В сградата на Биологическия факултет на СУ "Св. Климент Охридски"

Материалите по защитата на дисертацията са на разположение на интересующите се в канцеларията на Биологически Факултет при Софийски Университет „Св. Климент Охридски“, бул. Драган Цанков 8, гр. София, както и на интернет страницата на университета.

Дисертационният труд съдържа 110 страници и се състои от: въведение, шест глави, приноси, приложения, библиография, декларация за оригиналност и публикации. В обема му се съдържат 38 фигури и 8 таблици. Библиографията включва 173 литературни източника.

СЪДЪРЖАНИЕ

СПИСЪК НА ТАБЛИЦИТЕ	3
СПИСЪК НА ФИГУРИТЕ.....	4
ИЗПОЛЗВАНИ СЪКРАЩЕНИЯ.....	6
I. Увод.....	7
II. Изследователска хипотеза, цели и задачи.....	9
Изследователска хипотеза.....	9
Основна изследователска цел	9
Основни изследователски задачи	9
III. Материали и методи.....	10
Набори от данни.....	10
Извеждане на модела	10
Анализ на обогатяване на набор от гени	10
Cibersort	10
Анализ на лекарствената чувствителност	11
Интегриран анализ на ДНК метилиране.....	11
Общи статистически процедури	11
IV. Резултати.....	13
Идентифициране на проучвания за включване в анализа.....	13
Извеждане и първоначално оценка на двугенния прогностичен скор (2-ПС).....	14
Валидиране на 2-ПС.....	19
Анализ на обогатяването на списъци от гени в експресионни профили (GSEA)	23
Интегриран анализ на ДНК метилиране и генна експресия	27
Корелация с изчислени популации на инфилтриращи имунни клетки	32
Потенциална прогностична способност на 2-ПС.....	33
V. Обсъждане.....	36
VI. Заключение и изводи	41
VII. Приноси	42
Оригинални приноси.....	42
Приноси с потвърдителен характер.....	42
VIII. Публикации	43
Журнални статии	43
Участия в научни срещи.....	43

СПИСЪК НА ТАБЛИЦИТЕ

Таблица 1. Използвани омикс набори от данни за изграждане, валидиране и изследване на биологичната същност на 2-ПС.....	11
Таблица 2. Резултати от анализа за диференциално метилирани CpG острови от набора данни TCGA. За целите на анализа островът е дефиниран като поне два прилежащи CpG сайта, които са включени в Infinium 450K чипа на Illumina.....	28
Таблица 3. Резултати от интегрирания анализ на диференциално метилирани CpG острови и генна експресия от набора данни TCGA. За целите на анализа островът е дефиниран като поне два прилежащи CpG сайта, които са включени в Infinium 450K чипа на Illumina.	29

СПИСЪК НА ФИГУРИТЕ

Фигура 1. Брой записи в PubMed по години след търсене с термини „mesothelioma AND malignant“. Дата на последно търсене 10-12-2023 г.....	13
Фигура 2. PRISMA диаграма на процеса на подбор на проучвания за включване в анализа. Въз основа на систематичния подход, описан в диаграмата, стигнахме до селекция на 3 проучвания, съдържащи обществено достъпни мулти-омикс данни от пациенти с МПМ	14
Фигура 3. Диаграма на подбрания аналитичен подход.....	16
Фигура 4. Хистограма на разпределението на продължителния скор за пациентите от TCGA кохортата. Червената вертикална линия показва медианната стойност, която е използвана за дихотомното (бинарно) групиране на пациенти с висок и на пациенти с нисък скор.	17
Фигура 5. Унивариационен анализ на преживяемостта според 2-ПС в кохортата от TCGA. Пациентите са разделени на група с висок и група с нисък скор въз основа на медианата на продължителния скор. Стойността на p е от Cox-регресионен анализ.....	17
Фигура 6. Анализ на ROC кривата за прогностичната стойност на бинарния скор по отношение на общата преживяемост в TCGA кохортата. Изчислени са площта под кривата (AUC), както и чувствителността и специфичността при медианна стойност от 21,59, която е използвана за дихотомно разделяне.	18
Фигура 7. Мултивариационен анализ на преживяемостта според 2-ПС в кохортата от TCGA. Включени са ковариабилни възраст, пол, хистология, стадий и мутационен статус по отношение на 4 гена BAP1, TP53, SETD2 и NF2.....	18
Фигура 8. Хистограма на разпределението на продължителния скор за пациентите от Bueno кохортата. Червената вертикална линия показва медианната стойност, която е използвана за дихотомното (бинарно) групиране на пациенти с висок и на пациенти с нисък скор.	19
Фигура 9. Унивариационен анализ на преживяемостта според 2-ПС в кохортата от Bueno. Пациентите са разделени на група с висок и група с нисък скор въз основа на медианата на продължителния скор. Стойността на p е от Cox-регресионен анализ.....	20
Фигура 10. Анализ на ROC кривата за прогностичната стойност на бинарния скор по отношение на общата преживяемост в TCGA кохортата. Изчислени са площта под кривата (AUC), както и чувствителността и специфичността при медианна стойност от 11,78, която е използвана за дихотомно разделяне.	20
Фигура 11. Мултивариационен анализ на преживяемостта според 2-ПС в кохортата Bueno. Включени са ковариабилни възраст, пол, хистология, стадий и мутационен статус по отношение на 4 гена BAP1, TP53, SETD2 и NF2.....	21
Фигура 12. Хистограма на разпределението на продължителния скор за пациентите от Blum кохортата. Червената вертикална линия показва медианната стойност, която е използвана за дихотомното (бинарно) групиране на пациенти с висок и на пациенти с нисък скор.	22
Фигура 13. Унивариационен анализ на преживяемостта според 2-ПС в кохортата от Blum. Пациентите са разделени на група с висок и група с нисък скор въз основа на медианата на продължителния скор. Стойността на p е от Cox-регресионен анализ.....	22
Фигура 14. Анализ за на ROC кривата за прогностичната стойност на бинарния скор по отношение на общата преживяемост в Blum кохортата. Изчислени са площта под кривата (AUC), както и чувствителността и специфичността при медианна стойност от 6.56, която е използвана за дихотомно разделяне.	23
Фигура 15. Мултивариационен анализ на преживяемостта според 2-ПС в Blum кохортата. Включени са ковариабилни възраст, пол, хистология, стадий и мутационен статус по отношение на 4 гена BAP1, TP53, SETD2 и NF2.....	23
Фигура 16. Примерен резултат от GSEA анализа в набора от данни TCGA. (A) Enrichment plot за гените от списъка „Митотично вретено“; (B) Heatmap на експресията на гените същия списък. В сиво са показани пациентите с висок скор, а в жълто – тези с нисък.	24

Фигура 17. Примерен резултат от GSEA анализа в набора от данни Bueno. (A) Enrichment plot за гените от списъка „Митотично вретено“; (B) Heatmap на експресията на гените същия списък. В сиво са показани пациентите с висок скор, а в жълто – тези с нисък.....	25
Фигура 18. Примерен резултат от GSEA анализа в набора от данни Blum. (A) Enrichment plot за гените от списъка „Митотично вретено“; (B) Heatmap на експресията на гените същия списък. В сиво са показани пациентите с висок скор, а в жълто – тези с нисък	25
Фигура 19. Стълбови диаграми на сигнификантно по-често свръхекспресирани списъци от гени в пациентите с висок скор, които се откриват в поне една от трите кохорти (TCGA, Bueno, Blum), определени чрез GSEA анализ. Числата показват броя на списъците от гени във всяко от множествата.	26
Фигура 20. Venn диаграма на припокриващите се сигнификантно по-често свръхекспресирани списъци от гени в пациентите с висок скор от трите кохорти (TCGA, Bueno, Blum), определени чрез GSEA анализ. Числата показват броя на списъците от гени във всяко от множествата.	26
Фигура 21. Резултат от анализа на диференциално метилираните CpG сайтове в набора от данни TCGA. Heatmap на β -стойностите за диференциално експресираните CpG сайтове (по вертикала) спрямо пациентите (по хоризонтала)	28
Фигура 22. Двухмерен корелационен плот на метилиране на CpG острова chr2:113403001-113404079 спрямо експресията на гена SLC20A1 в набора от данни TCGA. Пациентите с висок и нисък скор са обозначени в различен цвят.	29
Фигура 23. Двухмерен корелационен плот на метилиране на CpG острова chr6:30654392-30654934 спрямо експресията на гена KIAA1949 в набора от данни TCGA. Пациентите с висок и нисък скор са обозначени в различен цвят.	30
Фигура 24. Резултат от анализа на диференциално метилираните CpG сайтове в набора от данни Blum. Heatmap на β -стойностите за диференциално експресираните CpG сайтове (по вертикала) спрямо пациентите (по хоризонтала).	31
Фигура 25. Venn диаграма на припокриващите се сигнификантно диференциално метилираните CpG сайтове в пациентите с висок спрямо нисък скор от две от кохортите (TCGA, Blum), определени чрез CONCAP анализ. Числата показват броя на списъците от гени във всяко от множествата.	31
Фигура 26. Корелационна матрица на продължителния скор с имуен скор за инфилтрация от определени популации клетки на имунната система според експресионните данни от TCGA кохортата. За изчисляването на имунните скорове е използван алгоритъмът CIBERSORT. Цветът на всяко квадратче от корелационната матрица отразява изчисленият корелационен коефициент.....	33
Фигура 27. Корелационна матрица на продължителния скор с имуен скор за инфилтрация от определени популации клетки на имунната система според експресионните данни от Bueno кохортата. За изчисляването на имунните скорове е използван алгоритъмът CIBERSORT. Цветът на всяко квадратче от корелационната матрица отразява изчисленият корелационен коефициент.....	33
Фигура 28. Корелационни двухмерни плотове между изчисленият 2-ПС скор и AUC за различни лекарства или съединения за МПМ клетъчни линии, анализирани като част от проекта GDSC1. Представени са само плотове за идентифицираните сигнификантни корелации.	35
Фигура 29. Корелационни двухмерни плотове между изчисленият 2-ПС скор и AUC за различни лекарства или съединения за МПМ клетъчни линии, анализирани като част от проекта GDSC2. Представени са само плотове за идентифицираните сигнификантни корелации.	35

ИЗПОЛЗВАНИ СЪКРАЩЕНИЯ

2-ПС	Двугенен прогностичен скор. Думата “скор” е директен превод на английския термин „score“, която означава резултат, точки и т.н. Предпочетено е използването на думата скор, тъй като тя е добила широка популярност в съвременната българска медицинска литература.
ДНК	Дезоксирибонуклеинова киселина
иРНК	Информационна рибонуклеинова киселина
МПМ	Малигнен плеврален мезотелиом
СЗО	Световна здравна организация
AIC	Akaike information criterion
ANOVA	Analysis of variance
AUC	Area under the curve
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
FDR	False discovery rate
GDC	Genomic Data Commons
GSEA	Gene set enrichment analysis
ROC	Receiver operating characteristics
RPKM	Reads per kilobase million
FRPKM	Фрагменти за килобаза на екзон на милион картирани фрагменти
PARP	Поли (АДФ-рибоза) полимераза
siRNA	Short interfering ribonucleic acid
scRNA-seq	Single-cell RNA sequencing, Small conditional RNA
TCGA	The Cancer Genome Atlas
TILs	Tumor infiltrating lymphocytes
TNM	TNM classification of malignant tumors

I. Увод

Понастоящем малигнените заболявания са водеща причина за смърт и загуба на трудоспособност и влошено качество на живот. Тежестта на тези заболявания е неравномерно разпределена в различните региони на света. По последни данни на Световната здравна организация (СЗО) за целия свят броят смъртни случаи през 2020 г., е около 10 милиона. Смъртността от ракови заболявания е неравномерно разпределена в различните региони на света, но е сред най-високите в някои от страните от Източна Европа, включително България.

Заболеваемостта и смъртност от ракови заболявания основно се дължи на няколко най-често срещани локализации със съответните полово обусловени различия. От друга страна, приблизително една четвърт от новодиагностицираните случаи на рак в Европа се свързват с различни редки видове рак (<https://www.rarecancerseurope.org/>). Обективната дефиниция на понятието рядък вид рак е изцяло епидемиологична и в известен смисъл се основава на заболеваемостта от съответния вид рак, а не на болестта. Понастоящем се приема, че като редки видове рак трябва да се означават такива нозологични единици, които се характеризират с годишна честота на новодиагностицирани случаи в дадена популация под 6 на 100 000 души население. Според обобщенията и препоръките на Европейската съвместната работна група за редки видове рак, представени в документа Rare Cancers Agenda 2030 (<https://www.esmo.org/content/download/294217/5832976/1/Rare-Cancer-Agenda-2030.pdf>), списъкът на редките видове рак се основава на така наречените „Tier-1“— нозологични единици с годишна заболеваемост под $6/100\ 000$, като са групирани в по-големи семейства според локализацията или хистологичния произход и изявата в детска възраст (педиатричните редки видове рак са обособени в отделно семейство).

Основните клинични проблеми, свързани с редките видове рак, произтичат от основните характеристики на редките заболявания, а именно: 1) клиничните решения са затруднени поради липса на медицински експертни познания и висококачествени доказателства от клиничните проучвания; 2) системата на здравеопазването трудно обслужва определени територии със специализирани грижи, от които се нуждаят тези пациенти; и 3) клиничните проучвания са затруднени и ограничени поради малкия брой пациенти и така се затруднява генерирането на висококачествени данни.

От друга страна по разнообразни причини делът на редките видове рак и смъртността от тях могат да варират значително дори между развитите страни. Така например средната 5-годишна преживяемост при пациентите с редки видове рак в САЩ е 54%, в ЕС – средно 48%, в Германия – 55%, а в България – 35%. Следва специално да се отбележи, че някои редки видове, като например тестикуларният карцином, се отличават с изключително висока преживяемост

при навременна и точна диагностика и проведено адекватно лечение. От друга страна, други редки видове рак като мезотелиомите се характеризират с 5-годишна преживяемост от порядъка на 5-10% независимо от страната и региона. Поради тази причина намаляването на общата смъртност от редки видове рак в целия свят се свързва както с навременна и адекватна грижа, така също и с провеждане на задълбочени биомедицински проучвания, които биха могли значително да подобрят диагностичните и терапевтичните възможности, а оттам и клиничните резултати. В тази връзка препоръките в документа Rare Cancers Agenda 2030 насърчават извършването на клинични, епидемиологични и транслиращи проучвания в областта на редките видове рак като се включват възможно най-голям брой центрове и пациенти.

Водени от очевидната най-голяма непосредствена медицинска нужда за подобряване на общата преживяемост при мезотелиомите с този труд ние насочихме нашите усилия именно към основната нозологична единица от тази група заболявания – малигнения плеврален мезотелиом (МПМ).

