

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М. В. ЛОМОНОСОВА

*На правах рукописи*

**Шафиков Радик Радикович**

**Структурно-функциональная характеристика  
лигандов маркера рака простаты GSPII и анализ  
регуляции экспрессии кодирующего его гена *FOLH1***

1.5.3 «молекулярная биология»

ДИССЕРТАЦИЯ

на соискание учёной степени  
кандидата химических наук

Научный руководитель:

к.х.н., доцент Скворцов Дмитрий Александрович

Москва – 2022

# Оглавление

Список сокращений.....	4
Введение .....	7
Актуальность проблемы.....	7
Цели и задачи исследования .....	7
Объект исследования.....	8
Предмет исследования.....	8
Научная новизна исследования .....	8
Научная и практическая значимость исследования .....	9
Методология диссертационного исследования .....	10
Основные положения, выносимые на защиту.....	11
Апробация работы .....	11
Публикации .....	12
Личный вклад автора.....	12
Структура и объём диссертации.....	13
Обзор литературы.....	14
GSPH – мишень для диагностики рака простаты .....	14
Ингибиторы GSPH.....	19
Элементы структуры GSPH .....	19
Взаимодействие частей ингибиторов GSPH с соответствующими им сайтами внутри фермента.....	22
Регуляторные последовательности и известные регуляторы экспрессии <i>FOLH1</i> .....	38
Регуляция транскрипции .....	38
Транскрипционная регуляция гена <i>FOLH1</i> .....	40
Последовательности гена <i>FOLH1</i> , содержащие цис-регуляторные элементы .....	41
Транскрипционные факторы – регуляторы транскрипции <i>FOLH1</i> .....	44
Экспериментальная часть .....	48
Материалы и методы .....	48
Реактивы.....	48
Клеточные линии и среды .....	48
МТТ тест .....	49
Измерение активности GSPH.....	50
Проточная цитометрия .....	51
Обработка поверхности культурального пластика полилизинном .....	51
Флуоресцентная микроскопия .....	51
Создание плазмид для постоянной экспрессии ТФ .....	52
Подготовка библиотека лентивирусов для нокаутов Gecko v2 и трансдукция .....	52

Отбор трансдуцированных Gecko v2 клеток LNCaP-cas9.....	54
Подготовка ампликонов, содержащих последовательности sgRNA, к секвенированию ....	54
РНК интерференция.....	55
Создание клеточных линий с конститутивной суперэкспрессией ТФ .....	55
Оценка совместного влияния суперэкспрессии или нокдауна ТФ и андрогена .....	56
Анализ транскриптомных данных для определения потенциальных регуляторов экспрессии гена <i>FOLH1</i> .....	56
Функции Python для анализа транскриптомных данных .....	57
Результаты и обсуждение .....	63
Характеристика ингибиторов GСРII и конъюгатов ингибиторов GСРII с флуоресцентными и цитотоксичными молекулами.....	63
Ингибиторы GСРII.....	63
Конъюгаты с флуоресцентными молекулами на основе В15 – ингибитора GСРII.....	70
Конъюгат с монометил ауристатином Е на основе В4.....	73
Обсуждение результатов .....	74
Потенциальные регуляторы экспрессии гена <i>FOLH1</i> .....	76
Потенциальные регуляторы экспрессии <i>FOLH1</i> по результатам корреляционного анализа транскрипционных данных .....	80
Вероятные регуляторы экспрессии <i>FOLH1</i> по результатам анализа данных транскриптомов клеток с нокаутами и нокдаунами транскрипционных факторов .....	83
Вероятные регуляторы экспрессии <i>FOLH1</i> по результатам скрининга библиотеки гидовых РНК Gecko v2.....	84
Проверка части потенциальных регуляторов экспрессии <i>FOLH1</i> .....	87
Обсуждение результатов .....	89
Заключение.....	93
Выводы .....	94
Публикации по теме диссертации в рецензируемых изданиях.....	95
Другие публикации по теме диссертации .....	95
Список литературы.....	96
Приложение.....	104

## Список сокращений

- [<sup>11</sup>C]DCMC - ингибитор GСРII
- [<sup>125</sup>I]DCIT - ингибитор GСРII
- [<sup>18</sup>F]DCFBC - ингибитор GСРII
- 2-PMРA – референсный ингибитор GСРII
- ABS – арен-связывающий сайт GСРII
- AKT – RAC-alpha serine/threonine-protein kinase
- AP-1 – activating protein-1
- AP-3 – activating protein-3
- AR – андрогеновый рецептор
- AR<sub>ABS1</sub> – арильный фрагмент 1 дипептидного фрагмента линкера ингибитора GСРII.
- AR<sub>ABS2</sub> – арильный фрагмент 2 дипептидного фрагмента линкера ингибитора GСРII.
- AR<sub>P1</sub> – арильный фрагмент ингибитора GСРII, связывающийся в S1 кармане фермента.
- ATPAF1 – ATP synthase mitochondrial F1 complex assembly factor 1
- BAD – BCL2 associated agonist of cell death
- BCL2L14 – BCL2 like 14
- BMРR1B – Bone morphogenetic protein receptor type 1B
- Cas9 – белок системы CRISPR-Cas9
- CCL-5 – C-C motif chemokine ligand 5
- CEBPB – CCAAT enhancer binding protein beta
- ChiP-seq – иммунопреципитация хроматина с секвенированием ДНК.
- CREB3L4 – cAMP responsive element binding protein 3 like 4, транскрипционный фактор
- CRISPR – clustered regularly interspaced short palindromic repeats
- CS – каталитический сайт GСРII
- DCFpyL – ингибитор GСРII
- DCIBzL – ингибитор GСРII
- DCL – ингибитор GСРII
- EGFR – Epidermal growth factor receptor, рецептор эпидермального фактора роста
- EPСА - ранний антиген рака простаты
- FAK – фокальная адгезионная киназа
- FAS – синтаза жирных кислот
- FOLH1* – Folate hydrolase 1, ген, кодирующий GСРII, объект исследования в данной работе.
- FOLH1B* – паралог *FOLH1*, у которого присутствует энхансер, аналогичный энхансеру *FOLH1*, но отсутствует аналогичный *FOLH1* промотор.
- FOXA1 – Forkhead box A, транскрипционный фактор.

GATA2 – GATA binding protein 2, транскрипционный фактор  
GCP II – глутаматкарбоксипептидаза II, объект исследования в данной работе.  
GPM6B – Glycoprotein M6B  
GRHL2 – Grainyhead like transcription factor 2  
HOXB13 – Homeobox B13  
IGF-1R – Insulin like growth factor 1 receptor, рецептор инсулиноподобного фактора роста 1  
IL-6 – интерлейкин 6  
KLK2 – калликреин 2  
KLK3 – калликреин 3, ген, кодирующий простатспецифический антиген  
LQ – лизин-глутамин  
LQE – лизин-глутамин-глутамат  
MAPK – mitogen-activated protein kinase  
MAZ – MYC associated zinc finger protein  
mGlu3 – Glutamate metabotropic receptor 3  
MOI – multiplicity of infection  
mTOR – Mechanistic target of rapamycin kinase  
MTT – 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide  
NAA – N-ацетил-аспартат  
NAAG – N-ацетил-аспартил-глутамат  
NFATc1 – Nuclear factor of activated T cells 1  
NF-κB – Nuclear factor kappa B  
NKG7 – Natural killer cell granule protein 7  
NKX3.1 – NK3 homeobox 1  
NMDA – N-метил-D-аспартат  
PAK – p21-активирующая киназа I  
PAP – кислая фосфатаза простаты  
PDB – protein data bank  
PI3K – фосфоинозитид-3-киназа  
PSA – простатспецифический антиген  
PSMA – простатспецифический мембранный антиген  
PSMA I&T – ингибитор GCP II  
RACK1 – Receptor for activated C kinase 1  
RBFOX2 – RNA binding fox-1 homolog 2  
SCUBE1 – Signal peptide, CUB domain and EGF like domain containing 1  
SERPINB5 – Serpin family B member 5

sgRNA – single guide RNA для системы CRISPR-Cas9

siRNA – small interfering RNA, малые интерферирующие РНК

SOX7 – SRY-box transcription factor 7

SPDEF – SAM pointed domain containing ETS transcription factor

SREAP2 – STEAP2 металлоредуктаза

TFAP4 – Transcription factor AP-4

TFIIB, TFIID, TFIIE, TFIIIF, TFIIH – основные транскрипционные факторы ДНК-зависимой РНК полимеразы II

TGF- $\beta$  - трансформирующий фактор роста бета

TMPRSS2-ERG – слитый белок из Transmembrane protease, serine 2 и ETS Transcription Factor

TOPORS – TOP1 binding arginine/serine rich protein, E3 убиквитин лигаза

ZBG – цинк-связывающая группа в составе ингибитора GСPII

ПЦР – полимеразная цепная реакция

ТФ – транскрипционный фактор

# Введение

## Актуальность проблемы

Рак простаты являлся вторым по распространенности типом рака среди мужчин в 2020 году. В диагностике и разработке терапии опухоли простаты используются различные маркеры [1]: кислая фосфатаза простаты PAP, простата-специфичный антиген PSA (KLK3), калликреин KLK2, трансформирующий фактор роста бета TGF- $\beta$ 1, интерлекин IL-6, синтаза жирных кислот FAS, ранний антиген рака простаты ERCA, глутаматкарбоксипептидаза II GCP II. Они различаются присутствием в простате и крови, методикой детекции, использованием для различных видов диагностики и применимостью для адресной доставки терапевтических молекул.

Среди рассматриваемых маркерных белков, фермент GCP II, кодируемый геном *FOLH1*, является мишенью для разработки новых методов интраоперационной онкодиагностики и адресной доставки терапевтических молекул в опухоль простаты, благодаря специфичности экспрессии и трансмембранной локализации в клетке [2]–[6]. Структура цинк-содержащего каталитического центра GCP II известна, и на основании этого в литературе описаны разные типы лигандов, содержащих цинк-связывающие группы. Особо интересны лиганды, содержащие малополярные и несклонные к окислению уреидо-группы, соединенные с остатками глутамата и лизина. Для них описан ряд модификаций различными группировками, способными взаимодействовать с близким к каталитическому центру GCP II гидрофобным карманом или дальним арил-связывающим участком фермента. Однако, систематического исследования лигандов с сочетаниями таких модификаций ранее не проводилось.

Высокая экспрессия гена *FOLH1* тканеспецифична: наибольший уровень его экспрессии наблюдается в опухоли простаты, в простате, в проксимальных канальцах почек и некоторых отделах двенадцатиперстной кишки. Установление белков-регуляторов экспрессии *FOLH1* позволит лучше понять каскады, участвующие в обеспечении опухоль- и тканеспецифичности, и рассматривать ключевые белки этих каскадов в качестве мишеней для диагностики и терапии заболеваний.

Структурно-функциональная характеристика лигандов белка GCP II и анализ возможных путей регуляции экспрессии кодирующего его гена *FOLH1* представляются важной задачей.

## Цели и задачи исследования

Целями данной работы являются оптимизация структуры лигандов белка GCP II и их конъюгатов с флуоресцентными и цитотоксичными молекулами, нацеленных на клетки рака

простаты, и обнаружение регуляторов, обеспечивающих высокую экспрессию гена *FOLH1* в простате и опухоли простаты. Для достижения этих целей были поставлены следующие задачи:

1) Исследовать структурно-функциональную зависимость новых низкомолекулярных лигандов GСРІІ для улучшения их взаимодействия с мишенью.

2) Оценить *in vitro* способность низкомолекулярных лигандов GСРІІ, конъюгированных с флуоресцентными молекулами, окрашивать клетки, экспрессирующие GСРІІ.

3) Исследовать цитотоксичность лигандов GСРІІ, конъюгированных с флуоресцентными молекулами или монометил ауристатином Е.

4) Идентифицировать потенциальные регуляторы экспрессии *FOLH1*.

## **Объект исследования**

Объектами исследования в настоящей работе являются белок GСРІІ человека и кодирующий его ген *FOLH1*.

## **Предмет исследования**

Предметами исследования в представленной работе являются свойства лигандов маркера опухолевых клеток простаты GСРІІ и регуляция экспрессии кодирующего его гена *FOLH1*.

## **Научная новизна исследования**

Исследованные в данной работе лиганды GСРІІ являются ингибиторами этого фермента, поэтому для оценки эффективности их взаимодействия с GСРІІ мы определяли значения константы ингибирования ( $K_i$ ). Были определены  $K_i$  фермента GСРІІ для более, чем сорока лигандов с различными комбинациями заместителей, синтезированных нашими коллегами с кафедры органической химии химического факультета МГУ им. М.В. Ломоносова. Для исследований, приведенных в настоящей работе, были использованы данные о структуре фармакофорного кармана фермента GСРІІ. Все соединения были разбиты на три "поколения" в зависимости от исследуемых модификаций (см. Рис. 1, таблицы 1-3). Для молекул поколения "А" было показано, что, как правило, наиболее выгодной конфигурацией дипептидного фрагмента линкера является (L,L), поскольку такие лиганды имеют наименьшее значение  $K_i$ . Для лигандов поколения "В" показано, что низкие значения  $K_i$  имеют лиганды, содержащие атом С1 в *мета*-положении арильного фрагмента Р1 и атомы С1 или Вг в *пара*-положении арильного фрагмента Р1. Лиганды, в которых эти заместители сочетаются с (L,L) конфигурацией дипептидного фрагмента линкера, более эффективны. В поколении "С"

наиболее эффективными оказались лиганды, содержащие карбоксильную группу в *para*-положении арильного фрагмента P1.

В настоящей работе были исследованы конъюгаты выбранных в первой части работы лигандов фермента GSPH с цитотоксическими и флуоресцентными молекулами. Для этих соединений были продемонстрированы специфичные взаимодействия с опухолевыми клетками простаты, экспрессирующими GSPH. В частности, для конъюгатов лигандов GSPH с монометил ауристатином E были определены значения  $CC_{50}$ , соответствующие концентрации лиганда, при которой гибнет половина клеток, а для флуоресцентных производных продемонстрирована высокая эффективность окрашивания GSPH-экспрессирующих клеток.

Несмотря на привлекательность белка GSPH как маркера опухоли простаты, регуляция его экспрессии изучена недостаточно. В литературе не представлено систематического исследования возможных транс-регуляторов экспрессии *FOLH1*. В данной работе проанализирован возможный регулон *FOLH1* с использованием нескольких экспериментальных и *in silico* подходов. Валидация транскрипционного фактора MAZ как активатора транскрипции *FOLH1* в клетках опухоли простаты 22Rv1 позволяет утверждать, что предложенная в данной работе стратегия применима для поиска регуляторов экспрессии *FOLH1*.

## **Научная и практическая значимость исследования**

На настоящий момент в числе лигандов белка GSPH, наиболее перспективных по совокупности ингибирующих и фармакофорных свойств, рассматриваются соединения, содержащие уреидо-группы, соединенные с остатками глутамата и лизина. В литературе описаны модификации этих лигандов с различными заместителями, способными взаимодействовать с близким к каталитическому центру GSPH гидрофобным карманом или дальним арил-связывающим участком фермента. Нами выполнено систематическое исследование новых лигандов, содержащих комбинации таких модификаций.

С использованием полученных в ходе настоящей работы данных о биологической активности соединений подана заявка на патент. Специфичность к GSPH является ключевой особенностью лигандов, синтезированных нашими коллегами с кафедры органической химии химического факультета МГУ им. М.В. Ломоносова, для направленной доставки молекулярных фрагментов для диагностики и терапии.

Создание эффективных лигандов GSPH необходимо не только в клиническом аспекте: конъюгаты этих лигандов с флуоресцентными молекулами могут помочь в поиске неизвестных ранее регуляторов транскрипции *FOLH1* в опухоли простаты. Окрашивание флуоресцентным

лигандом **B15-Cy5** экспрессирующих GСРII клеток позволило отбирать в ходе сортировки те клетки, где экспрессия гена *FOLH1* изменилась из-за нокаута потенциального регулятора. Полученные в настоящей работе с помощью различных *in silico* и экспериментальных подходов данные о вероятных регуляторах экспрессии *FOLH1* могут быть использованы для поиска регуляторов опухолевого процесса и для понимания тканеспецифичной регуляции в нормальных тканях простаты.

## Методология диссертационного исследования

Для решения поставленных задач в представленной работе применялись современные методы и подходы, а также использовались собственные разработки автора.

Исследование проводилось на клеточных линиях человека различного происхождения (LNCaP, 22Rv1, PC-3, A549, VA-13, MCF-7, HEK293T).

Исследование ингибирующих свойств лигандов GСРII и их конъюгатов с флуоресцентными молекулами проведено с помощью теста ингибирования GСРII *in vitro*. Значение константы ингибирования  $K_i$  использовалось в качестве параметра для оценки эффективности взаимодействия лигандов с GСРII. Оно определялось по уравнению Ченга-Пруссофа:  $K_i = IC_{50}/(1+S/K_m)$ , где  $IC_{50}$  – экспериментально полученная концентрация полумаксимального ингибирования GСРII в тесте ферментативной активности,  $S$  – концентрация NAAG, субстрата GСРII; значение  $K_m$  взято из литературы.

Цитотоксические свойства лигандов GСРII оценены в клеточных линиях человека различной этиологии посредством МТТ-теста.

Для оценки селективности флуоресцентных лигандов проведено окрашивание ими клеток, экспрессирующих и не экспрессирующих GСРII, с последующей проточной цитофлуориметрией или флуоресцентной микроскопией. Предложен метод отслеживания уровня экспрессии *FOLH1* по флуоресцентному сигналу от низкомолекулярного флуоресцентного лиганда, связанного с GСРII.

Потенциальные регуляторы экспрессии *FOLH1* выбраны посредством скрининга с использованием коммерчески доступной полногеномной библиотеки нокаутов Gecko v2, анализа баз данных нокадаунов и корреляционного анализа экспрессии генов в опухолевых клетках простаты. Для скрининга создана клеточная линия LNCaP-Cas9. Проводилась сборка лентивирусов в клетках HEK293T и трансдукция линии LNCaP-Cas9. После селекции на антибиотике клетки были инкубированы с флуоресцентным лигандом GСРII **B15-Cy5**, произведен отбор клеток с изменившейся экспрессией *FOLH1* на проточном цитометре. Для индивидуальной проверки генов регуляторов были осуществлены их нокадаун и

суперэкспрессия. Суперэкспрессия генов предполагаемых регуляторов проводилась с помощью системы транспозонов Sleeping Beauty в клеточных линиях рака простаты LNCaP, 22Rv1, PC-3. Влияние нокадауна или суперэкспрессии потенциальных регуляторов на экспрессию гена *FOLH1* оценивалось с помощью ПЦР в реальном времени.

### **Основные положения, выносимые на защиту**

1. Константы ингибирования фермента GСPII исследуемыми низкомолекулярными лигандами находятся в наномолярном диапазоне.
2. По сравнению с монометил ауристатином E, цитотоксичность его конъюгата с лигандом GСPII **V15** оказалась ниже для клеток, не экспрессирующих GСPII, при общем снижении цитотоксичности *in vitro*.
3. Флуоресцентный лиганд **V15-Cy5** имеет высокую специфичность к GСPII-экспрессирующим клеткам опухоли простаты и низкую цитотоксичность.
4. Флуоресцентный сигнал от лиганда **V15-Cy5**, связанного с GСPII, отражает уровень экспрессии *FOLH1* и может быть использован для поиска потенциальных регуляторов экспрессии *FOLH1* при скрининге библиотеки нокаутов.
5. Транскрипционный фактор MAZ является активатором транскрипции гена *FOLH1*.

### **Апробация работы**

Основные результаты диссертации докладывались на 5-ти конференциях и научных школах различного уровня (в том числе всероссийских и международных), а именно:

1. VII съезд биохимиков, молекулярных биологов и физиологов России (Сочи, 2022, сертификат победителя за доклад)
2. Всероссийская научная конференции "Марковниковские чтения: органическая химия от Марковникова до наших дней" Школа-конференция молодых ученых «Органическая химия: Традиции и Современность» (Сочи, 2022)
3. Кост-2021. Всероссийский конгресс по химии гетероциклических соединений (Сочи, 2021)
4. 26th EFMC International Symposium on Medicinal Chemistry (EFMC-ISMC-2021), (Виртуальная конференция, Швейцария, 2021)
5. Всероссийской научной конференции "Марковниковские чтения: органическая химия от Марковникова до наших дней", (Москва, 2020)

## Публикации

По теме диссертации опубликовано 8 печатных работ, в том числе 4 статьи в рецензируемых научных журналах, индексируемых в базах данных Scopus и Web of Science. Была подана совместная с группой химического синтеза (Мачулкин А.Э., Успенская А.А. и др.) заявка на патент, основанная в значительной степени на полученных в ходе настоящей работы данных.

## Личный вклад автора

Все экспериментальные данные, приведенные в рамках диссертационного исследования, за исключением *in vivo* исследований, получены лично автором или при его непосредственном участии.

Автором проведено исследование ингибирующих свойств лигандов GСРII и оценена их цитотоксичность на клеточных линиях различной этиологии. Совместно с Успенской А.А. выполнен анализ связи структура-активность лигандов GСРII. Автором выполнено окрашивание клеток флуоресцентными лигандами GСРII, проточная цитофлуориметрия и флуоресцентная микроскопия. Им предложен метод оценки уровня экспрессии *FOLH1* по флуоресцентному сигналу от связанного с GСРII флуоресцентного лиганда.

В части, посвященной поиску регуляторов экспрессии *FOLH1*, все экспериментальные процедуры выполнены автором лично либо при его непосредственном участии: ведение культур, проведение клонирований для суперэкспрессии транскрипционных факторов (совместно со Скворцовым Д.П.), создание клеточных линий с суперэкспрессией предполагаемых регуляторов экспрессии *FOLH1*, проверка их посредством ПЦР в реальном времени. Автор провел анализ баз данных нокдаунов, корреляционный анализ экспрессии в образцах опухоли простаты, нормальных тканях простаты и клеточных линиях. Подготовка вирусов с библиотекой последовательностей, кодирующих гидовые РНК для нокаута, заражение клеток, отбор и их окрашивание **B15-Cy5** выполнены автором самостоятельно. Сортировка клеток на проточном цитометре выполнена совместно с М.П.Рубцовой. Подготовка образцов для секвенирования и анализ результатов скрининга выполнены автором самостоятельно.

Личный вклад автора в проведённое исследование заключается также в сборе и анализе литературных данных, постановке задач, планировании экспериментов, анализе и оформлении полученных результатов, подготовке научных статей и в представлении результатов на научных конференциях.

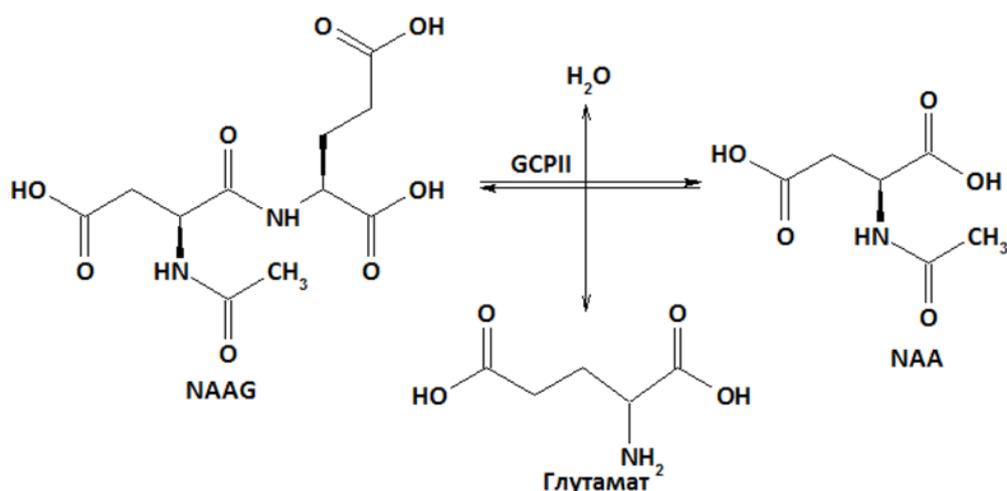
## **Структура и объём диссертации**

Диссертация изложена на 110 страницах и состоит из следующих разделов: «Введение», «Обзор литературы», «Экспериментальная часть», «Результаты», «Обсуждение результатов», «Выводы» и «Список литературы». Работа содержит 13 таблиц и 21 рисунок. Библиография включает 131 источник литературы.

# Обзор литературы

## ГСПИ – мишень для диагностики рака простаты

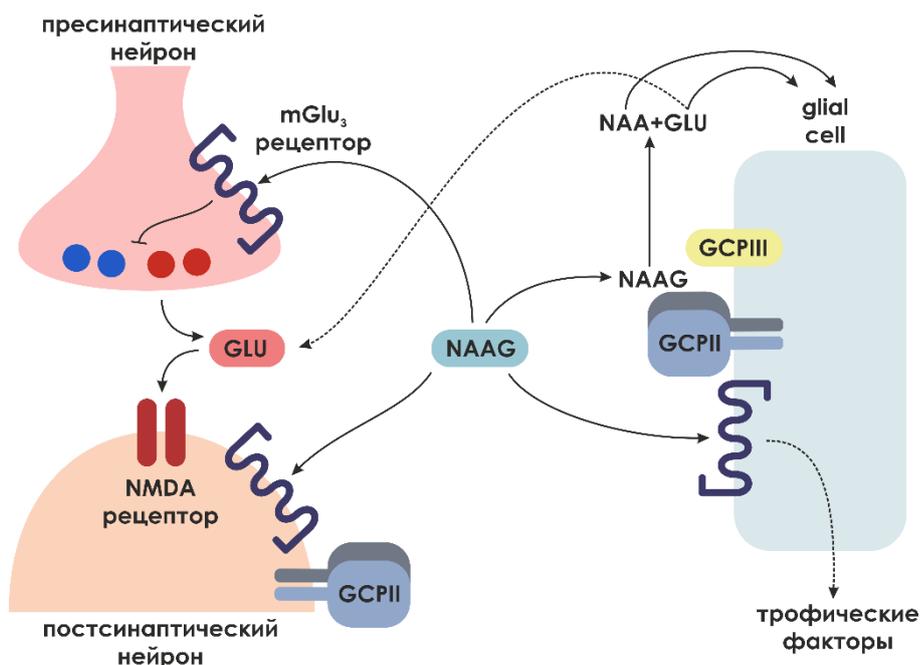
Глутамат-карбоксипептидаза ГСПИ является трансмембранным цинк-зависимым металлоферментом с активным сайтом, расположенным во внеклеточном домене. ГСПИ кодируется геном *FOLH1*, который экспрессируется тканеспецифически в опухоли простаты, простате, клетках проксимальных канальцев почек и щеточной каемки двенадцатиперстной кишки. Он выполняет различные функции в зависимости от тканей, в которых экспрессируется, и известен также под другими названиями: N-ацетил-L-аспартил-L-глутамат пептидаза I (NAALADase I) и простат-специфичный мембранный антиген (PSMA). ГСПИ катализирует гидролиз пептидного нейротрансмиттера N-ацетил-L-аспартил-L-глутамата (NAAG) до глутамата и N-ацетил-L-аспартата (NAA)[7] и расщепление  $\gamma$ -связанного глутамата в поли(глутамат) производных фолата (Рис.1) [7], [8] и некоторых коротких пептидах [9].



**Рис.1.** Схема реакции (KEGG: R10687), катализируемой ГСПИ.

Расщепляя N-ацетил-L-аспартил-L-глутамат, ГСПИ вовлечен в поддержание баланса между NAAG и глутаматом в нервной системе (Рис.2). NAAG и глутамат входят в число наиболее распространенных нейротрансмиттеров в центральной нервной системе [10]. Например, NAAG действует как агонист глутаматных метаботропных рецепторов III группы (mGlu<sub>3</sub>) в нейронах[11]. Активация рецепторов mGlu<sub>3</sub> посредством NAAG уменьшает концентрацию cAMP и cGMP в пресинаптических терминалях, ингибируя высвобождение различных нейротрансмиттеров, таких как гамма-аминомасляная кислота и глутамат [12], [13]. Стимуляция рецепторов mGlu<sub>3</sub> в астроцитах активирует синтез трансформирующего фактора роста  $\beta$  (TGF- $\beta$ ) [14], [15]. ГСПИ влияет на концентрацию глутамата как через гидролиз NAAG, так и через регуляцию высвобождения глутамата в синаптическую щель. Благодаря опосредованной NAAG защите нейронов от избыточного глутамата и низкому уровню экспрессии (и активности) ГСПИ в здоровых тканях, NAAG служит ингибитором

глутаматэргической передачи. Увеличение концентрации глутамата и/или излишняя активация таких белков, как N-метил-D-аспартат (NMDA) рецепторы или GCP II, приводит к повреждению и гибели нейронов при инсульте, ишемии, латеральном амиотрофическом склерозе [16]. При патологиях нервной системы повышенная активность GCP II резко усиливает гидролиз NAAG, обеспечивая чрезмерную концентрацию глутамата, который активирует глутаматергические рецепторы в нейронах и глиальных клетках [17], что может привести к повреждению нейронов [18].



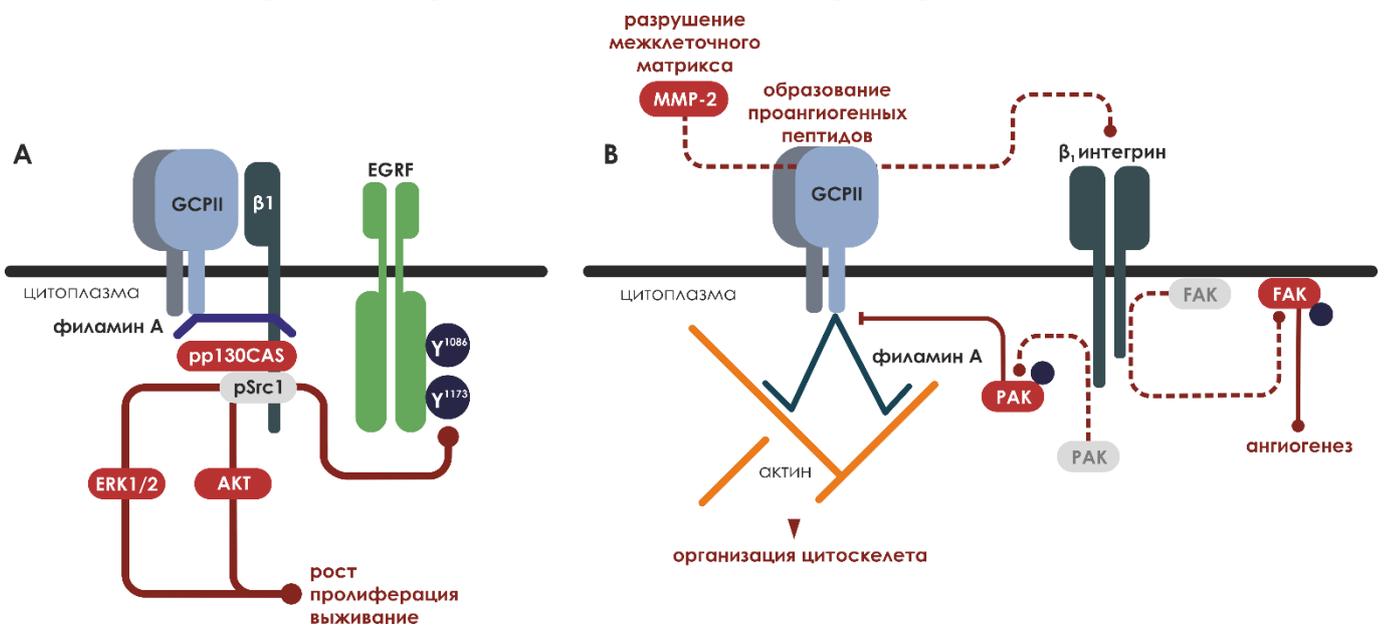
**Рис.2. Функции GCP II.** Участие GCP II в метаболизме NAAG в нейронах и глиальных клетках. GLU - глутамат, NMDA - N-метил-D-аспартат, mGlu3 - метаботропные глутаматные рецепторы III группы. Остатки глутамата и  $\gamma$ -аминомасляной кислоты представлены красными и синими кружками соответственно; когда рецепторы mGlu3 неактивны NAAG и глутамат высвобождаются в синаптическую щель.

Помимо экспрессии в клетках, где GCP II необходим для обеспечения физиологических процессов в норме, его уровень экспрессии увеличивается в клетках опухоли простаты, где он может быть использован как мишень для разработки лигандов для диагностики и терапии. Высокое содержание белка GCP II могут иметь как андроген-зависимые (LNCaP, VCaP), так и андроген-независимые клеточные линии опухоли простаты (C4-2)[6]. У пациентов с низким содержанием GCP II в андроген-зависимой опухоли простаты его экспрессия увеличивается в ходе андроген-депривационной терапии [19]. Его экспрессия была также обнаружена в циркулирующих опухолевых клетках, полученных из периферической крови больных раком простаты [5]. Анализ трех тысяч образцов опухолей показал наличие GCP II во многих других типах опухолей, но в заметно меньших количествах [6]. В частности, повышенная активность

фермента отмечается в новообразованных кровеносных сосудах внутри различных типов солидных опухолей, причем сами раковые клетки этих опухолей такой активностью не обладают [6], а в нормальных сосудах GCP II отсутствует.

Фолат-гидролизующая активность GCP II дает преимущество клеткам рака простаты в опухолевом микроокружении путем преобразования полиглутаматфолатов в глутаматфолат, который может быть поглощен клеткой. Показано, что клетки PC-3-GCP II пролиферируют лучше, чем PC-3 без GCP II в среде с небольшим количеством фолата [20]: это наблюдение подразумевает, что GCP II способствует передаче фолата, вероятно, путем прямого транспорта фолата и его глутамат-содержащих производных. Клеточные линии PC-3-GCP II и PC-3, выращенные в среде с фолат-(глутамат)<sub>5</sub>, показали, что только GCP II-содержащие клетки могут эффективно гидролизовать фолат-(глутамат)<sub>5</sub> до фолат-глутамата и поглощать его [20].

Недавние исследования пролили свет на возможную роль GCP II в росте и прогрессии рака простаты. Было показано, что GCP II собирается в макромолекулярный комплекс с  $\beta 1$  интегринами, филламином А, фосфо-p130CAS, фосфор-с-Src и рецептором эпидермального фактора роста (EGFR)[21] (Рис.3А). Внутри комплекса активация  $\beta 1$  интегринов сопровождается фосфорилированием с-Src и транс-фосфорилированием EGFR (в Y1086 и Y1173): это приводит к активации АКТ/mTOR/BAD и MAPK путей, которые защищают клетки от апоптоза в GCP II-позитивных клеточных линиях, таких как LNCaP, PC-3-GCP II [21]. Экспрессия GCP II, p130CAS и цитоплазматического филламина А положительно коррелируют с инвазивным и агрессивным фенотипом клеточных линий рака простаты [21].



**Рис.3. Функции GCP II.** А) Ассоциация GCP II с выживанием и пролиферацией клеток. GCP II собирает макромолекулярный комплекс с  $\beta 1$  интегринами, филламином А, фосфо-p130CAS, фосфо-с-Src и рецептором эпидермального фактора роста (EGFR). За активацией  $\beta 1$ -интегринов следует фосфорилирование с-Src и транс-фосфорилирование EGFR, что в

конечном итоге приводит к активации путей АКТ/mTOR/BAD и MAPK в GСPII-позитивных клеточных линиях. Фосфорилированные аминокислотные остатки показаны темно-синими кружками. В) Участие GСPII в организации цитоскелета и ангиогенезе в эндотелии внутри солидной опухоли приводит к направленной миграции эндотелиальных клеток. Фосфорилирование показано темно-синими кружками.

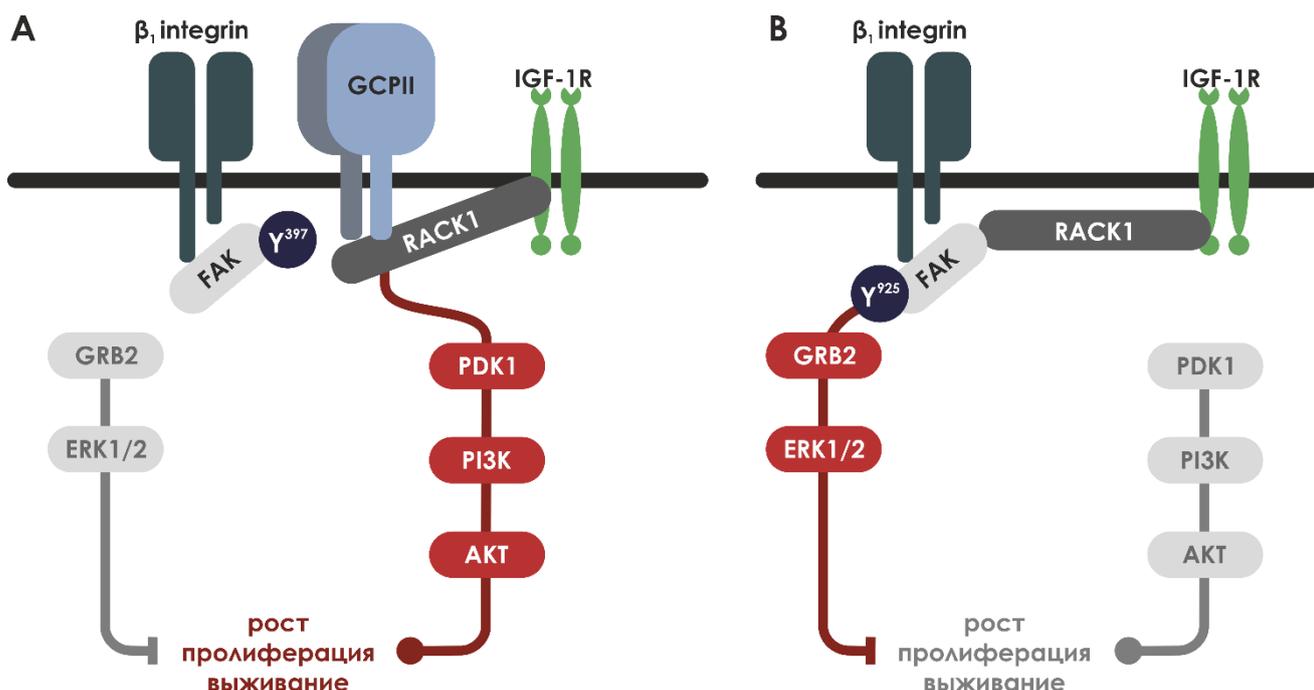
Инкубация клеток с антителами к GСPII усиливает передачу пролиферативных сигналов в LNCaP и PC-3-GСPII через кластеризацию GСPII, p130CAS и филамина А. В частности, кластеризация этих белков приводит к усилению активации NF-κB и его транслокации в ядро, где он регулирует экспрессию цитокинов и хемокинов, в том числе запуская экспрессию IL-6 и CCL-5. [22].

GСPII присутствует с новообразованных сосудах некоторых других типов опухолей, где его роль связана с регуляцией скорости роста сосудов (Рис.3В). Пептидазная активность GСPII приводит к образованию проангиогенного дипептида LQ из LQE посредством гидролиза белков внеклеточного матрикса, в частности ламинина, матриксными металлопротеиназами (ММР), такими как ММР-2, экспрессируемыми опухолевыми клетками. Было показано, что LQ облегчает адгезию эндотелия *in vitro* и ангиогенез *in vitro* дозо-зависимым образом [9].

Ингибирование или нокдаун GСPII уменьшает инвазию эндотелиальных клеток *in vitro* через его взаимодействие с β1-интегрином и снижение фосфорилирования p21-активирующей киназы I (РАК)[23] (Рис.3В). Повышение доли активной GСPII приводит к активации интегринов, а интегрины β1 и РАК, в свою очередь, негативно регулируют активность GСPII. Эти эффекты, опосредованные взаимодействием С-концевого цитоплазматического домена GСPII с филамином А, проявляются как *in vitro*, так и *in vivo* в эндотелиальных клетках и способствуют их миграции [23], [24]. Существует петля отрицательной обратной связи, в которой ингибирование активности РАК или β1-интегрин увеличивает взаимодействие GСPII с филамином А [23]. Интегрины представляют собой гетеродимеры, состоящие из α и β субъединиц, причем β1 субъединица может взаимодействовать с различными α субъединицами. Только небольшие пептидные субстраты, такие как LQ, обладают способностью связывать и активировать закрытую, неактивную форму интегринов, что делает возможным взаимодействие интегринов с более крупными продуктами гидролиза ламинина [9]. α2 и α3 субъединицы тесно связаны с LQ-опосредованным эффектом, а α6, экспрессирующийся в эндотелии на том же уровне [9], - нет. Передача сигнала через интегрины часто приводит к фосфорилированию и активации фокальной адгезионной киназы (ФАК), участвующей в клеточной адгезии [25]. Ингибирование GСPII снижает уровень фосфорилированной ФАК [26].

Активный уровень FAK повышается в клетках, выращенных в присутствии LQ: это доказывает участие LQ в активации интегрин-зависимой передачи сигнала [9].

GCP111 переключает сигналы с каскада MAPK на PI3K-AKT каскад [27] (Рис.4А): через взаимодействие с RACK1 GCP111 разрушает ассоциацию между интегрином и IGF-1R и активирует каскад PI3K-AKT, усиливая пролиферативную и антиапоптотическую сигнализацию [27]. В клетках опухоли простаты, в которых GCP111 отсутствует, комплекс из IGF-1R, интегрин и RACK1 запускает MAPK каскад (Рис.4В).



**Рис.4. Функции GCP111. Переключение между перекрестно взаимодействующими путями MAPK и PI3K-AKT в (А) GCP111-содержащих клетках рака простаты и (В) клетках рака простаты без GCP111.**

Несмотря на разнообразие выполняемых функций, нокаут GCP111 в мышах не имеет негативных последствий для онтогенеза [28], [29]. Благодаря своей высокой экспрессии и тканеспецифичности GCP111 является мишенью для создания ингибиторов, которые, являясь лигандами GCP111, могут использоваться для адресной доставки препаратов. Ингибирование GCP111, аналогично нокауту, вероятно не влияет на физиологические процессы в организме в норме, поэтому фермент рассматривается как маркер интраоперативной диагностики, проводимой с использованием конъюгатов ингибиторов с диагностическими молекулами.