II. Изследователска хипотеза, цели и задачи

Изследователска хипотеза

Въз основа на изложеното по-горе стигнахме до хипотезата, че чрез използване на вече наличните мулти-омикс данни е възможно да се изведе нов базиран на гена експресия скор, който да има прогностична и предиктивна стойност при пациенти с МПМ.

Основна изследователска цел

Да се изведе и валидира нов прогностичен скор, основаващ се на гена експресия, при пациенти с МПМ.

Основни изследователски задачи

1. Да се идентифицират публикувани проучвания при пациенти с МПМ с обществено достъпни транскриптомни данни и поне още един тип омикс данни (напр. геномни, епигеномни).
2. Да се изведе базиран на гена експресия скор като за трениращи данни се използват тези от най-обширното проучване (с най-голям брой омикс данни).
3. Да се определи прогностичната стойност на скората във връзка с други клинични данни от пациентите от трениращия набор данни.
4. Да се валидира полученият скор въз основа на транскриптомните данни от останалите идентифицирани проучвания.
5. Да се определи прогностичната стойност на скората във връзка с други клинични данни от пациентите от валидиращите набор данни.
6. Да се провери дали полученият скор дефинира специфични подгрупи от пациенти във основа на профила на гена експресия във всеки един от използваните набори от данни.
7. Да се провери дали полученият скор дефинира специфични подгрупи от пациенти въз основа на профила на ДНК метилиране във всеки един от използваните набори от данни с налични епигеномни данни.
8. Да се провери дали полученият скор корелира с профила на инфилтрация от специфични видове клетки на имунната система като за целта се използват техники за деконволиране на транскриптомни данни.
9. Да се провери дали полученият скор корелира с чувствителността към определени лекарства въз основа на налични обществено достъпни данни от *in vitro* проучвания с клетъчни линии от МПМ.

III. Материали и методи

Набори от данни

2. Данни от TCGA през GDC – Genomics Data Commons Portal (<https://portal.gdc.cancer.gov/>)
3. The European Genome-phenome Archive (EGA) (<https://ega-archive.org/>) (EGAD00001001915)
4. Trim Galore v. 0.6.3
5. FastQC v. 0.72
6. HISAT2 (v. 2.1)
7. featureCounts (v. 1.6.4)
8. пакета *limma* за R
 - RMA
9. Array Express (E-MTAB-6877)

Извеждане на модела

- DepMap (<https://depmap.org/portal/>)
- Пакета *rbsurv* за R статистическата среда
 - Akaike Information Criterion (AIC)
- *survival* пакета за R
 - Регресионен коефициент на Cox
- Пакета Cutoff Finder за R (https://molpathohidelberg.shinyapps.io/CutoffFinder_v1/)
 - Receiver operating characteristics (ROC) криви
- Пакета *survminer* за R

Анализ на обогатяване на набор от гени

- Анализ на обогатяване на набор от гени (GSEA) (<http://www.broadinstitute.org/gsea/index.jsp>) (самостоятелна версия 4.0.3)
- Онкогенни онтологични сигнатури от базата данни за молекулярни подписи (MSigDB) (<https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#H>)

Cibersort

- Имунния пейзаж на рака (The Immune Landscape of Cancer) (<https://gdc.cancer.gov/about-data/publications/panimmune>)
- CIBERSORTx (<https://cibersortx.stanford.edu/index.php>)
- *cor* функцията от *stats* пакета на R
- Пакета *corplot* за R

Анализ на лекарствената чувствителност

- RNA-Seq данни за МПМ в проекта за лекарствена чувствителност при рак (GDSC)
- ArrayExpress (E-MTAB-3983, <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-3983>)
- Площ под кривата (AUC)
- Моделите ANOVA
- Двустранна р-стойност

Интегриран анализ на ДНК метилиране

- Данни от TCGA през GDC
- ArrayExpress (E-MTAB-6884)
- β -стойности
- COHCAP алгоритъма, имплементиран чрез съответния пакет за R(COHCAP v. 1.48.0)
- р-стойност
- False discovery rate (FDR)
- Integrative Genomics Viewer (<https://www.igv.org/>)

Общи статистически процедури

- Хи-квадрат
- Двустранен t-тест
- Тестът на Wilcoxon-Mann-Whitney

Таблица 1. Използвани омикс набори от данни за изграждане, валидиране и изследване на биологичната същност на 2-ПС.

Набор от данни	Вид данни	Технология	Депозитен номер	Пациенти (n)	Предназначение
TCGA	RNA-Seq	Illumina HiSeq 2000	phs000178	87	Трениращ
TCGA	ДНК метилиране	Infinium HumanMethylation 450 BeadChip	phs000178	87	Експлориращ
Bueno	RNA-Seq	Illumina HiSeq 2000	EGAD00001001915 EGAS000010015631	211	Валидиращ
Blum	Експресионен микрочип	HG-U133 Plus 2.0 Chip	E-MTAB-6877	67	Валидиращ
Blum	ДНК метилиране	Infinium HumanMethylation 450 BeadChip	E-MTAB-6884	67	Експлориращ

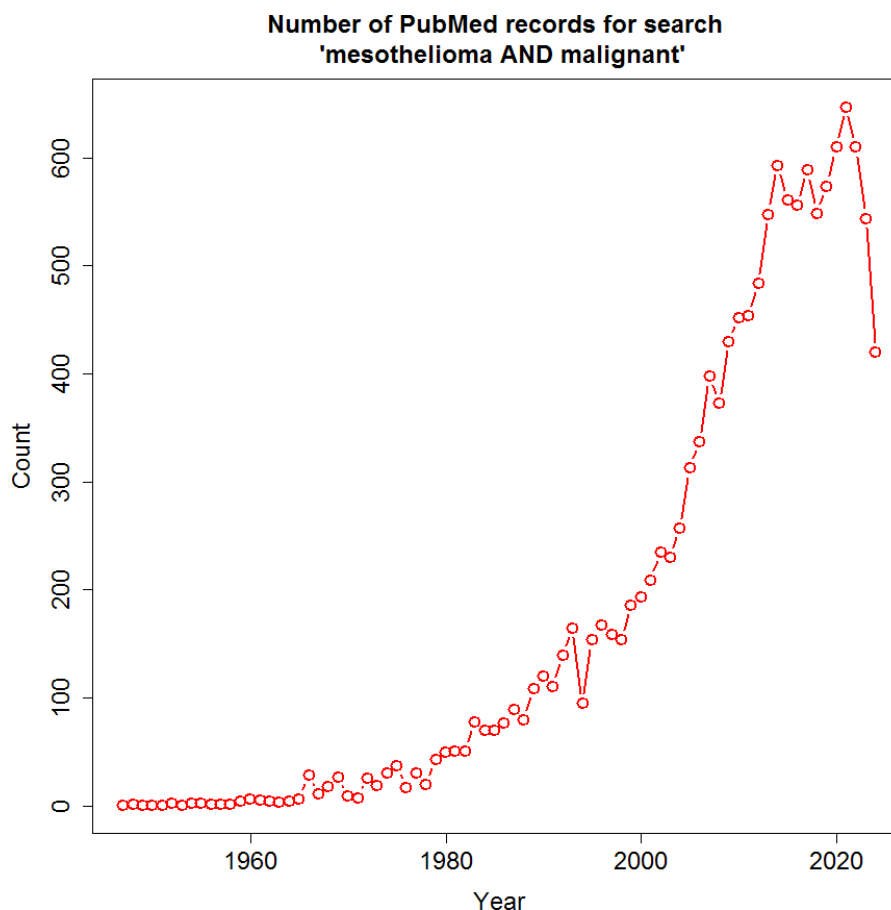
IV. Резултати

Идентифициране на проучвания за включване в анализа

Най-лесно достъпната база данни за научни публикации в областта на медицината и биологията е а Националната Библиотека по Медицина на САЩ, която е свободно достъпна он-лайн. Нейното подразделение за литературни източници се означава накратко PubMed. Както е показано на Фиг. 1. броят на публикациите, депозирани в PubMed, относно мезотелиом стремително нараства през последните 30 години, като в момента е около 600 на година.

Поради тази причина ние избрахме PubMed за единствен източник на търсене на литературни източници, които биха могли да съдържат описание на омикс данни от пациенти с МПМ. Поради бързото технологично развитие в областта на омикс

Фигура 1. Брой записи в PubMed по години след търсене с термини „mesothelioma AND malignant“. Дата на последно търсене 10-12-2023 г.



технологии ние решихме да ограничим нашето търсене темпорално за публикации от периода между 01-01-2015 и 31-12-2020 г. При идентифицирането и скринирането на записите от PubMed следвахме международния консенсус за извършване на систематични анализи на мета-анализи PRISMA, както е описано подробно на Фиг. 2.

Фигура 2. PRISMA диаграма на процеса на подбор на проучвания за включване в анализа. Въз основа на систематичния подход, описан в диаграмата, стигнахме до селекция на 3 проучвания, съдържащи обществено достъпни мулти-омикс данни от пациенти с МПМ

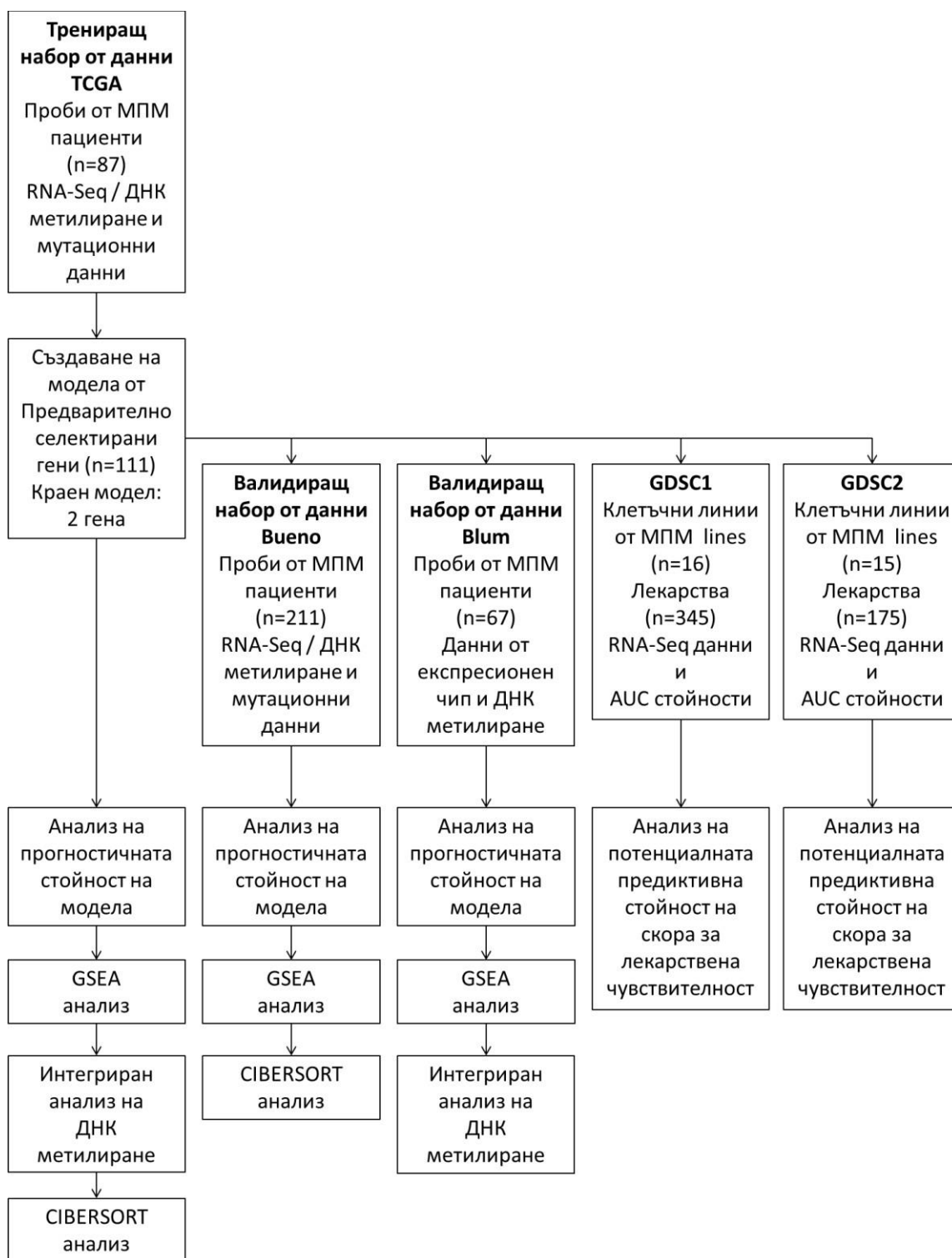


Извеждане и първоначално оценка на двугенния прогностичен скор (2-ПС)

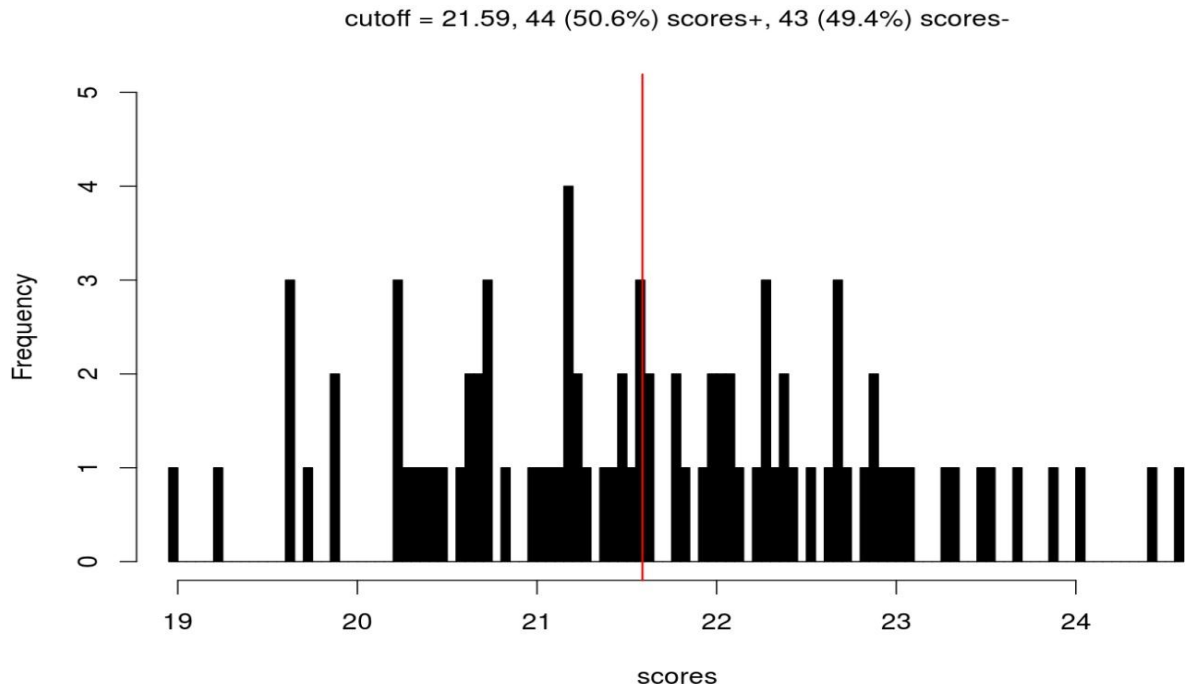
Приложихме надеждното вероятност-базирано моделиране на оцеляването с данни от микрочипове към обучителния набор от данни (TCGA набор от данни) с гени, към които е доказано, че МПМ клетъчните линии имат зависимост (Фиг. 3). Алгоритъмът избра най-ефективния прогностичен модел въз основа на най-ниската стойност на Akaike

Information Criterion (AIC). Избраният модел се състои от два гена – *GOLT1B* и *MAD2L1*. Изчислените коефициенти на регресия на Cox ($\ln(\text{HR})$) за *GOLT1B* и *MAD2L1* бяха съответно 1,403 и 0,945. Непрекъснатият скор за всяка проба във всеки набор от данни беше изчислен като сумата от стойностите на експресия за всеки от гените в модела, умножена по коефициента на регресия. При едновариантен анализ непрекъснатият скор е прогностичен за общата преживяемост (Фиг. 5). Дефинирахме и двоичен скор, използвайки за граница медианата на непрекъснатия скор за всички проби (Фиг. 4). При едновариантен анализ двоичният скор показва значителна прогностична стойност със стойност на AUC на ROC от 0,67 (Фиг. 6). Допълнително оценихме ефективността на двоичния скор в мултивариационния модел с възраст, пол, стадий, хистология и мутационен статус като ковариабилни, запазвайки независима прогностична способност (Фиг. 7).