## Ингибиторы GСРП

GСРП является мишенью для создания лигандов, способных к адресной доставке низкомолекулярных соединений в рак простаты. На настоящий момент в числе лигандов GСРП, наиболее перспективных по совокупности ингибирующих и фармакофорных свойств, рассматриваются производные, представляющие собой объединение уреидо-группы с остатками глутамата и лизина. Они содержат модификации по остатку лизина, направленными на взаимодействия с гидрофобным карманом GСРП или арил-связывающим участком фермента. Возможность конъюгации таких лигандов-ингибиторов GСРП с низкомолекулярными соединениями с минимальными потерями аффинности делает их привлекательными для создания на их основе диагностических и терапевтических конъюгатов. Некоторые такие соединения в данный момент проходят клинические исследования [<https://clinicaltrials.gov>][30], [31]. 123I-(MIP)-1072, 123I-(MIP)-1095 (H11), 131I-(MIP)-1095 (H11) уже внедрены в клинику [31]. 99mTc-MIP-1404 и некоторые 18F-меченые DCFBC (F5), DCFPyL (H20) и PSMA-1007 в настоящее время проходят клинические испытания для использования в качестве агентов визуализации при карциноме простаты [<https://clinicaltrials.gov>] [30], [31].

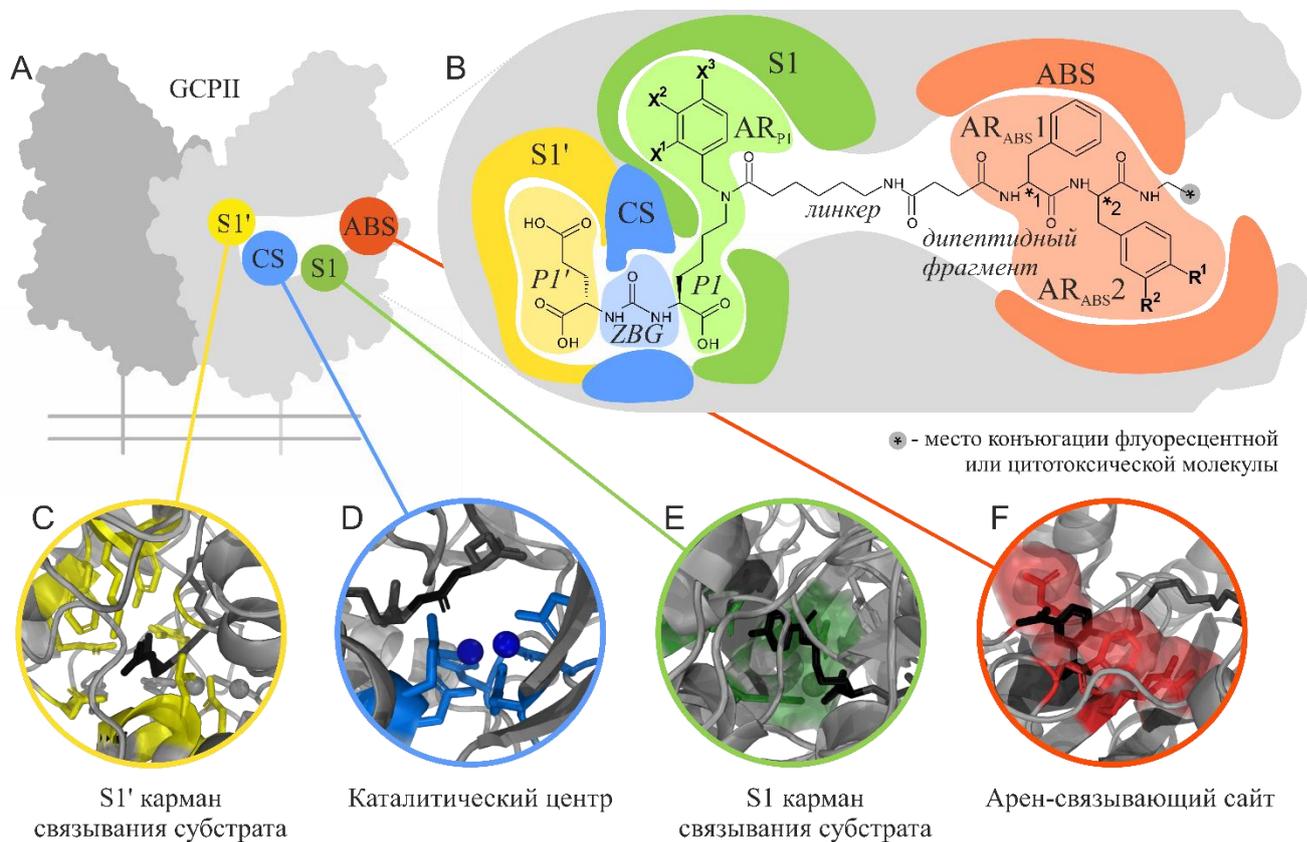
Данные рентгеноструктурного анализа GСРП в комплексе с разнообразными ингибиторами позволяют строить предположения о взаимодействии между фрагментами ингибиторов и участками структуры GСРП. Этот раздел обзора посвящен взаимосвязи структура-функция отдельных структурных элементов ингибиторов и их взаимодействию с соответствующими карманами GСРП.

## Элементы структуры GСРП

Структура GСРП получена для внеклеточной части фермента (остатки 44-750) [32], [33] и достаточно хорошо изучена. Цитоплазматический домен (остатки 1-18), вероятно, не имеет стабильной укладки, трансмембранная часть (остатки 19-43) была предсказана как  $\alpha$ -спираль [33]. Внеклеточная часть состоит из протеазоподобного домена I (содержащего 57-116 и 352-590 аминокислотные остатки), апикального домена II (117-351 остатки) и С-концевого домена III (591-750 остатки). Фермент активен только в виде гомодимера [34], где домен III одного мономера взаимодействует с доменами I и II другого. Сайт связывания субстрата образован остатками всех трех доменов. Участок связывания субстрата и воронка от поверхности GСРП к участку связывания вместе образуют карман фармакофора.

Карман фармакофора GСРП схематически представлен на Рис.5. Он содержит каталитический центр CS (Рис. 5С), участки связывания субстрата S1 (Рис. 5D) и S1' (Рис. 1Е),

а также арен-связывающий участок ABS (Рис. 5F). Большинство изученных низкомолекулярных лигандов являются миметиками природного субстрата – N-ацетил-аспартил-глутамата (NAAG). Известны ключевые структурные элементы ингибиторов, важные для взаимодействия с GСРII и его ингибирования (Рис. 5B): цинк-связывающая группа, мотивы P1' и P1 для некоторых ингибиторов. P1 фрагменты чаще всего представляют собой остаток лизина, боковая цепь которого рассматривается как фрагмент-"спейсер", который может быть дополнительно модифицирован арильным фрагментом и линкером, необходимым для конъюгирования ингибитора GСРII с терапевтическими или диагностическими молекулами. Цинк-связывающая группа — это негидролизуемый структурный аналог пептидной связи NAAG, взаимодействующий с каталитическим сайтом и содержащий атомы цинка. P1' часть большинства ингибиторов представляет собой глутамат, взаимодействующий с S1' карманом фермента, сайтом распознавания глутамата. Часть P1 ингибиторов может быть вариабельной: в основном она состоит из производных аминокислот, которые улучшают связывание с карманом S1 за счет своей  $\alpha$ -карбоксильной группы. Эти ингибиторы могут содержать спейсеры с дополнительным фрагментом ( $AR_{P1}$ ), распознающим гидрофобную поверхность сайта S1 [35]. При этом ключевым вопросом становится влияние заместителей, обеспечивающих дополнительное связывание в S1 и ABS сайтах и играющих роль линкера для используемой "нагрузки" диагностической или терапевтической молекулы.



**Рис. 5. Структурные элементы GCP II.** **А.** Четыре основных участка связывания ингибитора внутри GCP II. **В.** Связывание потенциального ингибитора внутри GCP II. Показаны четыре основные части ингибитора. PI' мотив обычно представляет собой глутамат и заполняет карман распознавания субстрата S1'. Негидролизуемая цинк-связывающая группа (ZBG) взаимодействует с каталитическим сайтом CS. Фрагмент PI ингибитора взаимодействует с сайтом S1, ароматическая часть AR<sub>PI</sub> может взаимодействовать с гидрофобным аргининовым трактом в S1. Ароматические фрагменты AR<sub>ABS1</sub> и AR<sub>ABS2</sub> необходимы для взаимодействия с ABS. \*1 и \*2 – хиральные атомы дипептидного фрагмента линкера; **С.** S1' - участок распознавания субстрата, связывающий глутамат или глутамат-подобный фрагмент субстрата (PDB ID: 3D7H); **Д.** CS - каталитический сайт (PDB ID: 3D7H); **Е.** S1 - участок распознавания субстрата, содержащий аргининовый тракт (PDB ID: 3D7H); **Ф.** ABS - арен-связывающий сайт (PDB ID: 2XEI). Приведены идентификаторы Protein Data Bank. Аминокислоты, значимые для связывания молекул ингибиторов, окрашены, а структуры ингибиторов показаны темно-серым цветом.

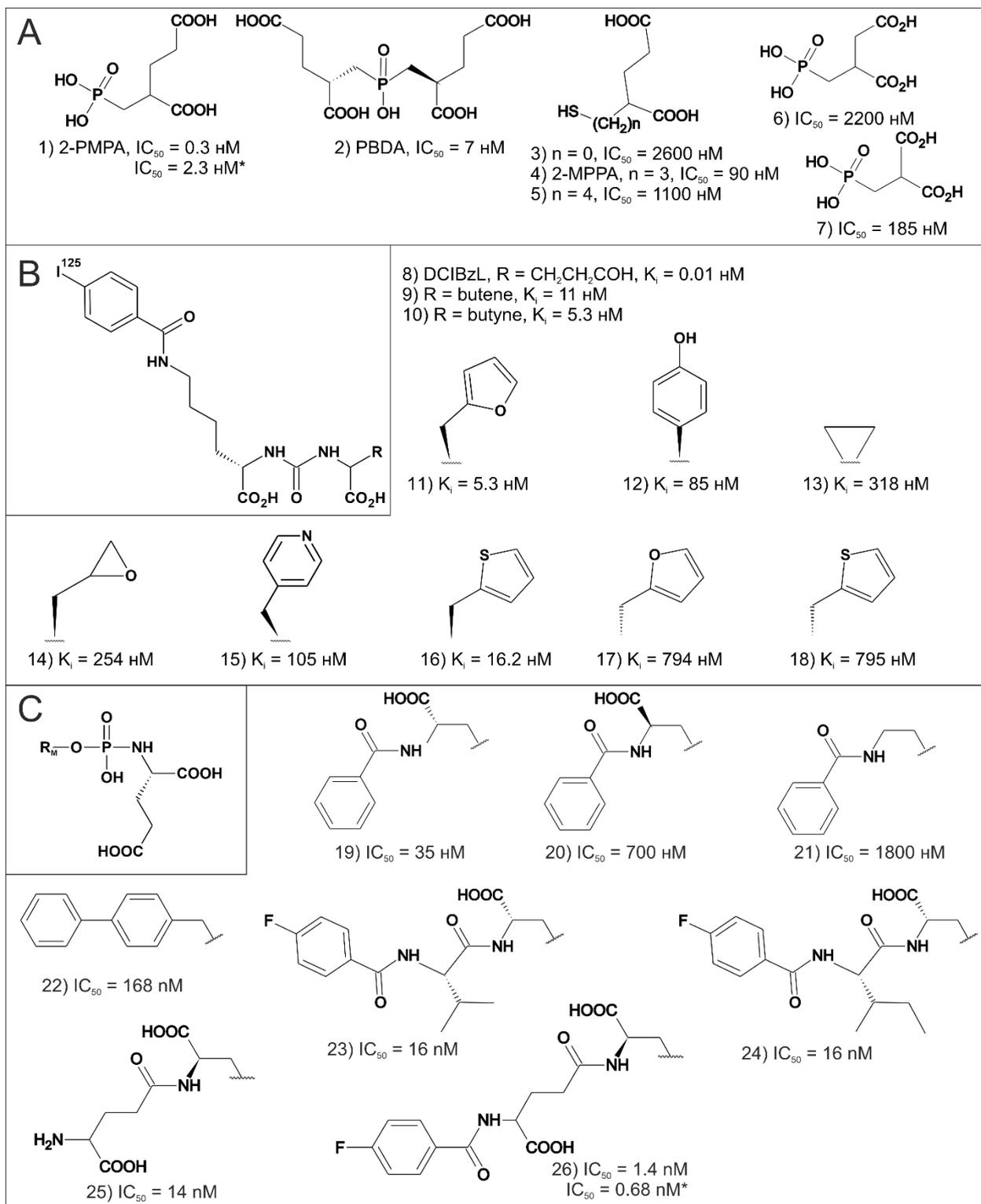
## Взаимодействие частей ингибиторов GСPII с соответствующими им сайтами внутри фермента

### Цинк-связывающие группы ингибиторов и каталитический центр GСPII

Ингибиторы GСPII классифицируются по типу цинк-связывающих групп. Большинство из них являются производными фосфонатов (фосфонатные эфиры со структурой  $RP(O)(OR_i)_n(OH)_{2-n}$ ) или фосфинатов ( $R_2P(O)OR'$ ), тиолов ( $R_2-C(S)-R_1$ ) и мочевины ( $R_2-C(O)-R_1$ ) (Рис.6). Другие производные включают сульфаниламид [36], пептиды [37], однако большинство из них неэффективны в наномолярных концентрациях и не имеют достаточно данных для анализа связи структура-активность. По сравнению с ингибиторами GСPII на основе тиолов и фосфатов, соединения на основе мочевины не обладают сильной полярностью и чувствительностью к окислению. Так как нами в экспериментальной части диссертации рассматриваются производные, представляющие собой объединение уреидо-группы с остатками глутамата и лизина, то далее будут описаны ингибиторы на основе уреидо-группы.

Класс ингибиторов с уреидо-группой основан на PBDA **2** с фосфатом, замененным на мочевины, что позволило получить (S)-Glu-C(O)-(S)-Glu с  $IC_{50}$  47 нМ (Рис.6, здесь и далее будет приведена сквозная нумерация соединений) [38]. Мочевина имитирует пептидную связь NAAG, но делает ингибитор устойчивым к гидролизу.

Кислород уреидо-группы взаимодействует с боковыми цепями Tyr552, His553 и мостиковой молекулой воды между ионами цинка [39]. Для соединений **3** и **5** (Рис.6) ингибирование GСPII *in vitro* не наблюдалось, но ингибиторы GСPII на основе фосфатов, такие как 2-РМРА (Рис.6, **1**), лишенные  $\alpha$ -карбоксила Р1, имеют  $IC_{50}$  в наномолярном диапазоне концентраций [40]. Находящаяся в Р1 части ингибиторов на основе мочевины  $\alpha$ -карбоксильная группа может служить для компенсации слабого взаимодействия мочевины с каталитическим сайтом. Уреидо-группа предположительно взаимодействует с Asn519, Arg534, Arg536, и это взаимодействие наблюдается во всех структурах [40].



**Рис.6.** Структуры ингибиторов GCP11 и значения их констант ингибирования GCP11. А. Ингибиторы на основе фосфатов и тиолов. В. Замена глутамата в P1' фрагменте ингибитора GCP11 значительно увеличивает значение  $K_i$ . С. Важность  $\alpha$ -карбоксовой группы в P1 фрагменте ингибитора GCP11, длины и состава спейсера ингибитора.

## Сайт узнавания глутамата (S1') GСРП и P1' фрагмент ингибиторов GСРП

Как природные субстраты GСРП, так и большинство ингибиторов имеют глутамат в P1' [41]. Взаимодействие S1' - P1' имеет решающее значение для связывания большинства ингибиторов GСРП, и замена глутамата в положении P1' резко снижает аффинность к GСРП.

Участок S1' занимает объем 8\*8\*8 Å<sup>3</sup> на дне полости GСРП. Распознавание глутамата в кармане S1' определяется сетью ионных и водородных связей [42]. Находящийся в P1' части ингибитора α-карбоксилатный мотив глутамата взаимодействует непосредственно с гуанидиновой группой Arg210 и гидроксильными группами Tyr552 и Tyr700. Контакты глутамата с Asn257 и Arg210 поддерживаются мостиковой молекулой воды. γ-карбоксилат S1'-связанного глутамата связан с амидной группой Asn257 водородной связью, образуя солевой мостик с ε-NH<sub>2</sub> группой Lys699; γ-карбоксилат также взаимодействует с Asn257 и Lys699 через молекулу воды. Позиционирование глутамата P1' обеспечивается за счет гидрофобных взаимодействий с боковыми цепями Phe209 и Leu428 [43]. В целом, пластичность кармана S1' ограничена, за исключением участка Tyr692-Ser704. Эти остатки могут быть перемещены на 4 Å<sup>3</sup>, увеличивая размер S1' [38], [42], [44].

Как и в природном субстрате NAAG, P1' фрагмент ингибитора GСРП представляет собой остаток глутамата. Модификации глутаматной боковой цепи или удаление α-карбоксильной группы в P1' части ингибиторов существенно снижают аффинность, поскольку карман S1' селективен для соединений с C-концевым глутаматом [44], [45] (Рис.6В). Мутации GСРП вблизи активного сайта оказывали слабое влияние при изменении аминокислот, взаимодействующих с γ-карбоксилатом, но сильное - при изменении аминокислот, взаимодействующих с α-карбоксилатами P1 и P1' [46], что указывает на их ключевое значение для связывания базового мотива глутамат-мочевина-лизин в кармане P1', активном центре и кармане P1. Сопоставление фосфонатного соединения 2-РМРА **1** с глутаматной боковой цепью и его производными **6** и **7** иллюстрирует ослабление ингибирования после модификации глутамата [47]. На примере 2-РМРА было обнаружено, что α-карбоксильная группа глутамата P1' участвует в ионном взаимодействии с Arg210 и образует водородные связи с Tyr552 и Tyr700; γ-карбоксильная группа 2-РМРА взаимодействует с Asn257 и Lys699 [47]. Аналогично, соединения на основе мочевины сильно зависят от глутамата в качестве фрагмента P1' [48], [49]. Связывание глутамата в кармане S1' GСРП одинаково для ингибиторов на основе мочевины и на основе фосфатов [40]. Сохранение естественной стереохимии на α-углероде P1' фрагмента ингибитора GСРП имеет решающее значение: соединения **11** с K<sub>i</sub> = 5,3 нМ и **16**, K<sub>i</sub> = 16,2 нМ, которые имеют (L)-конфигурацию на α-углероде, гораздо лучше соответствующих (D)-аналогов **17** и **18** с K<sub>i</sub> = 794 нМ и K<sub>i</sub> = 795 соответственно [49].

Замена  $\gamma$ -карбоксилата остатка глутамата P1' ингибитора приводит к несколько более слабому ингибированию GСРП (Рис.6В). Ингибиторы **13-18** с модификациями, замещающими остаток глутамата, были кристаллизованы в комплексе с GСРП[48]. Жесткость кармана S1' ограничивает ориентацию и размер боковых цепей, которые могут быть использованы во фрагменте P1' ингибитора. Небольшие ингибиторы GСРП с более длинными алифатическими боковыми цепями терминального звена, в отличие от коротких алкенов/алкинов, могут быть размещены в кармане S1' с конформационными изменениями, ограниченными только боковой цепью Lys699 [50].

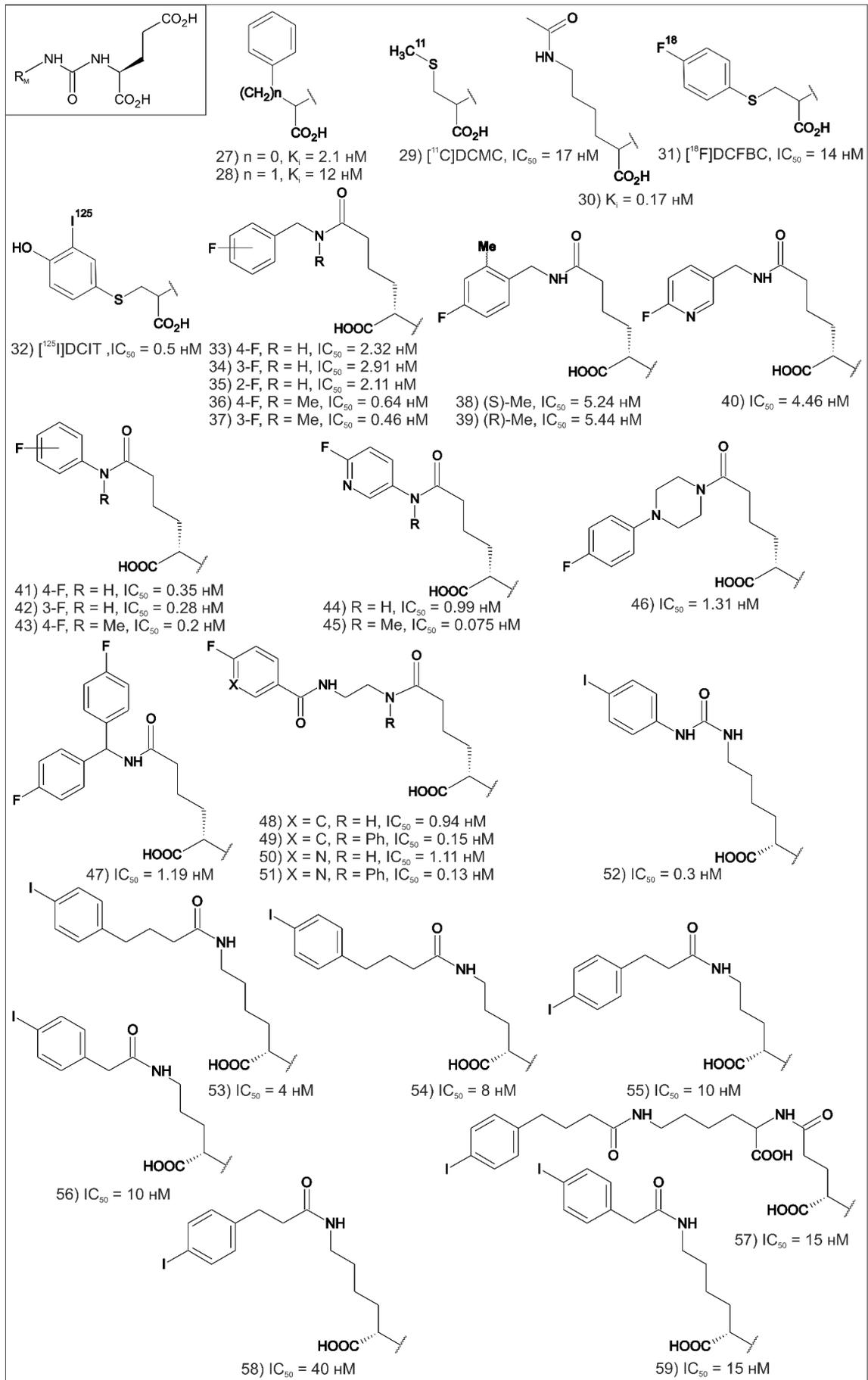
Модификации P1' часто направлены не на улучшение ингибирования GСРП, а на изменение фармакокинетических свойств:  $\gamma$ -карбоксилат был заменен на различные соединения для повышения его липофильности [49]. Такие модификации, как в **9-18** (Рис.6В), могут улучшить проникновение через гематоэнцефалический барьер путем пассивной диффузии, сохраняя средство связывания GСРП, достаточное для визуализации GСРП-экспрессирующих клеток глии в центральной нервной системе, но снижая средство ингибиторов к GСРП [49], [50]. DCIBzL **8**, например, слишком гидрофилен для проникновения через гематоэнцефалический барьер путем пассивной диффузии, в то время как одно из его производных, **9**, было обнаружено в мозге мышей с помощью однофотонной эмиссионной компьютерной томографии [49]. Это демонстрирует возможность использования ингибиторов GСРП в качестве средств диагностики и борьбы с нейропатологиями, связанными с повышенной экспрессией GСРП и глутамат-зависимым повреждением нейронов.

### **Связывание P1 фрагмента ингибитора в S1 кармане GСРП**

Связывание P1 части природных субстратов GСРП происходит в S1 кармане, который состоит из фрагмента поверхности воронки с аргининовым трактом. Аргининовый тракт S1 расположен в воронке между входной крышкой и каталитическим сайтом. Он состоит из гуанидиновых групп Arg534, Arg536 и Arg463, которые образуют положительно заряженную поверхность. Поэтому с GСРП лучше связываются соединения с кислотными группами, оказывающимися в S1-сайте. Все три остатка Arg укладываются друг на друга в отсутствие связанного ингибитора [51], но при наличии ингибитора происходит сдвиг гуанидиновой группы Arg536 на  $4,5 \text{ \AA}$  и смещение боковой цепи Arg463 [35]. Структура тракта частично поддерживается ионом хлора за счет стабилизации боковой цепи Arg534. Гибкость двух других аргининов позволяет связывать в S1 сайте белка фрагменты субстратов с широким разнообразием модификаций (Рис. 7).

## Карбоксильная группа P1 фрагментов ингибиторов GСPII

При конструировании высокоаффинных ингибиторов GСPII учитываются возможные ионные взаимодействия между аргининовым трактом и отрицательно заряженными функциональными элементами ингибиторов (в основном карбоксилатами). Показано, что включение в структуру P1 фрагмента ингибитора карбоксилата увеличивает аффинность ингибитора в несколько раз, а удаление  $\alpha$ -карбоксильной группы существенно снижает аффинность для молекул на основе фосфорамидата и мочевины. Иллюстрацией этого служит сопоставление  $\alpha$ -карбоксилат-содержащих соединений **19,20** с веществом **21** без  $\alpha$ -карбоксильной группы (Рис.6С). Сохранение естественной стереохимии на  $\alpha$ -углероде P1 является критическим для ингибирования GСPII.



**Рис.7.** Структуры ингибиторов GCP11 и значения их констант ингибирования GCP11.  
 Использование различных заместителей для диверсификации P1 фрагмента ингибитора.

## Модификации боковой цепи фрагмента P1 ингибиторов GСPII

Первыми соединениями для диагностики рака простаты с модификациями фрагмента S1 являются [<sup>11</sup>C]DCMC **29**, [<sup>18</sup>F]DCFBC **31** и [<sup>125</sup>I]DCIT **32** – производные тирозина и цистеина [52], [53] (Рис.7). Ароматическое замещение вблизи цинк-связывающей группы в соединениях **31** и **32** усилило ингибирование по сравнению с метил-замещенным цистеиновым производным **29**. Справедливо предположить, что из-за близости к цинк-связывающей группе эти ароматические группы расположены в S1, но их точная локализация неизвестна из-за отсутствия данных рентгеноструктурного анализа.

Дальнейшее замещение ε-аминогруппы лизина ингибиторов GСPII привело к производным с ароматическими фрагментами в мотиве P1. В этих соединениях боковая цепь лизина или его аналоги служат являются спейсером между карбоксильной и ароматической группами.

### Спейсеры между карбоксильными и ароматическими группами в фрагменте P1 ингибиторов GСPII

Большинство спейсеров основаны на производных лизина из-за его широкого применения в органическом синтезе и возможности использовать ε-аминогруппу как точку разветвления для конъюгации ингибиторов GСPII с терапевтическими или диагностическими молекулами (Рис.7). Соединения **41-45** с укороченным на один атом линкером P1 демонстрируют, что остаток аминокислоты в качестве основного мотива спейсера служит хорошей заменой традиционному лизину [54].

Участок ветвления в спейсере представляет собой амин, пептидную связь или мочевины, которая имитирует пептидную связь. Соединения с пептидной связью или мочевиной в спейсере имеют более высокое сродство к GСPII, чем аминокислотные производные, что видно из сравнения структур и ингибирующих свойств веществ **60** и **69**, **61** и **70**, **65** и **72**, и **68** и **73** (Рис.9). Исключением из этого правила являются соединения **65** и **71** с *para*-Cl-фенил-замещенной ε-аминогруппой. Вероятно, эти группы могут образовывать дополнительные водородные связи с S1 карманом GСPII.

Введение метильной группы на амидном азоте спейсера **33** и **34** увеличило аффинность этих ингибиторов к GСPII (соединения **36** и **37**)[54] (Рис.7). Как бензиламиды (**36**, **37**), так и анилиды (**43**, **45**) демонстрировали повышенное сродство, если они N-метилованы в пептидной связи в спейсере, по сравнению со структурно подобными N-H производными **41** и **44**. Более того, ароматическое соединение **45** входит глубже в карман, образованный Tug700, G206, K207 и V208, по сравнению с **44**[54].

Спейсер с четырьмя группами  $\text{CH}_2$  и точкой ветвления на основе мочевины в **52** (MP-1095) является наиболее эффективным среди соединений **52-59** (Рис.7). Замена аминогруппы спейсера в **52** на метиленовую группу в **59** привела к 50-кратному снижению  $\text{IC}_{50}$ . Другие изменения спейсера (соединения **53-58**) также не улучшили связывание с GCP II. Модификации P1 фрагмента ингибитора GCP II могут быть несущественными для ингибирования, но изменять фармакокинетические свойства соединения. Это наблюдение справедливо для DCFpyL (**79**,  $\text{IC}_{50} = 13$  нМ) и PSMA-1007 ( $\text{IC}_{50} = 4,2$  нМ): хотя ингибитор в 3 раза мощнее, PSMA-1007 продемонстрировал более медленную фармакокинетику с задержкой неспецифичной доставки молекулы-трейсера в желудочно-кишечный тракт, селезенку, почки, надпочечники и другие нецелевые ткани [55].

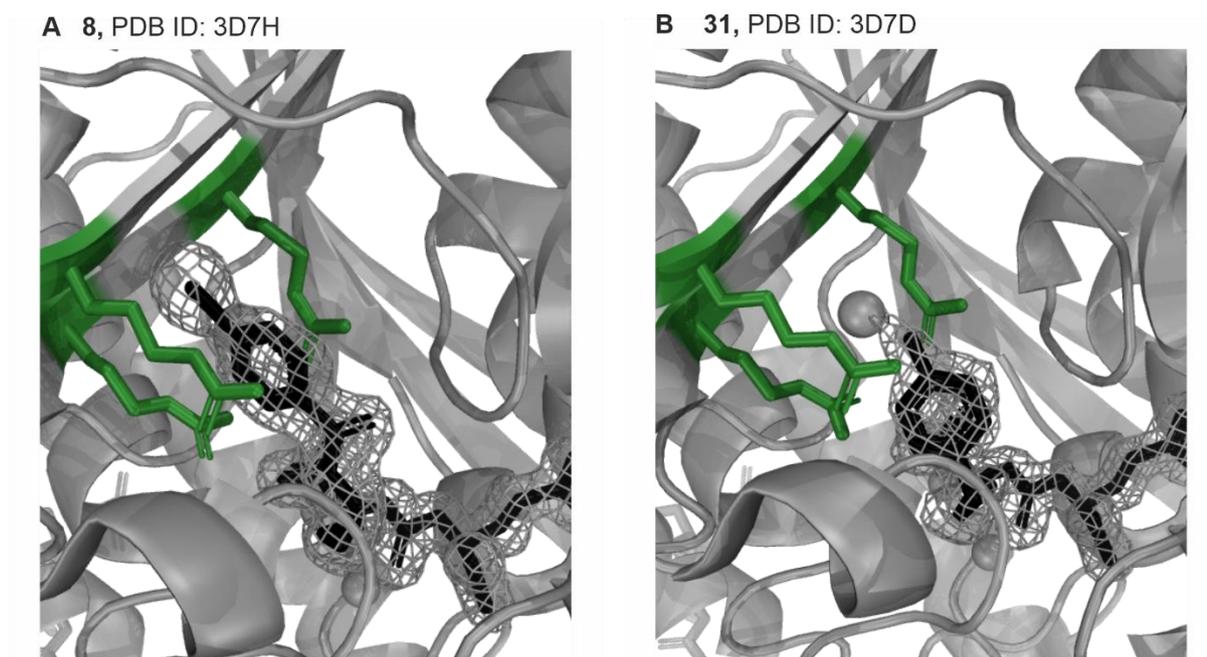
### Модификации ингибиторов GCP II ароматическими группами $\text{AR}_{\text{P1}}$

В природных фолат-содержащих ингибиторах GCP II часть P1 содержит ароматическую группу. P1 фрагмент ингибитора может быть модифицирован ароматическими группами  $\text{AR}_{\text{P1}}$  (соединения F1, F2) для усиления связывания с гидрофобным сайтом S1 [56] и увеличения липофильности соединения. Существует разница в способе взаимодействия с GCP II среди соединений **27-29**, **31**, **32**, которые имеют различные молекулы вместо остатка лизина (Рис.7), и **8**, который содержит лизин (Рис.6B). Замещенная ароматическая группа **8** полностью находится внутри кармана, образованного остатками аргининового тракта – Glu457, Arg463, Asp465, Arg534, Arg526 (PDB id: 3D7H, Рис.8A)[30]. Ароматическая группа **31** лишь частично расположена внутри кармана (PDB id: 3D7D, Рис.8B).

Рентгеновские структуры GCP II с ингибиторами, имеющими меньший размер группы, показывают, что при взаимодействии с такими соединениями карман S1 закрыт (PDB id: 3D7G для **29**, 3D7F для **32**)[46]. Достаточная длина линкера (как в Lys-C(O)-Glu) и оптимальная конформация замещенной ароматической группы **8,12** увеличивает способность ингибитора связываться с GCP II по сравнению с соединениями с меньшими линкерами. Аналогичная ситуация наблюдается и для других классов ингибиторов (**22-26**): гибкая P1 часть **25** и **26** с ароматическим фрагментом  $\text{AR}_{\text{P1}}$  увеличивает аффинность связывания с GCP II.

Аффинность анилидов **41-45** была в среднем в несколько раз выше, чем у соответствующих бензиламидов **33-40**, хотя соединение **40** с пиридиновым кольцом показало примерно в два раза меньшую аффинность к GCP II по сравнению с **33** [54] (Рис.7). Положение ароматического фрагмента в соединениях **41** и **44** существенно отличается от **33**. В структуре комплексов GCP II с ингибиторами **41** (PDB ID: 6H7Y) и **44** (PDB ID: 6HKJ) ароматический фрагмент  $\text{AR}_{\text{P1}}$  примыкает к боковой цепи Tyr700 и остаткам G206, K207 и V208, тогда как ароматический фрагмент **33** (PDB ID: 5OF0) полностью вставлен в карман S1'. Остатки F546,

S547, G548 и W541 входной крышки образуют напротив Tyr700 поверхность связывающего кармана вокруг сайта связывания **41** и **44**.



**Рис. 8.** Изменение положения фрагмента P1 ингибитора GCP II в гидрофобном кармане S1 фермента в зависимости от длины и природы спейсера. Ингибиторы и Arg534, Arg536 и Arg463 представлены в виде стержневой модели, причем атомы ингибиторов окрашены в черный цвет, а атомы аргининов - в зеленый. Карты Fo-Fc (серые) для отдельных ингибиторов имеют отсечку на уровне 1,6  $\sigma$ . А) Только AR<sub>P1</sub> **8** полностью помещается в карман. В) Ароматическая группа **31** частично находится внутри гидрофобного кармана S1.

Связь между заместителем в ароматическом фрагменте AR<sub>P1</sub> и аффинностью ингибитора может быть выведена из изученных модификаций лигандов **60-69** [57], [58], [59] и **74** (Рис.9). В случае соединений **60-68** на их сродство влияет природа заместителя и его расположение на арильном кольце [59][58]. Для *para*-замещенных соединений **61**, **64**, **65**, **68** значения IC<sub>50</sub> следуют тренду Cl<I<Br<<F=H, и *para*-положение в целом лучше для введения галогенных заместителей. Похоже, что наибольшее снижение константы ингибирования GCP II при замещении *para*-положения возможно при использовании в качестве заместителей AR<sub>P1</sub> более крупных и менее электроотрицательных атомов галогенов, за исключением **62**. В *ortho*-положении таким заместителем является атом I, в то время как в *meta*-положении – атом Cl. В соединениях **69-73** [58] сила ингибирования GCP II соответствует тренду Br=F=Cl<I=H и не имеет существенной разницы для ингибиторов с разными галогеновыми заместителями. Вероятно, на аффинность влияет образование дополнительных водородных связей за счет близко расположенной мочевины. Соединения **74-78** не демонстрируют предпочтений к заместителям из-за введения дополнительной группы в *ortho*-положение: группа, вероятно,

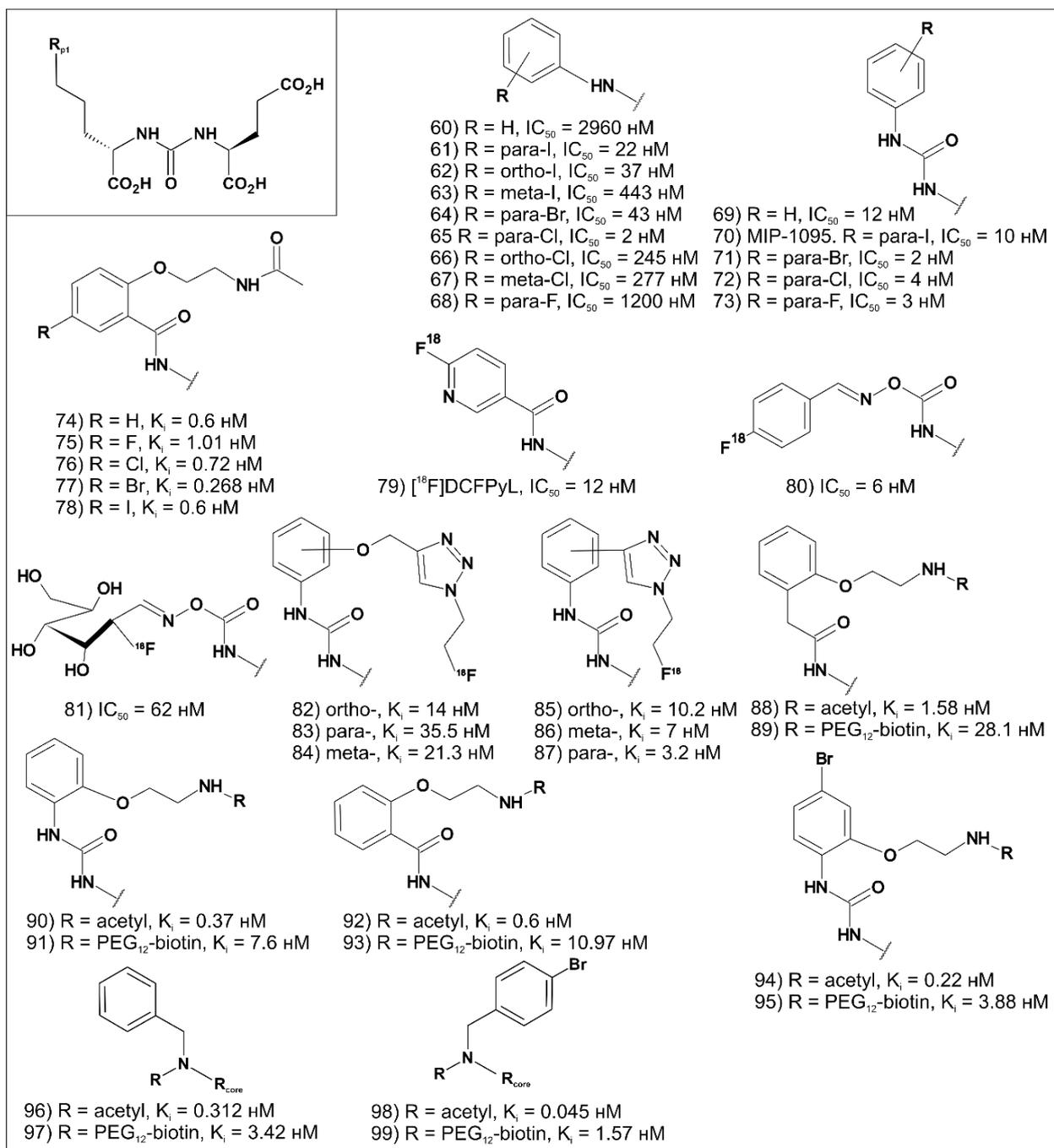
стерически нарушает связывание между ариловым мотивом и карманом S1 [60]. Расположение триазилольного заместителя имеет тот же эффект, что и атомов галогенов, и соответствует тренду аффинности *пара*-> *мета*-> *орто*- [61]. Более жесткий триазилольный заместитель в **85-87** предпочтительнее триазилилметоксигруппы в **82-84** [61]. Соединения **82-87** имеют лучшие характеристики *in vitro*, несмотря на то, что их константы ингибирования выше, чем у производных соединения **8**. Серия ингибиторов **85-87** лучше поглощается и удерживается в опухоли, имеет большее соотношение концентрации ингибитора в опухоли по сравнению с кровью, чем [<sup>125</sup>I]DCIBzL **8**, который в настоящее время проходит клинические испытания [61].

Положение атома фтора в арильном фрагменте P1 ингибиторов GCP II не влияет на ингибирование бензиламидными производными **33-35** [54] (Рис.7). Соединения **38** и **39** с метильной группой в *орто*-положении и фтором в *пара*-положении арильного фрагмента AR<sub>P1</sub> показали приблизительно 2- и 10-кратное снижение эффективности ингибирования GCP II по сравнению с **33** и **36**, соответственно. Метильная группа может препятствовать полному вхождению фтор-замещенного арильного фрагмента в S1' или создавать стерическое напряжение между кольцевым фрагментом и остатком лизина ингибитора при ротационном угле, наблюдаемом в кристаллической структуре комплекса GCP II с **33** (PDB ID: 5OF0) [54].