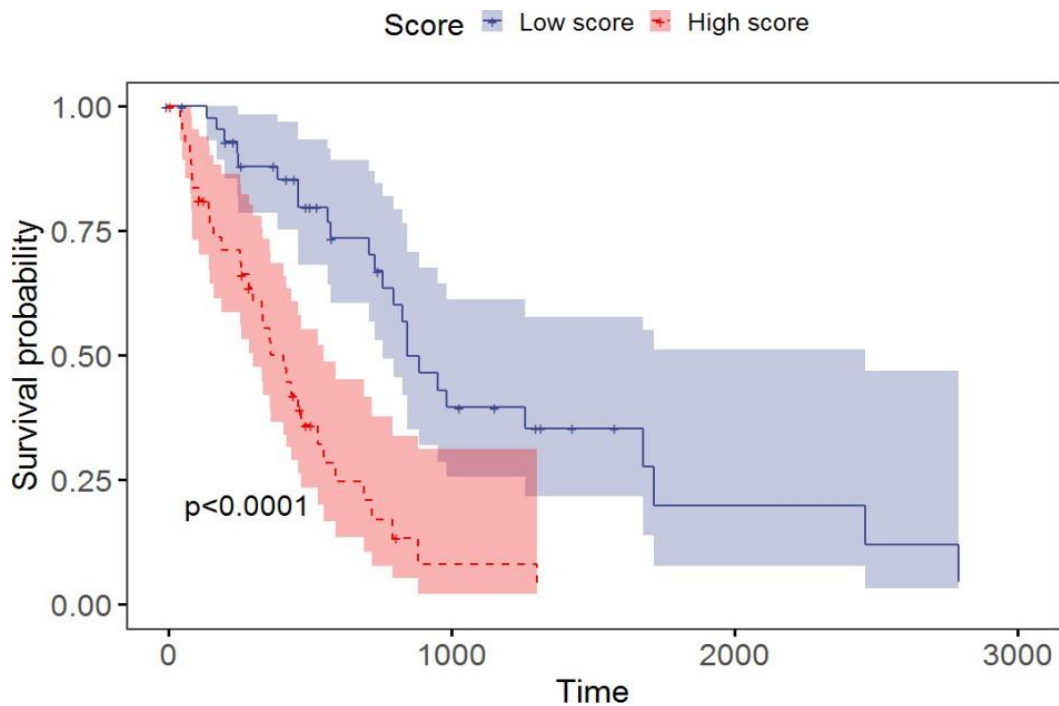
Фигура 3. Диаграма на подобрания аналитичен подход.



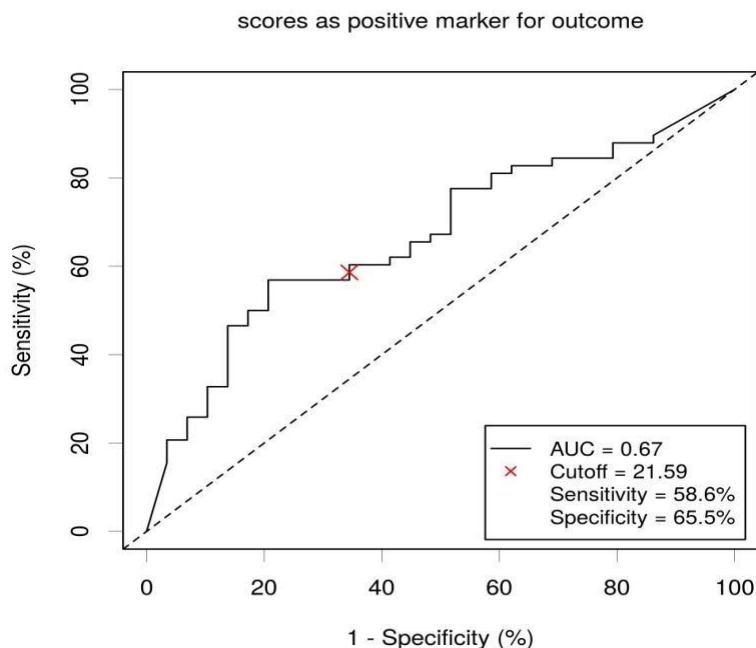
Фигура 4. Хистограма на разпределението на продължителния скор за пациентите от TCGA кохортата. Червената вертикална линия показва медианната стойност, която е използвана за дихотомното (бинарно) групиране на пациенти с висок и на пациенти с нисък скор.



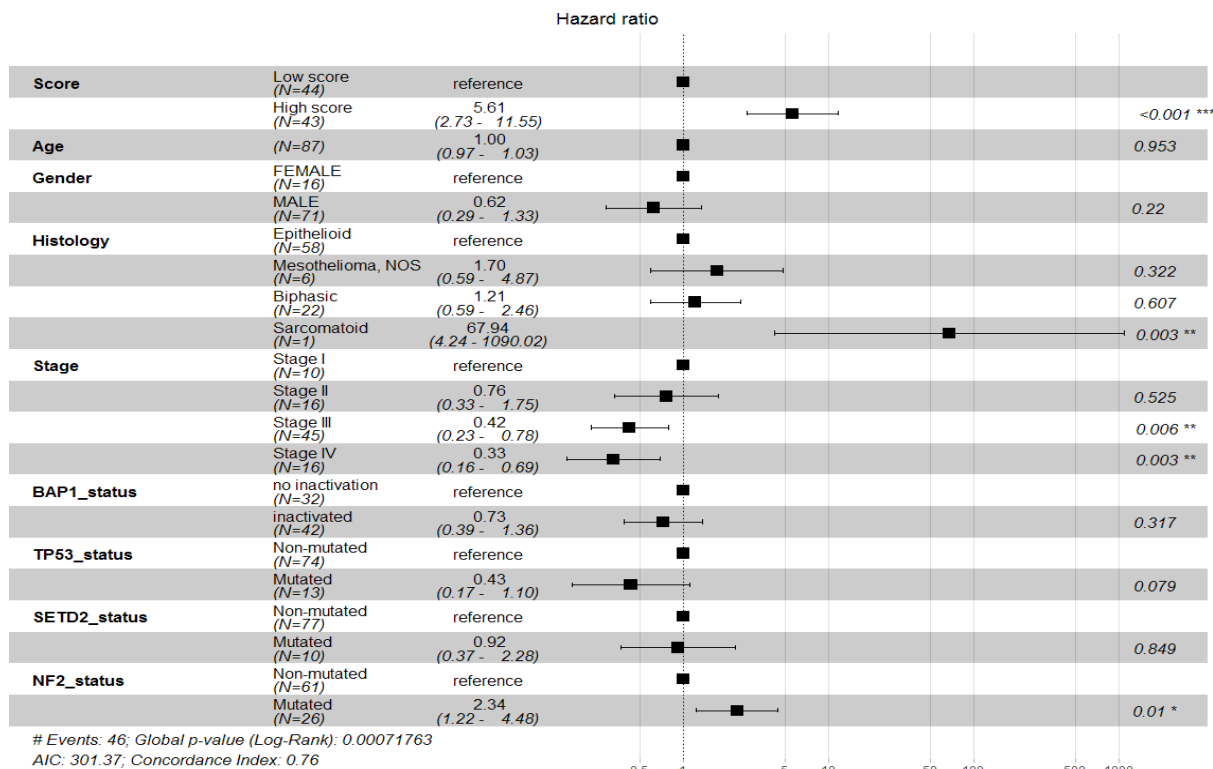
Фигура 5. Унивариационен анализ на преживяемостта според 2-ПС в кохортата от TCGA. Пациентите са разделени на група с висок и група с нисък скор въз основа на медианата на продължителния скор. Стойността на p е от Cox-регресионен анализ.



Фигура 6. Анализ на ROC кривата за прогностичната стойност на бинарния скор по отношение на общата преживяемост в TCGA кохортата. Изчислени са площта под кривата (AUC), както и чувствителността и специфичността при медианна стойност от 21,59, която е използвана за дихотомно разделяне.



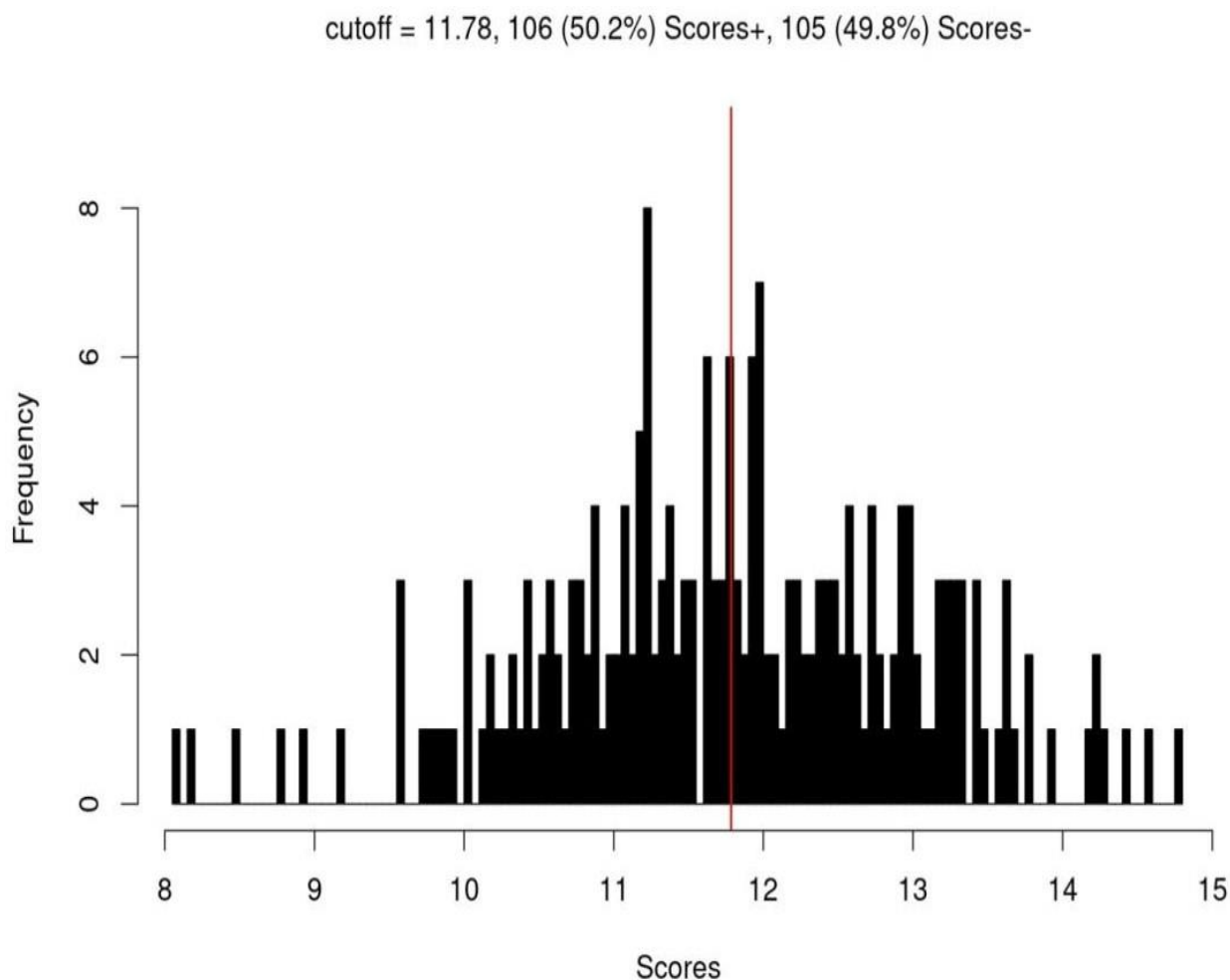
Фигура 7. Мултивариационен анализ на преживяемостта според 2-ПС в кохортата от TCGA. Включени са ковариабилни възраст, пол, хистология, стадий и мутационен статус по отношение на 4 гена BAP1, TP53, SETD2 и NF2.



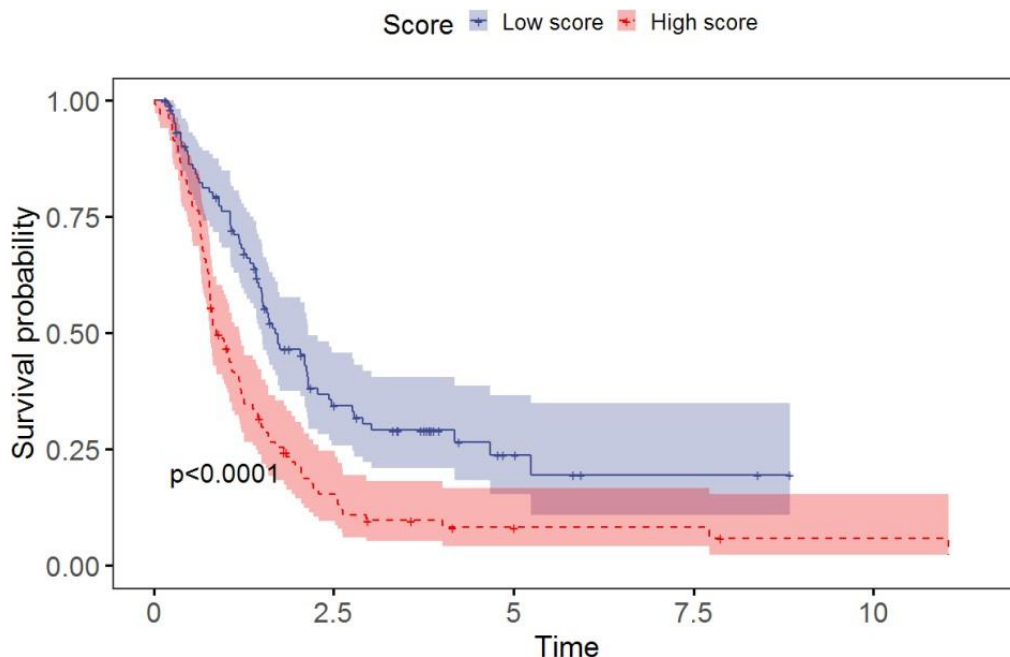
Валидиране на 2-ПС

За да валидираме нашия 2-генен прогностичен скор, използвахме обществено достъпните набора от данни, съдържащи данни за RNA-seq (n=211) на Bueno и планарен експресионен масив (n=67) на Blum. Изчисленият непрекъснат скор в набора от данни на Bueno показва ясна прогностична стойност и беше допълнително преобразуван в двоичен, като се използва медианата на непрекъснатия скор като граница. Аналогично на обучителния набор от данни, двоичният скор в набора от данни за валидиране също има прогностична стойност (Фиг. 8), с AUC на ROC анализа 0,75 (Фиг. 10). Подобно на разширения мултивариантен модел за набора от данни на Bueno, бинарният скор (Фиг. 9) е с независима прогностична стойност (Фиг. 11).

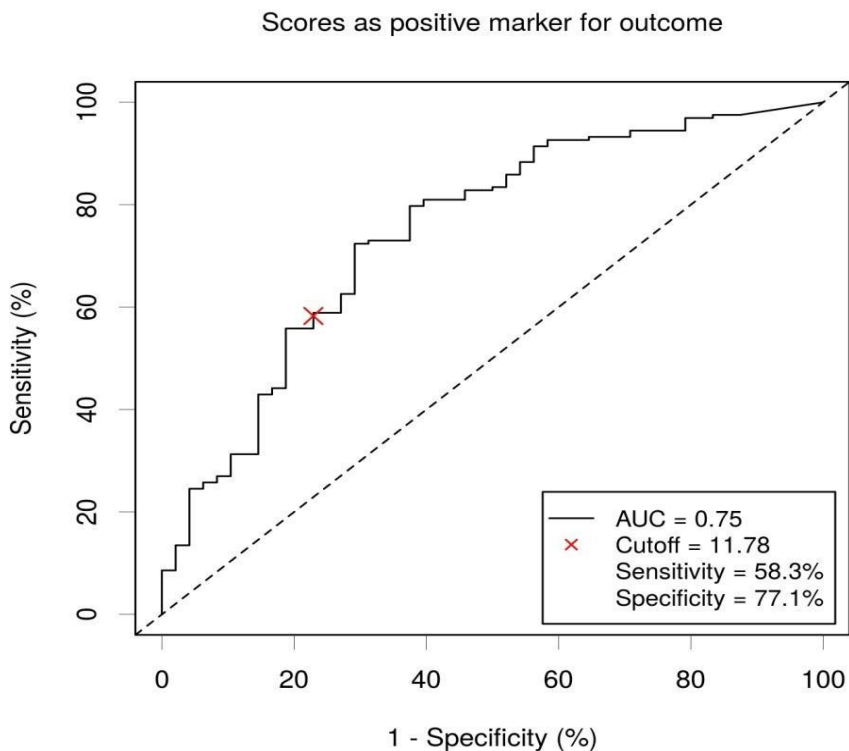
Фигура 8. Хистограма на разпределението на продължителния скор за пациентите от Bueno кохортата. Червената вертикална линия показва медианната стойност, която е използвана за дихотомното (бинарно) групиране на пациенти с висок и на пациенти с нисък скор



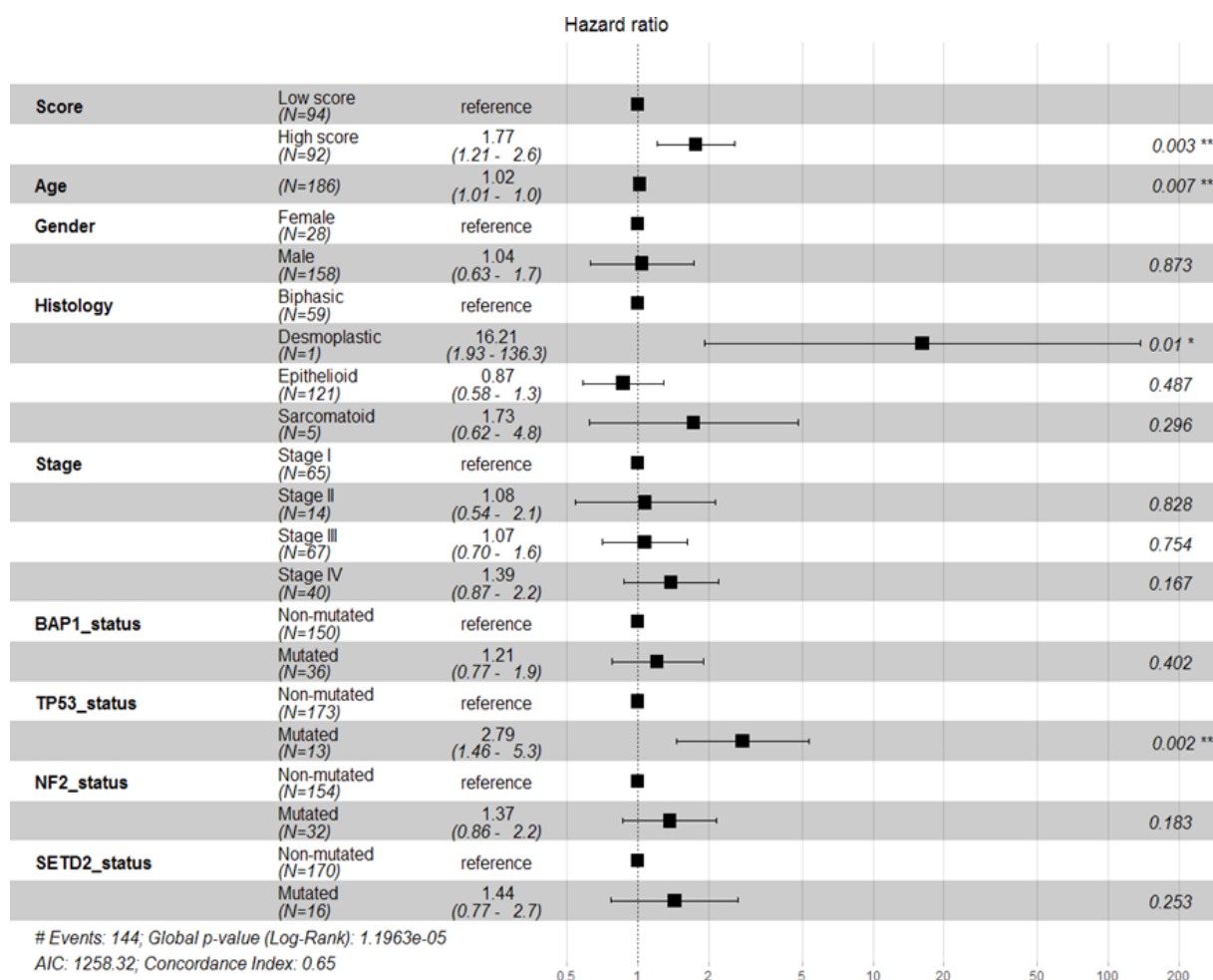
Фигура 9. Унивариационен анализ на преживяемостта според 2-ПС в кохортата от Вуено. Пациентите са разделени на група с висок и група с нисък скор въз основа на медианата на продължителния скор. Стойността на p е от Сох-регресионен анализ.



Фигура 10. Анализ на ROC кривата за прогностичната стойност на бинарния скор по отношение на общата преживяемост в Вуено кохортата. Изчислени са площта под кривата (AUC), както и чувствителността и специфичността при медианна стойност от 11,78, която е използвана за дихотомно разделяне.

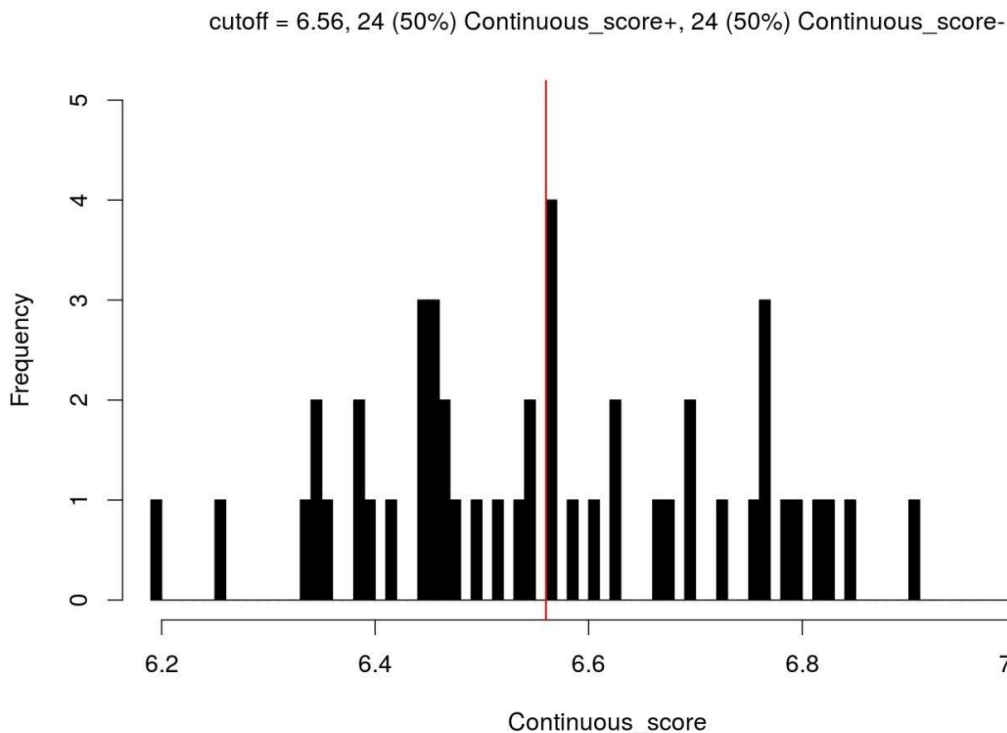


Фигура 11. Мултивариационен анализ на преживяемостта според 2-ПС в кохортата Виено. Включени са ковариабилни възраст, пол, хистология, стадий и мутационен статус по отношение на 4 гена BAP1, TP53, SETD2 и NF2.

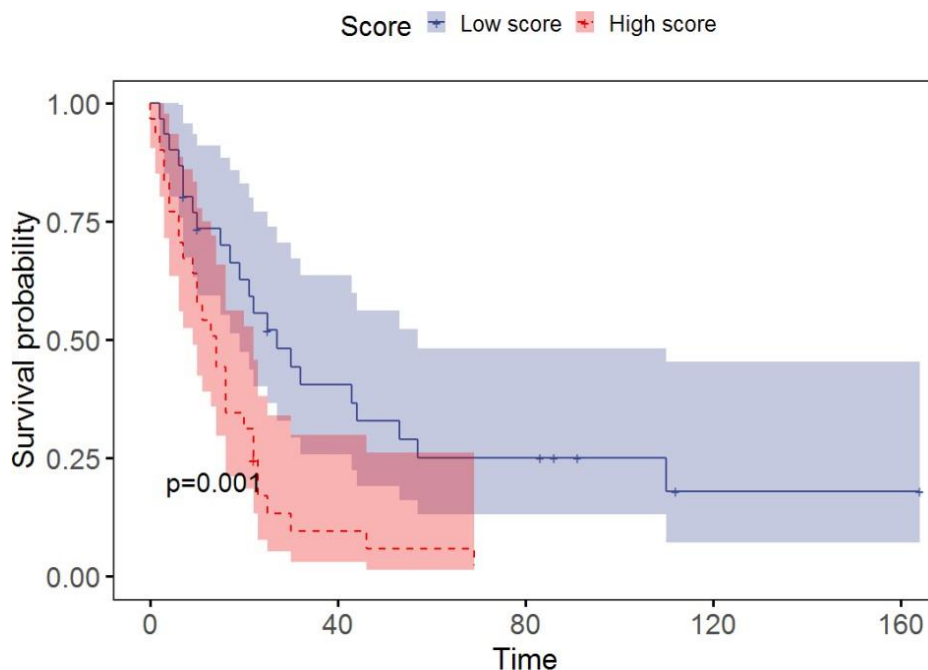


По същия начин беше анализирана и прогностичната стойност на непрекъснатия скор в набора от данни Vlum. При унивариационния анализ непрекъснатият и бинарният скор (Фиг. 12) дефиниран с прагова стойност от медианата за продължителния скор, се доказва като значим независим прогностичен фактор със стойност на AUC от анализа на ROC кривата от 0,85 (Фиг. 14). Мултивариационния модел беше построен с използване на бинарният скор (Фиг. 13), пола, възрастта, стадия и хистологичния подтип на заболяването. Мутационният статус не беше включен, тъй като тези данни в случая не са общодостъпни. Моделът демонстрира независимата прогностична стойност на бинарният скор в този валидиращ набор данни (Фиг. 15).