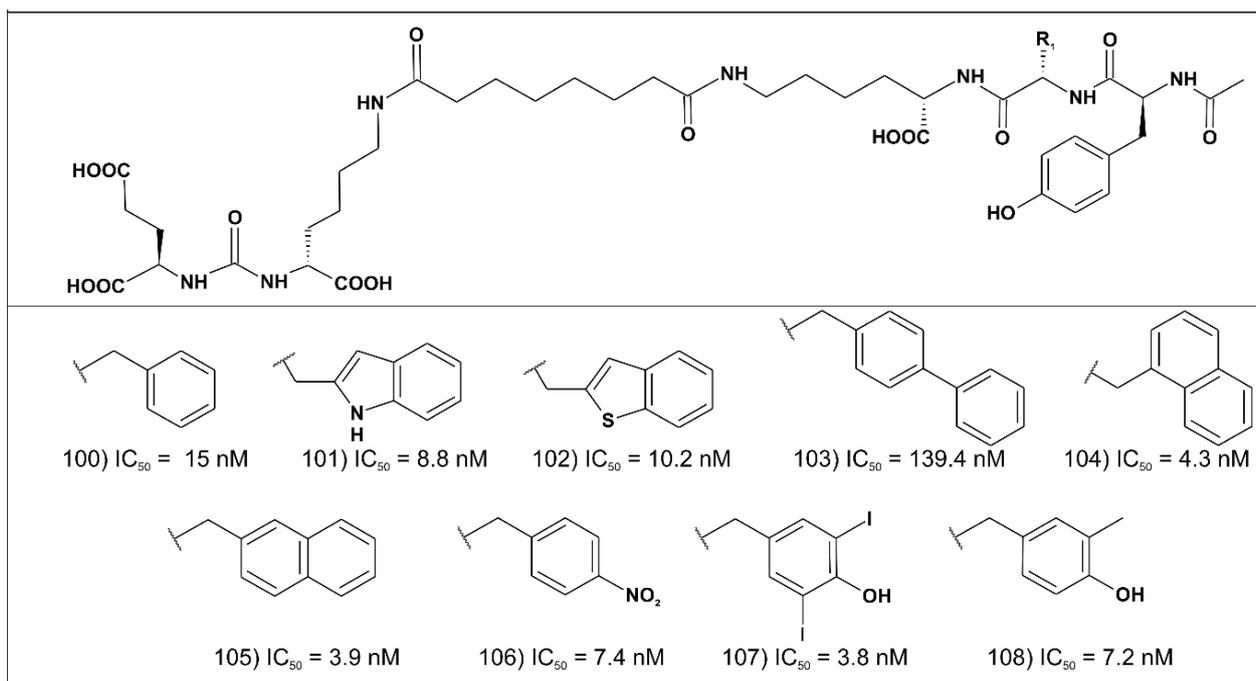
Возможны модификации спейсера P1 фрагмента ингибиторов GCP II за счет введения дополнительных звеньев между ароматическим фрагментом AR<sub>P1</sub> и точкой ветвления в ε-аминогруппе остатка лизина. Так, увеличение спейсера соединения **46** привело к 2-кратному уменьшению IC<sub>50</sub> по сравнению с **33** [54] (Рис.7). Добавление метиленового спейсера между мочевиной P1 и *мета*-I-замещенным ариловым кольцом почти не влияет на K<sub>i</sub>. На основании данных по ингибиторам со спейсерами разной длины можно сделать вывод, что спейсер должен содержать больше двух звеньев (как в **8**), а если спейсер намного длиннее (например, соединения **60-73** (Рис.9), то он должен быть достаточно гибким, чтобы позволить кольцу достичь кармана S1.

При замене ароматического фрагмента AR<sub>P1</sub> на гидрофильные фрагменты, такие как 2-дезоксигалактоза-2-[18F]фтор-D-глюкоза **81**, наблюдается увеличение константы ингибирования GCP II (Рис.9). По мнению авторов, такие фрагменты не образуют π-стекинг и π-катионные взаимодействия с S1-сайтом [62].

Большое влияние природы крупных заместителей арильного фрагмента продемонстрировано в **82-87**: в серии **82-84** *орто*-положение заместителя кажется наилучшим, тогда как в соединениях **85-87** *пара*-замещенный ингибитор **87** имеет самое низкое значение K<sub>i</sub> [61] (Рис.9).



**Рис.9.** Структуры ингибиторов GCP11 и значения их констант ингибирования GCP11 при использовании различных заместителей для диверсификации P1 фрагмента ингибитора. 60-80: использование галогеновых заместителей для модификации арильного фрагмента AR<sub>P1</sub>. 81-87: использование крупных заместителей для модификации арильного фрагмента AR<sub>P1</sub>. 88-99: влияние нагрузки в виде биотина, присоединенного к ε-аминогруппе лизина ингибитора GCP11, через линкер PEG<sub>12</sub>, на K<sub>i</sub>.



**Рис.10.** Структуры ингибиторов GCP II и значения их констант ингибирования GCP II при модификации линкера для усиления связывания ингибитора с GCP II за счет взаимодействия с арен-связывающим сайтом.

### Структурные особенности арен-связывающего сайта (ABS)

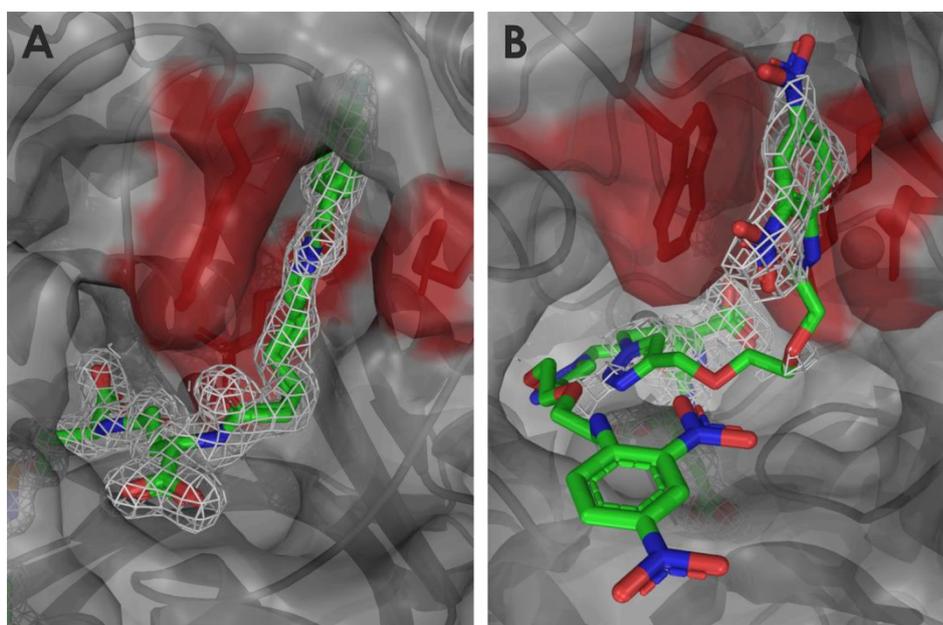
Каталитический сайт "спрятан" от внешней среды подвижной входной крышкой, образованной петлей Trp541-Gly548, способной приобретать открытую и закрытую конформацию [35]. Подвижность петли обусловлена пластичностью Gly548 и поворотом пептидной связи между Trp541 и Asn540. В закрытом состоянии петля образует один оборот  $\alpha$ -спирали, которая препятствует доступу к внутреннему пространству фермента. Открытая конформация может быть вызвана связыванием крупного лиганда, когда закрытая конформация не может быть сформирована из-за размера модификации P1-сайта соединения [35]. Поскольку GCP II гидролизует как маленькие, так и большие субстраты - как, например, поли(глутамат)-фолаты [63], - положение крышки не оказывает качественного влияния на активность GCP II. Trp541 открытой крышки вместе с Arg463 и Arg511 образует арен-связывающий сайт ABS, связывание с которым резко увеличивает сродство ингибитора к GCP II в случае, если лиганды имеют арильные заместители в части P1 [43].

### Заместители в AR<sub>ABS</sub>-части ингибиторов GCP II

Соединения **49** и **51** с фторсодержащим арильным фрагментом и N-фенильной группой, конъюгированных к  $\epsilon$ -аминогруппе лизина мотива глутамин-мочевина-лизин, проявляли в 6 и 8 раз большую ингибирующую активность в отношении GCP II по сравнению с вторичными амидами **48** и **50** (Рис.7) [54]. Это возможно благодаря тому, что N-метильная группа направлена

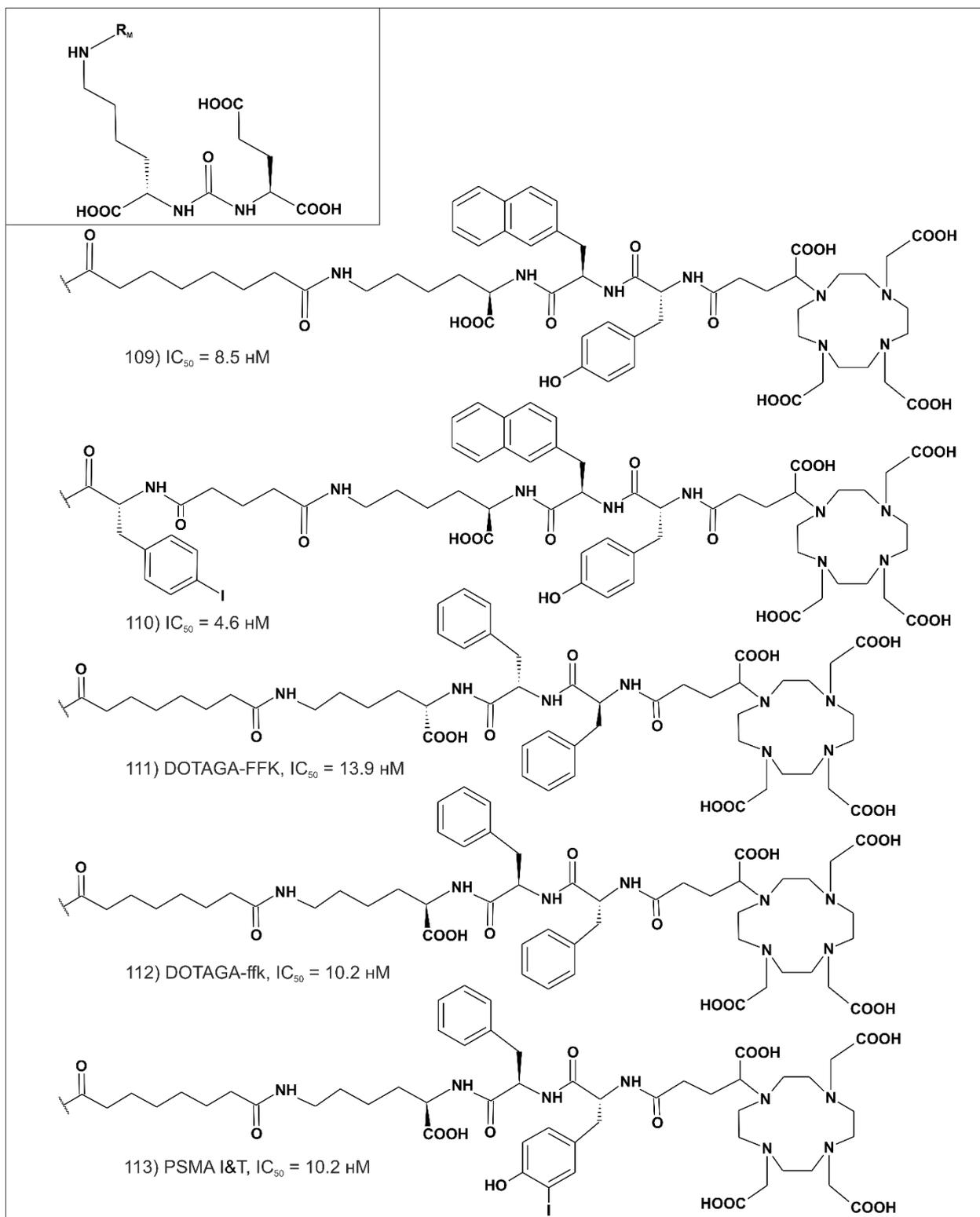
в сторону незанятого пространства входной воронки в структуре GСРІІ-45, что позволило ввести более крупный мотив [54].

Хотя структур GСРІІ с лигандом, имеющим модификации для взаимодействия с ABS, немного (Рис. 11А, В), из имеющихся данных можно сделать некоторые выводы. Соединения **111-113** с двумя арильными группами в AR<sub>ABS</sub> непосредственно участвуют во взаимодействии с ABS посредством одной из групп.



*Рис. 11. Положение арильных фрагментов аналогов глутамата в арен-связывающем сайте ABS (обозначен красным) GСРІІ: А) PDB ID: 2XEI [60]. В) PDB ID: 4JYW[64]. Карты Fo-Fc (серые) для отдельных ингибиторов отсечены при 1,6  $\sigma$ .*

Замещение одного из арильных фрагментов в дипептидном фрагменте линкера также влияет на K<sub>i</sub> ингибиторов. Модификация нафтилаланина привела к наибольшей аффинности к GСРІІ у **104, 105** среди **100-108**, тогда как остаток дифенилаланина увеличил константу ингибирования **103** (Рис.10) [65]. Изменение линкера может влиять на распределение исследуемых молекул в организме не только за счет изменения сродства к GСРІІ. Так, более высокая липофильность **109** и **110** по сравнению с "PSMA I&T" **113** оказалась благоприятной с точки зрения аффинности и интернализации и привела к более высокому поглощению опухолью (Рис.12) [65].



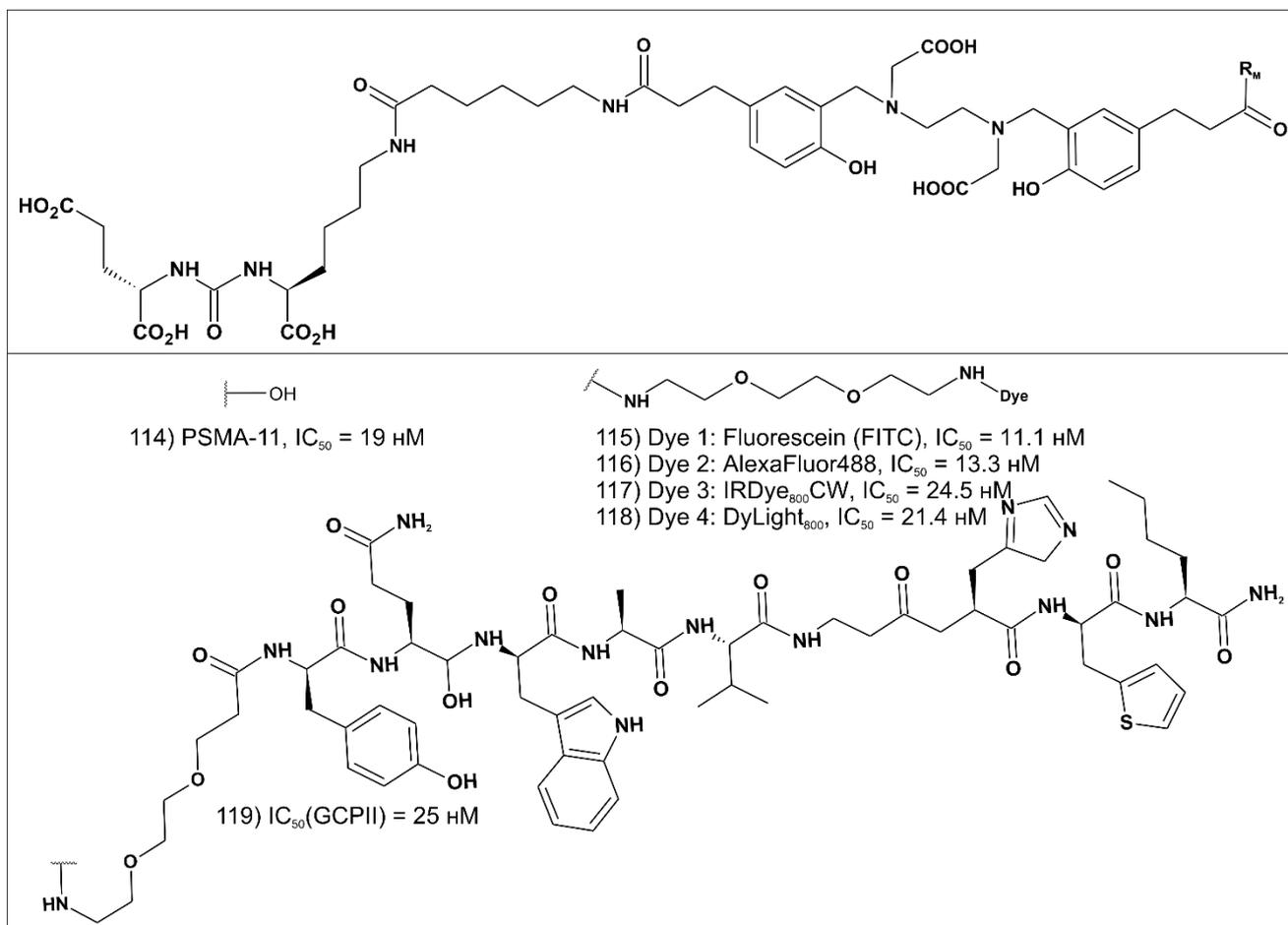
**Рис.12.** Структуры ингибиторов GPCII и значения их констант ингибирования GPCII при модификации линкера для усиления связывания ингибитора с GPCII за счет взаимодействия с ABS в присутствии хелатора в качестве нагрузки.

### Значение длины линкера ингибитора

Добавление линкера PEG<sub>12</sub> (Рис.9) приводит к 10-20-кратному снижению ингибирующих свойств используемого ингибитора [60]. Анализ взаимосвязи структура-активность соединений **88-99** демонстрирует необходимость разветвленного присоединения линкера и арильной группы к ε-аминогруппе [60]. В случае **88-93** линкер присоединен непосредственно к арильному мотиву, что меняет связывание этого мотива с карманом GCP II [60]. Наиболее эффективным способом присоединения молекулы биотина, использованной как "полезная нагрузка" оказался линкер, присоединенный к ε-аминогруппе ( $K_i(\mathbf{96}) = 0,312 \pm 0,032$  нМ,  $K_i(\mathbf{97}) = 3,42 \pm 0,47$  нМ). *Орто*-расположенный линкер (**90, 91**) и присоединение арильного кольца к ядру Glu-C(O)-Lys через амидную группу не приводили к резкому увеличению  $K_i$  биотинилированного ингибитора ( $K_i(\mathbf{90}) = 0,377 \pm 0,056$  нМ,  $K_i(\mathbf{91}) = 7,6 \pm 1,2$  нМ). Модификация соединений **90, 91, 96, 97** атомом Br в *para*-положении привела к созданию **94, 95, 98, 99**: это снизило  $K_i$  для биотинилированных соединений **95** и **99** примерно в два раза. Конъюгация P1 фрагмента ингибитора с линкером (в ε-аминогруппе лизина и в *орто*-положение AR<sub>P1</sub>) совместно с атомом Br, помещенным в *para*-положение ароматического кольца AR<sub>P1</sub>, показали синергический эффект и обеспечили наиболее сильное связывание ингибиторов **98** и **99** среди рассмотренной серии ингибиторов.

Ингибиторы GCP II могут быть конъюгированы с низкомолекулярными соединениями, обеспечивающими цитотоксичность [66] или возможность детекции опухоли во время хирургического вмешательства [67]. Конъюгированные с красителем лиганды **115-119** на основе ингибитора **114** позволяют использовать преимущества радиоактивных и флуоресцентных сигналов и демонстрируют высокую аффинность связывания с GCP II и хорошее поглощение клетками опухоли после связывания GCP II (Рис.13) [67]. Низкомолекулярные соединения незначительно влияют на IC<sub>50</sub>, изменяя ее не более чем в 2 раза, что может быть продемонстрировано сопоставлением соединений **115-119** и **114**. Конъюгация двух ингибиторов – GCP II и рецептора гастрин-рилизинг-пептида (GRPR) – не вызвала значительной потери ингибирующих свойств обоих ингибиторов [68]: IC<sub>50</sub>(GCP II) изменилась с 10,8 нМ для оригинального ингибитора до 25 нМ для химерного **119**, а IC<sub>50</sub>(GRPR) не изменилась.

На значение константы ингибирования влияют длина линкера и природа его модификаций, поэтому присутствие ароматических молекул может быть полезным для одних линкеров и ингибиторов в целом и бесполезным или даже вредным для других. Необходимы систематические исследования для выбора линкера, способствующего связыванию с GCP II, высокому поглощению опухолью, низкой фоновой активности и подходящей скорости выведения из организма.



**Рис.13.** Структуры ингибиторов GPCII и значения их констант ингибирования GPCII. Конъюгаты ингибитора **114** с флуоресцентными красителями и ингибитором GRPR.

На сегодняшний день наибольшее разнообразие модификаций было проверено с использованием Lys-C(O)-Glu - самого эффективного ядра. Следуя модульной стратегии конструирования ингибиторов GPCII, предполагается, что каждая часть молекулы ингибитора может быть диверсифицирована практически независимо. Однако в ходе выполнения данной работы было продемонстрировано, что это не совсем так, и систематическое исследование сочетаний описанных модификаций является перспективной стратегией поиска новых и эффективных ингибиторов GPCII.

# Регуляторные последовательности и известные регуляторы

## экспрессии *FOLH1*

В литературе принято рассматривать GСРII как маркер опухоли простаты и мишень для адресной доставки диагностических и терапевтических молекул на основе ингибиторов GСРII в опухоль. Однако его высокая экспрессия также наблюдается в двенадцатиперстной кишке, проксимальных канальцах почек, в глиальных клетках некоторых отделов мозга и в эндотелии сосудов, образованных в ходе опухолевого ангиогенеза. При этом большое количество фермента совпадает с высоким уровнем транскрипции его гена, *FOLH1*, и низкий уровень транскрипции гена соответствует низкой представленности GСРII на поверхности клеток как в образцах опухоли, так и в клеточных линиях. Такая связь позволяет предположить, что лимитирующей стадией на пути от ДНК к белку является транскрипция, и именно на ней сконцентрирована тканеспецифичная регуляция, определяющая уровень экспрессии *FOLH1*. Установление регуляторов экспрессии *FOLH1* позволит лучше понять каскады, участвующие в обеспечении опухоли- и тканеспецифичности, и рассматривать ключевые белки этих каскадов в качестве мишеней для диагностики и терапии заболеваний.

## Регуляция транскрипции

Транскрипция осуществляется ДНК-зависимой РНК-полимеразой, катализирующей синтез РНК, которая соответствует матрице ДНК. РНК-полимераза взаимодействует с промотором гена в процессе инициации транскрипции, расплетает ДНК в участке связывания, начинает синтез РНК и покидает промотор[69]. Элонгация РНК длится до тех пор, пока комплекс элонгации не встретит сигнал терминации транскрипции в последовательности гена, после которого происходит высвобождение РНК и ДНК. В клетках эукариот существует несколько РНК полимераз, осуществляющих транскрипцию разных групп генов, имеющих различные механизмы регуляции и ассоциированные факторы[70]: РНК полимеразы I синтезируют предшественник большой рибосомной РНК, РНК полимеразы II – матричные РНК (мРНК) и часть некодирующих РНК, РНК полимеразы III – транспортные РНК и малую рибосомную РНК. С учетом того, что продуктом интересующего нас гена *FOLH1* является белок, далее речь пойдет о транскрипции в клетках человека с участием РНК полимеразы II с особым вниманием к инициации транскрипции.

Транскрипция инициируется на 5'-конце гена – точке старта транскрипции, расположенной внутри минимального промотора, который захватывает примерно 50 п.о. последовательность до и 50 п.о. после точки старта транскрипции. Синтез пре-мРНК в клетке – это не непрерывный процесс. Транскрипция происходит вспышками, которые характеризуются частотой инициации, длительностью и амплитудой (количеством пре-мРНК)[71]–[73].

Минимальный промотор взаимодействует с общими транскрипционными факторами (ТФ) и РНК полимеразой II [74], [75]. К общим ТФ относятся TFIIВ, TFIIД, TFIIЕ, TFIIF, TFIIH [74]. В дополнение к общим ТФ и минимальным промоторам генов, тканеспецифичность регуляции транскрипции осуществляется цис- и транс-элементами. К цис-элементам относятся дистальные элементы промотора, последовательности-энхансерами и инсуляторы, а к транс-элементам – комплекс медиатора, специфичные ТФ, белки, обеспечивающие модификации гистонов и ДНК.

Промотор – последовательность неопределенной длины, т.к. нет ограничивающих его четких биологических критериев. Дистальные элементы промотора – регуляторные области, которые могут располагаться на удалении до точки старта транскрипции. Они содержат сайты посадки транскрипционных факторов, при взаимодействии с которыми ТФ способны влиять на динамику транскрипции.

Энхансеры регулируют тканеспецифичность и пространственно-временную динамику транскрипции [76]. Они могут быть значительно удалены от точки старта транскрипции, находиться по обе стороны от нее в пределах хромосомы или располагаться на другой хромосоме. На данный момент показана необходимость физического сближения энхансера и минимального промотора для усиления транскрипции, однако сам механизм усиления неясен [76]. Энхансеры содержат участки связывания транскрипционных факторов и после взаимодействия с ними усиливают транскрипцию с промотора гена или нескольких генов. Доступность участков связывания ТФ в разных клетках регулируется состоянием хроматина, обусловленного модификациями гистонов (эпигенетические факторы) и связыванием тканеспецифичных ТФ-пионеров, которые способны связываться со своими сайтами посадки внутри «недоступного» хроматина (условно – гетерохроматина) и приводить к его ремоделированию [77],[78].

Транскрипционные факторы связываются со своими участками посадки внутри регуляторных элементов и регулируют инициацию транскрипции. Эти участки описываются консенсусами – «усредненными» последовательностями ДНК, содержащими в каждой позиции нуклеотид, наиболее часто встречающийся у нескольких последовательностей, с которыми взаимодействует ТФ. Источником данных для моделирования консенсусов является иммунопреципитация хроматина с последующим секвенированием связанных фрагментов ДНК (ChIP-seq). Сайты посадки ТФ могут располагаться как поодиночке, так в составе гомотипических или гетеротипических кластеров: в первом случае рядом находятся участки связывания одного и того же ТФ, во втором – разных ТФ-партнеров. Время взаимодействия ТФ с их сайтами связывания зависит от силы сайта связывания и локальной концентрации ТФ, что

делает маловероятной возможность зафиксировать некоторые регуляторные взаимодействия методом ChIP-seq. Это, в свою очередь, отражается на качестве консенсусов сайтов связывания ТФ и точности предсказаний этих сайтов в регуляторных последовательностях.

Общепринятое в настоящее время представление об инициации транскрипции предполагает, что сначала образуется преинициаторный комплекс из общих ТФ и РНК полимеразы II. Транскрипция с минимального промотора активируется или усиливается дистальными элементами промотора и расположенными на удалении энхансерами: с доступными сайтами связывания внутри энхансеров взаимодействуют ТФ, а с ними – транскрипционные кофакторы, и при участии медиатора активирующий сигнал передается на преинициаторный комплекс. Взаимодействие ТФ с доступными им сайтами их посадки может регулироваться концентрацией ТФ, в том числе локальной концентрацией ТФ вблизи регуляторного элемента, пост-трансляционными модификациями ТФ в ответ на внутри- или внеклеточный сигнал, присутствием транскрипционных кофакторов и ТФ-партнеров, повышающих кооперативность связывания друг друга внутри гомо- или гетеротипических кластеров сайтов связывания ТФ внутри регуляторных элементов.

Увеличение локальной концентрации ТФ вблизи промотора возможно благодаря совместному действию медиатора, архитектурных белков, обеспечивающих пространственную организацию генома, и энхансеров: предполагают, что ТФ диффундируют с энхансера на промотор и насыщают его окрестности, увеличивая эффективность инициации транскрипции [79], [80].

### **Транскрипционная регуляция гена *FOLH1***

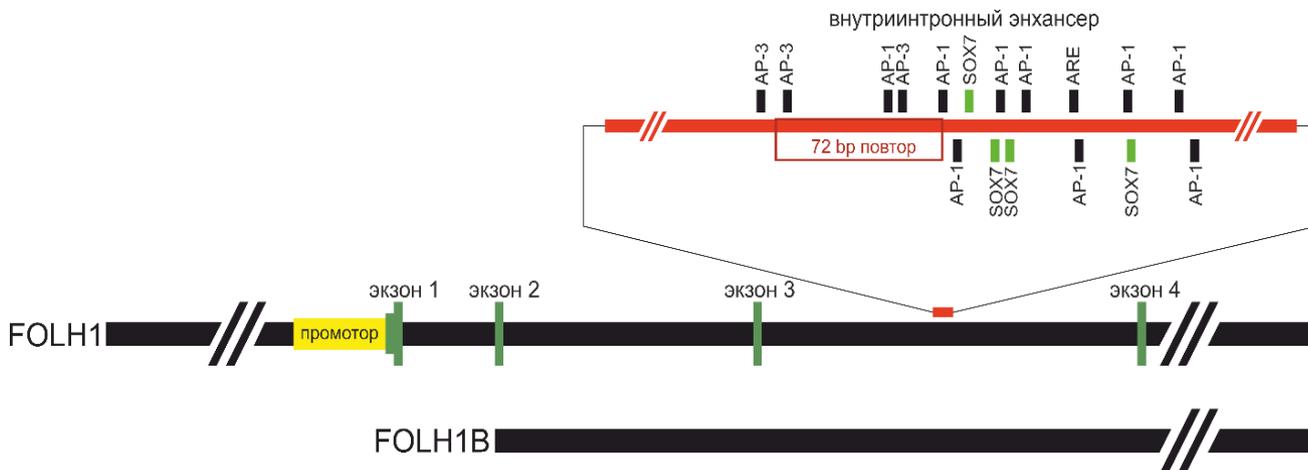
В тканях, как здоровых, так и опухолевых, где наблюдается экспрессия *FOLH1*, существует высокая корреляция между уровнем мРНК и белка, что позволяет судить о ключевой роли транскрипционной регуляции гена *FOLH1*. Однако транскрипционная регуляция экспрессии *FOLH1* изучена слабо и известно, что ключевой для простаты и опухоли простаты каскад андрогенового рецептора является репрессором транскрипции *FOLH1*. В данном разделе будут рассмотрены известные цис- и транс-регуляторы транскрипции *FOLH1*, а также будут упомянуты методы, с помощью которых показано влияние того или иного транскрипционного фактора на транскрипцию *FOLH1*. Большинство исследований проводилось авторами в клеточных линиях опухоли простаты, различающихся уровнем экспрессии гена *FOLH1*; клеточные линии неппростатного происхождения использовались в качестве контроля для определения простатспецифичности исследуемых регуляторов. Упоминаемые далее клеточные линии опухоли простаты LNCaP и VCaP – андроген-зависимые

линии, имеют высокий уровень экспрессии *FOLH1*, 22Rv1, одновременно андроген-чувствительная и нечувствительная клеточная линия, имеет сравнительно невысокую экспрессию *FOLH1*, а в PC-3 и DU-145, андроген-независимых линиях опухоли простаты, экспрессия *FOLH1* находится на очень низком уровне.

### **Последовательности гена *FOLH1*, содержащие цис-регуляторные элементы**

Экспериментально показаны две регуляторные последовательности, оказывающие влияние на транскрипцию *FOLH1* – промотор и внутриинтронный энхансер (Рис.14). На данный момент известно, что промотор и энхансер *FOLH1* в составе репортерных конструкций способны осуществлять специфичную экспрессию репортера. Стоит отметить, что оценка влияния регуляторных последовательностей на транскрипцию *FOLH1* проводилась посредством создания химерных конструкций вне исходного клеточного контекста, игнорируя влияние эпигенетических модификаций ДНК и гистонов, пространственную укладку генома в области инициации транскрипции и взаимное расположение цис- и транс- регуляторных элементов.

Впервые промоторная область была размечена при исследовании структуры гена *FOLH1* посредством клонирования и проверки специфичности экспрессии репортера как фрагмент 1244 пар оснований (п.о.), расположенный около 1 тысячи пар оснований (т.п.о.) до и 200 п.о. после точки старта транскрипции[81]. Важность 200 п.о. последовательности до точки старта транскрипции подтверждается и другими работами[82], [83]. Был выделен минимальный промотор 614 п.о., содержащий участок до точки старта транскрипции и 185 п.о. транскрибируемой последовательности, необходимый для транскрипции этого гена, но не регулирующий специфичность экспрессии [82]. Увеличение размера промотора до 2 т.п.о. приводило к увеличению специфичности по сравнению с 614 п.о.: немного повышалась экспрессия в клеточных линиях рака простаты LNCaP и DU-145, что говорит о том, что простата-специфичные регуляторные участки находятся за пределами минимального промотора[82]. Данные этого исследования подтверждаются иммуно-окрашиванием клеток LNCaP, PC-3 и DU-145. Однако отсутствует специфичность экспрессии репортера длиной 2 т.п.о. промотором в другой работе в некоторых линиях, в которых *FOLH1* в норме экспрессируется на высоком уровне: если сравнивать PC-3 и LNCaP, то в исследовании сигнал репортера в PC-3 был заметно выше, чем в LNCaP[82], что противоречит данным об экспрессии *FOLH1*. Есть противоречия и с предыдущим исследованием: в первой работе [81] сигнал репортера коррелирует с экспрессией *FOLH1* в клеточных линиях и заметно убывает в ряду LNCaP – PC-3 – DU-145 – MCF-7, во второй работе [82] среди линий рака простаты наблюдается противоположная зависимость и сигнал репортера убывает в ряду PC-3 – DU-145 – LNCaP и является высоким в непростатных клеточных линиях – даже выше, чем в LNCaP.



**Рис. 14.** Устройство 5'-конца гена *FOLH1* и его паралога *FOLH1B*. Энхансер *FOLH1* расположен внутри третьего интрона. Ранее были предсказаны участки посадки ТФ и проверены соответствующие потенциальные ТФ-регуляторы экспрессии *FOLH1*. На увеличенной схеме энхансера темно-зеленым выделены экзоны гена, светло-зеленым – участки посадки экспериментально подтвержденных регуляторов транскрипции *FOLH1*, черным – предсказанные участки посадки ТФ. 5'-конец гена *FOLH1* соотнесен с 5'-концом его паралога *FOLH1B*.

Энхансер *FOLH1* находится в третьем интроне гена, около 12 т.п.о. после сайта старта транскрипции[84]. Он был обнаружен посредством клонирования фрагментов гена *FOLH1* перед промоторным участком *FOLH1* длиной 1 т.п.о. в вектор с последующей трансфекцией в клетки и оценкой сигнала репортера[84]. Минимальный фрагмент энхансера имеет длину 331 п.о. и содержит прямой повтор длиной 72 п.о.[84]. Энхансер оказывает наибольшее влияние на транскрипцию репортера в клетках, экспрессирующих *FOLH1* (LNCaP), как в конструкции с промотором *FOLH1*, так и с гетерологичными промоторами[84]. Энхансер работает исключительно в линиях, экспрессирующих *FOLH1*, и не работает в непростатных линиях, причем сильнейшее усиление экспрессии репортера составляет более 250 раз в связке с 1 т.п.о. промотором *FOLH1* [84] и показано в клеточной линии LNCaP. В другом исследовании экспрессия репортера с конструкции, содержащей энхансер и 2 т.п.о. промотор *FOLH1*, увеличилась в 41 раз [85], что противоречит описанному выше исследованию промотора *FOLH1*[82]. Т.к. при увеличении длины промотора повышалась простатспецифичная экспрессия *FOLH1*, то сочетание 2 т.п.о. промотора с энхансером должно сильнее увеличивать экспрессию репортера, чем сочетание 1 т.п.о. промотора с энхансером. Существует значительная разница в разных исследованиях между оценкой эффекта энхансера с помощью репортерной конструкции и при сравнении эндогенной экспрессии *FOLH1* в клеточных линиях методом количественной ПЦР. При использовании репортерной системы в LNCaP в одном исследовании энхансер сильнее, чем в PC-3, в 10 раз[84], а в другом – только в 4 раза[85]. В

первом случае использовался 1 т.п.о. промотор, во втором – 2 т.п.о. В то же время уровень транскрипции *FOLH1* в этих линиях дикого типа различается более чем в 1000 раз. К сожалению, соотнести эти данные невозможно из-за разной природы сигнала – люминесценции и количества мРНК.

Особая важность двух участков – 90 п.о. на 5'-конце и 65 п.о. на 3'-конце энхансера – показана с помощью делеционного анализа его последовательности [86]. Делеция первого вызывает ~3-кратное уменьшение экспрессии репортера, что указывает на наличие сайтов связывания активаторов транскрипции, а делеция второго – увеличение экспрессии репортера в 3,5 раза, из чего можно предположить наличие сайтов связывания репрессоров транскрипции [86]. Только 90 п.о. фрагмент не способен обеспечивать высокой экспрессии репортера [86]. Важность фрагмента 35...69 п.о. показана посредством последовательной замены каждые 17 п.о. в 90 п.о. фрагменте на участок связывания фактора GAL4, т.к. замены в пределах 35...69 п.о. приводили к примерно трехкратному падению сигнала репортера [87]. Замена последовательности 69...86 п.о. слегка увеличивала сигнал репортера по сравнению с немутированным энхансером [87], однако представленные в работе данные в этом случае не представляются достоверными и скорее свидетельствуют об отсутствии эффекта, т.к. значения погрешностей велики и результаты не анализировались статистически. Подстановка фрагмента 35...69 п.о. на 5'-конец энхансера с делецией 90 п.о. на 5'-конце восстанавливала уровень экспрессии репортера, соответствующий энхансеру дикого типа [87], что позволило сделать вывод о наличии сайтов связывания активаторов транскрипции.

Прояснить некоторые аспекты транскрипционной регуляции могут паралоги и псевдогены, образовавшиеся в результате дубликации генов. Существует паралог *FOLH1B*, он отличается от *FOLH1* тем, что в нем отсутствуют первый экзон и 371 п.о. первого интрона, а также есть небольшое количество однонуклеотидных замен в экзонах [88]. Интронные участки обоих генов с энхансером *FOLH1* и его паралога идентичны на 99,3%, и энхансеры обоих генов способны обеспечивать специфичную экспрессию репортера в конструкции с промотором *FOLH1* [88]. Однако паттерны экспрессии *FOLH1* и *FOLH1B* по тканям не совпадают: если сильнейшая экспрессия *FOLH1* показана в клеточной линии рака простаты LNCaP, а далее следуют ткани мозга, то экспрессия *FOLH1B* наблюдается в тканях печени и почек [88]. Это не полностью согласуется с данными, полученными на конструкциях энхансера *FOLH1* с гетерологичными промоторами, где благодаря энхансеру достигалась специфичность экспрессии репортера в клетках рака простаты [77].

## Транскрипционные факторы – регуляторы транскрипции *FOLH1*

Известно несколько регуляторов транскрипции *FOLH1* – андрогеновый рецептор и некоторые его сплайс-изоформы, SOX7, NFATc1, TMPRSS2-ERG. В данном разделе описана связь между этими транскрипционными факторами и экспрессией *FOLH1*.

В раннем исследовании показано, что инкубация клеток линии LNCaP с дигидротестостероном уменьшает экспрессию *FOLH1* в 8-10 раз, а с тестостероном – в 3-4 раза. В другой работе демонстрируется куда менее значительное влияние дигидротестостерона на уровень мРНК в клетках LNCaP дикого типа: максимальное двукратное уменьшение количества мРНК *FOLH1* было обнаружено при максимальной протестированной концентрации дигидротестостерона (10 нМ)[85]. Инкубация клеток LNCaP, трансфецированных репортерной конструкцией под управлением 2 т.п.о. промотора *FOLH1*, с дигидротестостероном показала отсутствие влияния андрогена в концентрациях 10 мкМ и 100 мкМ [82]. Это несколько не соответствует другому исследованию, в котором для репортерных конструкций с 2 т.п.о. и 5,5 т.п.о. промоторами уменьшение сигнала репортера при добавлении в среду 10 нМ дигидротестостерона показано для 2 т.п.о. промотора и составило 21% [85]. Слабым оказалось влияние дигидротестостерона на экспрессию репортера в клетках LNCaP при транзientной и постоянной трансфекции конструкции с 2 т.п.о. промотором и энхансером: 42% (2 нМ дигидротестостерона, p-value = 0,0054) и 45% (10 нМ дигидротестостерона, p-value = 0,0037) соответственно[85]. В схожем исследовании показан куда больший эффект, но от синтетического ингибитора андрогенового рецептора, R1881[89], и разница в экспрессии репортера между образцами, обработанными и не обработанными R1881, была значительной (до ~ 20 раз для конструкции с энхансером без 65 п.о. на 3'-конце)[86]. Несмотря на расходящиеся данные, можно сделать вывод о важности энхансера в обеспечении андроген-зависимости экспрессии *FOLH1*, и косвенным подтверждением может являться тот факт, что в промоторном участке длиной 1244 п.о. сайты ответа на андрогены отсутствуют[81], но присутствуют в энхансере[84], в прямом повторе[86].

Существует несколько изоформ AR, образующихся в результате мутаций в гене AR[90], протеолитического процессинга AR[91] или в ходе альтернативного сплайсинга[92], [93]. При этом изоформы нуждаются в полноразмерном AR для регуляции экспрессии их мишеней и не могут действовать независимо[94]. Экспрессия AR-V1 и AR-V7 изоформ AR увеличивается при переходе от нормальной ткани простаты и андроген-чувствительной карциномы простаты к андроген-нечувствительной карциноме простаты, позволяя карциноме простаты расти при андрогенной депривации[95]. На основании данных секвенирования РНК (RNA-seq) клеток с экспрессией изоформ AR<sup>v567es</sup> и AR-V7 показано наличие уникального действия этих изоформ на экспрессию некоторых генов, в том числе маркера рака простаты KLK3, по сравнению с

полноразмерным AR[96], но для *FOLH1* разница отсутствует. Однако ингибитор-независимый С-концевой фрагмент AR, Q640X AR, оказался способен увеличивать экспрессию *FOLH1* в клетках LNCaP: при оверэкспрессии Q640X AR наблюдалось восстановление экспрессии *FOLH1* до уровня контроля по сравнению с оверэкспрессией AR дикого типа[97]. Т.е. эффект 640X AR и AR дикого типа разнонаправленный. Экспрессия изоформ AR не является специфичной для клеточных линий рака простаты LNCaP и VCaP и тканей простаты [92], что делает маловероятной специфичность зависимости экспрессии *FOLH1* от изоформ AR в клетках простаты и карциномы простаты.