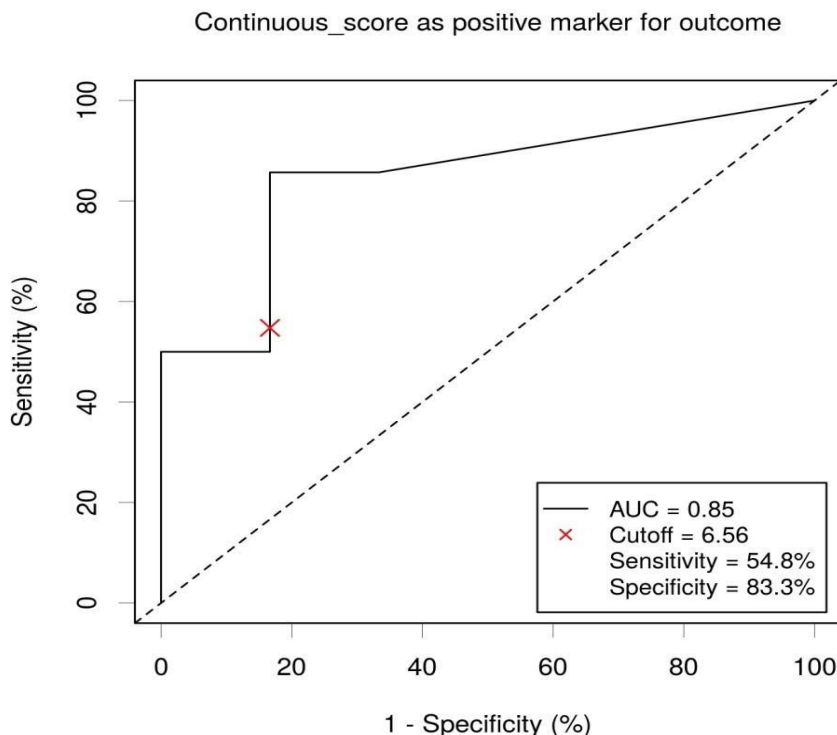
Фигура 12. Хистограма на разпределението на продължителния скор за пациентите от Blum кохортата. Червената вертикална линия показва медианната стойност, която е използвана за дихотомното (бинарно) групиране на пациенти с висок и на пациенти с нисък скор.



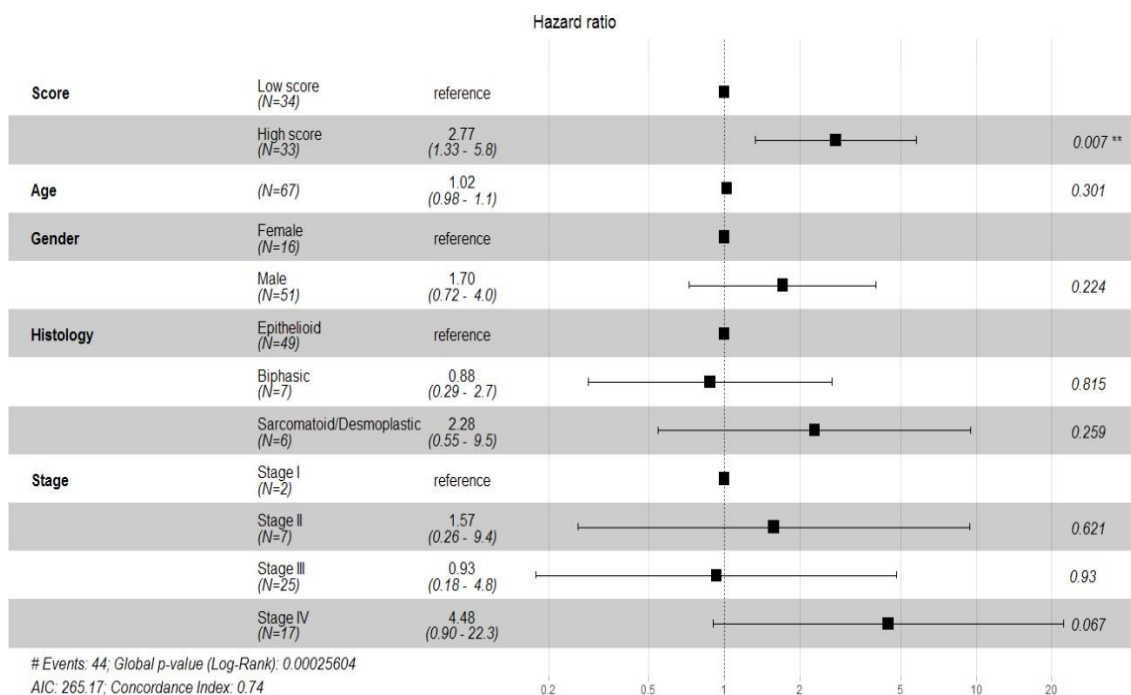
Фигура 13. Унивариационен анализ на преживяемостта според 2-ПС в кохортата от Blum. Пациентите са разделени на група с висок и група с нисък скор въз основа на медианата на продължителния скор. Стойността на p е от Cox-регресионен анализ.



Фигура 14. Анализ за на ROC кривата за прогностичната стойност на бинарния скор по отношение на общата преживяемост в Blum кохортата. Изчислени са площта под кривата (AUC), както и чувствителността и специфичността при медианна стойност от 6.56, която е използвана за дихотомно разделяне.



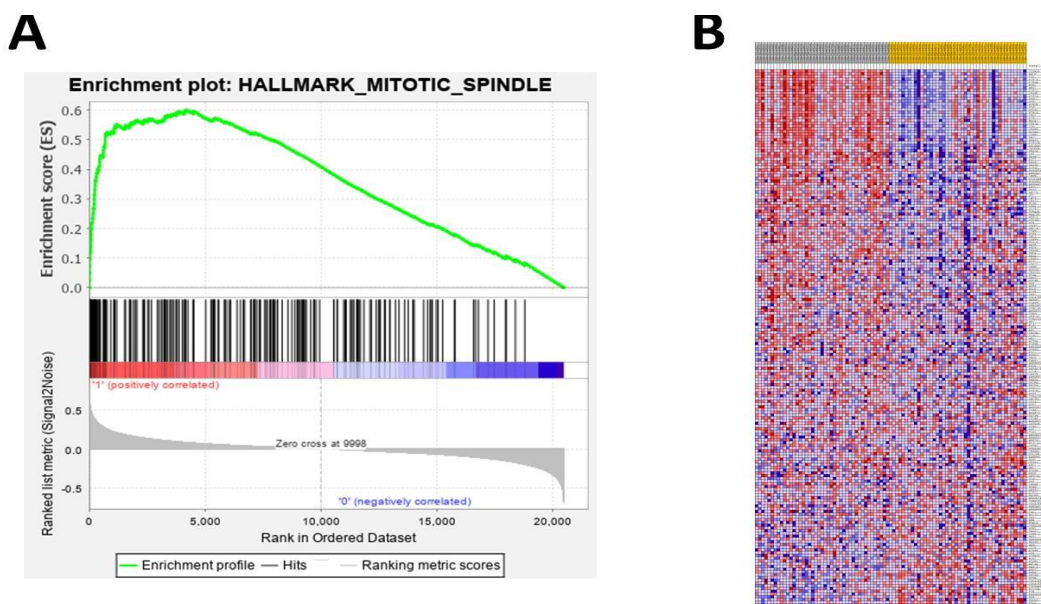
Фигура 15. Мултивариационен анализ на преживяемостта според 2-ПС в Blum кохортата. Включени са ковариабилни възраст, пол, хистология, стадий и мутационен статус по отношение на 4 гена ВАР1, TP53, SETD2 и NF2.



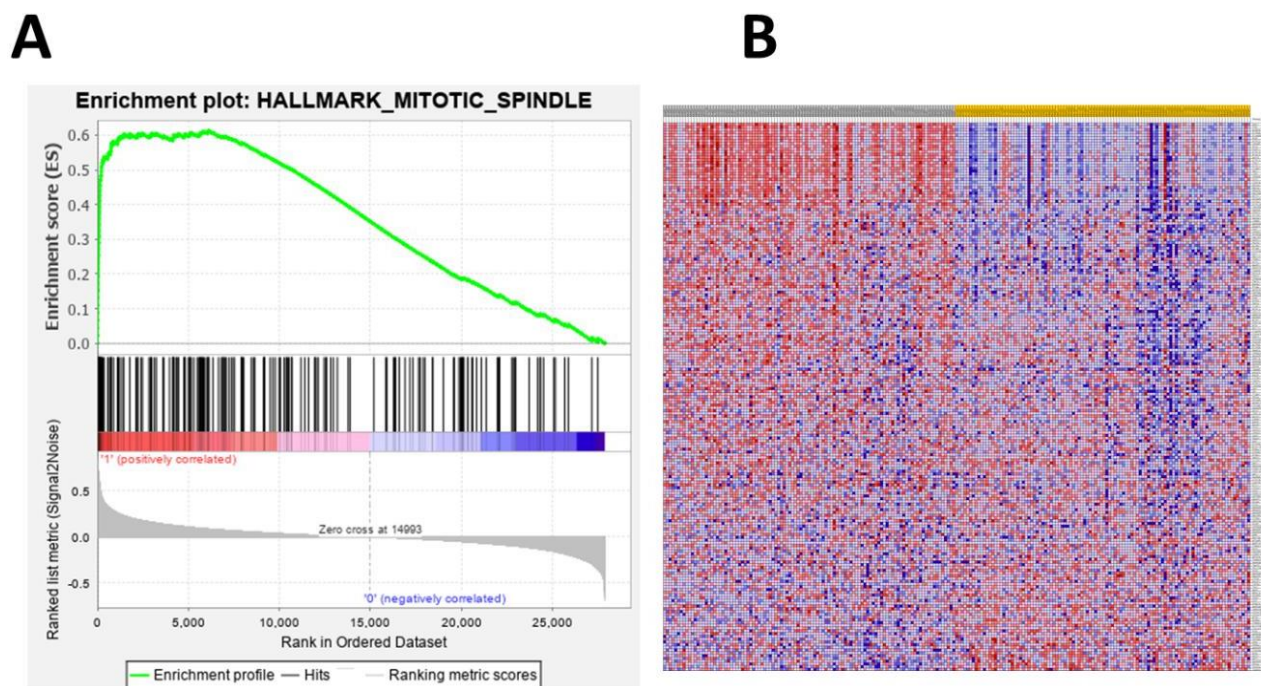
Анализ на обогатяването на списъци от гени в експресионни профили (GSEA)

Въз основа на наблюдението, че разработеният от нас нов 2-ПС показва сходна прогностична стойност както при трениращия, така и при валидиращите набори от данни скорът може да корелира със специфичен профил на генна експресия. Затова извършихме GSEA анализ с използване на предварително дефинирани списъци от гени, съответстващи на основните характеристики на рака, които са депозирани в базата данни MSig. За всеки от нашите набори от данни получихме определен брой профили от гени, които са обогатени в подгрупите с висок скор. Броят им беше както следва: TCGA (n=37), Вуено (n=34), Вlum (n=34). Примерни плотове, показващи обогатяването на профила „митотично вретено” и при пациентите с висок скор и при трите кохорти са представени на Фиг. 16, 17 и 18. Видно беше, че съществува по-значително припокриване между профилите между TCGA и Вуено, и малко по-слабо такова с набора данни Вlum. Въпреки това обаче общо 25 профила бяха свръхекспресирани в пациентите с висок скор от трите кохорти (Фиг. 19 и Фиг. 20). Повечето от тях бяха свързани с ДНК поправката и отговора към увреждане на ДНК и контрола на митотичното клетъчно делене.

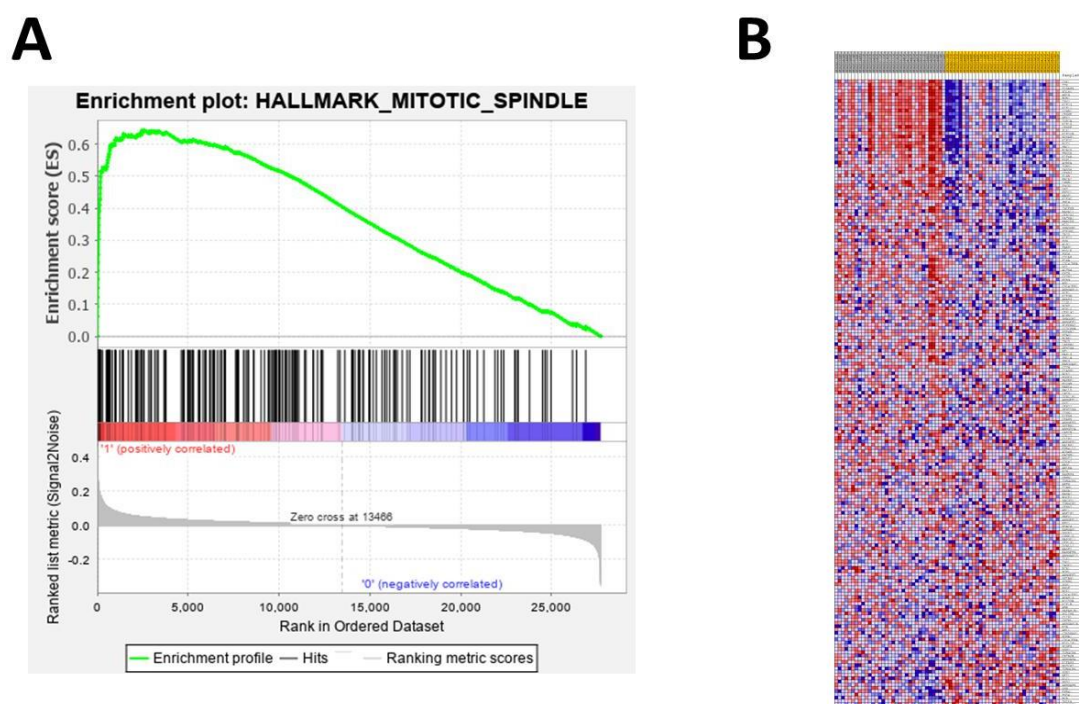
Фигура 16. Примерен резултат от GSEA анализа в набора от данни TCGA. (A) Enrichment plot за гените от списъка „Митотично вретено“; (B) Heatmap на експресията на гените същия списък. В сиво са показани пациентите с висок скор, а в жълто – тези с нисък.



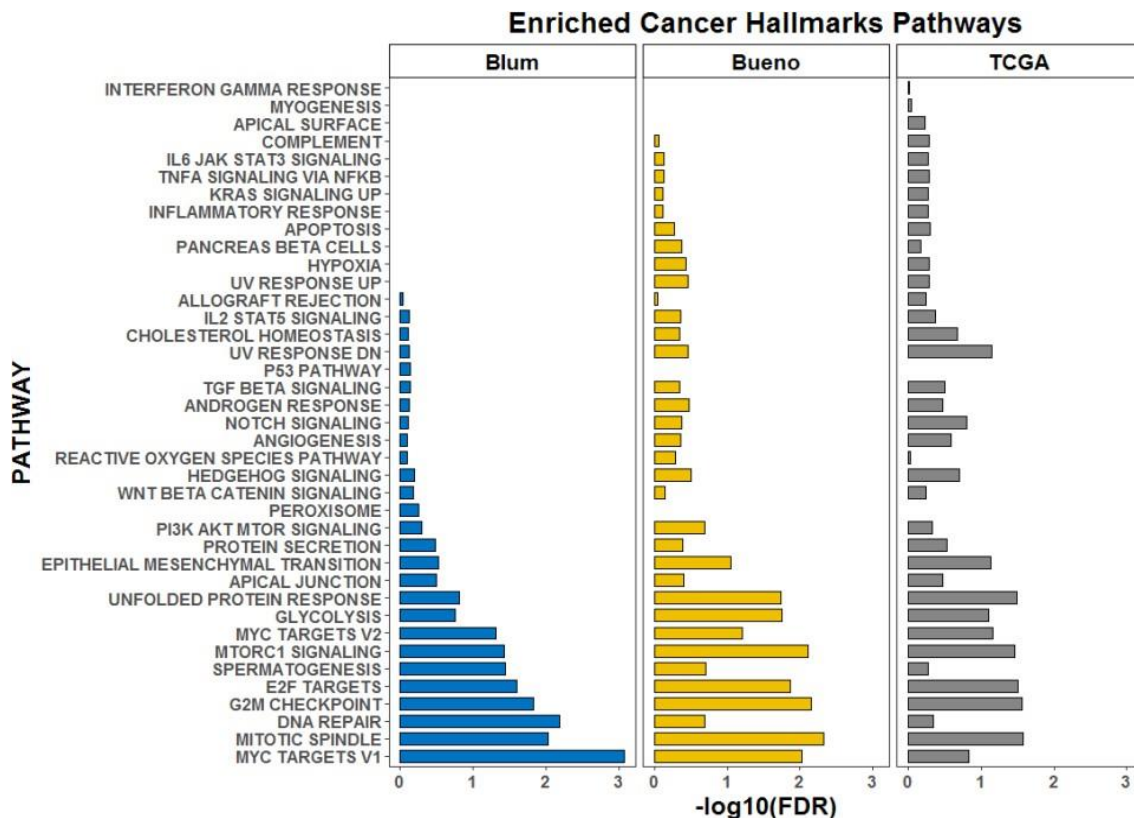
Фигура 17. Примерен резултат от GSEA анализа в набора от данни Bueno. (A) Enrichment plot за гените от списъка „Митотично вретено“; (B) Heatmap на експресията на гените същия списък. В сиво са показани пациентите с висок скор, а в жълто – тези с нисък



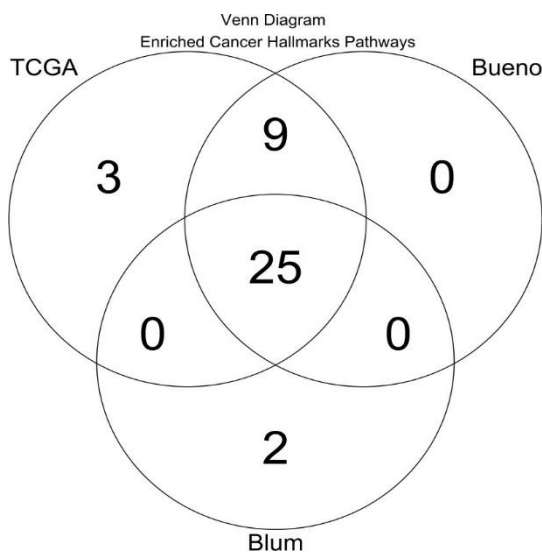
Фигура 18. Примерен резултат от GSEA анализа в набора от данни Blum. (A) Enrichment plot за гените от списъка „Митотично вретено“; (B) Heatmap на експресията на гените същия списък. В сиво са показани пациентите с висок скор, а в жълто – тези с нисък



Фигура 19. Стълбови диаграми на сигнификантно по-често свръхекспресирани списъци от гени в пациентите с висок скор, които се откриват в поне една от трите кохорти (TCGA, Bueno, Blum), определени чрез GSEA анализ. Числата показват броя на списъците от гени във всяко от множествата.



Фигура 20. Venn диаграма на припокриващите се сигнификантно по-често свръхекспресирани списъци от гени в пациентите с висок скор от трите кохорти (TCGA, Bueno, Blum), определени чрез GSEA анализ. Числата показват броя на списъците от гени във всяко от множествата.



Интегриран анализ на ДНК метилиране и генна експресия

Анализирахме дали скорът дефинира различни в епигенетично отношение подтипове на заболяването, като използвахме наличните данни за ДНК метилиране, получени чрез Infinium 450K bead chip (Illumina) за пациентите от TCGA и Blum кохортите. За анализа използвахме SONCAP алгоритъма, който позволява интегриране на данните от ДНК метилирането с тези за генна експресия, така че да се установят CpG сайтове и острови, които са диференциално метилирани, но нивото на метилиране корелира обратно с нивото на генна експресия на прилежащи гени, което помага за идентифициране на потенциално функционално значими сайтове за епигенетична регулация на генната експресия. Анализирани бяха 87 ДНК метилационни профила на пациенти от TCGA кохортата, при което бяха установени 428 диференциално метилирани CpG сайта между пациентите с висок и нисък скор (Фиг. 21), разположени в три диференциално метилирани CpG острова (Табл. 2), 2 от които бяха обратно корелиращи в генната експресия на прилежащите гени – *SLC20A1* и *KIAA1949* (Фиг. 22 и 23, Табл. 3).

Фигура 21. Резултат от анализа на диференциално метилираните CpG сайтове в набора от данни TCGA. Heatmap на β -стойностите за диференциално експресираните CpG сайтове (по вертикала) спрямо пациентите (по хоризонтала)

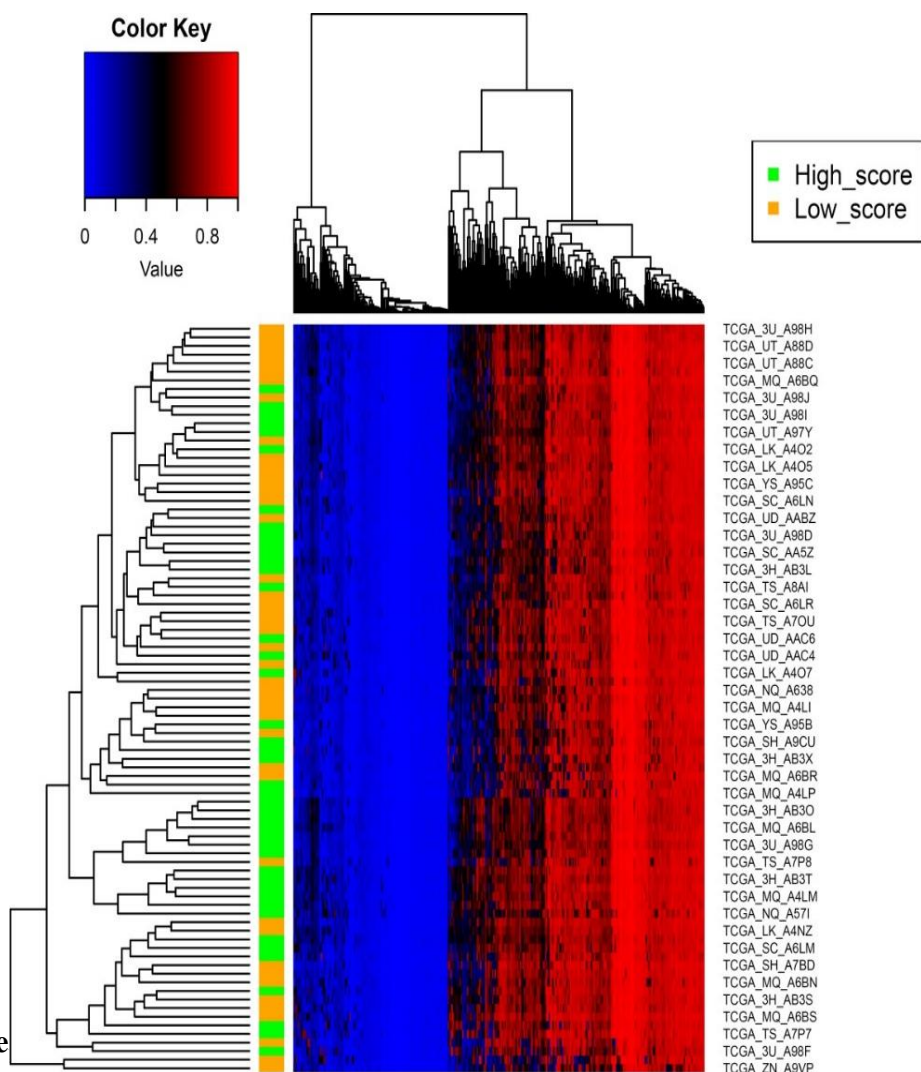


Таблица 2. Ресултати от анализа на CpG острови от набора данни TCGA. За целите на анализа островът е дефиниран като поне два прилежащи CpG сайта, които са включени в Infinium 450K чипа на Illumina

СрG остров координати	Гени	Средна β стойност за пациенти с висок скор	Средна β стойност за пациенти с нисък скор	Разлика в средните β стойности висок спр. Нисък скор	Номинална p-стойност за острова	FDR стойност за острова	Брой СрG сайтове
chr11:2923301-2923817	<i>SLC22A18</i> ; <i>SLC22A18AS</i>	0,166689202	0,342562598	-0,175873397	1,15E-06	3,46E-06	2
chr2:113403001-113404079	<i>SLC20A1</i>	0,217591513	0,364831843	-0,14724033	9,22E-05	0,00013824	2
chr6:30654392-30654934	<i>KIAA1949</i>	0,216549012	0,381515783	-0,164966771	0,000248645	0,00024864	13

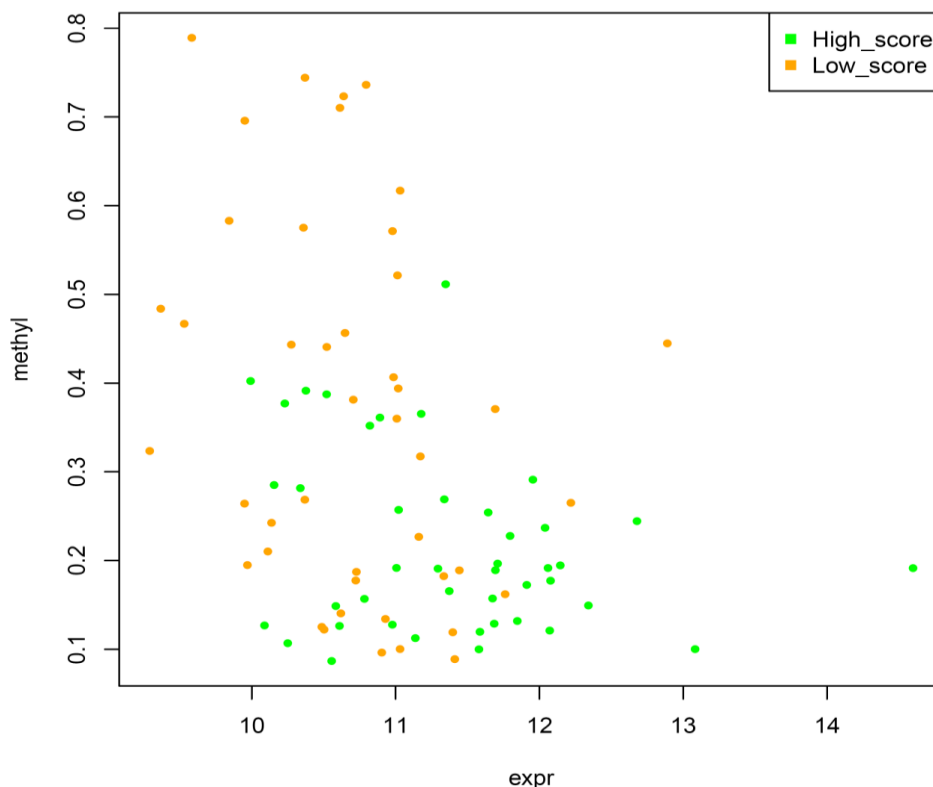
Таблица 3. Резултати от интегрирания анализ на диференциално метилирани CpG острови и гена експресия от набора данни TCGA. За целите на анализа островът е дефиниран като поне два прилежащи CpG сайта, които са включени в Infinium 450K чипа

СрG остров координати	Посока на метилирането	Гени	Корелационен коефициент	Номинална р-стойност	FDR стойност
chr2:113403001-113404079	Понижено метилиране	SLC20A1	-0,364768107	0,0005135	0,000513
chr6:30654392-30654934	Понижено метилиране	KIAA1949	-0,778921734	6,55E-19	1,31E-18

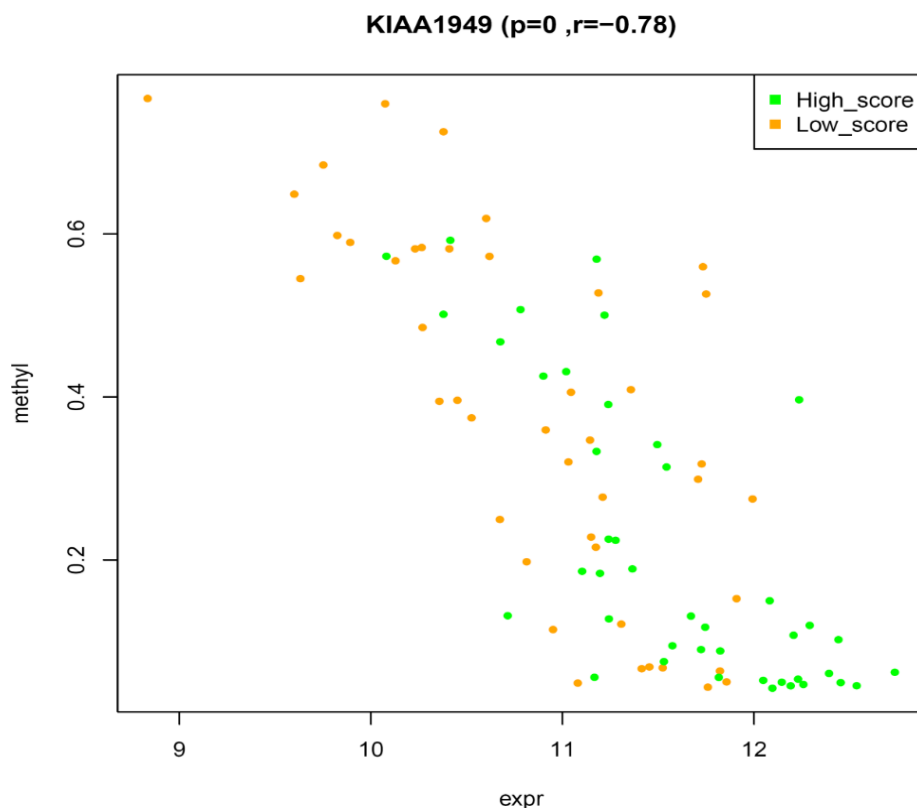
на Illumina.

Фигура 22. Двумерен корелационен плот на метилиране на CpG острова chr2:113403001-113404079 спрямо експресията на гена SLC20A1 в набора от данни TCGA. Пациентите с висок и нисък скор са обозначени в различен цвят.