Экспрессия AR и активность андрогенового сигнального пути показаны в разных тканях и видах опухолей [98], [99]. В опухоли простаты повышенная экспрессия *FOLH1* приходится на андроген-нечувствительные опухоли, однако в клеточных линиях опухоли простаты ситуация обратная: наибольшая экспрессия *FOLH1* показана в андроген-чувствительной клеточной линии LNCaP, наименьшая – в андроген-нечувствительной линии PC-3. В связи с этим едва ли можно говорить об определяющей роли AR и андрогенового сигнального пути в экспрессии *FOLH1*, чья экспрессия сильно специфична для простаты и карциномы простаты.

Влияние на экспрессию *FOLH1* оказывают изоформы химерного белка TMPRSS2-ERG, чья экспрессия в TMPRSS2-ERG-содержащих клеточных линиях, в свою очередь, андроген-зависима[100]. Добавление синтетического ингибитора AR, R1881, к VCaP увеличивает экспрессию TMPRSS2-ERG и значительно уменьшает экспрессию *FOLH1*[100]. Нокаунт ERG в VCaP увеличивает экспрессию *FOLH1* в 3 раза[100]. Экзогенная экспрессия разных изоформ TMPRSS2-ERG в LNCaP приводит к уменьшению экспрессии *FOLH1* до двух раз; добавление синтетического ингибитора AR, R1881, еще немного уменьшает экспрессию *FOLH1*[100]. Эффект в клетках LNCaP, дополнительно экспрессирующих одну из изоформ TMPRSS2-ERG, сильнее выражен вне исходного контекста – для конструкций с люциферазой, экспрессия которой регулируется промотором *FOLH1*, где присутствует сайт связывания ETS, с которым, по данным иммунопреципитации хроматина, взаимодействует ERG[100]. Влияние TMPRSS2-ERG на экспрессию *FOLH1* не является ключевым для обеспечения тканеспецифичности. Анализ образцов карциномы простаты, нормальной простаты и клеточных линий продемонстрировал: а) наличие TMPRSS2-ERG в опухолевых образцах[101], в то время как *FOLH1* имеет высокую экспрессию как в норме, так и в опухоли; б) отсутствие этой транслокации в клеточной линии LNCaP [100], [102], которая характеризуется значительной экспрессией *FOLH1*; в) отсутствие дифференциальной экспрессии *FOLH1* в зависимости от экспрессии TMPRSS2-ERG в опухолевых образцах[101].

Транскрипционные факторы, как правило, имеют несколько регулируемых генов, в том числе и гены других ТФ, и образуют транскрипционные сети с нестрогой иерархической структурой. Регуляторный каскад андрогенового рецептора необходим для нормального роста и развития простаты, а также для развития и роста опухоли простаты[103]. Мишенями транскрипционной регуляции андрогенового рецептора являются тысячи генов[104]. Несмотря на то, что сам AR экспрессируется во многих тканях (данные The Human Protein Atlas), специфичность выбора мишеней в простате достигается благодаря тканеспецифичности экспрессии ТФ-партнеров AR, таких как FOXA1[104], GATA2[104], [105], GRHL2[106], CREB3L4[107], HOXB13[108] и других.

Помимо андрогенового рецептора, энхансер *FOLH1* также содержит участки связывания, предсказанные для других ТФ. Анализ участка 35...69 п.о. в 90 п.о. фрагменте на 5'-конце энхансера *FOLH1* показал наличие сайтов связывания AP-1 и AP-3 (Adaptor Protein 3)[87]. Блокирующие связывание соответствующих регуляторов мутации показали отсутствие изменения экспрессии репортера для AP-1 и его наличие для AP-3[87]. Замена 90 п.о. участка на последовательность связывания AP-3 привела к увеличению экспрессии репортера примерно в 1,5 раза по сравнению с энхансером дикого типа; подстановка той же последовательности непосредственно к промотору не давала эффекта[87].

Взаимодействие транскрипционного фактора Nuclear Factor Of Activated T Cells 1 (NFATc1) с участком 35...69[87], вероятно, в сайтах AP-3, показано посредством иммунопреципитации хроматина. Суперэкспрессия NFATc1 приводила к увеличению экспрессии репортера в 2,5 раза, а нокдаун – к уменьшению до 50%[87]. К сожалению, нет информации о достоверности полученных данных ввиду отсутствия информации о репликах в описании экспериментов по суперэкспрессии и нокдауну. Активность NFATc1 зависит от внутриклеточной концентрации кальция: кальций-зависимое фосфорилирование остатков серина в мотиве SP на N-конце NFAT белков позволяет релоцировать их из цитоплазмы в ядро. Добавление ионофора иономицина, усиливающего транспорт ионов кальция в цитоплазму, и TPA (12-O-tetradecanoyl-phorbol 13-acetate), активатора протеинкиназы C, позволило увеличить сигнал репортера почти в 3 раза для конструкции, состоящей из AP-3 сайта и энхансера с делетированными 90 п.о. на 3'-конце[87]. Так как TPA необходим для активации NFAT белков посредством AP-1, то наличие сайтов связывания AP-1 в энхансере указывает на возможность кооперации NFATc1 и AP-1 и в LNCaP, что обеспечивает высокую активность энхансера[87]. К сожалению, в данном эксперименте не было контроля в виде фрагмента энхансера с удаленным участком связывания NFATc1, то есть нельзя утверждать, что именно NFATc1 повлиял на экспрессию репортера. Неясно, насколько полученные данные соответствуют тому, что происходит *in vitro*: при использовании химерных конструкций теряется исходный

хромосомный контекст, а показанные эффекты от суперэкспрессии и нокдауна ТФ незначительны по сравнению с разницей, наблюдаемой для *FOLH1* в простате или в карциноме простаты и в других тканях. Более того, паттерны экспрессии описанных факторов не совпадают с паттерном экспрессии *FOLH1*: NFATc1 не обладает выраженной тканеспецифичностью, не обладают ею и возможные компоненты AP-1 и AP-3 комплексов (согласно The Human Protein Atlas[99]). Это позволяет не рассматривать их в качестве ключевых тканеспецифичных регуляторов.

Было показано, что SOX7 может негативно влиять на экспрессию *FOLH1* и репортера, регулируя активность энхансера в составе химерных конструкций[109]. Транскрипция SOX7 имеет отрицательную корреляцию с *FOLH1*, а белок SOX7 обнаруживается в *FOLH1*-отрицательной линии PC-3, но отсутствует в *FOLH1*-экспрессирующих линиях LNCaP и 22Rv1[109]. Суперэкспрессия SOX7 в LNCaP приводит к двукратному ослаблению транскрипции *FOLH1*[109]. Коиммунопреципитация хроматина с SOX7 показала взаимодействие SOX7 с энхансерной последовательностью *FOLH1*[109], т.е. данное взаимодействие возможно. Однако без полногеномных данных, например CHIP-seq, сложно судить о вероятности этого взаимодействия. Паттерн экспрессии SOX7 в нормальных тканях (по данным GTEx portal <https://gtexportal.org>), опухолях (по данным ICGC Data Portal[110]) и клеточных линиях (по данным Cosmic Cell Line [111]) не соотносится с экспрессией *FOLH1*, что исключает тканеспецифичность влияния SOX7.

В одной из работ было предположено, что LyF-1 может участвовать в инициации транскрипции *FOLH1*, связываясь вблизи сайта старта транскрипции, однако гипотеза не подтвердилась[83].

Также выдвигалась идея, что одним из факторов, которые могут усиливать транскрипцию *FOLH1*, является уменьшение экспрессии микроРНК hsa-miR-101, -138, -186, -224, -26a, -26b, -374a, -410, -660 [112], однако экспериментальное подтверждение этого не было проведено.

Найденные регуляторы транскрипции *FOLH1* были обнаружены в ходе экспериментов вне клеточного контекста на химерных конструкциях и все вместе способны объяснить лишь малую часть повышенной экспрессии *FOLH1*. Паттерны их экспрессии в зависимости от предполагаемой роли этих транскрипционных факторов (активаторы/репрессоры) не соответствуют паттерну экспрессии *FOLH1* в тканях в норме и в опухоли по данным The Human Protein Atlas[92], GEPIA2[106]. В связи с этим необходимо провести поиск потенциальных регуляторов экспрессии *FOLH1* и их экспериментальную проверку.

# Экспериментальная часть

## Материалы и методы

### Реактивы

**Химмед:** хлороформ.

**Химстройснаб:** 2-пропанол ХЧ

**Sigma:** ампициллин, бромистый этидий, формальдегид.

**Helicon:** агар, агароза для электрофореза, борная кислота, дрожжевой экстракт, Triton X-100, изопропил-1-тио-β-D-галактопиранозид (ИПТГ), ЭДТА.

**Amresco:** ДМСО, среда LB, трис(оксиметил)аминометан (Tris), таблетки PBS 100 и 500 мл, бычий сывороточный альбумин

**Ферейн:** этанол

**Thermo Scientific:** генетицин G418, РНКаза А, эндонуклеазы рестрикции, GlycoBlue, смесь для ПЦР (PCR Master Mix), Maxima First Strand cDNA Synthesis Kit, Maxima Reverse Transcriptase, Т4 ДНК-лигаза, Т4 полинуклеотидкиназа, Trizol, липофектамин 3000, липофектамин RNAiMAX, Glycogen, hoeschst для окрашивания ядер, amplex red glutamic acid/glutamate oxidase assay kit (набор для теста ингибирования ферментативной активности GСPII), Phusion ДНК-полимераза, среда Opti-MEM™ для трансфекции, 0,5% трипсин-ЭДТА (10х), полибрэн, Днказа I,

**Calbiochem:** пурамицин дигидрохлорид

**NEB:** Q5 ДНК-полимераза.

**Евроген:** набор реагентов для выделения ДНК из агарозных гелей и реакционных смесей Cleanup mini, набор реагентов для выделения плазмид Plasmid miniprep, MMLV RT kit для обратной транскрипции.

**Lumiprobe:** олигодезоксирибонуклеотиды, LumiPure genome DNA extraction kit

**Roche:** Complete EDTA-free protease inhibitor cocktail.

**Gibco:** среда DMEM/F12, среда RPMI-1640, эмбриональная телячья сыворотка FBS, аланил-глутамин GlutaMAX (100х), смесь антибиотиков пенициллина (10000 ед/мл) и стрептомицина (10000мкг/мл).\

**MP:** Mineral Oil для ПЦР в реальном времени.

**ПанЭко:** среда RPMI-1640 с глутамином, поли-L-лизин, витамины для RPMI-1640 (100х).

### Клеточные линии и среды

**LNCaP.** Клеточная линия, созданная из образцов метастазов лимфатических узлов больных раком предстательной железы. Эта клеточная линия сохраняет цитологические

характеристики опухоли рака простаты, является андроген-зависимой. Имеют высокую экспрессию *FOLH1*. Эти клетки культивировались в среде RPMI-1640, 10% FBS, 100 U/mL Penicillin-Streptomycin, 1x GlutaMAX™ Supplement (100x).

**22Rv1.** Клеточная линия, полученная из ксенотрансплантата рака простаты человека CWR22R. Одновременно андроген-независимая линия и андроген-зависимая линия. Имеют невысокую экспрессию *FOLH1*. Эти клетки культивировались в среде RPMI-1640, 10% FBS, 100 U/mL Penicillin-Streptomycin, 1x GlutaMAX™ Supplement (100x).

**PC-3.** Клеточная линия, выделенная из костных метастазов рака простаты человека с низкой степенью дифференцировки. Это андроген-независимые клетки рака простаты, обладающие умеренным метастатическим потенциалом, но не имеющие эндогенных AR. Экспрессия *FOLH1* в клетках линии отсутствует. Эти клетки культивировались в среде DMEM/F-12, 10% FBS, 100 U/mL Penicillin-Streptomycin, 1x GlutaMAX™ Supplement (100x).

**HEK293T.** Клеточная линия, использованная для сборки лентивирусов с библиотекой нокаутов Gecko v2. Эти клетки культивировались в среде DMEM/F-12, 10% FBS, 100 U/mL Penicillin-Streptomycin, 1x GlutaMAX™ Supplement.

**VA-13, MCF-7, A549.** Клеточные линии непростатного происхождения, использованные для МТТ теста при оценке цитотоксичности лигандов GСPII. **VA-13** – иммортализованная клеточная линия фибробластов, выделенных из легкого. **MCF-7** – эпителиальная клеточная линия, выделенная из ткани молочной железы пациента, больного метастатической аденокарциномой. **A549** – альвеолярные базальные эпителиальные клетки, выделенные из легочной ткани больного раком легких. Все эти клетки культивировались в среде DMEM/F-12, 10% FBS, 100 U/mL Penicillin-Streptomycin, 1x GlutaMAX™ Supplement (100x).

Клетки культивировали CO<sub>2</sub>-инкубаторе (Sanyo) при температуре 37 °С и содержании CO<sub>2</sub> 5% во флаконах площадью 25 см<sup>2</sup> в 5 мл питательной среды. Клетки пересеивали от 1:4 до 1:10 каждые 3-4 дня по достижении 80-95% конfluence. Для посева убрали среду, промывали клетки 5 мл стерильного раствора PBS, инкубировали 5 минут при 37 °С в 5 мл раствора трипсина с ЭДТА, инактивировали полной питательной средой с 10%FBS затем разводили в полной среде до необходимой плотности клеток.

### МТТ тест

Для оценки цитотоксичности лигандов GСPII был использован колориметрический тест на панели клеточных линий различной этиологии[113]. МТТ – желтый тетразол, который восстанавливается в пурпурный формазан в живых клетках. Осадок формазана растворяется в ДМСО. Для предотвращения перероста клеток из-за разницы в скорости роста клеточных линий

сеяли клетки из расчета 2000 клеток РС-3, 3000 клеток 22Rv1 и LNCaP в 120 мкл среды в ячейку 96-луночной плашки (с обработкой для выращивания клеточных культур), инкубировали при 37°C с 5% CO<sub>2</sub> первые 16 часов без добавления препаратов. На следующий день добавили серию разведений исследуемых препаратов и серию разведений доксорубина (контроль) по 10 мкл, максимальная концентрация – 100 мкМ, минимальная – 45 нМ, шаг разведения был равен трем. Поместили на 72 часа в 37°C 5% CO<sub>2</sub>-инкубатор. На пятый день добавили по 10 мкл МТТ (5 мг/мл) в каждую ячейку, инкубировали 2-4 часа в CO<sub>2</sub>-инкубаторе. Заменяли среду с МТТ на 120 мкл 99% DMSO, перемешали на качалке 10-15 мин. Измерили оптическую плотность на длинах волн 555 и 670 нм с помощью планшетного спектрофотометра Victor. Обработку данных, включая расчет СС<sub>50</sub>, выполнили в Graphpad Prism 9.

### Измерение активности GСPII

Клетки промыли PBS, добавили Trypsin/EDTA и убрали в CO<sub>2</sub>-инкубатор на несколько минут. В зависимости от количества проверяемых ингибиторов GСPII, нужное количество клеток перенесли в пробирку (по 200.000 клеток на препарат + 200.000 клеток на контроли), центрифугировали 5 мин на 300g, дважды промыли PBS, каждый раз центрифугируя 5 мин на 300g. Прилили 100 мкл лизирующего буфера на 200.000 клеток (0,5% Triton X-100, 50 mM Tris-HCl (pH 7,5), 1x Proteinase Inhibition cocktail), обработали ультразвуковым зондом 7 раз по 7 сек с интервалами 20-30 секунд во льду во избежание перегрева. Центрифугировали 10 мин 1000g на 4°C.

Приготовили NAAG-смесь (в черной 384-луночной плашке): 4 мкл экстракта, 3 мкл 50 mM NAAG, 3 мкл разведенного препарата (100; 10; 2; 0,5; 0,125; 0,031; 0,08; 0,002 мкМ в 0,5% Triton X-100, 50 mM Tris-HCl (pH 7,5)).

При измерении активности GСPII использовались контроли двух видов: а) GСPII с субстратом NAAG без ингибиторов для определения максимального уровня сигнала – 4 мкл экстракта, 3 мкл 50 mM NAAG, 3 мкл буфера для разведения; б) GСPII без субстрата и без ингибиторов для определения минимального уровня сигнала – 4 мкл экстракта, 3 мкл mQ, 3 мкл буфера для разведения. Инкубировали 2 часа при 37°C.

Для визуализации глутамата, образующегося при гидролизе NAAG ферментом GСPII, использовали Amplex® Red Glutamic Acid/Glutamate Oxidase Assay Kit. Схема детекции ингибирования ферментативной активности GСPII представлена в приложении 6. Рабочий раствор Amplex Red (50 реакций): 5 мкл стокового раствора Amplex® Red, 1,25 мкл пероксидазы хрена HRP, 8 мкл L-глутамат оксидаза, 2,5 мкл L-глутамат-пируват трансминаза, 0,5 мкл L-alanine, 483 мкл 1X Reaction Buffer. Добавили 10 мкл рабочего раствора Amplex Red к 10 мкл NAAG-смеси. Инкубировали 30 мин при 37°C в фольге. Детекцию проводили на

планшетном спектрофотометре VICTOR Multilabel Plate Reader (длины волн: возбуждение – 570 нм, эмиссия – 585нм). Для каждого препарата сделано не менее трех биологических повтора теста ингибирования активности GСPII, в каждом эксперименте – по два технических повтора.

Значения IC<sub>50</sub> определены с помощью Graphpad Prism 9. K<sub>i</sub> посчитаны по уравнению Ченга-Пруссоффа:  $K_i = IC_{50}/(1+[S]/K_m)$ , где [S] – концентрация субстрата NAAG (10 мкМ), K<sub>m</sub> - константа Михаэлиса, равная 1,15 мкМ для фермента дикого типа[114].

### **Проточная цитометрия**

Промыли клетки LNCaP, 22Rv1 и PC-3 PBS, обработали Trypsin-EDTA. Для каждого образца брали по 200 тысяч клеток. Промыли дважды PBS (500g, 5 минут), добавили 0, 0.4, 4, 40, 100, 400, 1000 нМ тестируемого конъюгата (B15-Cy5) в 5% BSA в PBS. Инкубировали 1 час на +4°C во вращающемся штативе в фольге. Промыли дважды 5% BSA в PBS (500g, 5 минут). Развели клетки в 200 мкл 5% BSA в PBS. Для проточной цитофлуориметрии использовали проточный цитофлуориметр MACSQuant. Условия съемки подбирались так, чтобы на графике FFC-SSC (бокового-прямого светорассеивания) детектируемые события не концентрировались рядом с осями. На гистограмме распределения флуоресцентного сигнала от B15-Cy5 в биэкспоненциальной шкале сигнал от контроля без лиганда GСPII должен был находиться вокруг 0, а пик сигнала от максимальной концентрации не должен был выходить за границы измеряемой интенсивности. Подобранные значения напряжения на лазерах: FSC-A – 257, SSC-A – 400, APC-A – 480. Обработка и анализ данных выполнены в FlowJo X.

### **Обработка поверхности культурального пластика полилизинном**

Раствор полилизина в концентрации 50 мкг/мл добавили в ячейки планшета из расчета 40 мкл на 0,32 см<sup>2</sup>. Инкубировали 40 минут при комнатной температуре. Отобрали раствор, промыли ячейки PBS, высушили ячейки в стерильном ламинаре и обработали ультрафиолетом в течение часа.

### **Флуоресцентная микроскопия**

Клетки LNCaP и PC-3 (10000 клеток на лунку) были посеяны в 500 мкл среды на покровные стекла, обработанные поли-L-лизинном, в 24-луночные планшеты. На следующий день клетки промыли 3 раза PBS, инкубировали с исследуемыми соединениями в 5% BSA в PBS в течение 2 часов в темноте в 5% CO<sub>2</sub>, 37°C. Затем добавляли Hoechst 33342 для окрашивания ядер в концентрации, указанной производителем, инкубировали (5% CO<sub>2</sub>, 37°C) 10 минут. Покровные стекла промывали 3 раза PBS, фиксировали 4% раствором параформальдегида (разведенным в PBS), промывали 2 раза PBS, покрывали Mowiol, переносили на предметные стекла (клетками «вниз») и инкубировали в темноте при 4°C в течение ночи, чтобы Mowiol

застыл. Флуоресцентная микроскопия проводилась на микроскопе Nikon C2, производилась съемка z-стэков в двух каналах для определения характера сигнала внутри клеток. В первом канале регистрировался флуоресцентный сигнал от Neochst, во втором – от Cy5. Краситель ядер Neochst возбуждается ультрафиолетовым лазером, диапазон длин волн эмиссии – 460-490 нМ. Cy5: возбуждение – 650 нм, флуоресценция – 670 нм. Подготовка изображений для анализа выполнена в программе NIS Viewer.

### **Создание плазмид для постоянной экспрессии ТФ**

Векторы для постоянной экспрессии мРНК *CREB3L4*, *FOXA1*, *GRHL2* сделаны на основе плазмид pSBtet-Neo и pSBbi-GN. Это плазмиды транспозоновой системы Sleeping Beauty, позволяющие встроить транспозон с последовательностью интереса из плазмиды в случайное место в геноме. Отличаются наличием TRE промотора в pSBtet-Neo, индуцируемого 0,1 мкг/мл доксицилина, и конститутивного EF-1 $\alpha$  промотора в pSBbi-GN. Для синтеза комплементарной ДНК для мРНК выбранных транскрипционных факторов были использованы тотальная РНК LNCaP со специфичными праймерами для обратной транскрипции (последовательности праймеров даны в Приложении 1) и Maxima Reverse Transcriptase (200 U/ $\mu$ L, Thermo Fisher, EP0742). Синтез кДНК по протоколу производителя. Полученная кДНК ТФ амплифицирована праймерами с сайтами рестрикции Sfi I на концах (последовательности праймеров даны в Приложении 1), по которым осуществлялось клонирование в pSBtet-Neo или pSBbi-GN. Первоначально все ТФ были клонированы в pSBbi-GN (Приложение 7А), однако после трансфекции и отбора клеточных линий LNCaP, 22Rv1, PC-3 со вставкой суперэкспрессия *FOXA1* и *GRHL2* оказалась летальной для большинства клеточных линий. Для суперэкспрессии эти ТФ были переклонированы в pSBtet-Neo (Приложение 7В).

### **Подготовка библиотека лентивирусов для нокаутов Gecko v2 и трансдукция**

Лентивирусная библиотека GeCKO v2 (Genome-Scale CRISPR Knock-Out) состоит из более 100 000 уникальных sgRNA для нокаута генов в геноме человека. Библиотека представляет собой две полубиблиотеки (А и В). Обе библиотеки А и В содержат 1000 контрольных sgRNA, не имеющих мишени в геноме. Полубиблиотека А также содержит sgRNA на miRNA (4 sgRNA на miRNA).

Для подбора минимальной концентрации пурамицина, при которой гибнут все клетки LNCaP-Cas9 (для использования этой концентрации при отборе трансдуцированных клеток) посеяли по 5 тысяч клеток в ячейку 96-луночного планшета в среде RPMI без Penicillin-Streptomycin, на следующий день добавили серию разведений пурамицина (0 - 3 мкг/мл), меняли среду с антибиотиком каждые 3 дня в течение 1-2 недель до тех пор, пока клетки не

начнут погибать. После этого определяется минимальная эффективная концентрация антибиотика. Для LNCaP она составила 0,4 мкг/мл.

Для наработки лентивирусов использовалась линия НЕК293Т. Клетки НЕК293Т пересевали каждые 2 дня, не давая достичь более 70% конfluenceности. Для трансфекции перенесли по 18 млн НЕК293Т в 4 флакона Т175 (флакон культуральный, 175 см<sup>2</sup>, стерильный, Greiner bio-one, 660175) в 30 мл среды. На следующий день при достижении 70% конfluenceности провели липотрансфекцию. За час до трансфекции сменили среду на 13 мл Opti-MEM. Далее все расчеты на 1 флакон. Подготовили смеси 1) 33,3 мкл Plus Reagent в 666 мкл Opti-MEM, 3,33 мкг LentiCRISPR(A или B), 1,66 мкг pCMV pVSVg, 2,5 мкг psPAX2 и 2) 16,6 мкл Lipofectamine 3000 в 666 мкл Opti-MEM. Смешали 1 и 2, инкубировали 20 мин на комнатной температуре, прилили к клеткам. Инкубировали 8-10 часов в CO<sub>2</sub>-инкубаторе, сменили среду на 30 мл свежей, инкубировали при 37°C, 5% CO<sub>2</sub>. Через 2 дня после трансфекции среду с лентивирусами отфильтровали с помощью Millipore 0.45 мкм Stericup фильтра. Центрифугировали 1 час 100,000 g при температуре +4°C. Удалили супернатант, ресуспендировали в RPMI-1640, делили на аликвоты и хранили на -80°C.

Для измерения титра лентивирусов на каждую полубибблиотеку (А или В) посеяли 6 ячеек 12-луночного планшета по 1 млн клеток LNCaP-cas9 в 2 мл среды. На следующий день добавили Polybrene 8 мкг/мл. В каждую ячейку добавили 400/200/100/50/25/0 мкл среды с лентивирусами, центрифугировали 2 часа 1000 g на +33°C, убрали в CO<sub>2</sub> инкубатор. Через 24h определили концентрацию клеток в ячейке без лентивирусов. Для каждой концентрации лентивируса высеяли снятые клетки в 4 ячейки по 4000 клеток на ячейку 96-луночного планшета в 100 мкл RPMI, считая, что везде концентрация клеток такая же, как и в ячейке без лентивирусов, измеренная в предыдущем шаге. Добавили еще по 100 мкл среды с пуромицином в концентрации 0,4 мкг/мл для LNCaP в 2 ячейки, в остальные 2 - среду без пуромицина. После гибели клеток в ячейках без пуромицина определили долю выживших клеток во всех остальных ячейках. Посчитали MOI (multiplicity of infection, множественность заражения). Для трансдукции LNCaP-cas9 использовали MOI от 0,2 до 0,3.

Трансдукция и селекция LNCaP-Cas9 проводилась трижды для каждой из полубибблиотек. Посеяли LNCaP по 300 тысяч клеток в два 6-луночных планшета в 1 мл среды RPMI без антибиотиков. На следующий день при достижении 70% монослоя добавили 8 мкг/мл Polybrene, лентивирусы в количестве, необходимом для достижения MOI 0,2-0,3 во все ячейки, кроме контрольной, центрифугировали 2 часа 1000 g на +33°C, убрали в CO<sub>2</sub> инкубатор. Через 8-12 часов промыли клетки PBS, сняли Trypsin-EDTA, посадили в два флакона Т175 в 15 мл RPMI без антибиотиков. Контроль без лентивирусов оставили в 6-луночном планшете. На следующий

день добавили пуромицин во флаконы и в контроль до конечной концентрации 0,4 мкг/мл. Меняли среду и добавляли пуромицин каждые 4 дня, при рассеивая клетки при достижении 70% конфлюентности. После гибели клеток в контроле без лентивирусов сортировали трансдуцированные клетки.

### **Отбор трансдуцированных Gecko v2 клеток LNCaP-cas9**

Для сортировки использовано 30 млн клеток, столько же – в качестве контроля репрезентативности библиотеки. При достижении 80% конфлюентности промыли клетки PBS, обработали Trypsin-EDTA, посчитали (около 60 млн клеток). Клетки контроля промыли PBS дважды, заморозили на  $-80^{\circ}\text{C}$ . Оставшиеся клетки окрасили флуоресцентно меченым конъюгатом B15-Cy5.

Для окрашивания клетки промыли дважды PBS (500 rpm 5 min), добавили 100 нМ B15-Cy5 в 5% BSA в PBS. Инкубировали 1 час на  $+4^{\circ}\text{C}$  в темном месте, перемешивая. Промыли дважды 5% BSA в PBS. Ресуспендировали клетки в 5% BSA в PBS. Сортировали на клеточном сортере: отбирали 0,2% клеток с самым сильным сигналом от конъюгата B15-Cy5 и 0,2% клеток с самым слабым сигналом. После сортировки клетки высаживали в 48-луночный планшет, ячейки которого обработаны полилизин. Растили обе фракции до 30 млн клеток, проводили повторное окрашивание всех клеток B15-Cy5, отбор 0,2% клеток с самым сильным сигналом от конъюгата B15-Cy5 и 0,2% клеток с самым слабым сигналом. После сортировки клетки высаживали в 48-луночный планшет, ячейки которого обработаны полилизин. Растили до 10 млн клеток, промывали PBS, снимали клетки Trypsin-EDTA, промывали дважды PBS и хранили на  $-80^{\circ}\text{C}$ .

### **Подготовка ампликонов, содержащих последовательности sgRNA, к секвенированию**

Геномная ДНК выделялась набором LumiPure genome DNA extraction kit. Было использовано 5 млн дважды сортированных клеток и 15 млн клеток, содержащих полную библиотеку нокаутов (3 млн клеток на колонку). Ампликоны амплифицировались с помощью ПЦР с праймерами на фланкирующие участки sgRNA, содержащими баркоды (последовательности праймеров приведены в Приложении 1). Продукты ПЦР разделили в 2% агарозном геле, выделили ДНК из фрагментов геля, соответствующих 300 п.о. с помощью набора Monarch (New England Biolabs, T1020L), оценивали концентрацию ампликонов с помощью спектрофотометра QuBit по протоколу производителя. Все образцы смешивали в соотношениях, близких к соотношениям предполагаемых размеров библиотек (и необходимого числа ридов), чтобы обеспечить всем образцам достаточное покрытие: 4 млн ридов на дважды сортированные клетки, 8 млн ридов – на полную библиотеку. Секвенирование выполнено в

ООО «Геноаналитика». Полученные риды проанализированы с помощью MAGeCK tool [115] и Jupyter Notebook (python 3).

### **РНК интерференция**

Высеяли по 40 тыс. клеток в ячейку 24-луночный планшет в 500 мкл среды. На следующий день произвели трансфекцию siRNA (последовательности даны в Приложении 3) с помощью Lipofectamine RNAiMAX по протоколу производителя. Подобранные количества Lipofectamine RNAiMAX: для LNCaP – 2,25 мкл и 1 мкл, для PC-3 и 22Rv1 – 1,5 мкл и 0,5 мкл. На следующий день сменили среду. Через 48 часов после трансфекции промыли клетки PBS, добавили к ним 200 мкл QIAzol Reagent и выделили тотальную РНК по протоколу производителя. Полученную РНК обработали DNase I (Thermo Fisher, EN0521) по протоколу производителя, выполнили синтез кДНК (Maxima First Strand cDNA synthesis kit, Thermo Fisher, EN0521) по протоколу производителя и ПЦР в реальном времени для мРНК *FOLH1*, регуляторов, генов домашнего хозяйства *GAPDH* и *U2* (Thermo Scientific Maxima SYBR Green qPCR Master Mix, +ROX Solution, Thermo Fisher, K0253) в 384-луночном планшете. С помощью BioRad CFX и по анализу Cq определили относительные уровни экспрессии генов предполагаемых регуляторов и *FOLH1*, проанализировали изменения экспрессии *FOLH1*, определили значимость изменений с помощью t-теста Стьюдента.

### **Создание клеточных линий с конститутивной суперэкспрессией ТФ**

Высеяли по 40 тыс. клеток в ячейку 24-луночный планшет в 500 мкл среды. На следующий день произвели трансфекцию плазмид для суперэкспрессии с помощью Lipofectamine 3000 по протоколу производителя. Подобранные количества Lipofectamine 3000 и P3000 Reagent: для LNCaP – 2,25 мкл и 1 мкл, для PC-3 и 22Rv1 – 1,5 мкл и 0,5 мкл. На следующий день сменили среды, добавили антибиотик G-418 в концентрации 600 мкг/мл (рекомендации производителя для нашего набора клеточных линий). Меняли среду с антибиотиком каждые 4 дня, пока клетки в контрольной ячейке не погибли. Нарастили клетки, сделали стоки. Промыли часть клеток PBS, добавили 200 мкл QIAzol Reagent и выделили тотальную РНК по протоколу производителя. Полученную РНК обработали DNase I (Thermo Fisher, EN0521), выполнили синтез кДНК (Maxima First Strand cDNA synthesis kit, Thermo Fisher, EN0521) по протоколу производителя и ПЦР в реальном времени для мРНК *FOLH1*, регуляторов, генов домашнего хозяйства *GAPDH* и *U2* (Thermo Scientific Maxima SYBR Green qPCR Master Mix, +ROX Solution, Thermo Fisher, K0253) в 384-луночном планшете. С помощью BioRad CFX и Microsoft Excel и метода Cq определили относительные уровни экспрессии генов регуляторов и *FOLH1*, проанализировали изменения экспрессии *FOLH1*, определили значимость изменений с помощью t-теста Стьюдента.

## Оценка совместного влияния суперэкспрессии или нокдауна ТФ и андрогена

К клеткам (на следующий день после трансфекции при смене среды в случае нокдауна) в среду добавили дигидротестостерон до конечной концентрации 10 нМ. Инкубировали сутки (37°C, 5% CO<sub>2</sub>). Промыли клетки PBS, добавили 200 мкл QIAzol Reagent и выделили тотальную РНК по протоколу производителя с добавлением 0,5 мкл гликогена при переносе водной фазы в новую пробирку. Полученную РНК обработали DNase I по протоколу производителя, выполнили синтез кДНК (Maxima First Strand cDNA synthesis kit) по протоколу производителя и ПЦР в реальном времени для мРНК *FOLH1*, регуляторов, генов домашнего хозяйства *GAPDH* и *U2* (Thermo Scientific Maxima SYBR Green qPCR Master Mix) в 384-луночной планшете. BioRad CFX и Microsoft Excel и метода Cq определили относительные уровни экспрессии генов регуляторов и *FOLH1*, проанализировали изменения экспрессии *FOLH1*, определили значимость изменений с помощью t-теста Стьюдента.

### Анализ транскриптомных данных для определения потенциальных регуляторов экспрессии гена *FOLH1*

Корреляции экспрессии генов с *FOLH1* определены на основании транскриптомных данных пациентов с раком простаты (International Cancer Genome Consortium, ICGC, PRAD) в условной норме (образцы ткани вокруг опухоли), опухоли и изменений при переходе от нормы к опухоли. Также в простате в норме (Genotype-Tissue Expression portal, GTEx) и в клеточных линиях (Cancer Cell Line Encyclopedia, CCLE, и Catalogue of Somatic Mutations in Cancer, COSMIC). Фильтровались гены с модулем корреляции более 0,5.

Был проведен анализ уровня экспрессии генов транскрипционных факторов (ТФ) и протеинкиназ в клеточных линиях опухоли простаты LNCaP, VCaP, 22Rv1, PC-3, DU-145 по данным базы данных COSMIC cell lines. В ней уровень экспрессии генов описан категориально (повышенная экспрессия, пониженная экспрессия, экспрессия в норме). Те гены, чей уровень экспрессии был повышен в *FOLH1*-экспрессирующих клетках LNCaP, VCaP, 22Rv1, но находился в норме или был понижен в линиях PC-3, DU-145, в которых *FOLH1* не экспрессируется, рассматривались как активаторы экспрессии *FOLH1*. Из получившегося списка были исключены гены, чья экспрессия различается в клетках LNCaP и VCaP, где экспрессия *FOLH1* одинакова. Т.е. экспрессия отобранных генов различна между *FOLH1*-экспрессирующими линиями опухоли простаты и линиями, где *FOLH1* не экспрессируется (Рис. 4).

Были подсчитаны корреляции для пар ТФ-ТФ, ТФ-протеинкиназа и протеинкиназа-протеинкиназа. Списки ТФ взяты с <http://humantfs.ccbr.utoronto.ca/>, протеинкиназ – с <http://kinase.com/>. Все данные секвенирования РНК (RNA-seq) опухолевых образцов пациентов

(USA-PRAD, ICGC, размер выборки - 375 образцов опухоли простаты) были отсортированы по экспрессии *FOLH1*. Из первой и последней сотни пациентов были сформированы выборки, *FOLH1+* и *FOLH1-*, в которых экспрессия *FOLH1* значительно и достоверно отличалась (t-значение = 12,72, p-value = 5,34E-27). По результатам подсчета корреляций были отобраны пары регуляторов, которые достоверно (p-value < 0,01) коррелировали друг с другом в *FOLH1+* выборке и в *FOLH1-* выборке, при этом экспрессия обоих факторов достоверно (p-value < 0,05) отличалась между *FOLH1+* и *FOLH1-* выборками. Аналогичный анализ проведен для транскриптомных данных всех клеточных линий из CCLE (t-значение = 19,36, p-value = 1,50E-47, источник данных отличается от COSMIC, данные об экспрессии генов представлены в виде tpm - transcript per million). Применение этих операций к данным простаты в норме (GTEx) не позволило выделить потенциальных регуляторов экспрессии *FOLH1*, т.к. она в норме во всех образцах отличается незначительно.

## Функции Python для анализа транскриптомных данных

Для проведения расчетов в Jupyter Notebook (python 3) использовались библиотеки pandas, numpy, gzip, matplotlib.pyplot, scipy.stats, os, itertools.

1. Анализ дифференциальной экспрессии и корреляций генов из объединенного списка транскрипционных факторов и протеинкиназ в выборках образцов опухоли простаты (ICGC), различающихся по уровню экспрессии гена *FOLH1*. На вход подаются: file – расположение транскриптомных данных проекта из ICGC, separator – разделитель в file, goi – ген интереса, по экспрессии которого разбиваются данные на выборки (у нас – *FOLH1*), list\_of\_regulators – список генов, среди которых ищутся пары (у нас – протеинкиназы и ТФ). В процессе работы скрипта будет выведена гистограмма распределения экспрессии goi, размер выборки. По запросу необходимо ввести количество образцов (оно одинаковое) в каждой из выборок – goi-положительной и goi-отрицательной. Функция возвращает таблицу с корреляциями экспрессий генов и таблицу с результатами теста Стьюдента.

```
def icgc_rnaseq_data_analyzer(file, separator, goi, list_of_regulators):
    with gzip.open(file) as inp:
        icgc_rna = pd.read_csv(inp, sep = separator)
        columns =
['icgc_donor_id', 'submitted_sample_id', 'gene_id', 'normalized_read_count']
        icgc_rna = icgc_rna[columns]
        icgc_rna_list = list(icgc_rna.values)
        tumor_donors = {}
        #оставляю только те данные, что относятся к опухоли
        icgc_rna_tumor_list = []
        for donor in icgc_rna_list:
            if 'TCGA' in donor[1]:
                if int(donor[1].split('-')[3][:2]) < 10:
                    icgc_rna_tumor_list.append(donor)
                    tumor_donors[donor[0]] = []
```

```

tumor_donors = list(tumor_donors.keys())
icgc_rna = icgc_rna.drop('submitted_sample_id', axis=1)
icgc_rna = icgc_rna.loc[icgc_rna['icgc_donor_id'].isin(tumor_donors)]
#сортирую по экспрессии гена интереса, выбираю n доноров (вводится вручную
при запросе) с высокой и низкой экспрессией
icgc_rna = icgc_rna.loc[icgc_rna['gene_id'] ==
goi].sort_values('normalized_read_count')
sample_size = len(tumor_donors)
data_plt = icgc_rna['normalized_read_count']
data_plt = data_plt.astype(float).sort_values()
plt.hist(data_plt, 500, facecolor='green')
plt.tick_params(labelrotation=90, labelsz = 15)
plt.show()
plt.tight_layout()
group_size = input('The sample size is ' + str(sample_size) + ' donors.
Please, select the group size for further analysis. ')
group_size = int(group_size)
donors_low = list(icgc_rna['icgc_donor_id'].values[0:group_size])
donors_high = list(icgc_rna['icgc_donor_id'].values[(sample_size-
group_size):-1])

#отбираю данные в зависимости от высокой/низкой экспрессии гена интереса
data_high = {}
data_low = {}
for i in icgc_rna_tumor_list:
    if i[0] in donors_high:
        if i[0] not in data_high.keys():
            data_high[i[0]] = {}
            data_high[i[0]][i[2]] = i[3]
    if i[0] in donors_low:
        if i[0] not in data_low.keys():
            data_low[i[0]] = {}
            data_low[i[0]][i[2]] = i[3]

#отбираю простата-специфичные ТФ/ЕЗ/киназы
prostate_reg = list_of_regulators
high_el = {}
low_el = {}
for key1 in data_high.keys():
    high_el[key1] = {}
    for key2 in data_high[key1].keys():
        if key2 in prostate_reg:
            high_el[key1][key2] = data_high[key1][key2]
data_high = pd.DataFrame.from_dict(high_el)
for key1 in data_low.keys():
    low_el[key1] = {}
    for key2 in data_low[key1].keys():
        if key2 in prostate_reg:
            low_el[key1][key2] = data_low[key1][key2]
data_low = pd.DataFrame.from_dict(low_el)

#Подсчет корреляций с p-value, * за порог 10e-2, **10e-3, ***10e-4
data_high = data_high.astype(float)
rho = data_high.T.corr()
pval = data_high.T.corr(method=lambda x, y: pearsonr(x, y)[1]) -
np.eye(*rho.shape)
p = pval.applymap(lambda x: ''.join(['*'] for t in [0.01,0.001,0.0001] if
x<=t))
high_pvalue = p.to_dict()
high_corr = rho.to_dict()
data_low = data_low.astype(float)
rho = data_low.T.corr()
pval = data_low.T.corr(method=lambda x, y: pearsonr(x, y)[1]) -
np.eye(*rho.shape)

```

```

    p = pval.applymap(lambda x: ''.join(['*' for t in [0.01,0.001,0.0001] if
x<=t]))
    low_pvalue = p.to_dict()
    low_corr = rho.to_dict()
    temp_table = [['gene1','gene2','high'+goi+'_corr', 'high_sign',
'low'+goi+'_corr', 'low_sign']]
    for key1 in prostate_reg:
        empty_key = {}
        for key2 in prostate_reg:
            if key1 != key2:
temp_table.append([key1,key2,high_corr.get(key1,empty_key).get(key2),high_pvalue.
get(key1,empty_key).get(key2),low_corr.get(key1,empty_key).get(key2),
low_pvalue.get(key1,empty_key).get(key2)])
        temp_table = pd.DataFrame(temp_table)
        temp_table.columns = temp_table.loc[0]
        temp_table = temp_table[1::]
        t_test_data = [['gene','statistic','p-value']]
        for i in list(data_high.index):
            calc = ttest_ind(data_high.T[i], data_low.T[i])
            t_test_data.append([i, calc[0],calc[1]])
    return temp_table, t_test_data

```

2. Анализ дифференциальной экспрессии и корреляций генов из объединенного списка транскрипционных факторов и протеинкиназ в выборках клеточных линий (CCLE), различающихся по уровню экспрессии гена *FOLH1*. На вход подаются: file – расположение транскриптомных данных проекта из ICGC, goi – ген интереса, по экспрессии которого разбиваются данные на выборки (у нас – *FOLH1*), list\_of\_regulators – список генов, среди которых ищутся пары (у нас – протеинкиназы и ТФ). В процессе работы скрипта будет выведена гистограмма распределения экспрессии goi, размер выборки. По запросу необходимо ввести количество образцов (оно одинаковое) в каждой из выборок – goi-положительной и goi-отрицательной. Функция возвращает таблицу с корреляциями экспрессий генов и таблицу с результатами теста Стьюдента.

```

def ccle_analyzer(file, goi, list_of_regulators):
    ccle_csv = pd.read_csv(file, sep=',')
    ccle_csv.index = ccle_csv['Unnamed: 0']
    ccle_csv = ccle_csv.drop('Unnamed: 0', axis = 1)
    old_ccle_columns = ccle_csv.columns
    new_ccle_columns = []
    for i in old_ccle_columns:
        new_ccle_columns.append(i.split(' ')[0])
    ccle_csv.columns = new_ccle_columns

    data_plt = ccle_csv[goi]
    data_plt = data_plt.astype(float).sort_values()
    plt.hist(data_plt, 100, facecolor='green')
    plt.tick_params(labelrotation=90, labelszize = 15)
    plt.show()
    plt.tight_layout()
    group_size = input('The sample size is ' + str(sample_size) + ' donors.
Please, select the group size for further analysis. ')
    group_size = int(group_size)
    cells_low = ccle_csv[goi].sort_values()[0:group_size].index
    cells_high = ccle_csv[goi].sort_values()[-group_size:].index
    cell_lines_list = list(ccle_csv.index)
    ccle_dict = ccle_csv.to_dict(orient = 'index')

```

```

list_of_genes = list_of_regulators
list_of_genes.append(goi)

#отбираю данные в зависимости от высокой/низкой экспрессии гена интереса
data_high = {}
data_low = {}
for key in ccle_dict.keys():
    if key in cells_high:
        if key not in data_high.keys():
            data_high[key] = {}
        for tf in ccle_dict[key].keys():
            data_high[key][tf] = float(ccle_dict[key][tf])
    if key in cells_low:
        if key not in data_low.keys():
            data_low[key] = {}
        for tf in ccle_dict[key].keys():
            data_low[key][tf] = float(ccle_dict[key][tf])

data_high = pd.DataFrame.from_dict(data_high)
data_low = pd.DataFrame.from_dict(data_low)
data_high = data_high.loc[data_high.index.isin(list_of_genes)]
data_low = data_low.loc[data_low.index.isin(list_of_genes)]

#Подсчет корреляций с p-value, * за порог 10e-2, **10e-3, ***10e-4
data_high = data_high.astype(float)
rho = data_high.T.corr()
pval = data_high.T.corr(method=lambda x, y: pearsonr(x, y)[1]) -
np.eye(*rho.shape)
p = pval.applymap(lambda x: ''.join(['*'] for t in [0.01,0.001,0.0001] if
x<=t))
high_pvalue = p.to_dict()
high_corr = rho.to_dict()

data_low = data_low.astype(float)
rho = data_low.T.corr()
pval = data_low.T.corr(method=lambda x, y: pearsonr(x, y)[1]) -
np.eye(*rho.shape)
p = pval.applymap(lambda x: ''.join(['*'] for t in [0.01,0.001,0.0001] if
x<=t))
low_pvalue = p.to_dict()
low_corr = rho.to_dict()

temp_table = [['gene1', 'gene2', 'high'+goi+'_corr', 'high_sign',
'low'+goi+'_corr', 'low_sign']]
for key1 in list_of_genes:
    empty_key = {}
    for key2 in list_of_genes:
        if key1 != key2:
temp_table.append([key1, key2, high_corr.get(key1, empty_key).get(key2), high_pvalue.
get(key1, empty_key).get(key2), low_corr.get(key1, empty_key).get(key2),
low_pvalue.get(key1, empty_key).get(key2)])
temp_table = pd.DataFrame(temp_table)
temp_table = pd.DataFrame(temp_table)
temp_table.columns = temp_table.loc[0]
temp_table = temp_table[1::]
temp_table = temp_table.loc[temp_table['high_sign']== '***']
temp_table = temp_table.loc[temp_table['low_sign']== '***']

t_test_ccle = [['gene', 'statistic', 'p-value']]
for i in list(data_high.index):
    calc = ttest_ind(data_high.T[i], data_low.T[i])
    t_test_ccle.append([i, calc[0], calc[1]])
t_test_ccle = pd.DataFrame(t_test_ccle[1:], columns = ['gene', 't-stat', 'p-
value'])

```

```

t_test_ccle = t_test_ccle.sort_values('t-stat')
t_test_ccle = t_test_ccle.loc[t_test_ccle['p-value'] <= 0.01]
t_test_ccle = t_test_ccle.values.tolist()
return temp_table, t_test_ccle

```

3. Определение значений коэффициентов корреляции для списка генов по данным экспрессии в норме, условной норме, опухоли простаты, при переходе от нормы к опухоли, в клеточных линиях, а также по данным синтетической летальности и литературе (STRING). На вход подаются: path\_\*(icgc,ge,ccle – пути к файлам с данными ICGC, DEPMAP, CCLE, string1 – protein info, string2 – protein links); cancer\_type - название заболевания как в проекте (PRAD – рак простаты), genelist – список генов, для которых необходимо определить коэффициенты корреляции. Результатом является таблица со столбцами [gene1, gene2, tumor\_corr, normal\_corr, fc\_corr, GeneEffect\_corr, CCLE\_corr, neighborhood, fusion, cooccurrence, coexpression, experimental, database, textmining.combined\_score], где подчеркнутые названия столбцов относятся к STRING.

```

def measure_all_correlations(path_icgc, path_, cancer_type, genelist):
    #selection of normal and tumor samples from ICGC according to
    https://docs.gdc.cancer.gov/Encyclopedia/pages/TCGA_Barcode/
    with gzip.open(path_icgc) as inp:
        tumor_samples = {}
        normal_samples = {}
        next(inp)
        for line in inp:
            line = line.decode().split('\t')
            if cancer_type in line[1]:
                sample_id = line[4].split('-')
                if 'TCGA' == sample_id[0]:
                    type_number = int(sample_id[3][:-1])
                    if type_number < 10:
                        if line[0] not in tumor_samples.keys():
                            tumor_samples[line[0]] = {}
                            tumor_samples[line[0]][line[7]] = line[8]
                    if type_number > 9 and type_number < 20:
                        if line[0] not in normal_samples.keys():
                            normal_samples[line[0]] = {}
                            normal_samples[line[0]][line[7]] = line[8]

tumor_samples = pd.DataFrame.from_dict(tumor_samples)
normal_samples = pd.DataFrame.from_dict(normal_samples)
tumor_samples = tumor_samples[normal_samples.columns]
tumor_samples = tumor_samples.astype(float)
normal_samples = normal_samples.astype(float)
fc_samples = tumor_samples/normal_samples
tumor_corr = tumor_samples.T.corr(method='pearson')
normal_corr = normal_samples.T.corr(method='pearson')
fc_corr = fc_samples.T.corr(method='pearson')

#Gene_Effect_correlation
ge_csv = pd.read_csv(path_ge, sep=',')
ge_csv.index = ge_csv['DepMap_ID']
ge_csv = ge_csv.drop('DepMap_ID', axis = 1)
old_ge_columns = ge_csv.columns
new_ge_columns = []
for i in old_ge_columns:
    new_ge_columns.append(i.split(' ')[0])
ge_csv.columns = new_ge_columns
ge_corr = ge_csv.corr(method='pearson')

#CCLE cell lines correlation
ccle_csv = pd.read_csv(path_ccle, sep=',')

```

```

old_ccle_columns = ccle_csv.columns
new_ccle_columns = []
for i in old_ccle_columns:
    new_ccle_columns.append(i.split(' ')[0])
ccle_csv.columns = new_ccle_columns
ccle_csv.index = ccle_csv['Unnamed:']
ccle_csv = ccle_csv.drop('Unnamed:', axis = 1)
ccle_corr = ccle_csv.corr(method='pearson')

#STRING detailed and combined scores
with gzip.open(path_string1) as inp:
    string_annotation = {}
    for line in inp:
        line = line.decode().split('\t')
        string_annotation[line[0]] = line[1]
with gzip.open(path_string2) as inp:
    string_detailed = {}
    next(inp)
    for line in inp:
        line = line.decode().split(' ')
        line[-1] = line[-1][:-1]
        if string_annotation[line[0]] not in string_detailed.keys():
            string_detailed[string_annotation[line[0]]] = {}
string_detailed[string_annotation[line[0]]][string_annotation[line[1]]] =
[line[2:]]
ccle_corr = ccle_corr.to_dict()
tumor_corr = tumor_corr.to_dict()
normal_corr = normal_corr.to_dict()
fc_corr = fc_corr.to_dict()
list_of_tfs = select_tfs(normal_corr.keys())
list_of_tfs_and_markers = genelist
#Для маркеров и ТФ выводит их корреляции со всеми остальными маркерами и ТФ.
temp_table =
[['', '', 'ICGC', 'ICGC', 'ICGC', 'DEPMAP', 'DEPMAP', 'STRING', 'STRING', 'STRING', '
STRING', 'STRING', 'STRING', 'STRING', 'STRING']]
temp_table.append(['gene1', 'gene2', 'tumor_corr', 'normal_corr', 'fc_corr', 'Ge
neEffect_corr', 'CCLE_corr', 'neighborhood', 'fusion', 'cooccurrence',
'coexpression', 'experimental', 'database', 'textmining',
'combined_score'])
empty_key = {}
for key1 in list_of_tfs_and_markers:
    for key2 in list_of_tfs_and_markers:

        temp_table.append([key1, key2, tumor_corr.get(key1, empty_key).get(key2)
, normal_corr.get(key1).get(key2), fc_corr.get(key1).get(key2), ge_corr.
get(key1, empty_key).get(key2), ccle_corr.get(key1, empty_key).get(key2)
, string_detailed.get(key1, empty_key).get(key2)])
return(temp_table)

#Для выбранных генов выводит их корреляции со всеми остальными генами.
table_markers=['', '', 'ICGC', 'ICGC', 'ICGC', 'DEPMAP', 'DEPMAP', 'STRING', 'STRIN
G', 'STRING', 'STRING', 'STRING', 'STRING', 'STRING', 'STRING']

table_markers.append(['gene1', 'gene2', 'tumor_corr', 'normal_corr', 'fc_corr',
'GeneEffect_corr', 'CCLE_corr', 'neighborhood', 'fusion', 'cooccurrence',
'coexpression', 'experimental', 'database', 'textmining',
'combined_score'])
empty_key = {}
for key1 in list_of_tfs_and_markers:
    for key2 in list_of_tfs_and_markers:
        table_markers.append([key1, key2, tumor_corr.get(key1, empty_key).get(ke
y2), normal_corr.get(key1).get(key2), fc_corr.get(key1).get(key2), ge_corr.get
(key1, empty_key).get(key2), ccle_corr.get(key1, empty_key).get(key2), string_d
etailed.get(key1, empty_key).get(key2)])
return(table_markers)

```

# Результаты и обсуждение

## Характеристика ингибиторов GСРП и конъюгатов ингибиторов GСРП с флуоресцентными и цитотоксичными молекулами

### Ингибиторы GСРП

Все исследованные ингибиторы GСРП были поделены на три поколения в зависимости от исследуемых модификаций (Таблицы 1-3). Для оценки вклада взаимной ориентации AR<sub>ABS1</sub> и AR<sub>ABS2</sub> дипептидного фрагмента линкера, позволяющего усилить связывание ингибитора с GСРП, было проанализировано поколение А ингибиторов GСРП. Варьирование ориентации дипептидного фрагмента линкера, заместителей и позиций замещения в AR<sub>P1</sub> и AR<sub>ABS2</sub> в ингибиторах GСРП рассматривается в поколении В. В поколении С на основе ингибитора GСРП с (L,L) конфигурацией дипептидного фрагмента линкера оценен эффект нитро-, карбокси- и алкокси-групп в качестве заместителей в AR<sub>P1</sub> и AR<sub>ABS2</sub> фрагментах лиганда. Для соединений были определены значения констант ингибирования, для части из них – показатель цитотоксичности, СС<sub>50</sub>, концентрация, при которой гибнет половина клеток. Данные представлены в таблицах 1-4.

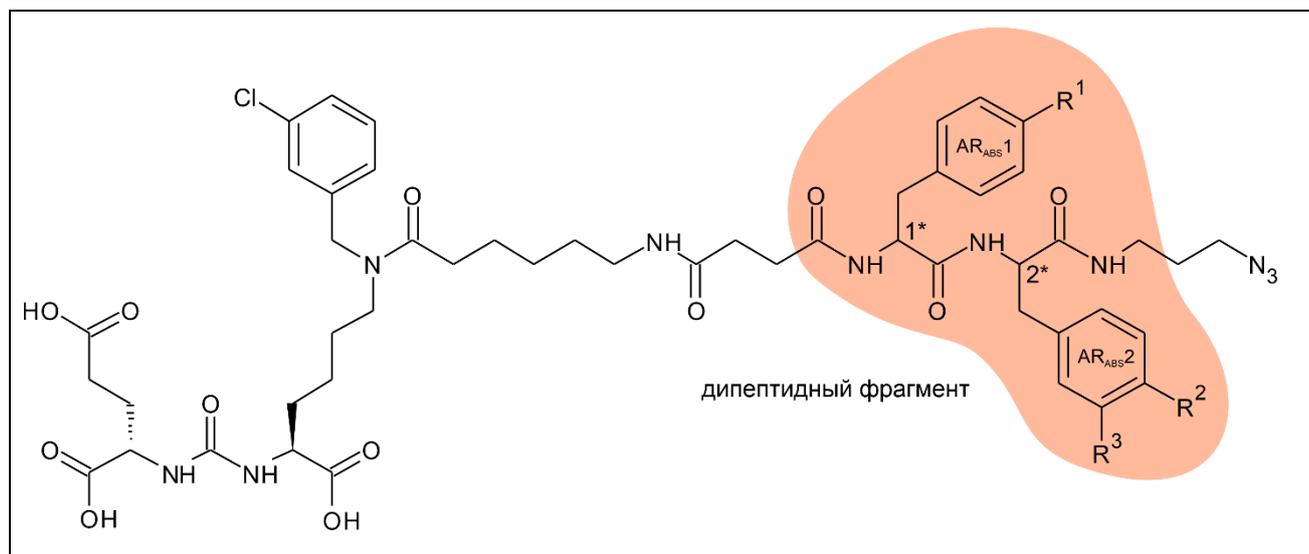
### Поколение А – влияние конфигурации дипептидного фрагмента на K<sub>i</sub> ингибиторов GСРП

Для оценки вклада взаимной ориентации арильных групп AR<sub>ABS1</sub> и AR<sub>ABS2</sub> дипептидного фрагмента линкера, позволяющей усилить связывание лиганда с GСРП, было проанализировано синтезированное нашими коллегами поколение А лигандов GСРП (Табл. 1).

В серии энантиомеров **A2-A5** ингибитор **A2** с конфигурацией (D,D) ингибирует GСРП немного сильнее, чем **A3** с конфигурацией (L,L), однако оба этих ингибитора имеют K<sub>i</sub> значительно ниже, чем **A4** и **A5** с (L,D) и (D,L) конфигурациями. Замещение *para*-позиции AR<sub>ABS2</sub> гидроксильной группой меняет паттерн K<sub>i</sub> для разных конфигураций дипептидного фрагмента. В серии с гидроксигруппой в *meta*-положении AR<sub>ABS2</sub>: **A1**, **A6**, **A7**, **A8**, в которой реализованы все комбинации энантиомеров без иных различающихся модификаций, конфигурации (D,L) в **A8** (K<sub>i</sub> = 137±31 нМ), (L,D) в **A7** (K<sub>i</sub> = 111±18 нМ) и (D,D) в **A6** (K<sub>i</sub> = 85±45 нМ) являются значительно менее выгодными, чем (L,L) в **A1** (K<sub>i</sub> = 2,3±0,6 нМ). В отличие от небольшого различия K<sub>i</sub> для **A2** и **A3**, где **A2** с конфигурацией (D,D) немного лучше, чем **A3** с (L,L), внесение гидроксигруппы, с одной стороны, уменьшило K<sub>i</sub> для ингибитора **A1** с конфигурацией (L,L), с другой – значительно ухудшило ингибитор **A6** с (D,D). Внесение Br в *meta*-положение AR<sub>ABS2</sub> в дополнение к *para*-гидрокси-заместителю позволило уменьшить K<sub>i</sub> ингибитора с (D,D) конфигурацией дипептидного фрагмента (**A6** имеет K<sub>i</sub> = 85±45 нМ, а его

*para*-Br-замещенное производное **A10** имеет  $K_i = 28 \pm 16$  нМ), но значительно ухудшило ингибитор с (L,L)-конфигурацией (**A1** имеет  $K_i = 2,3 \pm 0,6$  нМ, а его *para*-Br-замещенное производное **A12** имеет  $K_i = 143 \pm 35$  нМ). Это указывает на отсутствие явной аддитивности вклада конфигурации дипептидного фрагмента ингибитора и заместителей его ароматических фрагментов, что усложняет предсказание влияния отдельной модификации на  $K_i$ .

**Таблица 1.** Поколение А – влияние конфигурации дипептидного фрагмента на  $K_i$  ингибиторов GCPH.



#	AR <sub>ABS1</sub>		AR <sub>ABS2</sub>			K <sub>i</sub> , нМ
	*1	R <sup>1</sup>	*2	R <sup>3</sup>	R <sup>4</sup>	
A1	L	H	L	OH	H	2,4±0,6
A2	D		D	H		6±2
A3	L		L	H		9±3
A4	L		D	H		29±4
A5	D		L	H		54±11
A6	D		D	OH		85±45
A7	L		D	OH		111±18
A8	D		L	OH		137±31
A9	D	OH	L	H		76±13
A10	D	H	D	OH	Br	28±16
A11	L		L	H	NO <sub>2</sub>	99±11
A12	L		L	OH	Br	143±35
A13	L		L	OH	OH	621±117

## Поколение В – ингибиторы GСРП с галоген- и гидроксигруппами $AR_{P1}$ и $AR_{ABS2}$ .

Для дальнейшего изучения ингибиторов GСРП было проанализировано поколение ингибиторов В, синтезированное с учетом наблюдаемой выше зависимости вклада дипептидного фрагмента ингибитора и заместителей его ароматических частей. Поколение В позволило оценить комбинаторный вклад нескольких типов модификаций: замещения  $AR_{P1}$  галогенами, пространственной изомерии  $AR_{ABS1}$  и  $AR_{ABS2}$ , замещений  $AR_{ABS1}$  и  $AR_{ABS2}$  галогенами или гидроксильной группой. Для точности соотнесения с уже известными ингибиторами GСРП оценивались в тех же условиях значения  $K_i$  для 2-РМРА и DCL.

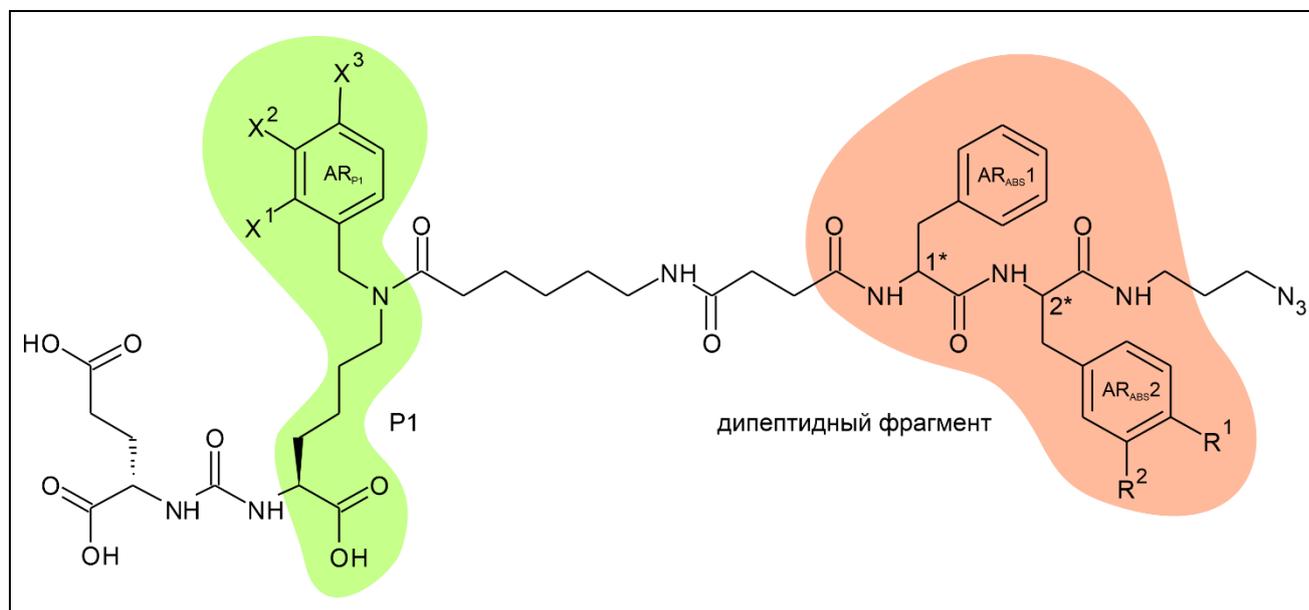
Среди ингибиторов GСРП с замещенными галогенами ароматическими фрагментами  $AR_{P1}$  низкие константы ингибирования наблюдаются у ингибиторов с Cl в *meta*-положении  $AR_{P1}$  (**V12**, **V13**, **V15**) и Cl (**V3**, **V5**, **V6**) или Br (**V8**, **V11**) в *para*-положении  $AR_{P1}$ . Ингибиторы GСРП с заместителями различной природы подтверждают литературные данные о предпочтении S1 карманом более крупных и менее электроотрицательных заместителей в *para*-положении  $AR_{P1}$ [59].

Взаимодействие GСРП с ингибиторами, содержащими дипептидный фрагмент, зависит от конфигурации дипептидного фрагмента. Однако строгого предпочтения той или иной конфигурации – (L,L), (D,D), (L,D), (D,L) – не наблюдается, что известно как по результатам анализа серии А, так и по опубликованным другими коллективами данным: **112** и **113** практически не демонстрируют различий между стереоизомерами. Тем не менее, взаимная ориентация заместителей в **V2**, **V3**, **V5**, **V7**, **V11**, **V13**, **V15**, **V17**, **V18**, **V21** наиболее благоприятна, когда повторяет структуру **V2** ( $K_i = 2,9 \pm 0,7$  нМ), т.е. L-  $AR_{ABS1}$ , L- $AR_{ABS2}$ : **V11** ( $K_i = 1,8 \pm 0,6$  нМ) значительно лучше, чем **V9** ( $K_i = 325 \pm 287$  нМ), **V15** ( $K_i = 0,9 \pm 0,3$  нМ) - чем **V14** ( $K_i = 11 \pm 4$  нМ), **V21** ( $K_i = 6 \pm 2$  нМ) - чем **V20** ( $K_i = 11 \pm 7$  нМ), **112** ( $IC_{50} = 10,2$  нМ) - чем **111** ( $IC_{50} = 13,9$  нМ).

Введение гидроксильной группы в *para*-положение  $AR_{ABS2}$  имеет разнонаправленные эффекты. В некоторых случаях улучшается связывание ингибитора с GСРП: при добавлении гидроксильной группы в *para*-положение  $AR_{ABS2}$  ингибитора **V4** получен ингибитор **V6**,  $K_i$  которого почти в девять раз меньше. Схожий эффект наблюдается для **V19** и **V20** – трехкратное уменьшение  $K_i$ . Обе пары имеют (D,D) конфигурацию дипептидного фрагмента. Наблюдается и обратный эффект для (D,D) конфигурации дипептидного фрагмента: добавление гидроксила в *para*-положение  $AR_{ABS2}$  ингибитора **V8** и **V12** с образованием **V9** и **V14** приводит к увеличению  $K_i$  в десять раз и в два раза соответственно. Положительный эффект также наблюдается и при замещении  $AR_{ABS2}$  гидроксильной группой при (L,L) конфигурации дипептидного фрагмента: в паре **V13** и **V15** добавление гидроксила к **V13** уменьшило  $K_i$  в четыре

раза. В парах **B18** и **B21**, **B3** и **B5** видно отсутствие влияния этой модификации – их значения  $K_i$  почти не меняются.

**Таблица 2.** Поколение В – ингибиторы GCP II с галоген- и гидроксигруппами-замещенными  $AR_{P1}$  и  $AR_{ABS2}$ .



#	$AR_{P1}$			$AR_{ABS}$ 1	$AR_{ABS2}$		$K_i$ , нМ			
	$X^1$	$X^2$	$X^3$	1*	2*	$R^1$		$R^2$		
B1	$AR_{P1} = H$			L	L	H	H	74±18		
B2	H	H	H	L	L			2,9±0,7		
B3			Cl		L	L			6±3	
B4					D	D			23±10	
B5					L	L	OH		5±2,6	
B6					D	D	OH		2,8±1,4	
B7				Br		L	L	H		5,2±1,4
B8						D	D	H		33±10
B9						D	D	OH		325±287
B10					D	D	OH	Br	9±5	
B11			L		L	OH		1,8±0,6		
B12		Cl	H			D	D	H		4,7±2,5
B13						L	L	H	H	3,9±1,2
B14						D	D	OH		11±4
B15					L	L	OH		0,9±0,3	
B16					D	D	OH	Br	8±5	

B17	F	H	Cl	L	L	H	H	4±2
B18	H	F		L	L	H		7±2
B19				D	D	H		30 ±10
B20				D	D	OH		11±7
B21				L	L	OH		6±2
Референсные соединения								
2-PMPA								8±3
DCL								56±40

Увеличение  $K_i$ , наблюдаемое при замещении  $AR_{ABS2}$  гидроксилем в ингибиторе **B8** образованием **B9**, может быть обращено: при дальнейшем замещении **B9** атомом Br в *meta*-положении  $AR_{ABS2}$  у получившегося ингибитора **B10** показано многократное снижение значения  $K_i$  ( $325 \pm 287$  нМ и  $9,3 \pm 5,0$  нМ соответственно) по сравнению с **B9** и четырехкратное снижение  $K_i$  по сравнению с **B8** – предшественником **B9**. Слабое улучшение также наблюдается у **B16** ( $K_i = 8,3 \pm 5,2$  нМ) по сравнению с **B14** ( $K_i = 11,0 \pm 3,9$  нМ), однако полученный ингибитор GCP II хуже, чем **B12** без заместителей  $AR_{ABS2}$ .

Само наличие  $AR_{P1}$  уменьшает  $K_i$ , что свидетельствует об усилении связывания ингибитора с GCP II: при сравнении **B1** и **B2** видно, что модификация  $\epsilon$ -аминогруппы лизина незамещенным метилбензолом позволила добиться снижения  $K_i$  более чем в 20 раз. Значения  $K_i$  у ингибиторов с Cl в *meta*-положении, как у **B12** и **B13**, и Cl или Br в *para*-положении фрагмента  $AR_{P1}$ , как у **B3**, **B4** и **B7**, **B8**, хоть и не меньше, чем у исходного **B2**, однако близки к нему.

Вклад галогеновых заместителей  $AR_{P1}$  раскрывается при их сочетании с различными заместителями  $AR_{ABS2}$  и конфигурацией дипептидного фрагмента линкера. Например, разница в  $K_i$  между ингибиторами GCP II с (L,L) и (D,D) конфигурациями мала или отсутствует для *para*-Cl-замещенного  $AR_{P1}$  (**B3** и **B4**, **B5** и **B6**), однако значительна при замене Cl на Br (**B7** и **B8**, **B9** и **B11**). Сложность расчета вклада отдельных модификаций и их совместное влияние на  $K_i$  хорошо иллюстрируется на примере пар **B9**, **B10** и **B14**, **B16**. В паре **B9**, **B10** при наличии Br-замещенного  $AR_{P1}$  и гидроксил-замещенного  $AR_{ABS2}$  при конфигурации дипептидного фрагмента (D,D) добавление Br в *para*-положение  $AR_{ABS2}$  позволяет добиться снижения  $K_i$  в более чем тридцать раз. Однако при замещении той же позиции атомом Br в **B14**, где, по сравнению с предыдущей парой, в  $AR_{P1}$  вместо Br в *para*-положении присутствует Cl в *meta*-положении,  $K_i$  полученного ингибитора **B16** меняется незначительно.

При анализе сочетания заместителей в AR<sub>P1</sub> показано, что добавление F в *мета*-положение и *пара*-положение к C1 в *пара*-положении практически не меняло K<sub>i</sub> (**V17-V21** в Таблице 2). F- и C1-замещенные арильные фрагменты также предотвращали потерю ингибирующей активности **V20** (L,L) по сравнению с **V21** (D,D), его стереоизомером.

Для того, чтобы ингибиторы GСPII можно было использовать в адресной доставке молекул в опухоль в рамках диагностики и терапии, необходимо минимальное количество побочных эффектов и высокая селективность к GСPII. В ходе первичной оценки *in vitro* все проверенные ингибиторы GСPII оказались нетоксичны для клеточных линий разного происхождения HEK293T, VA-13, MCF-7, A549 в концентрациях до 50-100 мкМ. Оба лидера по значениям K<sub>i</sub>, **V11** и **V15**, оказались лучше референсных 2-РМРА и DCL, а также нетоксичными при концентрациях до 100 мкМ. Ингибитор **V15** с наилучшими ингибирующими свойствами оказался также нетоксичным для GСPII-экспрессирующих клеток LNCaP и GСPII-негативных РС-3 в концентрациях до 100 мкМ (Таблица 4).

#### **Поколение С – ингибиторы GСPII, в которых проведена вариация заместителей AR<sub>P1</sub> и AR<sub>ABS2</sub> в сочетании с (L,L) конфигурацией дипептидного фрагмента.**

Анализ поколения С ингибиторов GСPII демонстрирует возможность уменьшить K<sub>i</sub> ингибиторов с конфигурацией дипептидного фрагмента (L,L) путем замещения AR<sub>P1</sub> и AR<sub>ABS2</sub> нитро-, карбокси- и алкокси-группами, которые также позволяют увеличить гидрофильность ингибиторов. В ингибиторах GСPII этого поколения наименьшую способность ингибировать GСPII показали соединения с алкокси-заместителями в *пара*-положении **C6** и **C7**, а также **C8** с метокси-группами в *пара*- и *мета*-положении и **C11** с NO<sub>2</sub> в качестве *орто*-заместителя. Самыми эффективными ингибиторами являются **C4** (K<sub>i</sub> = 0,7±0,2 нМ) и **C13** (K<sub>i</sub> = 0,1±0,1 нМ) с карбокси-группами в *пара*-положении. Введение гидроксила в *пара*-положение AR<sub>ABS2</sub> **C4** дало **C12** с K<sub>i</sub> = 2,0±0,5 нМ, однако дополнительная модификация *мета*-положения AR<sub>ABS2</sub> группой NO<sub>2</sub> позволила нивелировать падение K<sub>i</sub> и получить **C13** с минимальной K<sub>i</sub> в поколении С.



Таблица 4. Цитотоксичность ( $CC_{50}$ ) ингибиторов GСPII на клеточных линиях человека.

Клеточные линии человека различного происхождения				
#	CC <sub>50</sub> , мкМ			
	HEK293T	VA-13	MCF-7	A549
B2	114±51	>100	>100	>100
B7	>100	>100	>100	>100
B8	88±40	>100	>100	>100
B11	>100	>100	>100	>100
B15	>100	>100	>100	>100
B17	>100	>100	>100	>100
B18	71±18	>100	>100	>100
B19	51±11	>100	>100	>100
B21	>100	>100	>100	>100
DCL	>100	>100	>100	>100

### Конъюгаты с флуоресцентными молекулами на основе B15 – ингибитора GСPII

Лучший ингибитор GСPII из поколения В (B15) и был конъюгирован нашими коллегами с флуоресцентной молекулой SulfoCy5 или SulfoCy7. Для конъюгатов оценены константы ингибирования, цитотоксичность и эффективность окрашивания *in vitro* (с помощью проточной цитофлуориметрии) и *in vivo* (выполнено коллегами из РОНЦ им. Н.Н. Блохина на мышах с привитой опухолью из 22Rv1[116]).

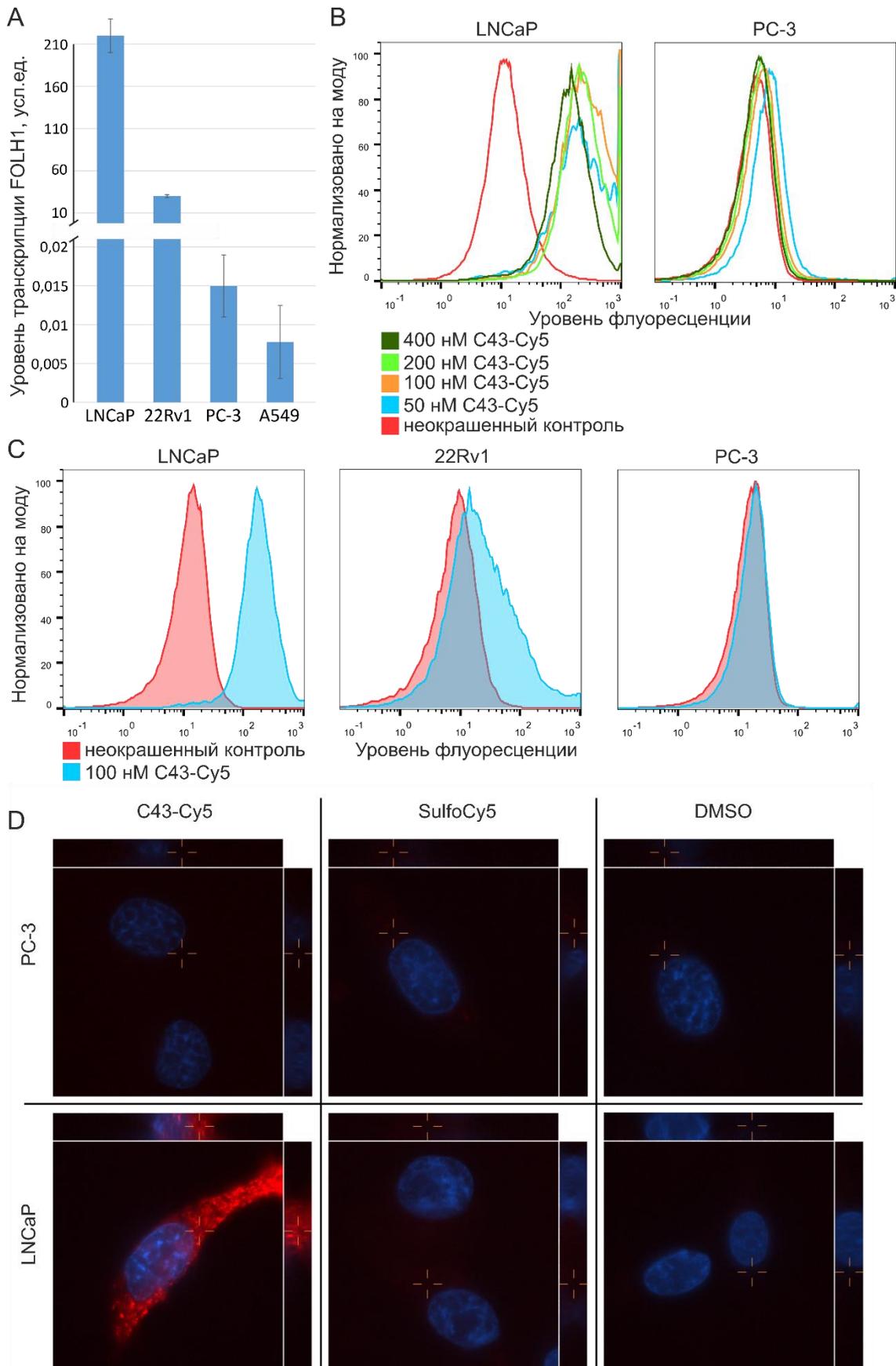
Значение  $K_i$  слабо изменилось для конъюгата B15-Cy5 по сравнению с исходным ингибитором B15, что отражает небольшое ослабление связывания с GСPII, и увеличилось в двадцать раз для конъюгата B15-Cy7. B15-Cy7 оказался цитотоксичным для линий рака простаты LNCaP и PC-3 со значениями  $CC_{50}$  12,2 и 18,7 мкМ соответственно (Таблица 5).

**Таблица 5.** Цитотоксичность ( $CC_{50}$ ) ингибиторов GСPII на клеточных линиях человека,  $K_i$  и  $CC_{50}$  для конъюгатов В15 с флуоресцентными молекулами.

#	$K_i$ , нМ	$CC_{50}$ , мкМ	
		LNCaP	PC-3
<b>В15</b>	0,9±0,3	>100	>100
<b>В15-Сy5</b>	2,6±0,6	>100	>100
<b>В15-Сy7</b>	18±8	12,2±8,1	18,7±4,5
SulfoСy5	-	>100	>100
SulfoСy7	-	>100	>100

Специфичность конъюгата В15-Сy5 к GСPII-экспрессирующим клеткам была показана с помощью проточной цитофлуориметрии на панели из клеточных линий рака простаты LNCaP, 22Rv1, PC-3: сильнейшее окрашивание наблюдалось для LNCaP с наибольшей экспрессией GСPII[5] (Рис.15А), что соответствует литературным данным, слабее – для 22Rv1, где экспрессия GСPII в восемь-десять раз ниже, чем в LNCaP[5], а PC-3, где экспрессия GСPII в тысячи раз ниже, чем в LNCaP, не окрашивались (Рис.15В). Среди проверенных концентраций В15-Сy5, 100 нМ концентрация оказалась оптимальной для окрашивания клеток LNCaP (Рис.15С), т.к. при дальнейшем увеличении концентрации конъюгата гистограмма распределения флуоресценции окрашенных клеток не менялась, указывая на отсутствие GСPII, доступного для связывания.

Специфичность взаимодействия В15-Сy5 с GСPII была подтверждена флуоресцентной микроскопией наличием яркого сигнала в LNCaP и его отсутствием в PC-3 (Рис.15D). Флуоресцентный сигнал от SulfoСy5 в обеих линиях отсутствовал.

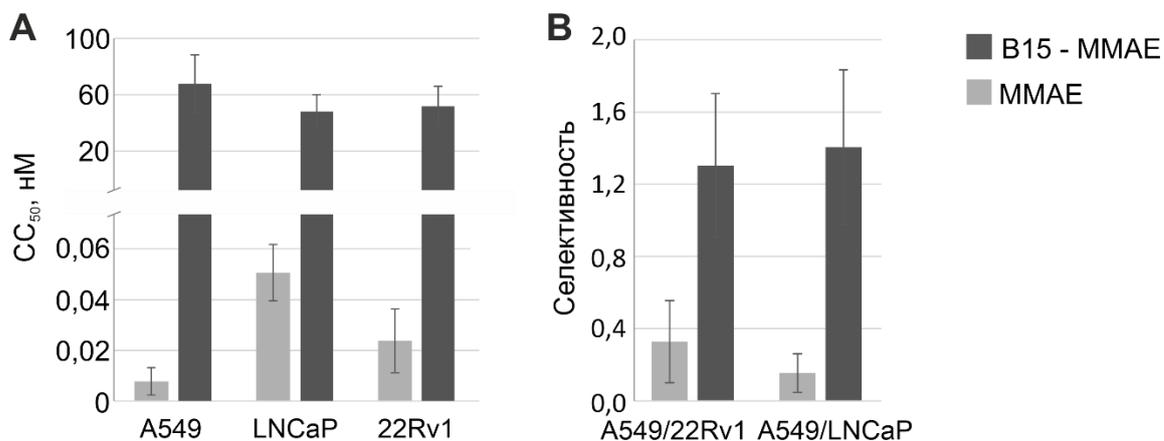


**Рис. 15.** Окрашивание клеток линий рака простаты **B15-Cy5**. А) Сравнение уровня транскрипции *FOLH1* в клеточных линиях, определенный с помощью RT-qPCR. В) Клетки *LNCaP* демонстрируют окрашивание, зависящее от концентрации конъюгата. Клетки *PC-3*,

не экспрессирующие GCP II, не окрашиваются и при самой высокой концентрации – 400 нМ **B15-Cy5**. С) Уровень флуоресценции при проточной цитофлуориметрии отражает уровень экспрессии GCP II в клеточных линиях рака простаты. D) Распределение флуоресцентного сигнала от **B15-Cy5** в клетках LNCaP и PC-3 на z-стеках. Окрашивание конъюгатом специфично увеличивает сигнал флуоресценции в GCP II-положительных клетках LNCaP.