SLC20A1 ($p=0$, $r=-0.36$)

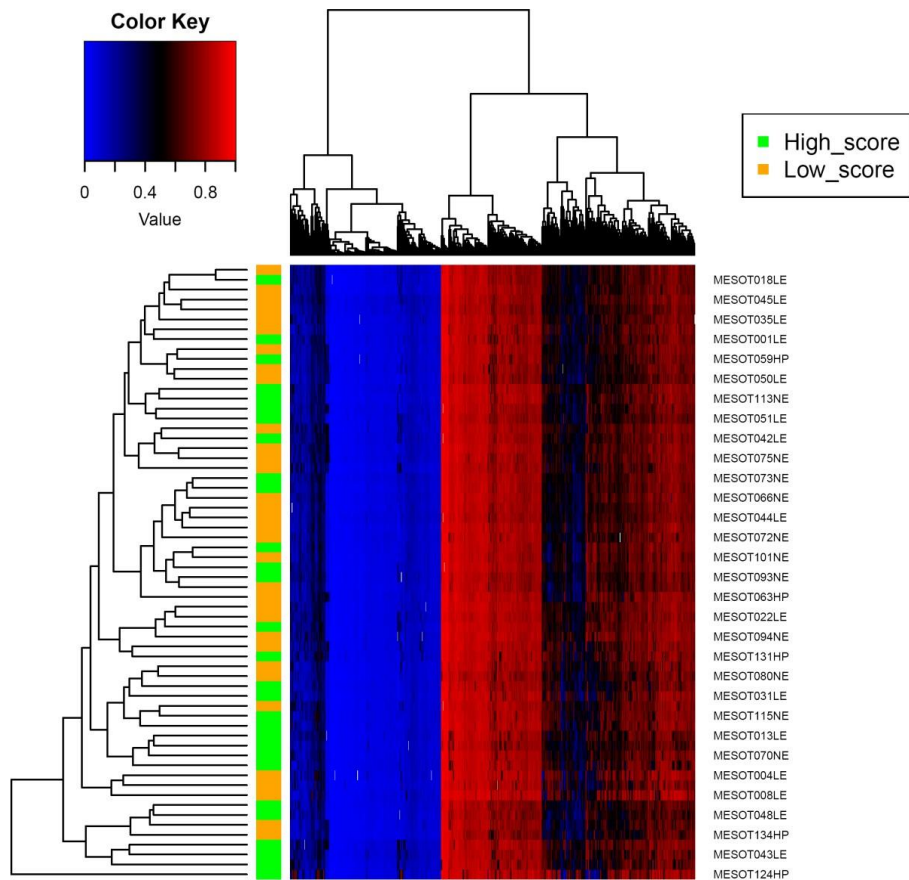


Фигура 23. Двумерен корелационен плот на метилиране на CpG острова chr6:30654392-30654934 спрямо експресията на гена KIAA1949 в набора от данни TCGA. Пациентите с висок и нисък скор са обозначени в различен цвят.



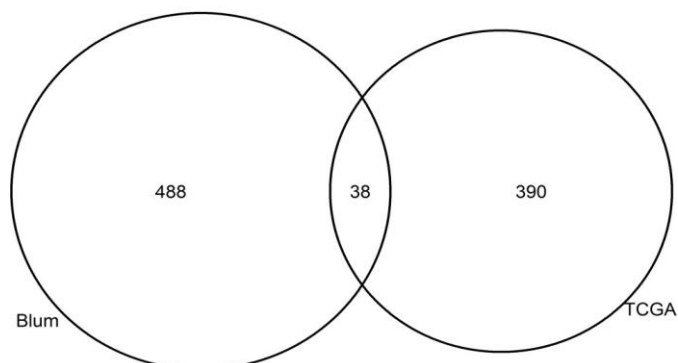
По същия метод анализирахме 62 проби от набора данни Vlum. Така бяха установени 526 диференциално метилирани CpG сайта (Фиг. 24), като само малка част от тях (n=38) бяха идентични с тези от TCGA набора данни (Фиг. 25). Бяха определени 39 CpG острова, които показаха диференциално метилиране и обратна корелация с нивата на генна експресия на прилежащите гени.

Фигура 24. Резултат от анализа на диференциално метилираните CpG сайтове в набора от данни Blum. Heatmap на β -стойностите за диференциално експресирани CpG сайтове (по вертикала) спрямо пациентите (по хоризонтала).



Фигура 25. Venn диаграма на припокриващите се сигнификантно диференциално метилираните CpG сайтове в пациентите с висок спрямо нисък скор от две от кохортите (TCGA, Blum), определени чрез CONCAP анализ. Числата показват броя на списъците от гени във всяко от множествата.

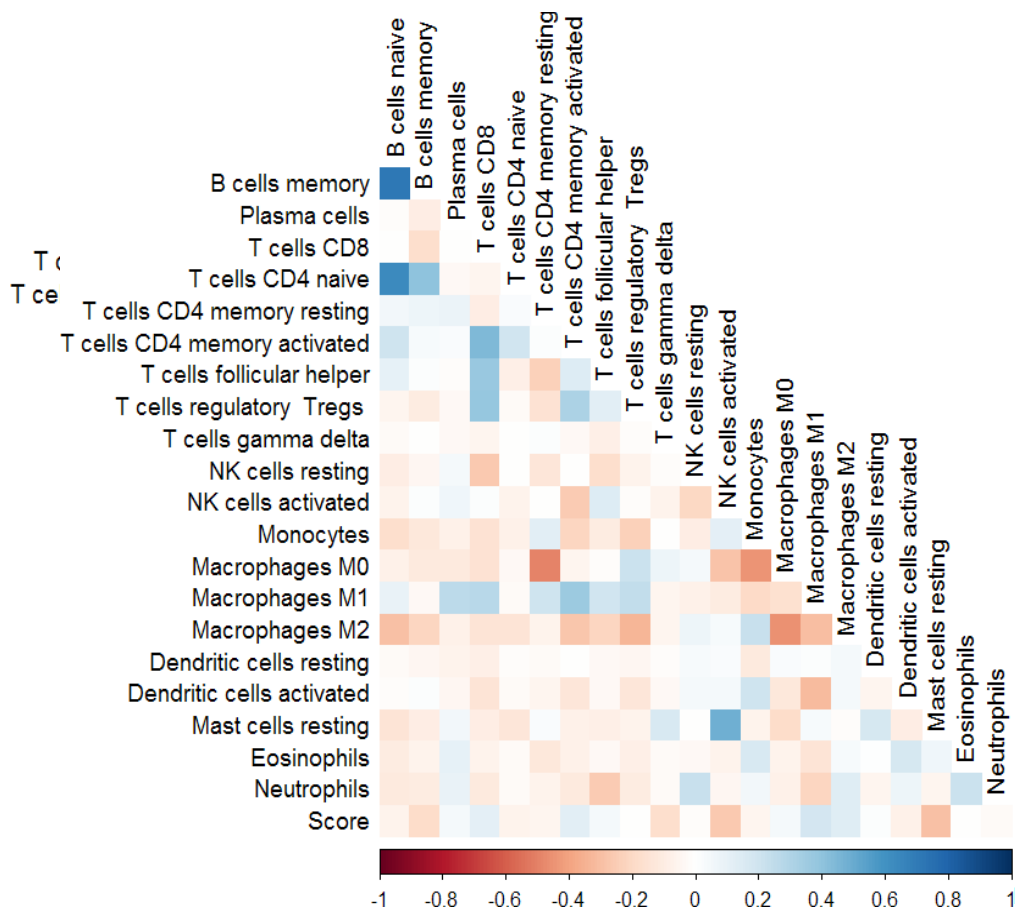
Differentially methylated CpGs



Корелация с изчислени популации на инфилтриращи имунни клетки

Инфилтриращите имунни клетки са основен играч в имунния отговор срещу рак и могат да се използват като прогностичен и предиктивен маркер. Затова анализирахме дали 2-ПС корелира с инфилтрация на подтип на специфични имунни клетки в МПМ. Използвахме изведените фракции на инфилтриращи имунни клетки, използвайки алгоритъма CIBERSORT върху наборите от данни TCGA (Фиг. 26) и Виено (Фиг. 27). Специалното и за двата набора от данни е, че непрекъснатият прогностичен скор показва положителна корелация с CD8⁺ Т-клетъчната фракция, както и с фракциите на макрофагите M1 и M2.

Фигура 26. Корелационна матрица на продължителния скор с имуен скор за
 Фигура 27. Корелационна матрица на продължителния скор с имуен скор за
 инфилтрация от определени популации клетки на имунната система според
 експресионните данни от Вуено кохортата. За изчисляването на имунните скорове е
 използван алгоритъмът CIBERSORT. Цветът на всяко квадратче от корелационната
 матрица отразява изчисленият корелационен коефициент.

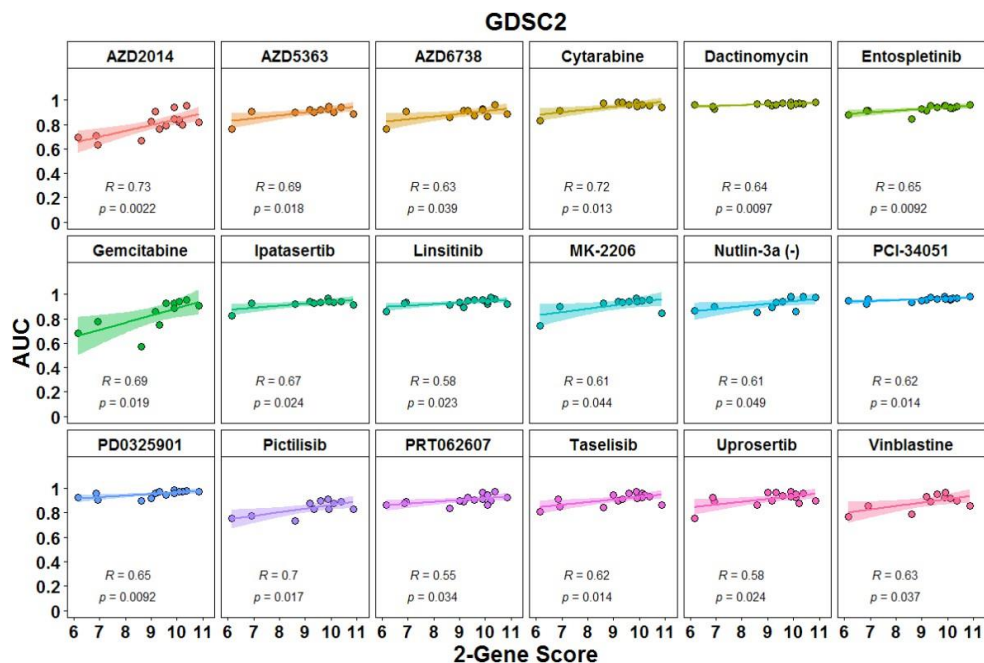


Потенциална прогностична способност на 2-ПС

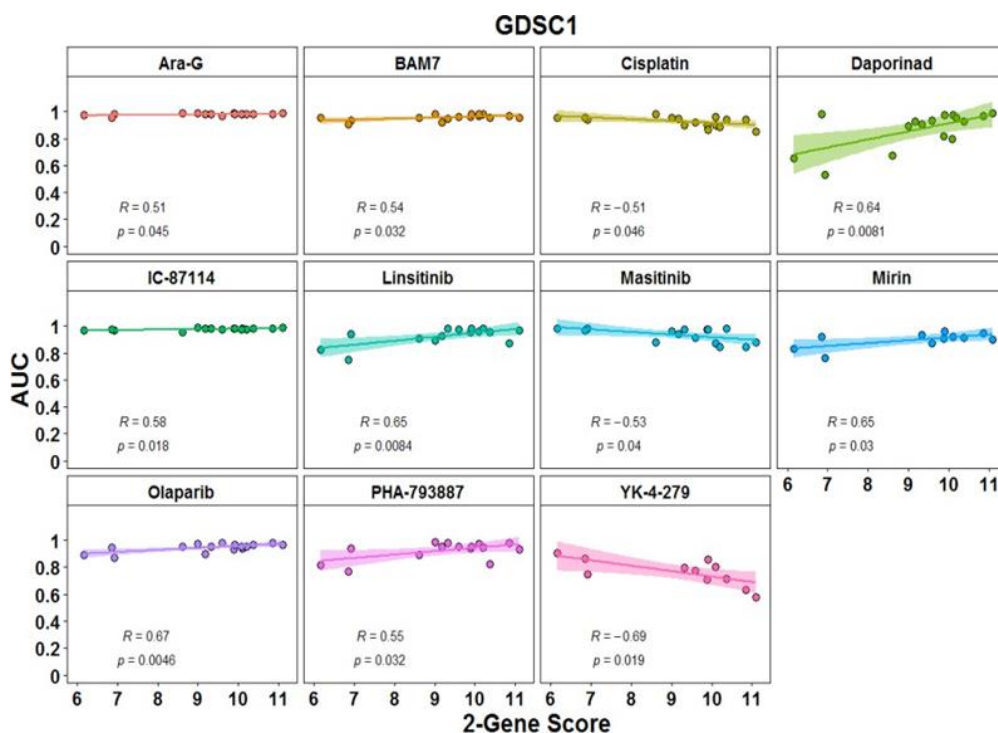
Изчислихме 2-ПС за всяка от МПМ клетъчните линии, включени в проекта за геномика на лекарствената чувствителност при рака (GDSC) - GDSC1 (n=16) и GDSC2 (n=15). След това търсихме корелация между 2-ПС и чувствителността към всяко от лекарствата, тествани в проекта, използвайки стойности на AUC, чрез която стойност определихме отговора към 11 лекарства от набора GDSC1, което показва значителна корелация с 2-ПС на тестваните клетъчни линии на мезотелиом (Фиг. 28), докато за GDSC2 броят на такива

значими корелации е 18 (Фиг. 29). Този анализ разкри корелация на 2-ПС с отговора на често използвани лекарства при лечение на мезотелиом, като цисплатин ($R=-0.51$, $p=0.046$), гемцитабин ($R=0.69$, $p=0.019$) и винбластин ($R=0.63$, $p=0.037$).

Фигура 28. Корелационни двуразмерни плотове между изчисленият 2-ПС скор и AUC за различни лекарства или съединения за МПМ клетъчни линии, анализирани като част от проекта GDSC1. Представени са само плотове за идентифицираните сигнификантни корелации.



Фигура 29. Корелационни двуразмерни плотове между изчисленият 2-ПС скор и AUC за различни лекарства или съединения за МПМ клетъчни линии, анализирани като част от проекта GDSC2. Представени са само плотове за идентифицираните сигнификантни корелации.