### Конъюгат с монометил ауристатином E на основе **B4**

Конъюгирование **B15** с монометил ауристатином E (ММАЕ) приводило к снижению цитотоксичности по сравнению с исходным ММАЕ – цитостатиком, ингибирующим полимеризацию тубулина и влияющим на все делящиеся клетки. При этом снижении цитотоксичности лиганда, конъюгированного с цитостатиком, было сильнее для клеток, не экспрессирующих GCP II (Рис.16). Полученное изменение цитотоксичности позволяет предположить возможность снижения неспецифического действия препарата в организме и уменьшение побочных эффектов при его приеме. В экспериментах наших коллег из НМИЦ онкологии им. Н.Н. Блохина конъюгат **B15-ММАЕ** успешно ингибировал рост ксенографтной опухоли в мышцах на основе GCP II-экспрессирующей клеточной линии 22Rv1, слабо влияя на опухоли из клеток линии PC-3, где GCP II отсутствует[117].



**Рис.16.** Сопоставление ММАЕ с конъюгатом **B15-ММАЕ** по действию на опухолевые клетки. А. Конъюгация с **B15** уменьшила цитотоксический эффект ММАЕ. В. Отношение цитотоксичностей для клеточных линий человека, экспрессирующих (22Rv1, LNCaP) и не экспрессирующих GCP II (A549). CC50 - концентрация соединения, необходимая для уменьшения жизнеспособности клеток на 50%. LNCaP, 22Rv1 - GCP II-экспрессирующие линии, GCP II отсутствует в A549.

## Обсуждение результатов

Создание новых синтетических ингибиторов GСPII продолжается ввиду важности не только их константы ингибирования GСPII, но и возможности варьирования части их структурных элементов при сохранении ингибирующих свойств, что упрощает дальнейшую оптимизацию фармакологических характеристик. Вместе с коллегами с Химического Факультета МГУ (Мачулкин А.Э., Успенская А.А. и др.), проводившими моделирование и синтез соединений, мы продолжили анализировать взаимосвязь структуры и активности ингибиторов GСPII. Значение  $K_i$  в нашей работе служит характеристикой силы связывания ингибитора с GСPII, однако цель создания новых молекул – именно усиление связывания GСPII, а не ингибирование его ферментативной активности.

Для оценки вклада взаимной ориентации  $AR_{ABS1}$  и  $AR_{ABS2}$  дипептидного фрагмента линкера, позволяющего усилить связывание ингибитора с GСPII, было создано поколение А ингибиторов GСPII (Таблица 1). Оказалась важна комбинация всех модификаций, т.к. разные заместители арильных групп  $AR_{ABS2}$  дипептидного фрагмента и арильной группы  $AR_{P1}$  фрагмента P1 в силу природы заместителей и положения замещения меняют паттерн  $K_i$ . Так, в серии энантиомеров **A2-A5** ингибитор **A2** с конфигурацией (D,D) немного лучше, чем **A3** с конфигурацией (L,L), и оба этих ингибитора имеют  $K_i$  значительно ниже, чем **A4** и **A5** с (L,D) и (D,L) конфигурациями. Если заместить *para*-положение  $AR_{ABS2}$  гидроксигруппой, то паттерн меняется: (L,L)-конфигурация становится самой выгодной, а  $K_i$  ингибиторов GСPII с остальными видами энантиомерии увеличиваются, и особенно сильное четырнадцатикратное увеличение наблюдается для наиболее выгодной ранее конфигурации (D,D). Анализ  $K_i$  ингибиторов GСPII поколения А не позволил выявить строгой взаимосвязи между  $K_i$  и ориентацией обоих пептидных остатков. Энантиомеры (D,L) и (L,D), тем не менее, имеют большие значения  $K_i$  вне зависимости от рассмотренных заместителей  $AR_{ABS2}$ , что может быть связано с нарушением  $\pi$ -стэкинга с Trp541 арен-связывающего сайта в случае, когда оба ароматических фрагмента  $AR_{ABS1}$  и  $AR_{ABS2}$  направлены в пространстве в одну сторону. По этой причине было решено отказаться от использования этих энантиомеров в следующих поколениях ингибиторов GСPII.

Варьирование ориентации дипептидного фрагмента линкера, заместителей и позиций замещения в  $AR_{P1}$  и  $AR_{ABS2}$  в ингибиторах GСPII поколения В (Таблица 2) продемонстрировало невозможность строгого предсказания вклада конкретной модификации в значение  $K_i$ . Тем не менее, существуют тренды, согласующиеся с литературными данными. Ингибиторы GСPII с (L,L) конфигурацией дипептидного фрагмента линкера продемонстрировали меньшие значения  $K_i$ , чем с (D,D), и выбранный для конъюгации ингибитор имеет (L,L)-конфигурацию. Для (L,L)

ингибиторов GСРII С1 в *para*-положении AR<sub>P1</sub> и С1 или Вг в *meta*-положении AR<sub>P1</sub> позволяют снизить значение K<sub>i</sub>. Поколение С на основе ингибитора GСРII с (L,L) конфигурацией дипептидного фрагмента линкера продемонстрировало возможность добиться снижения K<sub>i</sub> путем использования в качестве заместителей AR<sub>P1</sub> и AR<sub>ABS2</sub> нитро-, карбокси- и алкокси- групп.

Создание новых ингибиторов GСРII, чьи значения K<sub>i</sub> близки или меньше, чем у опубликованных ингибиторов, важно из-за разницы совместного вклада разных наборов заместителей AR<sub>P1</sub>, AR<sub>ABS1</sub> и AR<sub>ABS2</sub> в поведение ингибиторов на уровне организма. Так, использование в качестве заместителей галогенов, гидроксо-группы и карбокси-группы позволяет уменьшить липофильность и неспецифическую цитотоксичность, т.е. возможность внесения модификаций практически без изменения константы ингибирования дает возможность улучшить фармакологические характеристики высокоаффинных ингибиторов GСРII.

Конъюгирование **V15** с монометил ауристатином Е позволило добиться снижения цитотоксичности по сравнению с чистым монометил ауристатином Е – цитостатиком, ингибирующим полимеризацию тубулина и влияющим на все делящиеся клетки. Полученное снижение цитотоксичности позволяет предположить возможность снижения неспецифического действия препарата в организме и уменьшение побочных эффектов от его приема. Паттерн токсичности **V15-ММАЕ** смещен в сторону клеток, экспрессирующих GСРII, что говорит о его большей селективности, чем у ауристатина. Конъюгат успешно ингибировал рост ксенографтной опухоли в мышцах на основе GСРII-экспрессирующей клеточной линии 22Rv1, слабо влияя на опухоли из клеток линии РС-3, где GСРII отсутствует[117].

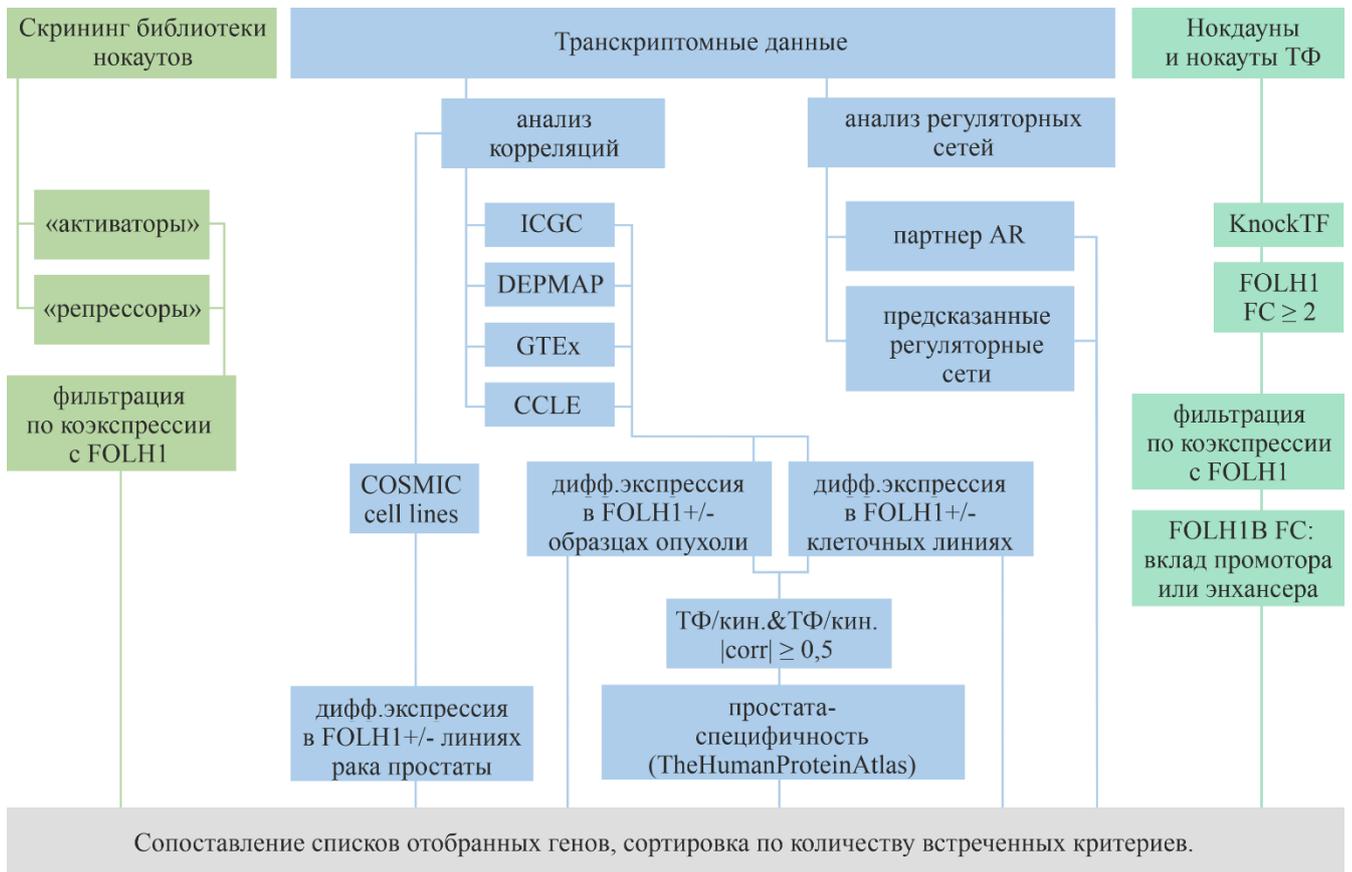
Конъюгирование ингибиторов GСРII с флуоресцентными молекулами слабо повлияло на их способность связываться с GСРII. Несмотря на цитотоксичность *in vitro* конъюгата **V15-Cy7** (CC<sub>50</sub> = 12,2±8,1 мкМ), его K<sub>i</sub> (18±8 нМ) значительно ниже, поэтому его было решено использовать для визуализации в мышечной модели рака простаты. Конъюгаты **V15-Cy5** и **V15-Cy7** оказались эффективными для визуализации ксенографтной опухоли в мышцах с очень низкой токсичностью[116]. Благодаря высокой специфичности **V15-Cy5** к GСРII, конъюгат было решено использовать для скрининга библиотеки нокаутов Gecko v2 в одном из подходов по идентификации регуляторов экспрессии гена *FOLH1* в LNCaP.

## Потенциальные регуляторы экспрессии гена *FOLH1*

Тканеспецифичная экспрессия гена *FOLH1* достигается за счет его транскрипционной регуляции, которая мало изучена. Для отбора потенциальных регуляторов экспрессии *FOLH1* использовано несколько подходов: корреляционный анализ экспрессии генов в образцах нормы и опухоли, анализ баз данных транскриптомов нокадаунов потенциальных регуляторов в клеточных линиях различной этиологии, скрининг полногеномной библиотеки гидовых РНК для нокаутов функциональных генов в экспрессирующих GSPII клетках.

Источниками транскриптомных и функциональных данных выступали ICGC (International Cancer Genome Consortium, транскриптомные данные пациентов), GTEx (Genotype-Tissue Expression portal, транскриптомные данные здоровых тканей), CCLE (Cancer Cell Line Encyclopedia, транскриптомные данные для клеточных линий), COSMIC (Catalogue of Somatic Mutations in Cancer, содержит обработанные данные для клеточных линий, в которых указана категория для экспрессии гена: повышенная экспрессия, нормальная экспрессия, пониженная экспрессия), The Human Protein Atlas (агрегатор информации об экспрессии на уровне мРНК и белка, функциональной роли генов, включает анализ тканеспецифичности экспрессии генов), GEPIA 2 (аналог The Human Protein Atlas), KnockTF (агрегатор транскриптомных данных для нокаутов и нокадаунов транскрипционных факторов в клеточных линиях), DEPMAP (Cancer Dependency Map, данные синтетической летальности в виде Gene Effect, GE).

Краткая схема проведенного в работе и описанного далее отбора потенциальных регуляторов экспрессии *FOLH1* представлена на Рис.17. Сводная таблица потенциальных регуляторов представлена в Табл.6.



**Рис.17.** Схема поиска потенциальных регуляторов экспрессии гена *FOLH1*. Нами использовано три подхода, основанных на скрининге библиотеки последовательностей генов *РНК Gesc* v2, анализе транскриптомных данных и анализе базы данных нокаутов и нокдаунов транскрипционных факторов.

**Таблица 6.** Сводная таблица потенциальных регуляторов экспрессии гена *FOLH1*. Указаны критерии, по которым происходил отбор регуляторов. Цвета заголовков соответствуют цветам на схеме на Рис. X. В таблице представлены данные только для тех регуляторов экспрессии гена *FOLH1*, которые имеют «+» в двух и более колонках. Числа, приведенные во второй строке, определены по полной таблице, в которой также присутствуют регуляторы, соответствующие одному критерию. Fc – изменение экспрессии. \* - по опубликованным экспериментальным данным и предсказаниям регуляторных сетей; \*\* - по данным COSMIC; \*\*\* - по данным The Human Protein Atlas. \*\*\*\* - для подробного разбора регуляторов экспрессии гена *FOLH1*, найденных при скрининге полногеномной библиотеки нокаутов *Gecko v2*, см. таблицу 9.

Потенциальные регуляторы \ Критерии выбора регуляторов	Fc ( <i>FOLH1</i> ) > 2 при нокауне регулятора	Fc ( <i>FOLH1B</i> ) > 2 при нокауне регулятора	участник регуляторных событий в простате и раке простаты *	дифференциальная экспрессия в образцах опухоли (ICGC)	дифференциальная экспрессия в клеточных линиях (CCLE)	дифференциальная экспрессия в клеточных линиях рака простаты**	повышенная экспрессия в простате***	найден при скрининге <i>Gecko v2</i> ****
<b>Всего:</b>	<b>25</b>	<b>23</b>	<b>10</b>	<b>58</b>	<b>59</b>	<b>28</b>	<b>10</b>	<b>45</b>
BMPR1B				+	+	+	+	
HOXB13			+	+	+		+	
AR	+		+			+		
CREB3L4			+			+	+	
GGT7				+	+			+
GLIS1				+	+		+	
GPM6B				+	+			+
HIC1		+		+	+			
IRX4				+	+		+	
KREMEN1				+	+			+
SOX7		+		+	+			
SPDEF	+	+					+	
ZMPSTE24				+	+			+
AHCTF1					+			+
AKNA				+	+			
ARX				+	+			
ASCL1	+	+						
ATPAF1				+				+
AXL				+	+			
BRSK2				+	+			
C17orf70				+				+
C1orf115					+			+
CDK4				+	+			
CDK6				+	+			

CREB1	+	+						
DAPK1				+	+			
DDR1				+	+			
EHHADH					+			+
EPHA6					+			+
EPHB2				+	+			
ESRRA	+	+						
FAM32A				+				+
FGR				+	+			
FOXA1			+				+	
GATS				+				+
GRHL2			+			+		
HOXC4				+	+			
HOXC6				+	+			
HOXD8				+	+			
HSF1	+	+						
ID1	+	+						
IRF4	+	+						
KLF17				+	+			
KLF2	+	+						
KLHL4					+			+
MAP3K14				+	+			
MAZ			+			+		
MESP1						+	+	
NCOA1				+	+			
NFYA	+		+					
NKG7					+			+
NKX3-1						+	+	
NKX6-1				+	+			
NPAS4				+	+			
NPR2				+	+			
OSR1				+	+			
PCGF2				+	+			
PITX1				+	+			
POU2AF1				+	+			
POU2F2				+	+			
PRKD1				+	+			
PYGO1					+			+
SCUBE1				+				+
SOX8				+	+			
SP140L				+	+			
SPIB				+	+			
SREBF2				+	+			
SRF				+	+			
TESK2				+	+			
TFAP4	+	+						
TOPORS						+		+

TP63	+	+						
ULK3				+	+			
VPS8				+				+
WNK2				+	+			
ZEB1				+	+			
ZNF181				+	+			
ZNF233				+	+			
ZNF467				+	+			
ZNF469				+	+			
ZNF488				+	+			
ZNF613						+	+	
ZNF680				+	+			
ZNF77				+	+			
ZSCAN1				+	+			

### Потенциальные регуляторы экспрессии *FOLH1* по результатам корреляционного анализа транскрипционных данных

Была оценена корреляция экспрессии *FOLH1* и генов транскрипционных факторов и протеинкиназ. Списки ТФ взяты с <http://humantfs.ccb.utoronto.ca/>, протеинкиназ – с <http://kinase.com/>. Корреляции экспрессии генов с *FOLH1* определены на основании данных транскриптомов пациентов (International Cancer Genome Consortium, ICGC) в условной норме (образцы ткани вокруг опухоли), опухоли и изменений при переходе от нормы к опухоли, а также в простате (Genotype-Tissue Expression portal, GTEx) и в клеточных линиях (Cancer Cell Line Encyclopedia, CCLE, и Catalogue of Somatic Mutations in Cancer, COSMIC). Несколько потенциальных регуляторов экспрессии *FOLH1* с порогом для модуля корреляции  $> 0,5$  – FOXA1, STAT1, FOSL1, AR – были проверены с помощью нокдауна их мРНК (Рис. 21), но привели лишь к незначительным изменениям экспрессии интересующего нас гена. Поэтому было решено добавить дополнительные шаги выбора вероятных регуляторов.

На следующем шаге отбора потенциальных регуляторов экспрессии *FOLH1* был проведен анализ уровня экспрессии генов транскрипционных факторов (ТФ) и протеинкиназ в клеточных линиях опухоли простаты LNCaP, VCaP, 22Rv1, PC-3, DU-145 по данным COSMIC cell lines. В этой базе данных приведена информация об уровне экспрессии гена в клеточной линии (повышенная экспрессия, пониженная экспрессия, экспрессия в норме). Те гены, чей уровень транскрипции был повышен в *FOLH1*-экспрессирующих клетках LNCaP, VCaP, 22Rv1, но находился в норме или был понижен в линиях PC-3, DU-145, в которых *FOLH1* не экспрессируется, рассматривались как активаторы транскрипции *FOLH1*. Из получившегося списка были исключены гены, чья экспрессия различается в клетках LNCaP и VCaP, где экспрессия *FOLH1* одинакова. То есть, транскрипция отобранных 28 генов различна между

*FOLH1*-экспрессирующими линиями опухоли простаты и линиями, где *FOLH1* не экспрессируется (Табл.6).

В представленной работе было предположено, что, как в случае с андрогеновым рецептором и его партнерами, регуляция экспрессии *FOLH1* может зависеть от одновременного присутствия двух транскрипционных факторов (ТФ) или ТФ и протеинкиназы. Были подсчитаны корреляции для пар ТФ-ТФ, ТФ-протеинкиназа и протеинкиназа-протеинкиназа (скрипт 1). Для этого все транскриптомные данные 375 опухолевых образцов пациентов (USA-PRAD, ICGC) были отсортированы по экспрессии *FOLH1*. Из транскриптомных данных первой и последней сотни пациентов были сформированы выборки, *FOLH1+* и *FOLH1-*. По результатам подсчета корреляций были отобраны пары регуляторов, которые достоверно коррелировали друг с другом в *FOLH1+* выборке и в *FOLH1-* выборке, при этом экспрессия обоих факторов достоверно отличалась между *FOLH1+* и *FOLH1-* выборками. Было выбрано 59 генов, образующих такие пары, и среди этих пар 13 образованы специфически экспрессирующимися в тканях простаты генами (Табл.7) согласно The Human Protein Atlas. Аналогичный анализ проведен для транскриптомных данных всех клеточных линий из CCLE, отобрано 58 генов.

*Таблица 7. Пары регуляторов, дифференциально экспрессирующихся в FOLH1-положительных и FOLH1-отрицательных образцах опухоли простаты. P-value - вероятность получить для данной вероятностной модели распределения значений случайной величины такое же или более экстремальное значение статистики, по сравнению с ранее наблюдаемым, при условии, что нулевая гипотеза верна. Значения p-value: \* < 10E-2, \*\* < 10E-3, \*\*\* < 10E-4. Белым выделены пары предполагаемых активаторов транскрипции FOLH1, серым – пары репрессоров. Показаны только те пары, экспрессия которых достоверно коррелирует также в норме в простате или в клеточных линиях.*

регулятор 1	регулятор 2	корреляция в образцах с высокой экспрессией <i>FOLH1</i>	p-value	корреляция в образцах с низкой экспрессией <i>FOLH1</i>	p-value	корреляция в простате (GTEx)	p-value	корреляция GE (DERMAP)	p-value	корреляция в клеточных линиях	p-value
HOXD11	HOXD13	0,84	***	0,83	***	0,51	***	0,32	***	0,54	***
EVX2	HOXD13	0,69	***	0,71	***	0,88	***	0,09	*	0,43	***
BMPR1B	FOXA1	0,42	***	0,79	***	0,76	***	0,00		0,18	***
HOXD13	TBX4	0,29	**	0,66	***	0,55	***	0,09	*	0,05	

FOXA1	HOXB13	0,29	**	0,64	***	0,81	***	0,16	***	0,08	*
HOXD13	ISL2	0,68	***	0,74	***	0,43	***	0,04		0,16	***
ISL2	HOXD11	0,65	***	0,61	***	0,50	***	-0,03		0,13	***
GLIS1	HOXD13	0,61	***	0,85	***	0,41	***	0,01		0,09	*
ISL2	GLIS1	0,50	***	0,63	***	0,33	***	0,01		0,03	
HOXD11	GLIS1	0,49	***	0,64	***	0,31	***	0,04		0,19	***
TBX4	GLIS1	0,37	***	0,71	***	0,27	***	0,06		0,05	
ISL2	TBX4	0,31	**	0,41	*	0,08		0,02		-0,01	
BMPR1B	HOXB13	0,13		0,49	***	0,66	***	0,04		0,04	

Пересечение списков генов, полученных при анализе дифференциальной экспрессии генов в транскриптомных данных пациентов из ICGC и клеточных линий из CCLE, позволило выделить 5 простатспецифических генов (согласно The Human Protein Atlas) – репрессоры HIC1 и GLIS1, активаторы IRX4, BMPR1B, HOXB13 (Табл.8).

*Таблица 8. Дифференциальная экспрессия потенциальных регуляторов транскрипции FOLH1 в FOLH1+ выборке и в FOLH1- выборке образцов опухоли. FOLH1 выделен жирным шрифтом, простатспецифические активаторы – серым, репрессоры – светло-серым.*

ген	t-значение, CCLE	p-value, CCLE	t-значение, ICGC	p-value, ICGC
<b>FOLH1</b>	<b>19,36</b>	<b>1,50E-47</b>	<b>12,72</b>	<b>5,34E-27</b>
SOX8	6,05	7,04E-09	2,98	3,35E-03
HOXB13	5,99	9,86E-09	5,69	6,61E-08
HOXD8	5,60	7,09E-08	3,27	1,35E-03
HOXC6	5,46	1,43E-07	4,69	6,12E-06
PCGF2	4,20	4,05E-05	2,71	7,53E-03
ZNF77	4,17	4,53E-05	2,65	8,90E-03
ZNF233	4,17	4,60E-05	3,58	4,64E-04
WNK2	4,09	6,34E-05	2,87	4,67E-03
TESK2	4,08	6,47E-05	3,30	1,23E-03
ULK3	3,95	1,07E-04	2,63	9,55E-03
BMPR1B	3,91	1,27E-04	3,06	2,60E-03
ZNF680	3,86	1,56E-04	3,67	3,33E-04
PRKD1	3,84	1,68E-04	3,79	2,23E-04
ZNF467	3,62	3,81E-04	3,76	2,41E-04
NKX6-1	3,40	8,15E-04	2,64	9,13E-03
NCOA1	3,21	1,56E-03	3,50	6,15E-04
DDR1	3,20	1,58E-03	2,69	7,94E-03
IRX4	3,16	1,83E-03	3,61	4,17E-04
BRSK2	3,12	2,09E-03	2,91	4,21E-03
DAPK1	3,05	2,59E-03	2,81	5,65E-03
ARX	2,99	3,14E-03	3,04	2,81E-03
ZNF181	2,98	3,25E-03	2,97	3,48E-03
ZSCAN1	2,88	4,37E-03	2,81	5,61E-03
SREBF2	2,85	4,87E-03	3,70	2,99E-04

ген	t-значение, CCLE	p-value, CCLE	t-значение, ICGC	p-value, ICGC
HOXC4	2,81	5,52E-03	4,69	6,07E-06
CDK4	2,74	6,64E-03	2,72	7,32E-03
ZMPSTE24	2,70	7,62E-03	4,79	3,97E-06
FGR	-2,67	8,29E-03	-3,73	2,72E-04
SRF	-2,76	6,28E-03	-3,57	4,82E-04
ZNF488	-2,80	5,61E-03	-3,89	1,51E-04
GLIS1	-2,84	5,02E-03	-4,73	5,18E-06
POU2AF1	-2,84	5,00E-03	-2,73	7,13E-03
NPAS4	-2,84	5,00E-03	-3,88	1,59E-04
SOX7	-3,00	3,08E-03	-3,14	2,06E-03
NPR2	-3,04	2,70E-03	-4,54	1,16E-05
SP140L	-3,10	2,22E-03	-3,52	5,76E-04
OSR1	-3,26	1,30E-03	-3,05	2,73E-03
SPIB	-3,30	1,16E-03	-2,80	5,72E-03
ZNF469	-3,32	1,07E-03	-4,07	7,76E-05
EPHB2	-3,64	3,51E-04	-2,95	3,71E-03
AKNA	-3,66	3,28E-04	-3,10	2,34E-03
KLF17	-3,82	1,80E-04	-2,78	6,10E-03
MAP3K14	-3,92	1,22E-04	-4,08	7,34E-05
CDK6	-3,95	1,10E-04	-3,59	4,49E-04
AXL	-3,96	1,04E-04	-3,18	1,79E-03
PITX1	-4,69	5,00E-06	-3,78	2,27E-04
HIC1	-5,02	1,17E-06	-3,81	2,04E-04
ZEB1	-5,57	8,19E-08	-2,64	9,26E-03
POU2F2	-5,76	3,11E-08	-4,84	3,20E-06

Список потенциальных регуляторов, отобранных по корреляционному анализу экспрессий, был дополнен транскрипционными факторами, регулирующими экспрессию простатспецифических генов и являющихся при этом партнерами андрогенового рецептора. К ним относятся FOXA1[104], GATA2[104], [105], GRHL2[106], CREB3L4[107], HOXB13[108], NKX3.1[118], GABPA[119], ранее на рассматривавшиеся в качестве регуляторов экспрессии гена *FOLH1*.

Белок FOXA1 является ТФ-пионером, который усиливает связывание и активность андрогенового рецептора (AR), является регулятором GATA2 [104]. GATA2 является ТФ-пионером, который усиливает связывание и активность андрогенового рецептора на ранних и поздних стадиях заболевания[105]. Сверхэкспрессия GATA2 при раке простаты увеличивает клеточную подвижность и инвазивность, пролиферацию, устойчивость к стандартным методам лечения[105]. Оба ТФ регулируют экспрессию гена маркера рака простаты PSA (prostate-specific antigen), KLK3, и других калликреинов KLK2 [104]. GRHL2 также является ко-регулятором AR, напрямую с ним взаимодействующим[106]. Паттерны их экспрессии в клеточных линиях рака простаты VCaP, PC-3, DU145 совпадают, причем показано, что AR регулирует транскрипцию GRHL2[106]. Нокдаун GRHL2 не влияет на транскрипцию *FOLH1* [106], комбинации нокаута GRHL2 и нокаута AR либо добавления дигидротестостерона не проводилось. Другой партнер AR, HOXB13, рекрутирует AR к части мишеней[106], его сайты обогащены также сайтами FOXA1 [120]. Экспрессия HOXB13 необходима для метастазирования кастрационно-резистентного рака простаты [108]. HOXB13 экспрессируется как в андроген-чувствительных клеточных линиях LNCaP и VCaP, так и в андроген-нечувствительных C4-2B и 22Rv1[108], [121]. Экспрессия CREB3L4 андроген-зависима, нокдаун CREB3L4 приводит к ухудшению роста андроген-зависимых клеток рака простаты и негативно влияет на экспрессию ряда генов-мишеней AR, включая PSA, FKBP5, HPGD, KLK2 и KLK4[107]. CREB3L4 напрямую взаимодействует с AR и увеличивает связывание AP с элементами ответа на андрогены[107]. NKX3.1 также взаимодействует с AR и является важным ТФ, определяющим образование тканей простаты, и его потеря является одним из ключевых факторов развития рака простаты из-за его роли как супрессора опухоли [118]. GABPA способствует прогрессии рака простаты в сторону резистентности к депривации андрогенов, является партнером AR[119].

### **Вероятные регуляторы экспрессии *FOLH1* по результатам анализа данных транскриптомов клеток с нокаутами и нокдаунами транскрипционных факторов**

Для поиска вероятных регуляторов была использована база данных KnockTF. Она содержит транскриптомные данные для опубликованных нокдаунов и нокаутов ТФ в различных

клеточных линиях (Приложение 4, 5). Были выбраны те ТФ, нокаут или нокадаун которых приводил к изменению экспрессии *FOLH1* в два и более раза (Приложение 5). В качестве потенциальных регуляторов транскрипции *FOLH1* выбраны *GABPA*, *TFAP4*, *AR*, *CEBPB*, *NFYA*, *RBFOX2*, *SPDEF*. Эти белки, за исключением *AR*, ранее не рассматривались в литературе как регуляторы *FOLH1*. Самым интересным ТФ, значительно влияющим на транскрипцию *FOLH1* в клетках LNCaP (fold change = 66) и C4-2B (fold change = 43) по данным нокадауна, является *GABPA*, выступающий в роли «репрессора» транскрипции *FOLH1*. Любопытно, что *GABPA* слабо влияет на транскрипцию *FOLH1B* – паралога *FOLH1*, в котором присутствует тот же самый внутриинтронный энхансер, что и у *FOLH1*, но отсутствует промотор исходного гена. Нокадаун *SPDEF* в MCF7 привел к 3,4-кратному увеличению уровня транскрипции *FOLH1* и примерно трехкратному уменьшению экспрессии *FOLH1B*. Нокадаун другого гена, *TFAP4*, в клетках нейробластомы приводит к четырехкратному падению экспрессии *FOLH1*. Этот же нокадаун также приводит к пятикратному падению экспрессии *FOLH1B*, что может указывать на то, что эффект опосредован энхансером *FOLH1*. Наличие влияния нокадауна другого потенциального регулятора экспрессии *FOLH1* – *CEBPB* – на экспрессию *FOLH1* в эмбриональных стволовых клетках человека (данные в клетках простаты отсутствуют) тоже заслуживает внимания, поскольку *CEBPB* в опухоли простаты позволяет клеткам пережить депривацию андрогенов при терапии. Нокадаун *SPDEF* в MCF7 привел к 3,4-кратному увеличению уровня транскрипции *FOLH1* и, что любопытно, примерно трехкратному падению *FOLH1B*. *SPDEF* имеет повышенную экспрессию в раке простаты, положительно ассоциирован со стадией опухоли, является маркером негативного прогноза в раке простаты[122]. Нокадаун гена другого потенциального регулятора экспрессии *FOLH1*, *RBFOX2*, в печени приводит к пятикратному падению транскрипции *FOLH1*; сам *RBFOX2* интересен ключевым участием в тканеспецифичной регуляции мезенхимального сплайсинга в том числе и в клетках рака простаты[123].

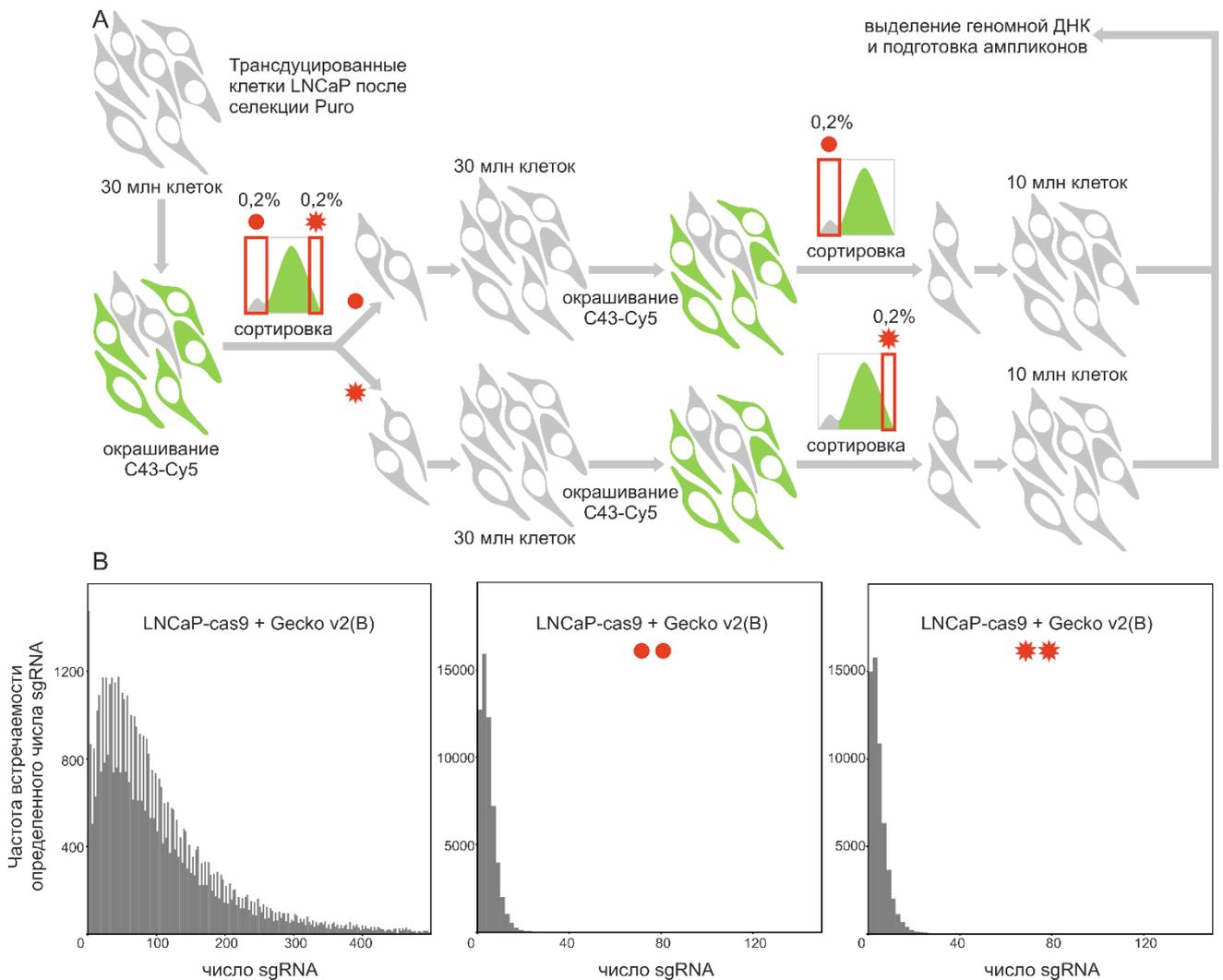
### **Вероятные регуляторы экспрессии *FOLH1* по результатам скрининга библиотеки гидовых РНК Gecko v2**

Библиотека Gecko v2 (genome-scale CRISPR-Cas9 knockout) – лентивирусная полногеномная библиотека нокаутов, которая включает в себя гидовые РНК (sgRNA) против всех активно транскрибируемых генов человека. По увеличению представленности sgRNA против тех или иных генов в модифицированных клетках LNCaP, экспрессирующих Cas9 (LNCaP-Cas9), можно судить об их участии в процессе, изучаемом в рамках скрининга. Благодаря **V15-Cy5** было возможно отслеживать изменение экспрессии *FOLH1* по флуоресцентному сигналу от конъюгата, связавшегося с GCPII.

Следуя идее, предложенной в работе по поиску мишени моноклональных антител, мы провели двойной отбор как несветящихся после окрашивания клеток LNCaP-Cas9, потерявших GСРІІ, так и самых ярких клеток, где количество GСРІІ могло увеличиться. Схема эксперимента представлена на Рис.18А. Чтобы попытаться нивелировать влияние клеточного цикла и стохастических факторов, а также обогатить популяцию клеток теми нокаутами, что важны для экспрессии *FOLH1*, была выбрана двойная сортировка. В ходе сортировки собирались как клетки с наименьшим уровнем флуоресцентного сигнала, так и с наибольшим. В первом случае предполагали нокауты генов-«активаторов», во втором – генов-«репрессоров».

Как видно из Рис.18В,С, в клетках линии LNCaP присутствовала популяция клеток, которые не окрашиваются **V15-Cy5**. Это не обособленная популяция с четко выраженным пиком, соответствующим неокрашенному контролю, но ее можно условно выделить на графике как пересечение областей под гистограммами распределения сигнала контроля и окрашенных клеток (Рис.15В,С). Увеличение концентрации конъюгата не позволяло окрасить клетки этой популяции, а популяция хорошо окрашиваемых клеток, отобранная однократно на клеточном сортере, при культивировании и последующем окрашивании вновь «распадалась» до исходного состояния, что может указывать на влияние клеточного цикла. Чтобы попытаться нивелировать влияние клеточного цикла и стохастических факторов, а также обогатить популяцию клеток LNCaP-Cas9 теми нокаутами, что важны для экспрессии *FOLH1*, была использована двойная последовательная сортировка с отбором наиболее и наименее флуоресцирующих клеток, трансдуцированных библиотекой последовательностей гидовых РНК Gecko v2.

Клетки LNCaP-Cas9 после трансдукции и селекции продемонстрировали хорошую репрезентативность библиотеки – от 96,8% до 99,58% генов в зависимости от реплики, что соответствует потерям от 3,2% до 0,42% генов (Рис.18В). Для полубιβлиотеки В во всех трех репликах потеряно 44 общих гена, в двух -130. Есть несколько объяснений подобных потерь: низкое значение MOI (multiplicity of infection) при трансдукции, летальность нокаута важных генов, неспецифичная активность Cas9.



**Рис.18.** Скрининг полногеномной библиотеки нокаутов *Gecko v2* в клетках линии *LNCaP-Cas9*. А) Схема двойного отбора клеток с интересующими нас параметрами флуоресценции. В) Пример гистограммы распределения ридов на одной из реплик. Слева направо: клетки с библиотекой до сортировки, дважды отсортированная несветящаяся фракция клеток, дважды отсортированная светящаяся фракция клеток. Репрезентативность исходной библиотеки в отобранных после заражения клетках – от 96,8% до 99,58% генов в зависимости от реплики.

Так как в рамках скрининга полногеномной библиотеки желательно, чтобы в одну клетку попало не более одной *sgRNA*, значение *MOI* (количество вирусных частиц на 1 клетку при трансдукции) подбиралось в диапазоне 0,1-0,3, дабы уменьшить вероятность попадания двух и более *sgRNA* в клетку. Низкое значение *MOI* приводит к уменьшению покрытия библиотеки, что приводит к уменьшению репрезентативности. Показано, что репрезентативность библиотеки снижается также из-за летальности нокаута генов, важных для клеточного метаболизма, деления и выживания[124]. Другим фактором, влияющим на репрезентативность, является неспецифическая активность *Cas9*. К ней можно в том числе отнести появление Non-

targeting control в лидерах среди sgRNA фракций светящихся и несветящихся клеток после двойной сортировки, это встречается и в других работах [125].

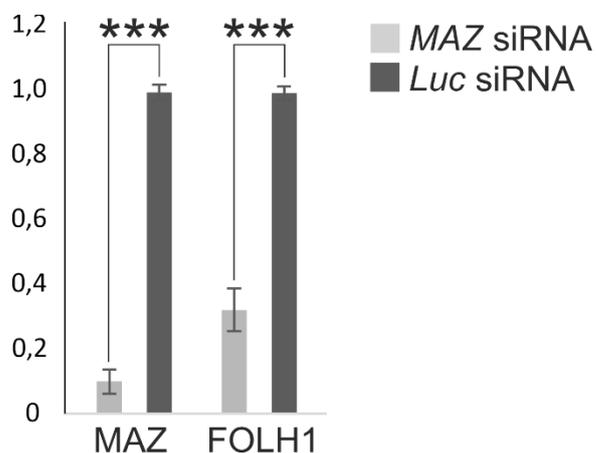
Геномная ДНК из отобранных фракций клеток была использована для амплификации участков, содержащих последовательности sgRNA. Полученные таким образом ампликоны были отсеквенированы. Результаты секвенирования были обработаны с помощью Mageck Tool и на основании анализа обогащения sgRNA было выбрано порядка 40 генов с p-value менее 0.05, присутствовавших в двух и более репликах из трех, которые далее фильтровались по следующим критериям:

- 1) Известные из литературы или баз данных функции белка как активатора или репрессора экспрессии генов не противоречат предполагаемой функции белка при скрининге библиотеки гидовых РНК для нокаутов Gesco2. Если предполагаемый регулятор найден в популяции несветящихся клеток, то его функция условно определяется как «активатор», т.к. его экспрессия положительно коррелирует с экспрессией *FOLH1*. Если регулятор найден в популяции светящихся клеток, то его функция условно определяется как «репрессор», т.к. его экспрессия отрицательно коррелирует с экспрессией *FOLH1*.
- 2) Наличие достоверной корреляции с экспрессией *FOLH1*: положительной, если речь идет о популяции несветящихся клеток, и отрицательной, если нокаут гена обнаружен в популяции светящихся клеток. Допускается низкое значение коэффициента корреляции.
- 3) Наличие достоверной разницы в уровне экспрессии гена в *FOLH1*<sup>+</sup> и *FOLH1*<sup>-</sup> выборках образцов опухоли простаты с направлением изменения экспрессии, соответствующим изменению экспрессии *FOLH1* для «активаторов», и с обратным направлением для «репрессоров». Использовались скрипты 1 и 2.