V. Обсъждане

МППМ представлява значително медицинско предизвикателство поради тежката прогноза на заболялите и поради това, че е пример за рядко раково заболяване, предизвикано от експозиция на облигатен канцероген във външната (най-често работна) среда. Специалното при МППМ е, че цикълът на екстремна диверсификация на туморната тъкан се провокира от хронично тумор-промотиращо възпаление, което се инициира и поддържа от проникналите и персистиращи в плевралната кухина азбестови фибри. По-нататък в патогенетичните механизми на развитието на МППМ се включват допълнителни характерни черти на рака като геномна нестабилност и екстремна мутационна изменчивост, репликационно безсмъртие, загуба на туморни супресори, продължително пролиферативно сигнализиране, избягване на клетъчната смърт, неоангиогенеза, придобиване на способност за метастазиране чрез епителна-мезенхимна транзиция. Способността на трансформираните мезотелни клетки да придобиват разнообразни основни характеристики на раковия растеж определят агресивния клиничен ход на заболяването и ограниченото терапевтично повлияване с конвенционалните хирургични и химиотерапевтични подходи.

В такива случаи от особено значение е правилното прогностично стратифициране на пациентите към диагнозата и евентуално използване на прогностични и предиктивни биомаркери. Стандартното стадиране на МППМ според TNM класификацията не дефинира групи с голяма разлика в преживяемостта, тъй като тя зависи от подлежащите биологични характеристики на заболяването, по-точно хистологичният субтип.

Поради тези причини откриването на нови биомаркери и разработването на нови прогностични и предиктивни модели при МППМ чрез използване на данни, които по-адекватно отразяват подлежащите патологични характеристики на индивидуалното заболяване са от основно значение. Най-рано приложената в прогностичната оценка на онкологичните пациенти омикс технология беше транскрипционният анализ чрез използване на планарни микрочипове. През последното десетилетие допълнително се използва и определяне на генната експресия чрез РНК секвениране. След 2000 г. тези технологии се прилагат и за молекулярно профилиране на МППМ. Редица проучвания предлагат базирани на генна експресия прогностични модели при МППМ. Те се различават значително в подхода за избор на отличителни черти (гени), набори от данни за обучение и валидиране, броя на гените, включени в крайния модел, както и в

качеството на различни МПМ кохорти, в резултат на което са се извеждали 3-генен и 5-генен прогностичен скор.

При нашето проучване въведохме нов оригинален подход при избора на изходни маркери за модела, като ограничихме броя на гените, които да се тестват за включване в МПМ прогностичния модел само до такива, за които се знае, че МПМ клетъчните линии са чувствителни към тяхното изключване (knock-down) чрез siRNA или CRISPR/Cas9 редакция. След това приложихме *rbsurv* подхода към набора от данни на TCGA и изградихме двугенен прогностичен модел, който показва умерена прогностична способност като непрекъснат или двоичен скор както в едновариантни, така и в мултивариантни модели в три различни МПМ кохорти. Тази умерена прогностична способност е в компромис с минималния брой гени, включени в прогностичния скор, като се избягва прекомерното напасване на модела чрез включване на по-голям брой гени. Ограниченият брой гени в нашия 2-ПС може да позволи по-нататъшно просто валидиране с помощта на техники с ниска разделителна способност, като количествен RT-PCR или имунохистохимия. Двата гена, включени в нашия модел не са широко изследвани в МПМ, макар че такива данни са налични поне за *MAD2L1*.

Генът *MAD2L1* кодира съответния протеин, който е неразделна част от контролния пункт за асемблирането на митотичното вретено и гарантира, че всички хромозоми са правилно подредени в метафазната пластинка преди клетката да може да премине към анафаза. Наскоро беше установено, че *MAD2L1* гена е свръхекспресиран в няколко МПМ клетъчни линии на ниво иРНК и протеин. Това изследване съответства на предишно, което демонстрира по-висока експресия на *MAD2L1* белтък (както ядрена, така и цитоплазмена) в МПМ клетъчни линии в сравнение с нормалния мезотел. По-високата ядрена експресия на *MAD2L1*, определена с помощта на имунохистохимия, корелира с по-кратка обща преживяемост. Скорозно проучване показва, че *BRCA1* при мезотелиом води до съвместно изчерпване на *MAD2L1* иРНК и протеин. Освен това, загубата на *BRCA1/MAD2L1* се свързва с резистентност към винорелбин *ex vivo* и преживяемостта на пациентите. Преживяемостта е по-кратка при пациенти с липса на *BRCA1/MAD2L1* експресия в сравнение с тези двойно положителни с тумори по двата гена. Това наблюдение може да обясни факта, че нашият 2-ПС корелира с резистентност към винбластин (инхибитор на сглобяването на митотичното вретено) и олапариб (инхибитор на поли (АДФ-рибоза) полимераза, PARP) в МПМ клетъчни линии. Освен това, сред най-обогатените пътища в анализа на

GSEA и на трите кохорти има пътища, включващи директно механизми на репликация на ДНК, като: „Митотично вретено”, „G2M контролен пункт” и „поправка на ДНК“

Генът *GOLT1B* кодира човешкия везикуларен транспортен протеин (Golgi Transport 1B) GOT1B белтък. *GOLT1B* може да бъде свръхекспресирани в различни тумори поради амплификация в хромозомна област 12p. Последните проучвания показват, че свръхекспресията на *GOLT1B* при рак на гърдата и рак на дебелото черво може да бъде свързана с по-лоши скорове поради промотиране на избягването на имунната система. В съответствие с това открихме, че при пациенти с висок скор 2-ПС МПМ от всичките три кохорти в нашия анализ „Епително-мезенхимна транзиция—, „Апикална връзка— и „Пътища на протеинова секреция— са значително обогатени в профилите на генната експресия.

Нашия GSEA анализ показва, че и при трите кохорти има значително обогатяване на свръхекспресирани гени в пациентите с висок скор, които имат значение за основните характеристики на рака, които са интегрална част от биологията на МПМ. Освен това става ясно, че поради голямата възпроизводимост на профилите и в трите кохорти вероятно изведеният от нас 2-ПС отразява истински подлежащи особености на пациентите с висок скор спрямо тези с нисък. Не толкова възпроизводими обаче се оказаха профилите на ДНК метилиране като имахме възможност да използваме данни само от две кохорти, които обаче бяха от една и съща платформа за анализ на ДНК метилиране. При нашият анализ обаче това може да се дължи на разлики в биологията на МПМ, включени в отделните кохорти, тъй като пролиферативният индекс на различните пациенти може да е много различен и това да води до по-различни нива на ДНК метилиране поради високата клетъчна кинетика. Освен това открихме много малък брой гени в TCGA кохортата, чиято експресия обратно корелира с нивото на метилиране на прилежащи CpG острови, открити при набора от данни Blum.

Споменатите по-горе съобщения относно ролята на *GOLT1B* в избягването на имунния надзор допълнително ни подтикнаха да проучим дали нашият 2-ПС корелира с оценените фракции на имунните клетки в туморната тъкан. Използвахме вече стандартния алгоритъм за деконволуция и извеждане на имуноен скор за инфилтрация от имунни клетки в туморна тъкан, за да покажем, че 2-ПС корелира със съдържанието на CD8+ Т-клетки и M1/2 макрофаги. Корелацията на нашия скор с инфилтрация от макрофаги е в съответствие с вече известния патогенетичен механизъм на поддържано високо ниво на хронично възпаление при МПМ поради неефективна фагоцитоза на

азбестовите фибри. Известно е, че макрофажната поляризация към M2 фенотип може да е свързана с по-лоша прогноза при МПМ. По-високата инфилтрация от CD8+ Т лимфоцити може да е израз на по-мошен адаптивен имунен отговор при пациентите с по-висок 2-ПС, респективно с по-агресивно и по-високо пролиферативно заболяване. Това се основава на предишни наблюдения, че по-високата инфилтрация на МПМ от CD8+ TILs е свързана с по-добра прогноза след резекция. В условията на хронично възпаление, голяма част от CD8+ TILs придобиват експресионен профил на изтощени лимфоцити с характерна повишена експресия на рецептори за инхибиращи сигнали като PD-1. Използвайки подобен на нашия подход за оценка на инфилтриращите МПМ фракции имунни клетки, Blum et al. демонстрираха, че епителоподобна морфология и транскриптомният профил корелират с изчислената фракция на CD8+ Т клетки. Nguen et al. също показва, че изчислените фракции на инфилтриращи имунни клетки могат да бъдат комбинирани с геномни параметри с цел разработване на прогностични модели в МПМ. Друго скорошно проучване показва, че маркерите за по-високи нива на системно възпаление корелират с по-кратка обща преживяемост при пациенти с МПМ. Нашите наблюдения в контекста на тези проучвания предполагат, че имунните маркери трябва да бъдат включени в прогностичните схеми за пациенти с МПМ. Освен това е смислено да се очаква, че те могат да имат предиктивна способност за успеха на терапията, базирана на инхибитори на имунни контролни пунктове (ICIs) при МПМ, заедно с други имуногенетични маркери.

В ерата на имунотерапията с ICIs, комбинациите с конвенционална химиотерапия или таргетна терапия, могат да доведат до допълнителна клинична полза при МПМ. Затова ние допълнително оценихме нашия 2-ПС като възможен маркер за прогнозиране на чувствителността на МПМ клетъчните линии към различни нискомолекулни лекарства. Интересно е, че 2-ПС корелира обратно със стойностите на AUC за цисплатина, което предполага, че може да предскаже по-висока чувствителност към него. Обратното наблюдение беше направено за два други често срещани химиотерапевтика като гемцитабин и винбластин, което предполага, че нашият 2-ПС може да предвиди резистентност към тях.

Въпреки че нашето проучване успешно показва мощта на интегрирането на различни омикс данни от различни платформи и кохорти пациенти при МПМ, особеностите на раковия растеж, като интратуморна хетерогенност ограничават този подход. В тази насока показателно е едно ново мулти-омикс проучване при МПМ

което се основава на scRNA-Seq и генотипни данни, което идентифицира три типа фенотипа клетки в МПМ, а именно с изразена способност за туморна пролиферация, клетки, способни да избягват имунния отговор, и такива с ациерна морфология и загуба на експресия на *BAP1* гена като клиничното протичане при различните пациенти зависи от преобладаващия подтип малигнени мезотелни клетки.

Всичко това показва, че бъдещите насоки в прогностичното моделиране при МПМ ще се фокусират върху интегрирането на нови биомаркери, методи за изображение и технологии за молекулярно профилиране, за да се подобри прогностичната точност и да се усъвършенства стратификацията на пациентите. Мулти-омикс подходите, комбиниращи геномни, транскриптомни и протеомни данни, вероятно ще идентифицират все по-успешно прогностични профили и терапевтични мишени при МПМ. Освен това, напредъкът в изкуствения интелект и анализа на данни ще позволява разработването на сложни предиктивни модели, способни да предсказват индивидуалния риск и оптимизират лечението. Непрекъснатите усилия в разработването, валидирането и транслирането в клиничната практика на подобни модели в крайна сметка ще са от съществено значение за подобряване на клиничните резултатите при пациентите с МПМ, както и с други редки видове рак.

VI. Заключение и изводи

Като постигнахме поставената изследователска цел чрез точно изпълнение на поставените изследователски задачи можем да заключим следното:

1. Изведен и валидиран е олигогенен прогностичен скор при МПМ чрез първична селекция на гени въз основа на проведен скрининг за зависимост;
2. Изведеният скор определя подгрупи пациенти с МПМ, които имат много сходен експресионен профил при всички изследвани кохорти – тренираща и валидираща;
3. Профилът на ДНК метилиране не е свързан с възпроизводими диференциални профили на ДНК метилиране;
4. Прогностичният скор е свързан със специфични възпроизводими профили на инфилтрация от имунни клетки, която може да бъде обяснена с патогенезата на заболяването;
5. Изведеният прогностичен скор вероятно има и предиктивна стойност по отношение на чувствителността или резистентността към често използвани химиотерапевтични средства за лечение на МПМ.

Въз основа на тези изводи лесно могат да бъдат дефинирани примерни рационални насоки за бъдещи изследвания:

1. Конвертиране на изведения скор в лесен за клинично приложение скор, използващ имунохистохимия или RT-PCR върху архивни хистологични материали от пациенти с МПМ;
2. Проспективно валидиране на прогностичната стойност 2-ПС в клинични условия като се използва опростен имунохистохимичен или RT-PCR-базиран експресионен скор;
3. Проспективна оценка на предиктивната стойност 2-ПС в клинични условия за отговор към конвенционална химиотерапия и/или имунотерапия с ICIs като се използва опростен имунохистохимичен или RT-PCR-базиран експресионен скор;
4. Прилагане на описания от нас подход за разработване на прогностични и предиктивни експресионни скорове при други видове неопластични заболявания, вкл. редки видове рак.

VII. Приноси

Оригинални приноси

1. За първи път е демонстрирано възможното разработване на прогностичен скор въз основа на генна експресия при МПМ чрез първоначална селекция на гени, от чиято експресия вероятно зависи преживяемостта на клетъчната популация при МПМ.
2. За пръв път е демонстрирана и дискутирана прогностичната стойност на експресията на *GOLT1B* гена.
3. За пръв път е демонстриран олигогенен прогностичен модел при МПМ, включващ само 2 гена.
4. За пръв път е показано, че разработеният прогностичен модел има и вероятно предиктивна стойност за отговор към лечението с различни конвенционални химиотерапевтични средства.

Приноси с потвърдителен характер

1. Потвърдена е прогностичната стойност на експресията на *MAD2L1*.
2. Демонстрирано е, че прогностичните скорове при МПМ неминуемо корелират с определените основни характеристики на рака, особено свързани с ДНК поправката и митозата, което е очаквано от гледна точка на пролиферативния и агресивен характер на това заболяване.
3. Демонстрирано е, че възможно интегрирането на различни омикс платформи при МПМ, и че платформите за транскриптомен анализ имат добра възпроизводимост независимо от използвания метод и мястото на използването му.
4. Потвърдена е слабата възпроизводимост на профилите на ДНК метилиране при МПМ и тяхната слаба корелация с прогностично значими подгрупи на заболяването.

VIII. Публикации

Журнални статии

SHIVAROV, Velizar; BLAZHEV, Georgi; YORDANOV, Angel. A Novel Two-Gene Expression-Based Prognostic Score in Malignant Pleural Mesothelioma. *Diagnostics*, 2023, 13.9: 1556.

SHIVAROV, Velizar; BLAZHEV, Georgi. Bringing Together the Power of T Cell Receptor Mimic and Bispecific Antibodies for Cancer Immunotherapy: Still a Long Way to Go. *Monoclonal Antibodies in Immunodiagnosis and Immunotherapy*, 2021, 40.2: 81-85.

Участия в научни срещи

BLAZHEV, Georgi; SHIVAROV, Velizar. Development of a Novel Gene Expression-Based Prognostic Score in Malignant Pleural Mesothelioma. Климентови дни 2020, София, България.