По этим критериям было выбрано 7 генов – *ATPAF1*, *TOPORS*, *BCL2L14*, *GPM6B*, *NKG7*, *SCUBE1*, *SERPIN5*.

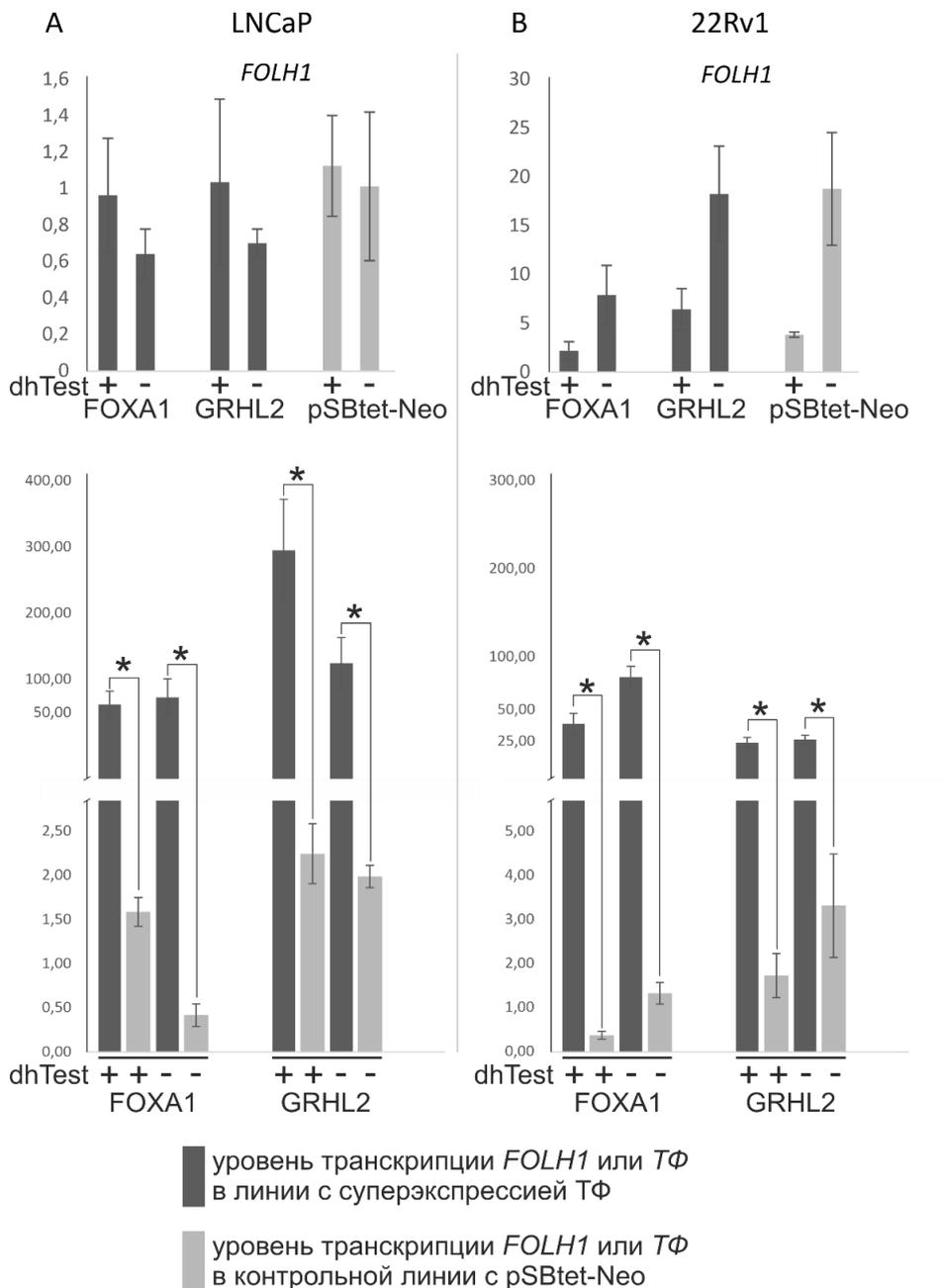
### **Проверка части потенциальных регуляторов экспрессии *FOLH1***

Были подобраны условия для нокадауна генов в клеточных линиях LNCaP, 22Rv1, PC-3 с помощью РНК интерференции. Нокадауны *FOSL1*, *FOS*, *STAT1*, *NKX3.1*, *BMPR1B*, *FOXA1*, *NFYA*, *HOXB13*, *STEAP2*, *SPDEF*, *TFAP4*, *RBFOX2*, *CEBPB* в линиях рака простаты LNCaP, 22Rv1, PC-3 не приводили к значимому изменению уровня экспрессии *FOLH1*. Нокадаун *MAZ* привел к уменьшению экспрессии *FOLH1* в клетках 22Rv1 (Рис.19).



**Рис.19.** Влияние нокдауна гена *MAZ* на экспрессию *FOLH1* в линии *22Rv1*. \*\*\*  $p$ -value <  $10^{-3}$ . Люциферазная (*Luc*) *siRNA* использована в качестве контроля.

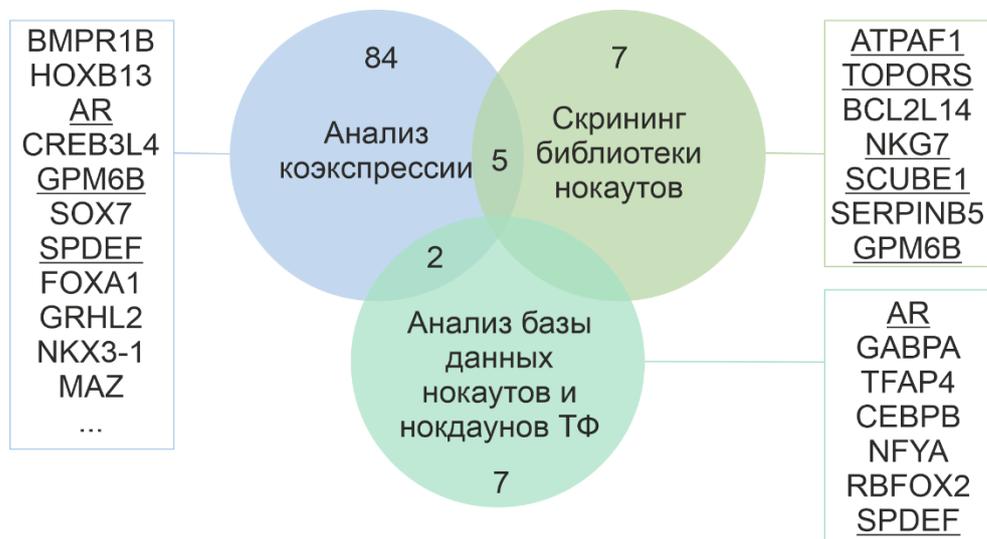
Были сделаны генноинженерные конструкции для суперэкспрессии транскрипционных факторов *FOXA1* и *GRHL2* на основе системы транспозонов *Sleeping Beauty*. Получены клеточные линии с суперэкспрессией их мРНК под контролем индуцируемого промотора, а также контрольные клеточные линии со вставкой исходного фрагмента плазмиды *pSBtet-Neo*. Суперэкспрессия *FOXA1* и *GRHL2* не оказывала значимого влияния на экспрессию *FOLH1* в клеточных линиях *LNCaP*, *22Rv1* (Рис.20).



**Рис.20.** Суперэкспрессия транскрипционных факторов *FOXA1* и *GRHL2* не приводила к изменению экспрессии гена *FOLH1* в клеточных линиях опухоли простаты *LNCaP* (A) и *22Rv1*(B). Для A и B сверху – уровень экспрессии *FOLH1* в линиях с суперэкспрессией *TF*, снизу – уровень экспрессии *TF* в клеточной линии соответствующего *TF*. *pSBtet-Neo* – исходная плазмида, контроль. +/- - с добавлением дигидротестостерона или без него.

## Обсуждение результатов

Потенциальные регуляторы экспрессии *FOLH1* отобрано на основании скрининга полногеномной библиотеки нокаутов, корреляционного анализа и анализа баз данных, опубликованных предсказаний регуляторных сетей в раке простаты. Объединение потенциальных регуляторов экспрессии *FOLH1*, найденных с использованием разных подходов, дает пул из нескольких десятков генов (Рис.21, Табл.6).



**Рис.21.** Диаграмма Венна для потенциальных регуляторов экспрессии *GCP11*. Синий круг – результаты анализа дифференциальной экспрессии в клеточных линиях, образцах опухолей простаты и литературы. Зеленый круг – результаты анализа базы данных нокаутов и нокадаунов *KnockTF*. Хаки – результат скрининга полногеномной библиотеки нокаутов. Подчеркнуты белки, отобранные с использованием разных подходов.

Проверка влияния нокадаунов шести (*FOSL1*, *FOS*, *FOXA1*, *STAT1*, *AR*, *MAZ*) и суперэкспрессии двух (*FOXA1*, *GRHL2*) генов потенциальных регуляторов на экспрессию *FOLH1* в клеточных линиях LNCaP, 22Rv1, PC-3 показала, что нокадаун *MAZ* привел к уменьшению экспрессии *FOLH1* в клетках 22Rv1 (Рис.19). Проверка таких регуляторов, как *FOXA1* и *AR*, ключевых для рака простаты, не продемонстрировала значительной связи их уровня транскрипции с уровнем транскрипции *FOLH1*. При этом данные для *AR* сходятся с описанной в литературе слабой зависимостью транскрипции *FOLH1* от *AR*. Ввиду связи *FOXA1* с *AR* было также изучено их совместное влияние на экспрессию *FOLH1* с помощью добавления дигидротестостерона – андрогена, взаимодействующего с *AR*, однако кумулятивного эффекта обнаружено не было. Учитывая достоверную, но слабую репрессию транскрипции *FOLH1*, наблюдаемую при добавлении в среду к клеткам дигидротестостерона, это позволяет предположить, что высокая экспрессия *FOLH1* не обусловлена повышенной активностью андрогенового сигнального пути. Проверенные вероятные регуляторы экспрессии *FOLH1* были выбраны при анализе транскрипционных данных из COSMIC и корреляций экспрессии генов с *FOLH1* (CCLE). Несмотря на обнаружение одного из регуляторов транскрипции *FOLH1*, *MAZ*, что подтверждает возможность их поиска по коэкспрессии, эффективность этого подхода недостаточна. Поэтому далее обсуждаются следующие подходы: анализ дифференциальной экспрессии генов в образцах опухоли и клеточных линиях, различающихся уровнем транскрипции *FOLH1*, а также транскриптомных данных нокаутов и нокадаунов

транскрипционных факторов и скрининг библиотеки последовательностей гидовых РНК Gecko v2.

Разбиение транскриптомных данных образцов опухоли простаты и клеточных линий на группы с высокой и низкой экспрессией *FOLH1*, определение дифференциально экспрессирующихся генов транскрипционных факторов и протеинкиназ в этих группах и значений коэффициента корреляции между ними позволило идентифицировать как отдельные потенциальные регуляторы, так и их пары, в которых белки могут действовать совместно или заменять друг друга. *BMPR1B* и *HOXB13* отобраны в качестве пары активаторов экспрессии *FOLH1* как в *FOLH1*-экспрессирующих клеточных линиях человека, так и в образцах опухоли простаты с высокой экспрессией *FOLH1* (Табл.7 и 8). Пересечение результатов анализа образцов опухолей простаты и клеточных линий человека дало несколько простатспецифичных генов (Табл.7) - *HIC1*, *GLIS1*, *IRX4*, *BMPR1B*, *HOXB13*.

При анализе транскриптомных данных нокаутов и нокдаунов транскрипционных факторов в клеточных линиях рассматривалось влияние регуляторов не только на *FOLH1*, но и на *FOLH1B*, его паралог, имеющий такой же внутриинтронный энхансер, что и у *FOLH1*. На основании такого сопоставления можно было строить предположения о том, действует тот или иной регулятор на промотор *FOLH1* или на его энхансер. Так, нокдаун *GABPA* значительно влияет на экспрессию *FOLH1*, но не на *FOLH1B*, нокдаун *TFAP4* приводит к одинаковому уменьшению уровня транскрипции *FOLH1* и его паралога, а влияние нокдауна *SPDEF* приводит к изменению экспрессии *FOLH1* и *FOLH1B* в разных направлениях. Можно предположить, что *GABPA* взаимодействует с промотором *FOLH1*, *TFAP4* – с внутриинтронным энхансером, а влияние *SPDEF* может зависеть как от того, связывается ли он с промотором или энхансером, так и от модуляции его кофакторами.

При скрининге библиотеки нокаутов мишенью найденных «активаторов» и «репрессоров» может выступать не только транскрипция *FOLH1*, но и сплайсинг, пост-трансляционные модификации как продукта гена *FOLH1*, так и транскрипционных факторов (ТФ), которые могут быть ответственны за инициацию транскрипции *FOLH1*.

Известно, что нокдаун *ATPAF1* (ATP synthase mitochondrial F1 complex assembly factor 1) серьезно ухудшает рост LNCaP в условиях депривации андрогенов [126]. Экспрессия гена *ATPAF1* положительно коррелирует с экспрессией *FOLH1* в опухоли простаты, его экспрессия достоверно отличается в *FOLH1+* и *FOLH1-* выборках образцов опухоли простаты. TOPORS (TOP1 Binding Arginine/Serine Rich Protein, E3 убиквитин лигаза) убиквитинирует белок-супрессор опухоли NKX3.1, что приводит к протеасомной деградации NKX3.1 [127]. Другой регулятор, BCL2L14 (Apoptosis facilitator Bcl-2-like protein 14), теряется в результате делеции в

опухоли простаты: отсутствует в 11-40% образцов на ранних стадиях заболевания и в 35-95% образцов при прогрессирующей опухоли простаты [128]. Экспрессия *GPM6B* (Glycoprotein M6B) уменьшается в опухоли простаты по сравнению с нормой простаты и негативно коррелирует со стадиями прогрессии опухоли [129]. Белок GPM6B может связываться с SERT (транспортер серотонина) и приводить к значительному снижению поглощения серотонина [130], что замедляет рост клеток опухоли простаты DU-145 и 22Rv1 [129]. Экспрессия другого потенциального регулятора экспрессии *FOLH1* – SCUBE1 (Signal Peptide, CUB Domain And EGF Like Domain Containing 1) в фибробластах, ассоциированных с опухолью простаты, приводит к двукратному уменьшению скорости роста опухоли и к трехкратному уменьшению ее размера [131]. Экспрессия «репрессора» SERPINB5 (Serpin Family B Member 5) ассоциирована с положительным прогнозом; SERPINB5 является супрессором опухоли, теряется в опухоли простаты [132]. Найденные белки ранее не описывались как вероятные участники регуляции транскрипции *FOLH1*. Среди генов, чьи sgRNA были утеряны после трансдукции клеточной линии LNCaP-Cas9 ввиду, например, летальности их нокаутов, KLF2, согласно данным KnockTF, является потенциальным репрессором транскрипции гена *FOLH1*. Однако критериям, связанным с анализом транскриптомных данных, примененным в данной работе, KLF2 не соответствовал, поэтому в качестве потенциального регулятора транскрипции *FOLH1* не рассматривался. Некоторые регуляторы из этого пула оказались и в списке регуляторов, отобранных с использованием иных критериев (Табл.6, Рис.): это ATPAF1, TOPORS, NKG7, SCUBE1, GPM6B. Большинство из этих белков не являются ни транскрипционными факторами, ни протеинкиназами, однако известно их влияние на рост клеток опухоли простаты.

Обнаружение новых регуляторов экспрессии *FOLH1* позволит уточнить схему транскрипционной регуляции в раке простаты и обнаружить новые гены и белки-мишени для диагностики и терапии рака простаты.

## Заключение

В ходе данной работы впервые были определены ингибирующие свойства свыше сорока синтезированных нашими коллегами лигандов GСРП с различными комбинациями заместителей, способных связываться с несколькими участками в структуре GСРП. Это позволило определить зависимость активности лигандов GСРП от их структуры.

Анализ взаимосвязи структура-функция лигандов GСРП позволил отобрать лиганды с  $K_i$  в наномолярном диапазоне и создать их эффективные конъюгаты с цитотоксическими и флуоресцентными молекулами.

Лиганд GСРП **V15-MMAE**, конъюгированный с монометил ауристатином E, был менее цитотоксичен, чем монометил ауристатин E, и при его использовании наблюдалось изменение цитотоксичности в сторону GСРП-экспрессирующих клеток. Конъюгация лиганда GСРП с флуоресцентными молекулами SulfoCy5 и SulfoCy7 привела к небольшому увеличению значений  $K_i$ . С помощью проточной цитофлуориметрии и флуоресцентной микроскопии показано эффективное окрашивание GСРП-экспрессирующих клеток лигандом, конъюгированным с SulfoCy5. Возможность окрашивания конъюгатами с SulfoCy5 и SulfoCy7 также показана нашими коллегами на мышах с ксенографтной опухолью из GСРП-экспрессирующих клеток 22Rv1. Специфичность **V15-Cy5** к GСРП позволила использовать его для поиска потенциальных регуляторов экспрессии гена *FOLH1* в комбинации с библиотекой последовательностей гидовых РНК для системы CRISPR-Cas9.

На основании анализа транскрипционных данных, в том числе транскриптомов образцов с нокаутами и нокдаунами транскрипционных факторов, и скрининга полногеномной библиотеки нокаутов Gecko v2 сформирован пул генов, кодирующих потенциальные белки-регуляторы экспрессии *FOLH1* (Табл.6). Количество регуляторов, идентифицированных с использованием разных подходов, превышает ожидаемое число пересечений при случайном выборе белков. Так, найденные в ходе скрининга библиотеки нокаутов регуляторы ATPAF1, TOPORS, NKG7, SCUBE1, GPM6B также соответствовали критериям, использованным при анализе корреляций экспрессии. Также, нокдауны транскрипционных факторов AR и SPDEF приводили к изменению экспрессии *FOLH1* согласно транскриптомным данным, и при этом результатах анализа корреляции экспрессии генов, так и данных образцов с нокдаунами. Только часть пула регуляторов связана с известными регуляторными каскадами в опухоли простаты: FOXA1, GATA2, GRHL2, CREB3L4, HOXB13, NKX3.1, GABPA являются партнерами андрогенового рецептора AR и могут действовать в рамках каскада андрогенового рецептора. Среди проверенных регуляторов транскрипционный фактор MAZ являлся слабым активатором транскрипции *FOLH1* по данным нокдауна. Нокдауны генов *FOSL1*, *FOS*, *NKX3.1*, *FOXA1*,

*NFYA, STEA2, SPDEF, TFAP4, RBFOX2, CEBPG* в линиях опухоли простаты LNCaP, 22Rv1, PC-3 не приводили к изменению уровня экспрессии *FOLH1*. Суперэкспрессия FOXA1 и GRHL2 не оказывала значимого влияния на экспрессию *FOLH1* в клеточных линиях LNCaP, 22Rv1, PC-3.

Анализ регуляторов экспрессии *FOLH1* и установление их принадлежности к регуляторным каскадам в простате и опухоли простаты позволят использовать их в качестве мишеней для диагностики и терапии заболеваний в будущем.

## Выводы

1. В результате проведенного структурно-функционального анализа низкомолекулярных лигандов глутаматкарбоксипептидазы II (GCP II) выявлены соединения с  $K_i$  в наномолярном диапазоне. Показано, что при варьировании части их структурных элементов сохранялись ингибирующие свойства, что упрощает дальнейшую оптимизацию фармакологических характеристик.
2. По сравнению с монометил ауристатином E, цитотоксичность его конъюгата с лигандом GCP II **V15** уменьшалась для клеток, не экспрессирующих GCP II, при общем снижении цитотоксичности *in vitro*.
3. Лиганд **V15**, конъюгированный с флуорофором SulfoCy5, окрашивал в наномолярных концентрациях GCP II-экспрессирующие клетки опухоли простаты, а также имел низкую цитотоксичность.
4. Предложен метод поиска регуляторов экспрессии гена *FOLH1* с помощью сортировки по интенсивности окрашивания лигандом **V15-Cy5** пула клеток с GCP II, трансдуцированных вирусами с гидовыми РНК для нокауту всех функциональных генов.
5. Идентифицировано *in silico* несколько десятков вероятных регуляторов экспрессии *FOLH1*, один из которых, транскрипционный фактор MAZ, валидирован как активатор транскрипции *FOLH1* в клетках опухоли простаты 22Rv1.

## Публикации по теме диссертации в рецензируемых изданиях

- [1] Synthesis and biological evaluation of psma ligands with aromatic residues and fluorescent conjugates based on them / A. E. Machulkin\*, **R. R. Shafikov\***, A. A. Uspenskaya et al. // *Journal of Medicinal Chemistry*. — 2021. — Vol. 64, no. 8. — P. 4532–4552. Impact Factor = 7.446 (Web of Science)  
\*равный вклад авторов.
- [2] Influence of dipeptide linker configuration on the activity of psma ligands / A. Uspenskaya, A. E. Machulkin, E. A. Nimenko, **R.R. Shafikov** et al. // *Mendeleev Communications*. — 2020. — Vol. 30. — P. 756–759. Impact Factor = 1.77(Web of Science)
- [3] Synthesis and initial in vitro evaluation of psma-targeting ligands with a modified aromatic moiety at the lysine e-nitrogen atom / N. Y. Zyk, A. P. Ber, E. A. Nimenko, **R.R. Shafikov** et al. // *Bioorganic and Medicinal Chemistry Letters*. — 2022. — P. 128840. Impact Factor = 2.823 (Web of Science)
- [4] Synthesis, characterization and preclinical evaluation of small-molecule prostate-specific membrane antigen targeted monomethyl auristatin e conjugate / A. E. Machulkin, A. A. Uspenskaya, N. U. Zyk, E.A. Nimenko, A.P. Ber, S.A. Petrov, V.I. Polshakov, **R.R. Shafikov** et al. // *Journal of Medicinal Chemistry*. — 2021. — Vol. 64, no. 23. — P. 17123–17145. Impact Factor =7.446 (Web of Science)

## Другие публикации по теме диссертации

- [1] Synthesis and biological evaluation of psma targeted fluorescent conjugates / A. Machulkin, **R. Shafikov**, A. Uspenskaya et al. // 26th EFMC International Symposium on Medicinal Chemistry (EFMC-ISMС-2021). — 2021. — P. V002.
- [2] Синтез и оптимизация структуры линкеров лигандов ПСМА и получение диагностических и терапевтических конъюгатов / А. А. Успенская, Е. А. Нищенко, **Р. Р. Шафиков** и др. // МАРКОВНИКОВСКИЕ ЧТЕНИЯ: ОРГАНИЧЕСКАЯ ХИМИЯ ОТ МАРКОВНИКОВА ДО НАШИХ ДНЕЙ. — 2022. — С. 92–93.
- [3] Конъюгат монометил ауристатина Е на основе лиганда ПСМА: синтез и доклиническое исследование / А. Э. Мачулкин, А. А. Успенская, **Р. Р. Шафиков** и др. // Всероссийский конгресс по химии гетероциклических соединений (KOST2021) Сборник тезисов. — 2021. — С. 235–235.
- [4] Синтез и биологические испытания конъюгатов ПСМА с диагностическими и терапевтическими агентами / А.А. Успенская, А.Э. Мачулкин, С.А. Петров, А.П. Бер, Е.А. Нищенко, Н.Ю. Зык, Д.А. Скворцов, **Р.Р. Шафиков** и др. // МАРКОВНИКОВСКИЕ ЧТЕНИЯ: ОРГАНИЧЕСКАЯ ХИМИЯ ОТ МАРКОВНИКОВА ДО НАШИХ ДНЕЙ (WSOC-2020). — 2020. — С. 91–91.

## Список литературы

- [1] S. Pentylala *et al.*, “Prostate cancer markers: An update (Review),” *Biomedical Reports*, vol. 4, no. 3. Spandidos Publications, pp. 263–268, Mar. 01, 2016. doi: 10.3892/br.2016.586.
- [2] S. Bravaccini *et al.*, “PSMA expression: A potential ally for the pathologist in prostate cancer diagnosis,” *Sci Rep*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41598-018-22594-1.
- [3] M. C. Hupe *et al.*, “Expression of prostate-specific membrane antigen (PSMA) on biopsies is an independent risk stratifier of prostate cancer patients at time of initial diagnosis,” *Front Oncol*, vol. 8, no. DEC, 2018, doi: 10.3389/fonc.2018.00623.
- [4] J. S. Batra, K. J. Pienta, M. G. Pomper, M. A. Gorin, and S. P. Rowe, “Can the interplay between androgen signaling and PSMA expression be leveraged for theranostic applications?,” *Translational Andrology and Urology*, vol. 8. AME Publishing Company, pp. S263–S264, 2019. doi: 10.21037/tau.2019.03.13.
- [5] T. M. Gorges *et al.*, “Heterogeneous PSMA expression on circulating tumor cells-a potential basis for stratification and monitoring of PSMA-directed therapies in prostate cancer.” [Online]. Available: [www.impactjournals.com/oncotarget](http://www.impactjournals.com/oncotarget)
- [6] P. Mhawech-Fauceglia *et al.*, “Prostate-specific membrane antigen (PSMA) protein expression in normal and neoplastic tissues and its sensitivity and specificity in prostate adenocarcinoma: An immunohistochemical study using mutiple tumour tissue microarray technique,” *Histopathology*, vol. 50, no. 4, pp. 472–483, Mar. 2007, doi: 10.1111/j.1365-2559.2007.02635.x.
- [7] M. B. Robinson\$, R. D. Blakelys, R. Couto, and J. T. Coyleenii, “Hydrolysis of the Brain Dipeptide N-Acetyl-L-aspartyl-L-glutamate,” 1987.
- [8] B. Fowler, “The folate cycle and disease in humans.”
- [9] R. E. Conway *et al.*, “Prostate-specific membrane antigen (PSMA)-mediated laminin proteolysis generates a pro-angiogenic peptide,” *Angiogenesis*, vol. 19, no. 4, pp. 487–500, Oct. 2016, doi: 10.1007/s10456-016-9521-x.
- [10] H. Betz, “NAAG as a neurotransmitter,” vol. 1965, pp. 317–325, 2001.
- [11] J. H. Neale, “N-Acetylaspartylglutamate is an agonist at mGluR3 in vivo and in vitro,” *Journal of Neurochemistry*, vol. 119, no. 5. pp. 891–895, Dec. 2011. doi: 10.1111/j.1471-4159.2011.07380.x.
- [12] J. Zhao, E. Ramadan, M. Cappiello, B. Wroblewska, T. Bzdega, and J. H. Neale, “NAAG inhibits KCl-induced [3H]-GABA release via mGluR3, cAMP, PKA and L-type calcium conductance,” *European Journal of Neuroscience*, vol. 13, no. 2, pp. 340–346, 2001, doi: 10.1046/j.1460-9568.2001.01396.x.
- [13] E. R. G. Sanabria, K. M. Wozniak, B. S. Slusher, and A. Keller, “GCP II (NAALADase) inhibition suppresses mossy fiber-CA3 synaptic neurotransmission by a presynaptic mechanism.,” *J Neurophysiol*, vol. 91, no. 1, pp. 182–93, 2004, doi: 10.1152/jn.00465.2003.
- [14] A. G. Thomas *et al.*, “Neuroprotection mediated by glutamate carboxypeptidase II (NAALADase) inhibition requires TGF-??,” *Eur J Pharmacol*, vol. 430, no. 1, pp. 33–40, 2001, doi: 10.1016/S0014-2999(01)01239-0.
- [15] V. Bruno, G. Battaglia, G. Casabona, A. Copani, F. Caciagli, and F. Nicoletti, “Neuroprotection by Glial Metabotropic Glutamate Receptors is Mediated by Transforming Growth Factor-β,” *J Neurosci*, vol. 18, no. 23, pp. 9594–9600, 1998.
- [16] J. Zhou, J. H. Neale, M. G. Pomper, and A. P. Kozikowski, “NAAG peptidase inhibitors and their potential for diagnosis and therapy.,” *Nat Rev Drug Discov*, vol. 4, no. 12, pp. 1015–1026, 2005, doi: 10.1038/nrd1903.

- [17] A. A. Kritis, E. G. Stamoula, K. A. Paniskaki, and T. D. Vavilis, “Researching glutamate – induced cytotoxicity in different cell lines: A comparative/collective analysis/study,” *Front Cell Neurosci*, vol. 9, Mar. 2015, doi: 10.3389/fncel.2015.00091.
- [18] J. Zhou, J. H. Neale, M. G. Pomper, and A. P. Kozikowski, “NAAG peptidase inhibitors and their potential for diagnosis and therapy,” *Nature Reviews Drug Discovery*, vol. 4, no. 12. pp. 1015–1026, Dec. 2005. doi: 10.1038/nrd1903.
- [19] U. Sommer *et al.*, “Impact of Androgen Receptor Activity on Prostate-Specific Membrane Antigen Expression in Prostate Cancer Cells,” *Int J Mol Sci*, vol. 23, no. 3, Feb. 2022, doi: 10.3390/ijms23031046.
- [20] V. Yao, C. E. Berkman, J. K. Choi, D. S. O’Keefe, and D. J. Bacich, “Expression of prostate-specific membrane antigen (PSMA), increases cell folate uptake and proliferation and suggests a novel role for PSMA in the uptake of the non-polyglutamated folate, folic acid,” *Prostate*, vol. 70, no. 3, pp. 305–316, Feb. 2010, doi: 10.1002/pros.21065.
- [21] M. E. Perico *et al.*, “Prostate-specific membrane antigen (PSMA) assembles a macromolecular complex regulating growth and survival of prostate cancer cells ‘in vitro’ and correlating with progression ‘in vivo.’” [Online]. Available: [www.impactjournals.com/oncotarget](http://www.impactjournals.com/oncotarget)
- [22] M. Colombatti *et al.*, “The prostate specific membrane antigen regulates the expression of IL-6 and CCL5 in prostate tumour cells by activating the MAPK pathways,” *PLoS One*, vol. 4, no. 2, 2009, doi: 10.1371/journal.pone.0004608.
- [23] R. E. Conway, N. Petrovic, Z. Li, W. Heston, D. Wu, and L. H. Shapiro, “Prostate-Specific Membrane Antigen Regulates Angiogenesis by Modulating Integrin Signal Transduction,” *Mol Cell Biol*, vol. 26, no. 14, pp. 5310–5324, Jul. 2006, doi: 10.1128/mcb.00084-06.
- [24] G. Anilkumar, S. A. Rajasekaran, S. Wang, O. Hankinson, N. H. Bander, and A. K. Rajasekaran, “Prostate-specific Membrane Antigen Association with Filamin A Modulates Its Internalization and NAALADase Activity 1,” 2003. [Online]. Available: <http://aacrjournals.org/cancerres/article-pdf/63/10/2645/2503488/ch1003002645.pdf>
- [25] E. R. Horton, J. D. Humphries, J. James, M. C. Jones, J. A. Askari, and M. J. Humphries, “The integrin adhesome network at a glance,” *J Cell Sci*, vol. 129, no. 22, pp. 4159–4163, 2016, doi: 10.1242/jcs.192054.
- [26] R. E. Conway *et al.*, “Prostate specific membrane antigen produces pro-angiogenic laminin peptides downstream of matrix metalloprotease-2,” *Angiogenesis*, vol. 16, no. 4, pp. 847–860, Oct. 2013, doi: 10.1007/s10456-013-9360-y.
- [27] L. A. Caromile *et al.*, “PSMA redirects cell survival signaling from the MAPK to the PI3K-AKT pathways to promote the progression of prostate cancer,” *Sci Signal*, vol. 10, no. 470, Mar. 2017, doi: 10.1126/scisignal.aag3326.
- [28] L. A. Caromile *et al.*, “PSMA redirects cell survival signaling from the MAPK to the PI3K-AKT pathways to promote the progression of prostate cancer,” *Sci Signal*, vol. 10, no. 470, Mar. 2017, doi: 10.1126/scisignal.aag3326.
- [29] C. L. Grant *et al.*, “Prostate specific membrane antigen (PSMA) regulates angiogenesis independently of VEGF during ocular neovascularization,” *PLoS One*, vol. 7, no. 7, Jul. 2012, doi: 10.1371/journal.pone.0041285.
- [30] A. E. Machulkin *et al.*, “Small-molecule PSMA ligands. Current state, SAR and perspectives,” *Journal of Drug Targeting*, vol. 24, no. 8. Taylor and Francis Ltd, pp. 679–693, Sep. 13, 2016. doi: 10.3109/1061186X.2016.1154564.

- [31] K. Rahbar, A. Afshar-Oromieh, H. Jadvar, and H. Ahmadzadehfar, "PSMA Theranostics: Current Status and Future Directions," *Molecular Imaging*, vol. 17. SAGE Publications Inc., Jun. 01, 2018. doi: 10.1177/1536012118776068.
- [32] M. Navrátil *et al.*, "Structural and biochemical characterization of the folyl-poly- $\gamma$ -L-glutamate hydrolyzing activity of human glutamate carboxypeptidase II," *FEBS Journal*, vol. 281, no. 14, pp. 3228–3242, 2014, doi: 10.1111/febs.12857.
- [33] N. D. Rawlings and A. J. Barrett, "Structure of membrane glutamate carboxypeptidase," 1997.
- [34] N. Schü *et al.*, "The homodimer of prostate-specific membrane antigen is a functional target for cancer therapy," 2003. [Online]. Available: [www.pnas.org](http://www.pnas.org).
- [35] C. Barinka *et al.*, "Structural Basis of Interactions between Human Glutamate Carboxypeptidase II and Its Substrate Analogs," *J Mol Biol*, vol. 376, no. 5, pp. 1438–1450, Mar. 2008, doi: 10.1016/j.jmb.2007.12.066.
- [36] C. J. Choy *et al.*, "Rationally designed sulfamides as glutamate carboxypeptidase II inhibitors," *Chem Biol Drug Des*, vol. 82, no. 5, pp. 612–619, Nov. 2013, doi: 10.1111/cbdd.12174.
- [37] D. Shen, F. Xie, and W. B. Edwards, "Evaluation of Phage Display Discovered Peptides as Ligands for Prostate-Specific Membrane Antigen (PSMA)," *PLoS One*, vol. 8, no. 7, Jul. 2013, doi: 10.1371/journal.pone.0068339.
- [38] A. P. Kozikowski *et al.*, "Design of remarkably simple, yet potent urea-based inhibitors of glutamate carboxypeptidase II (NAALADase) [1]," *Journal of Medicinal Chemistry*, vol. 44, no. 3, pp. 298–301, Feb. 01, 2001. doi: 10.1021/jm000406m.
- [39] C. Barinka *et al.*, "Interactions between human glutamate carboxypeptidase II and urea-based inhibitors: Structural characterization," *J Med Chem*, vol. 51, no. 24, pp. 7737–7743, Dec. 2008, doi: 10.1021/jm800765e.
- [40] D. v Ferraris, K. Shukla, and T. Tsukamoto, "Structure-Activity Relationships of Glutamate Carboxypeptidase II (GCPII) Inhibitors," 2012. [Online]. Available: <http://pymol.org>
- [41] J. J. Vornov *et al.*, "Still NAAG'ing After All These Years. The Continuing Pursuit of GCPII Inhibitors.," in *Advances in Pharmacology*, vol. 76, Academic Press Inc., 2016, pp. 215–255. doi: 10.1016/bs.apha.2016.01.007.
- [42] C. Bařinka *et al.*, "Structural insight into the pharmacophore pocket of human glutamate carboxypeptidase II," *J Med Chem*, vol. 50, no. 14, pp. 3267–3273, Jul. 2007, doi: 10.1021/jm070133w.
- [43] A. X. Zhang *et al.*, "A Remote arene-binding site on prostate specific membrane antigen revealed by antibody-recruiting small molecules," *J Am Chem Soc*, vol. 132, no. 36, pp. 12711–12716, Sep. 2010, doi: 10.1021/ja104591m.
- [44] J. R. Mesters *et al.*, "Structure of glutamate carboxypeptidase II, a drug target in neuronal damage and prostate cancer," *EMBO Journal*, vol. 25, no. 6, pp. 1375–1384, Mar. 2006, doi: 10.1038/sj.emboj.7600969.
- [45] J. Pavlíek, J. Ptáek, and C. Baina, "Glutamate Carboxypeptidase II: An Overview of Structural Studies and Their Importance for Structure-Based Drug Design and Deciphering the Reaction Mechanism of the Enzyme," 2012.
- [46] C. Barinka *et al.*, "Interactions between human glutamate carboxypeptidase II and urea-based inhibitors: Structural characterization," *J Med Chem*, vol. 51, no. 24, pp. 7737–7743, Dec. 2008, doi: 10.1021/jm800765e.

- [47] P. F. Jackson *et al.*, “Design, Synthesis, and Biological Activity of a Potent Inhibitor of the Neuropeptidase N-Acetylated r-Linked Acidic Dipeptidase,” 1996. [Online]. Available: <https://pubs.acs.org/sharingguidelines>
- [48] J. Pavlicek *et al.*, “Structural characterization of P1'-diversified urea-based inhibitors of glutamate carboxypeptidase II,” *Bioorg Med Chem Lett*, vol. 24, no. 10, pp. 2340–2345, May 2014, doi: 10.1016/j.bmcl.2014.03.066.
- [49] H. Wang *et al.*, “Bioisosterism of urea-based GCPII inhibitors: Synthesis and structure-activity relationship studies,” *Bioorg Med Chem Lett*, vol. 20, no. 1, pp. 392–397, Jan. 2010, doi: 10.1016/j.bmcl.2009.10.061.
- [50] A. Plechanovová *et al.*, “Novel substrate-based inhibitors of human glutamate carboxypeptidase II with enhanced lipophilicity,” *J Med Chem*, vol. 54, no. 21, pp. 7535–7546, Nov. 2011, doi: 10.1021/jm200807m.
- [51] C. Barinka, J. Starkova, J. Konvalinka, and J. Lubkowski, “A high-resolution structure of ligand-free human glutamate carboxypeptidase II,” *Acta Crystallogr Sect F Struct Biol Cryst Commun*, vol. 63, no. 3, pp. 150–153, Feb. 2007, doi: 10.1107/S174430910700379X.
- [52] Y. Chen *et al.*, “Radiohalogenated prostate-specific membrane antigen (PSMA)-based ureas as imaging agents for prostate cancer,” *J Med Chem*, vol. 51, no. 24, pp. 7933–7943, Dec. 2008, doi: 10.1021/jm801055h.
- [53] C. Bařinka, C. Rojas, M. Pomper, and R. H. Morgan, “Glutamate Carboxypeptidase II in Diagnosis and Treatment of Neurologic Disorders and Prostate Cancer,” 2012.
- [54] R. Nakajima, Z. Nováková, W. Tueckmantel, L. Motlová, C. Bařinka, and A. P. Kozikowski, “2-Amino adipic Acid-C(O)-Glutamate Based Prostate-Specific Membrane Antigen Ligands for Potential Use as Theranostics,” *ACS Med Chem Lett*, vol. 9, no. 11, pp. 1099–1104, Nov. 2018, doi: 10.1021/acsmchemlett.8b00318.
- [55] S. Robu *et al.*, “Synthesis and preclinical evaluation of novel 18F-labeled Glu-urea-Glu-based PSMA inhibitors for prostate cancer imaging: a comparison with 18F-DCFpyl and 18F-PSMA-1007,” *EJNMMI Res*, vol. 8, 2018, doi: 10.1186/s13550-018-0382-8.
- [56] A. P. Kozikowski *et al.*, “Synthesis of Urea-Based Inhibitors as Active Site Probes of Glutamate Carboxypeptidase II: Efficacy as Analgesic Agents,” *J Med Chem*, vol. 47, no. 7, pp. 1729–1738, Mar. 2004, doi: 10.1021/jm0306226.
- [57] A. E. Machulkin *et al.*, “Synthesis and biological evaluation of PSMA-targeting paclitaxel conjugates,” *Bioorg Med Chem Lett*, vol. 29, no. 16, pp. 2229–2235, Aug. 2019, doi: 10.1016/j.bmcl.2019.06.035.
- [58] K. P. Maresca *et al.*, “A series of halogenated heterodimeric inhibitors of prostate specific membrane antigen (PSMA) as radiolabeled probes for targeting prostate cancer,” *J Med Chem*, vol. 52, no. 2, pp. 347–357, Jan. 2009, doi: 10.1021/jm800994j.
- [59] K. P. Maresca *et al.*, “A series of halogenated heterodimeric inhibitors of prostate specific membrane antigen (PSMA) as radiolabeled probes for targeting prostate cancer,” *J Med Chem*, vol. 52, no. 2, pp. 347–357, Jan. 2009, doi: 10.1021/jm800994j.
- [60] J. Tykvart *et al.*, “Rational design of urea-based glutamate carboxypeptidase II (GCPII) inhibitors as versatile tools for specific drug targeting and delivery,” *Bioorg Med Chem*, vol. 22, no. 15, pp. 4099–4108, Aug. 2014, doi: 10.1016/j.bmc.2014.05.061.
- [61] J. Kelly *et al.*, “Synthesis and pre-clinical evaluation of a new class of high-affinity 18F-labeled PSMA ligands for detection of prostate cancer by PET imaging,” *Eur J Nucl Med Mol Imaging*, vol. 44, no. 4, pp. 647–661, Apr. 2017, doi: 10.1007/s00259-016-3556-5.

- [62] V. Bouvet *et al.*, “Targeting Prostate-Specific Membrane Antigen (PSMA) with F-18-Labeled Compounds: the Influence of Prosthetic Groups on Tumor Uptake and Clearance Profile,” *Mol Imaging Biol*, vol. 19, no. 6, pp. 923–932, Dec. 2017, doi: 10.1007/s11307-017-1102-x.
- [63] V. Yao, C. E. Berkman, J. K. Choi, D. S. O’Keefe, and D. J. Bacich, “Expression of prostate-specific membrane antigen (PSMA), increases cell folate uptake and proliferation and suggests a novel role for PSMA in the uptake of the non-polyglutamated folate, folic acid,” *Prostate*, vol. 70, no. 3, pp. 305–316, Feb. 2010, doi: 10.1002/pros.21065.
- [64] T. Ganguly *et al.*, “A high-affinity [18F]-labeled phosphoramidate peptidomimetic PSMA-targeted inhibitor for PET imaging of prostate cancer,” *Nucl Med Biol*, vol. 42, no. 10, pp. 780–787, Oct. 2015, doi: 10.1016/j.nucmedbio.2015.06.003.
- [65] M. Wirtz *et al.*, “Synthesis and in vitro and in vivo evaluation of urea-based PSMA inhibitors with increased lipophilicity,” *EJNMMI Res*, vol. 8, 2018, doi: 10.1186/s13550-018-0440-2.
- [66] Y. A. Ivanenkov *et al.*, “Synthesis and biological evaluation of Doxorubicin-containing conjugate targeting PSMA,” *Bioorg Med Chem Lett*, vol. 29, no. 10, pp. 1246–1255, May 2019, doi: 10.1016/j.bmcl.2019.01.040.
- [67] A. C. Baranski *et al.*, “PSMA-11-Derived dual-Labeled PSMA inhibitors for preoperative PET imaging and precise fluorescence-Guided surgery of prostate cancer,” *Journal of Nuclear Medicine*, vol. 59, no. 4, pp. 639–645, Apr. 2018, doi: 10.2967/jnumed.117.201293.
- [68] M. Eder, M. Schäfer, U. Bauder-Wüst, U. Haberkorn, M. Eisenhut, and K. Kopka, “Preclinical evaluation of a bispecific low-molecular heterodimer targeting both PSMA and GRPR for improved PET imaging and therapy of prostate cancer,” *Prostate*, vol. 74, no. 6, pp. 659–668, 2014, doi: 10.1002/pros.22784.
- [69] D. B. Nikolov and S. K. Burley, “Review RNA polymerase II transcription initiation: A structural view,” 1997. [Online]. Available: [www.pnas.org](http://www.pnas.org).
- [70] A. Sentenac, “Eukaryotic RNA polymerase,” *Crit Rev Biochem Mol Biol*, vol. 18, no. 1, pp. 31–90, 1985, doi: 10.3109/10409238509082539.
- [71] A. J. M. Larsson *et al.*, “Genomic encoding of transcriptional burst kinetics,” *Nature*, vol. 565, no. 7738, pp. 251–254, 2019, doi: 10.1038/s41586-018-0836-1.
- [72] D. Nicolas, B. Zoller, D. M. Suter, and F. Naef, “Modulation of transcriptional burst frequency by histone acetylation,” *Proc Natl Acad Sci U S A*, vol. 115, no. 27, pp. 7153–7158, 2018, doi: 10.1073/pnas.1722330115.
- [73] P. Quintero-Cadena, T. L. Lenstra, and P. W. Sternberg, “RNA Pol II Length and Disorder Enable Cooperative Scaling of Transcriptional Bursting,” *Mol Cell*, vol. 79, no. 2, pp. 207-220.e8, 2020, doi: 10.1016/j.molcel.2020.05.030.
- [74] D. B. Nikolov and S. K. Burley, “RNA polymerase II transcription initiation: A structural view,” 1997. [Online]. Available: [www.pnas.org](http://www.pnas.org).
- [75] P. Cramer, “Organization and regulation of gene transcription,” *Nature*, vol. 573, no. 7772, pp. 45–54, 2019, doi: 10.1038/s41586-019-1517-4.
- [76] A. Panigrahi and B. W. O’Malley, “Mechanisms of enhancer action: the known and the unknown,” *Genome Biology*, vol. 22, no. 1. BioMed Central Ltd, Dec. 01, 2021. doi: 10.1186/s13059-021-02322-1.
- [77] M. Iwafuchi-Doi, “The mechanistic basis for chromatin regulation by pioneer transcription factors,” *Wiley Interdiscip Rev Syst Biol Med*, vol. 11, no. 1, p. e1427, 2019, doi: 10.1002/wsbm.1427.

- [78] J. Soutourina, “Transcription regulation by the Mediator complex,” *Nature Reviews Molecular Cell Biology*, vol. 19, no. 4. Nature Publishing Group, pp. 262–274, Apr. 01, 2018. doi: 10.1038/nrm.2017.115.
- [79] L. el Khattabi *et al.*, “A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers,” *Cell*, vol. 178, no. 5, pp. 1145-1158.e20, 2019, doi: 10.1016/j.cell.2019.07.011.
- [80] D. S. O’keefe *et al.*, “Mapping, genomic organization and promoter analysis of the human prostate-specific membrane antigen gene.”
- [81] D. Good *et al.*, “Cloning and Characterization of the Prostate-Specific Membrane Antigen Promoter,” 1999.
- [82] L. Han, D. Lee Wong, G. Tsai, Z. Jiang, and J. T. Coyle, “Promoter analysis of human glutamate carboxypeptidase II,” *Brain Res*, vol. 1170, no. SUPPL.: COMPLETE, pp. 1–12, Sep. 2007, doi: 10.1016/j.brainres.2007.07.017.
- [83] F. Watt *et al.*, “A tissue-specific enhancer of the prostate-specific membrane antigen gene, FOLH1,” *Genomics*, vol. 73, no. 3, pp. 243–254, May 2001, doi: 10.1006/geno.2000.6446.
- [84] K. R. Noss, S. A. Wolfe, and S. R. Grimes, “Upregulation of prostate specific membrane antigen/folate hydrolase transcription by an enhancer.” [Online]. Available: [www.elsevier.com/locate/gene](http://www.elsevier.com/locate/gene)
- [85] S. J. Lee *et al.*, “Novel prostate-specific promoter derived from PSA and PSMA enhancers,” *Molecular Therapy*, vol. 6, no. 3, pp. 415–421, Sep. 2002, doi: 10.1006/mthe.2002.0682.
- [86] S. J. Lee *et al.*, “NFATc1 with AP-3 site binding specificity mediates gene expression of prostate-specific-membrane-antigen.,” *J Mol Biol*, vol. 330, no. 4, pp. 749–760, 2003, doi: 10.1016/S0022-2836(03)00640-5.
- [87] D. S. O’Keefe, D. J. Bacich, and W. D. W. Heston, “Comparative Analysis of Prostate-Specific Membrane Antigen (PSMA) Versus a Prostate-Specific Membrane Antigen-Like Gene,” *Prostate*, vol. 58, no. 2, pp. 200–210, Feb. 2004, doi: 10.1002/pros.10319.
- [88] Culig and al, “Switch from antagonist to agonist of the androgen receptor blocker bicalutamide is associated with prostate tumour progression in a new model system,” 1999.
- [89] J. Céraline *et al.*, “Constitutive activation of the androgen receptor by a point mutation in the hinge region: A new mechanism for androgen-independent growth in prostate cancer,” *Int J Cancer*, vol. 108, no. 1, pp. 152–157, Jan. 2004, doi: 10.1002/ijc.11404.
- [90] S. J. Libertini, C. G. Tepper, V. Rodriguez, D. M. Asmuth, H. J. Kung, and M. Mudryj, “Evidence for calpain-mediated androgen receptor cleavage as a mechanism for androgen independence,” *Cancer Res*, vol. 67, no. 19, pp. 9001–9005, Oct. 2007, doi: 10.1158/0008-5472.CAN-07-1072.
- [91] D. G. Hu *et al.*, “Identification of Androgen Receptor Splice Variant Transcripts in Breast Cancer Cell Lines and Human Tissues,” *Horm Cancer*, vol. 5, no. 2, pp. 61–71, 2014, doi: 10.1007/s12672-014-0171-4.
- [92] S. Cao, Y. Zhan, and Y. Dong, “Emerging data on androgen receptor splice variants in prostate cancer,” *Endocrine-Related Cancer*, vol. 23, no. 12. BioScientifica Ltd., pp. T199–T210, Dec. 01, 2016. doi: 10.1530/ERC-16-0298.
- [93] P. A. Watson *et al.*, “Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor,” *Proc Natl Acad Sci U S A*, vol. 107, no. 39, pp. 16759–16765, Sep. 2010, doi: 10.1073/pnas.1012443107.

- [94] E. Hörnberg *et al.*, “Expression of androgen receptor splice variants in prostate cancer bone metastases is associated with castration-resistance and short survival,” *PLoS One*, vol. 6, no. 4, 2011, doi: 10.1371/journal.pone.0019059.
- [95] H. Nagandla, M. J. Robertson, V. Putluri, N. Putluri, C. Coarfa, and N. L. Weigel, “Isoform-specific Activities of Androgen Receptor and its Splice Variants in Prostate Cancer Cells,” *Endocrinology (United States)*, vol. 162, no. 3, Mar. 2021, doi: 10.1210/endo/bqaa227.
- [96] A. ben Jemaa, S. Sallami, J. Céraline, and R. Oueslati, “A novel regulation of PSMA and PSA expression by Q640X AR in 22Rv1 and LNCap prostate cancer cells,” *Cell Biol Int*, vol. 37, no. 5, pp. 464–470, May 2013, doi: 10.1002/cbin.10055.
- [97] K. de Gendt and G. Verhoeven, “Tissue- and cell-specific functions of the androgen receptor revealed through conditional knockout models in mice,” *Molecular and Cellular Endocrinology*, vol. 352, no. 1–2, pp. 13–25, Apr. 16, 2012. doi: 10.1016/j.mce.2011.08.008.
- [98] M. Uhlén *et al.*, “Tissue-based map of the human proteome,” *Science (1979)*, vol. 347, no. 6220, Jan. 2015, doi: 10.1126/science.1260419.
- [99] L. Yin, P. Rao, P. Elson, J. Wang, M. Ittmann, and W. D. W. Heston, “Role of TMPRSS2-ERG gene fusion in negative regulation of PSMA expression,” *PLoS One*, vol. 6, no. 6, 2011, doi: 10.1371/journal.pone.0021319.
- [100] J. C. Brase *et al.*, “TMPRSS2-ERG -specific transcriptional modulation is associated with prostate cancer biomarkers and TGF- $\beta$  signaling,” *BMC Cancer*, vol. 11, Dec. 2011, doi: 10.1186/1471-2407-11-507.
- [101] S. A. Tomlins *et al.*, “Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer.” [Online]. Available: <https://www.science.org>
- [102] P. Lonergan and D. Tindall, “Androgen receptor signaling in prostate cancer development and progression,” *J Carcinog*, vol. 10, 2011, doi: 10.4103/1477-3163.83937.
- [103] M. C. A. Melnick L. Sheng, “FOXA1 acts upstream of GATA2 and AR in hormonal regulation of gene expression,” *Physiol Behav*, vol. 176, no. 1, pp. 100–106, 2016, doi: 10.1038/onc.2015.496.
- [104] V. Rodriguez-Bravo, M. Carceles-Cordon, Y. Hoshida, C. Cordon-Cardo, M. D. Galsky, and J. Domingo-Domenech, “The role of GATA2 in lethal prostate cancer aggressiveness,” *Nat Rev Urol*, vol. 14, no. 1, pp. 38–48, 2017, doi: 10.1038/nrurol.2016.225.
- [105] S. Paltoglou *et al.*, “Novel androgen receptor coregulator GRHL2 exerts both oncogenic and antimetastatic functions in prostate cancer,” *Cancer Res*, vol. 77, no. 13, pp. 3417–3430, 2017, doi: 10.1158/0008-5472.CAN-16-1616.
- [106] T. H. Kim, J. M. Park, M. Y. Kim, and Y. H. Ahn, “The role of CREB3L4 in the proliferation of prostate cancer cells,” *Sci Rep*, vol. 7, no. March, pp. 1–11, 2017, doi: 10.1038/srep45300.
- [107] J. Yao *et al.*, “The Homeobox gene, HOXB13, Regulates a Mitotic Protein-Kinase Interaction Network in Metastatic Prostate Cancers,” *Sci Rep*, vol. 9, no. 1, pp. 1–18, 2019, doi: 10.1038/s41598-019-46064-4.
- [108] W. Peng *et al.*, “Sox7 negatively regulates prostate-specific membrane antigen (PSMA) expression through PSMA-enhancer,” *Prostate*, vol. 79, no. 4, pp. 370–378, Mar. 2019, doi: 10.1002/pros.23743.
- [109] Junjun Zhang *et al.*, “The International Cancer Genome Consortium Data Portal,” *Nature Biotechnology*, vol. 37, no. 4. Nature Publishing Group, pp. 358–367, Apr. 01, 2019. doi: 10.1038/s41587-019-0055-9.
- [110] F. Iorio *et al.*, “A Landscape of Pharmacogenomic Interactions in Cancer,” *Cell*, vol. 166, no. 3, pp. 740–754, Jul. 2016, doi: 10.1016/j.cell.2016.06.017.

- [111] K. Erdmann *et al.*, “Elevated expression of prostate cancer-associated genes is linked to down-regulation of microRNAs,” *BMC Cancer*, vol. 14, no. 1, Feb. 2014, doi: 10.1186/1471-2407-14-82.
- [112] T. Mosmann, “Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays,” 1983.
- [113] P. Mlčochová *et al.*, “Mapping of the active site of glutamate carboxypeptidase II by site-directed mutagenesis,” *FEBS Journal*, vol. 274, no. 18, pp. 4731–4741, Sep. 2007, doi: 10.1111/j.1742-4658.2007.06021.x.
- [114] W. Li *et al.*, “MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens,” *Genome Biol*, vol. 15, no. 12, p. 554, 2014, doi: 10.1186/s13059-014-0554-4.
- [115] A. E. Machulkin *et al.*, “Synthesis and Biological Evaluation of PSMA Ligands with Aromatic Residues and Fluorescent Conjugates Based on Them,” *J Med Chem*, vol. 64, no. 8, pp. 4532–4552, Apr. 2021, doi: 10.1021/acs.jmedchem.0c01935.
- [116] A. E. Machulkin *et al.*, “Synthesis, Characterization, and Preclinical Evaluation of a Small-Molecule Prostate-Specific Membrane Antigen-Targeted Monomethyl Auristatin e Conjugate,” *J Med Chem*, vol. 64, no. 23, pp. 17123–17145, Dec. 2021, doi: 10.1021/acs.jmedchem.1c01157.
- [117] D. P. Labbé and M. Brown, “Transcriptional regulation in prostate cancer,” *Cold Spring Harb Perspect Med*, vol. 8, no. 11, pp. 1–14, 2018, doi: 10.1101/CSHPERSPECT.A030437.
- [118] N. L. Sharma *et al.*, “The ETS family member GABP $\alpha$  modulates androgen receptor signalling and mediates an aggressive phenotype in prostate cancer,” *Nucleic Acids Res*, vol. 42, no. 10, pp. 6256–6269, Jun. 2014, doi: 10.1093/nar/gku281.
- [119] K. M. Jozwik, I. Chernukhin, A. A. Serandour, S. Nagarajan, and J. S. Carroll, “FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3,” *Cell Rep*, vol. 17, no. 10, pp. 2715–2723, Dec. 2016, doi: 10.1016/j.celrep.2016.11.028.
- [120] H. Brechka, R. R. Bhanvadia, C. VanOpstall, and D. J. vander Griend, “HOXB13 mutations and binding partners in prostate development and cancer: Function, clinical significance, and future directions,” *Genes Dis*, vol. 4, no. 2, pp. 75–87, 2017, doi: 10.1016/j.gendis.2017.01.003.
- [121] J. Meiners *et al.*, “Upregulation of SPDEF is associated with poor prognosis in prostate cancer,” *Oncol Lett*, vol. 18, no. 5, pp. 5107–5118, 2019, doi: 10.3892/ol.2019.10885.
- [122] J. P. Venables *et al.*, “RBFOX2 Is an Important Regulator of Mesenchymal Tissue-Specific Splicing in both Normal and Cancer Tissues,” *Mol Cell Biol*, vol. 33, no. 2, pp. 396–405, Jan. 2013, doi: 10.1128/mcb.01174-12.
- [123] O. Shalem *et al.*, “Genome-scale CRISPR-Cas9 knockout screening in human cells,” *Science (1979)*, vol. 343, no. 6166, pp. 84–87, 2014, doi: 10.1126/science.1247005.
- [124] A. Zotova, I. Zotov, A. Filatov, and D. Mazurov, “Determining antigen specificity of a monoclonal antibody using genome-scale CRISPR-Cas9 knockout library,” *J Immunol Methods*, vol. 439, pp. 8–14, Dec. 2016, doi: 10.1016/j.jim.2016.09.006.
- [125] E. N. Schinke *et al.*, “A novel approach to identify driver genes involved in androgen-independent prostate cancer,” *Mol Cancer*, vol. 13, no. 1, May 2014, doi: 10.1186/1476-4598-13-120.
- [126] B. Guan *et al.*, “Ubiquitination by TOPORS regulates the prostate tumor suppressor NKX3.1,” *Journal of Biological Chemistry*, vol. 283, no. 8, pp. 4834–4840, Feb. 2008, doi: 10.1074/jbc.M708630200.
- [127] J. R. Schoenborn, P. Nelson, and M. Fang, “Genomic profiling defines subtypes of prostate cancer with the potential for therapeutic stratification,” *Clinical Cancer Research*, vol. 19, no. 15, pp. 4058–4066, Aug. 2013, doi: 10.1158/1078-0432.CCR-12-3606.

- [128] S. He *et al.*, “GPM6B inhibit PCA proliferation by blocking prostate cancer cell serotonin absorptive capacity,” *Dis Markers*, vol. 2020, 2020, doi: 10.1155/2020/8810756.
- [129] A. W. Fjorback, H. K. Müller, and O. Wiborg, “Membrane glycoprotein M6B interacts with the human serotonin transporter,” *Journal of Molecular Neuroscience*, vol. 37, no. 3, pp. 191–200, Mar. 2009, doi: 10.1007/s12031-008-9092-4.
- [130] B. Orr *et al.*, “Reduction of pro-tumorigenic activity of human prostate cancer-associated fibroblasts using Dlk1 or SCUBE1,” *DMM Disease Models and Mechanisms*, vol. 6, no. 2, pp. 530–536, Mar. 2013, doi: 10.1242/dmm.010355.
- [131] S. S. Y. Teoh, J. C. Whisstock, and P. I. Bird, “Maspin (SERPINB5) is an obligate intracellular serpin,” *Journal of Biological Chemistry*, vol. 285, no. 14, pp. 10862–10869, Apr. 2010, doi: 10.1074/jbc.M109.073171.

## Приложение

### Приложение 1. Последовательности олигонуклеотидов

Название	Последовательность
<b>Праймеры для обратной транскрипции мРНК ТФ</b>	
mCREB3L4_RT	TCCTGAAGTGATTTGGACC
mGRHL2_RT	ACCTGCAGAGTCCTGG
mFOXA1_RT	ТААСАГТСТТТТГСАСТГГ
<b>Праймеры для амплификации кДНК ТФ с сайтами Sfi I</b>	
FOXA1_OE_F	tggcctctgaggccaccatgTTAGGAACTGTGAAGATGGAAGG
FOXA1_OE_R	tggcctgacaggcctCTAGGAAGTGTTTAGGACGGGTCTG
CRHL2_OE_F	tggcctctgaggccaccatgTCACAAGAGTCGGACAATAATAAAAGACTAG
CRHL2_OE_R	tggcctgacaggcctCTAGATTTCCATGAGCGTGACC
CREB3L4_OE_F	tggcctctgaggccaccatgGATCTCGGAATCCCTGACC
CREB3L4_OE_R	tggcctgacaggcctTCACATCTCATCTGCATGCAGC
<b>Праймеры для NGS</b>	
NGS-f1	ACACTCTTCCCTACACGACGCTCTCCGATCTTCCTTGGTTCTTGTGGAAAGGACGAAACACC
NGS-f2	ACACTCTTCCCTACACGACGCTCTCCGATCTACAGGATTCTTGTGGAAAGGACGAAACACC
NGS-f3	ACACTCTTCCCTACACGACGCTCTCCGATCTAGGTAAGGTCTTGTGGAAAGGACGAAACACC
NGS-f4	ACACTCTTCCCTACACGACGCTCTCCGATCTAACAATGGTCTTGTGGAAAGGACGAAACACC
NGS-f5	ACACTCTTCCCTACACGACGCTCTCCGATCTACTGTATCTCTTGTGGAAAGGACGAAACACC
NGS-f6	ACACTCTTCCCTACACGACGCTCTCCGATCTAGGTCGCATCTTGTGGAAAGGACGAAACACC
NGS-f7	ACACTCTTCCCTACACGACGCTCTCCGATCTAGGTTATCTCTTGTGGAAAGGACGAAACACC
NGS-f8	ACACTCTTCCCTACACGACGCTCTCCGATCTCAACTCTCTCTTGTGGAAAGGACGAAACACC
NGS-f9	ACACTCTTCCCTACACGACGCTCTCCGATCTCCAACATTTCTTGTGGAAAGGACGAAACACC
NGS-f10	ACACTCTTCCCTACACGACGCTCTCCGATCTCTAACTCGTCTTGTGGAAAGGACGAAACACC
NGS-f11	ACACTCTTCCCTACACGACGCTCTCCGATCTATTCCCTCTCTTGTGGAAAGGACGAAACACC
NGS-f12	ACACTCTTCCCTACACGACGCTCTCCGATCTCTACCAGGTCTTGTGGAAAGGACGAAACACC
NGS-f13	ACACTCTTCCCTACACGACGCTCTCCGATCTGGTACTCATCTTGTGGAAAGGACGAAACACC
NGS-f14	ACACTCTTCCCTACACGACGCTCTCCGATCTGCAGTCATTCTTGTGGAAAGGACGAAACACC
NGS-f15	ACACTCTTCCCTACACGACGCTCTCCGATCTGTCTAGCATCTTGTGGAAAGGACGAAACACC
NGS-f16	ACACTCTTCCCTACACGACGCTCTCCGATCTGTGCCATCTCTTGTGGAAAGGACGAAACACC
NGS-rev	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTGTCTACTATTCTTCCCTGCCTG
<b>Олигонуклеотиды с sgRNA и сайтами рестрикции Bbs I для нокаутов энхансеров FOLH1</b>	

EE_T1_F	GAAGACGGCACCCATTTGTATTACTGCACTGTTTATGTCTTC
EE_T1_R	GAAGACATAAACAGTGCAGTAATACAAATGGGTGCCGTCTTC
EE_T2_F	GAAGACATCACCCACCTTGTAATCCACCCGCCTGTTTCAGTCTTC
EE_T2_R	GAAGACTGAAACAGGGCGGGTGGATTACAAGGTGGTGATGTCTTC
EP_T1_F	GAAGACATCACCCCTAAGAGTGAAGCACACCAGTTTCAGTCTTC
EP_T1_R	GAAGACTGAAACTGGTGTGCTTCCACTCTTAGGGTGATGTCTTC
EP_T2_F	GAAGACATCACCCCTTATATGTGAGAGTCCTGGGTTCAGTCTTC
EP_T2_R	GAAGACTGAAACCCAGGACTCTCACATATAAGGGTGATGTCTTC

**Приложение 2.** Примеры для оценки экспрессии *FOLH1* и его потенциальных регуляторов в ОТ-кПЦР.

<b>Праймеры для количественной ПЦР в реальном времени</b>	
<i>FOXA1_qF</i>	GGGGGTTTGTCTGGCATAGC
<i>FOXA1_qR</i>	GCACTGGGGGAAAGGTTGTG
<i>HOXB13_qF</i>	GCACCGTCAAGGCTGAGAAC
<i>HOXB13_qR</i>	TGGTGAAGACGCCAGTGGA
<i>GRHL2_qF</i>	TCAATACCCGAAGAGCCTACA
<i>GRHL2_qR</i>	CTTGGCTGTCACTTGCTTTGC
<i>CREB3L4_qF</i>	TCTGGACTCTACCACTTGGGTCTCC
<i>CREB3L4_qR</i>	TGCAGACGCTAATTGCTCAAACCTCC
<i>FOLH1_qF</i>	ATTTGAACCACCTCCTCCAGG
<i>FOLH1_qR</i>	CAATTGATTTTCATGTCCCGTTCC
<i>FOSL1_qF</i>	GGAGGAAGGAACTGACCGACTT
<i>FOSL1_qR</i>	CTCTAGGCGCTCCTTCTGCTTC
<i>FOS_qF</i>	GCCTCTCTTACTACCACTCACC
<i>FOS_qR</i>	AGATGGCAGTGACCGTGGGAAT
<i>ANDR_qF</i>	ATGGTGAGCAGAGTGCCCTATC
<i>ANDR_qR</i>	ATGGTCCCTGGCAGTCTCCAAA
<i>MAZ_qF</i>	GGATCACCTCAACAGTCACGTC
<i>MAZ_qR</i>	GGCACTTTCTCCTCGTGTCGTA
<i>BMPR1B_qF</i>	CTGTGGTCACTTCTGGTTGCCT
<i>BMPR1B_qR</i>	TCAATGGAGGCAGTGTAGGGTG
<i>STEAP2_qF</i>	CCTCTGCTTACCGATGAGAAGG
<i>STEAP2_qR</i>	CAGGAGGGAAAGTAAGCCAAGG
<i>GABPA_qF</i>	CTGCTGCACTGGAAGGCTATAG
<i>GABPA_qR</i>	GGTGAGGTCTATATCGGTCATGC
<i>RBFOX2_qF</i>	CCAGCTTTCAAGCAGATGTGTCC
<i>RBFOX2_qR</i>	CAAATGGGCTCCTCTGAAAGCG
<i>TFAP4_qF</i>	GGCAAATCTGGACACCATCGTG
<i>TFAP4_qR</i>	GACAGGCTTCACGATGACAGCT
<i>FOXP1_qF</i>	CAAAGAACGCCTGCAAGCCATG
<i>FOXP1_qR</i>	GGAGTATGAGGTAAGCTCTGTGG
<i>SPDEF_qF</i>	CGAAGTGCTCAAGGACATCGAG
<i>SPDEF_qR</i>	CGGTATTGGTGCTCTGTCCACA
<i>FOXL1_qF</i>	GCTGCTCTAGCTGCCTCGG
<i>FOXL1_qR</i>	TGCCGTTGAGCGTGACC
<i>NKX3-1_qF</i>	CGCAGAACGACCAGCTGAGCA

<i>NKX3-1_qR</i>	CCTGAAGTGTTTTTCAGAGTCCAAC
<i>NFYA_qF</i>	GGCAGACCATCGTCTATCAACC
<i>NFYA_qR</i>	ATCTGTGCTCCTGCCAAACTGG
<i>CEBPG_qF</i>	GCTTACAGCAGGTTTCCTCAGCT
<i>CEBPG_qR</i>	CGTTGCCGATACTCGTCACTGT
<i>GAPDH_qF</i>	GTCTCCTCTGACTTCAACAGCG
<i>GAPDH_qR</i>	ACCACCCTGTTGCTGTAGCCAA
<i>FOSL1_qF</i>	GGAGGAAGGAACTGACCGACTT
<i>FOSL1_qR</i>	CTCTAGGCGCTCCTTCTGCTTC
<i>FOS_qF</i>	GCCTCTCTTACTACCACTCACC
<i>FOS_qR</i>	AGATGGCAGTGACCGTGGGAAT

**Приложение 3.** Последовательности siRNA для нокдауна потенциальных регуляторов экспрессии *FOLH1*.

<b>Последовательности siRNA</b>	
<i>NKX3-1_s1</i>	AAGAGGAGUGCUUCUCCAAGUCUCC
<i>NKX3-1_as1</i>	GGAGACUUGGAGAAGCACUCCUCUU
<i>HOXB13_s1</i>	CGCAUUCCGUACAGCAAGGdTdT
<i>HOXB13_as1</i>	CCUUGCUGUACGGAAUGCGdTdT
<i>HOXB13_s3</i>	UUCAUGAAUUGAGCUAAUUAUGA
<i>HOXB13_as3</i>	UCAUAAUUAGCUCAAUUCAUGAA
<i>BMPR1B_s1</i>	GGAACGAAUGUAAUAAAGAdTdT
<i>BMPR1B_as1</i>	UCUUUAUUACAUCGUUCCdTdT
<i>BMPR1B_s2</i>	GGACGAGAGCUUGAACAGAdTdT
<i>BMPR1B_as2</i>	UCUGUUCAAGCUCUCGUCCdTdT
<i>STEAP2_s1</i>	CCAUUCGACUUAUUAGAUGdTdT
<i>STEAP2_as1</i>	CAUCUAAUAAGUCGAAUGGdTdT
<i>STEAP2_s2</i>	GCAACAAUAUUC AAGCGCGdTdT
<i>STEAP2_as2</i>	CGCGCUUGAAUAUUGUUGCdTdT
<i>NFYA_s3</i>	CCAUCGUCUAUCAACCAGUUAdTdT
<i>NFYA_as3</i>	UACUGGUUGAUAGACGAUGGdTdT
<i>NFYA_s1</i>	GGUCAAAACCAUCAUGCAAGdTdT
<i>NFYA_as1</i>	CUUGCAUGAUGGUUUGACCCdTdT
<i>CEBPG_s1</i>	GGAACAACAUGGCUGUGAAAdTdT
<i>CEBPG_as1</i>	UUCACAGCCAUGUUGUUCdTdC
<i>CEBPG_s2</i>	GGCUAUUUUCUGGGAUCAGdTdT
<i>CEBPG_as2</i>	CUGAUCCCAGAAAUAAGCCdTdC
<i>GABPA_s2</i>	GGAGCUGAUAGAAAUUGAGAUUGAU
<i>GABPA_as2</i>	AUCAAUUCUCAAUUUCUAUCAGCUCC
<i>GABPA_s3</i>	GCAGAGUGCACAGAAGAAAGCAUUG
<i>GABPA_as3</i>	CAAUGCUUUCUUCUGUGCACUCUGC
<i>RBFOX2_s1</i>	GCAUCAACACUACAUCUCCUCUAAU
<i>RBFOX2_as1</i>	AUUAGAGGGAUGUAAGUGUUGAUGC
<i>RBFOX2_s2</i>	CCCAAGCGACUACAUGUCUCUAAUA
<i>RBFOX2_as2</i>	TATTAGAGACATGTAGTCGCTTGGG
<i>TFAP4_s2</i>	GGACAAGGACGAAGGCAUA
<i>TFAP4_as2</i>	UAUGCCUUCGUCCUUGUCC

<i>TFAP4_s4</i>	GACGCAUGCAGAGCAUCAAdTdT
<i>TFAP4_as4</i>	UUGAUGCUCUGCAUGCdTdT
<i>FOXP1_s1</i>	UCAGUGGUAACCCUCCCCUUA
<i>FOXP1_as1</i>	UAAGGGAAGGGUUAACCACUGA
<i>SPDEF_s3</i>	GACGAGAAAGCCAAUAAAA
<i>SPDEF_as3</i>	UUUUAUUGGCUUUCUCGUC
<i>FOXL1_s1</i>	UCCGCCACAACCUCUCGCUdTdT
<i>FOXL1_as1</i>	AGCGAGAGGUUGUGGCGGAdTdT
<i>FOXL1_s2</i>	GUCCAAGAGCUUCAGCAUAdTdT
<i>FOXL1_as2</i>	UAUGCUGAAGCUCUUGGACdTdT
<i>FOXL1_s3</i>	GAGCUCCGACAAGUCCAAGdTdT
<i>FOXL1_as3</i>	CUUGGACUUGUCGGAGCUCdTdT
<i>hsFOXA1-1</i>	GcAcuGcAAuAcucGccuuTsT
	AAGGCGAGuAUUGcAGUGCTsT
<i>hsFOXA1-2</i>	GAAcAGcuAcuAcGcAGAcTsT
	GUCUGCGuAGuAGCUGUUCTsT
<i>hsFOXA1-3</i>	ccuAAAcAcuuccuAGcucTsT
	GAGCuAGGAAGUGUUuAGGTsT
<i>hsFOXA1-4</i>	AcAccuAcAuGAccAuGAATsT
	UUcAUGGUcAUGuAGGUGUTsT
<i>hsFOXA1-5</i>	ccAuGAAcccGuGcAuGAGTsT
	CUcAUGcACGGGUcAUGGTsT
<i>hsFOXA1-6</i>	ccucGGAGcAGcAGcAuAATsT
	UUcAUGCUGCUGCUCcCGAGGTsT

**Приложение 4.** Данные о нокаутах и нокадаунах TF в клеточных линиях рака простаты, полученные из базы данных KnockTF 2.0 (<http://www.licpathway.net/KnockTFv2/index.php>).

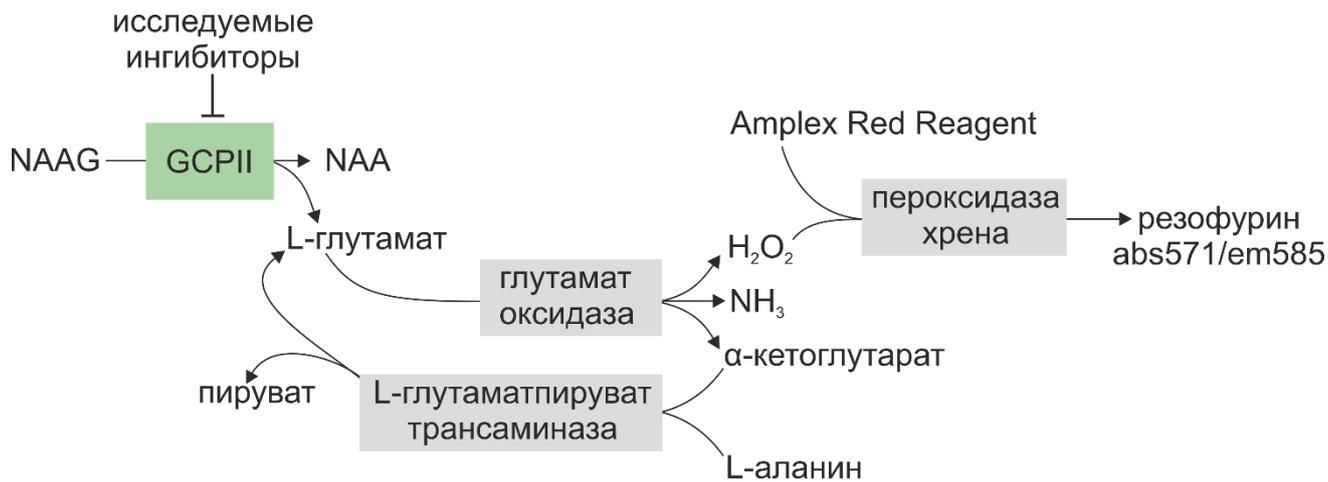
TF/ клеточная линия	VCaP	LNCaP	22Rv1	PC-3	PPC1	DU145	C4-2B	RI-D567	M12	CWR22Pc	abl cells	Immortalized prostate
ERG	sh											
CBFB					sh							
STAT3						si						
NR2F2				si								
FOXA1		si	sh									
ETV1		sh										
HSF1				si								
HSF2				si								
GABPA		sh					sh					
AR	si	si						si		si/sh	si	
GATA2		si										
REST		si										
EGR3									sh			
ERG	si											
FOXP1		si										
STAT5										sh		
SOX9	si											
HNF4G			sh									
HOXC6		si										
ETV4				sh								

ZFX						si					
AOF2		si									
BAG1		si									
CAV1											
CTCF										si/sh	
CTNNB		si									
EP300		si									
FHL2		si									
HIC1				CRISP							
HTATIP		si									
HTATIP		si									
JUND				si							
MEP50		si									
MYB	si	si		si		si					
NCOA1		si									
NCOA2		si									
NCOA3		si									
NCOR1		si									
PARK7		si									
PRKCL		si									
RCHY1		si									
SMARC		si									
SMARC		si									
SRF		si									
STAT3		si	si			si					
TET2		si				si					

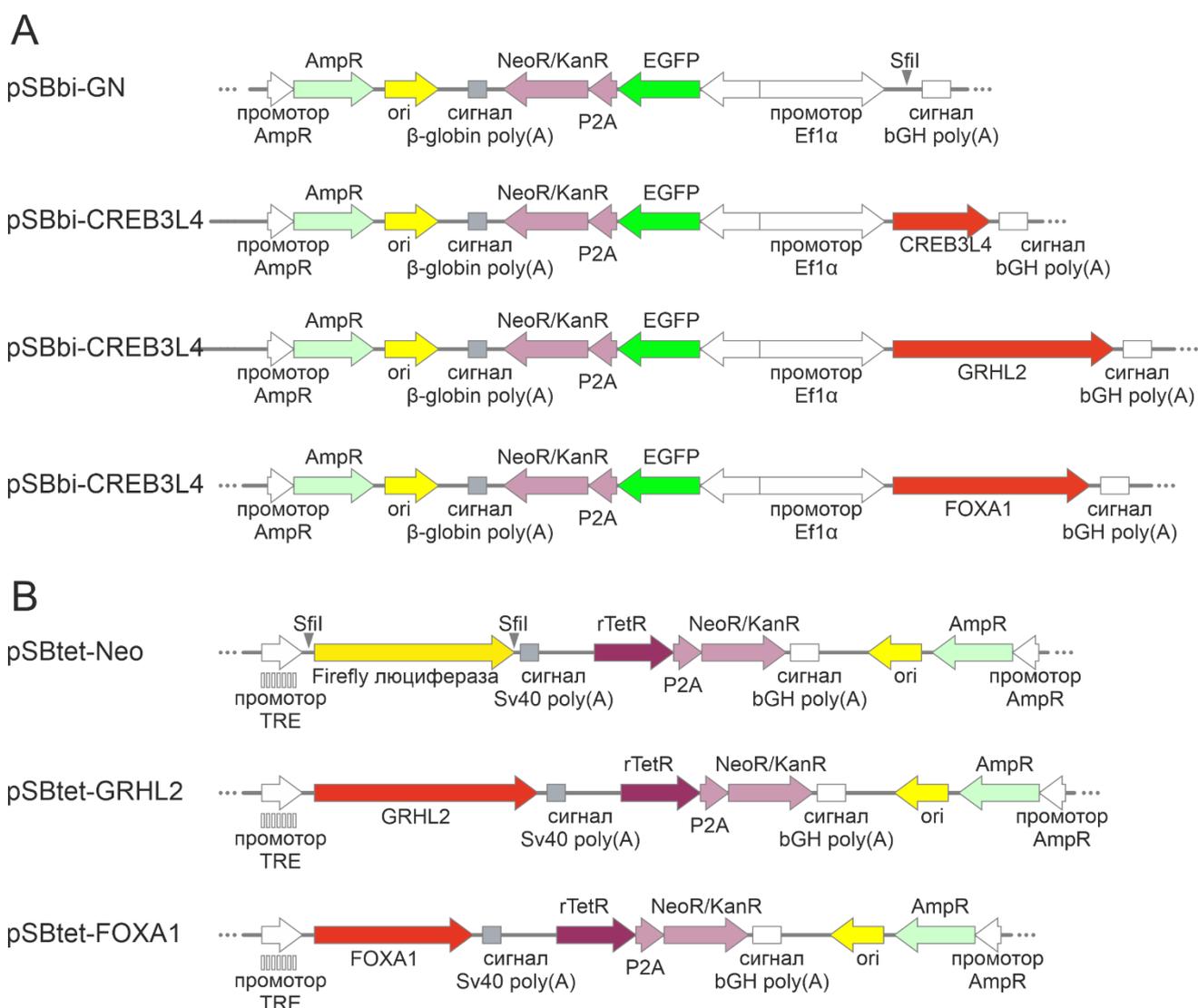
*Приложение 5. Данные о нокаутах и нокадаунах ТФ, повлиявших на экспрессию FOLH1, в клеточных линиях любого происхождения. Секция с клеточными линиями рака простаты выделена светло-серым. Интересные гены выделены темно-серым. AR – жирным шрифтом. Данные получены из базы KnockTF 2.0 (<http://www.licpathway.net/KnockTFv2/index.php>), порог изменения экспрессии = 1,5 (0,66 в случае уменьшения экспрессии в образце с нокадауном по сравнению с контролем).*

ТФ	Метод	Тип ткани	Клеточная линия	Средняя экспрессия, контроль	Средняя экспрессия, образец	Изменение экспрессии
KLF2	shRNA	Bone_marrow	RPMI8226	3,62	8,47	2,34
TFAP4	siRNA	Brain	BE(2)-C	7,77	1,78	0,23
POSTN	shRNA	Brain	GSC272	59,71	154,45	2,59
ZNF746	siRNA	Colon	HCT116	1,44	0,74	0,62
MYB	siRNA	Cord_Blood	CD34+ progenitor cells	2,95	5,48	1,86
CEBPB	siRNA	Embryo	hESC	3,81	0,65	0,17
STAT3	CRISPR	Embryo	hESC	0,01	0,02	1,51
ESRRA	shRNA	Embryo_kidney	293T	14,07	40,75	2,90
IKZF1	shRNA	Haematopoietic and lymphoid tissue	JURKAT	16,67	9,70	0,58
STAT1	shRNA	Haematopoietic and lymphoid tissue	JURKAT	12,96	6,09	0,47
SP3	siRNA	Haematopoietic and lymphoid tissue	JURKAT	5,51	9,57	1,74
RUNX1	siRNA	Haematopoietic and lymphoid tissue	Kasumi-1	96,18	189,38	1,97

FOXP1	siRNA	Haematopoietic and lymphoid tissue	OCI-Ly10	4,21	2,36	0,56
FOXP1	siRNA	Haematopoietic and lymphoid tissue	OCI-Ly7	4,81	2,19	0,45
TP63	siRNA	Head_and_neck	SCC-1	27,30	66,20	2,42
TP63	siRNA	Immortalized keratinocyte	HaCaT RG	70,80	29,90	0,42
FLCN	siRNA	Kidney	UOK257-2	11,69	2,49	0,21
HIF3A	siRNA	Liver	Hep3B	6,82	5,72	0,47
HIF3A	siRNA	Liver	Hep3B	6,82	5,72	0,47
RBFOX2	shRNA	Liver	HepG2	0,14	0,03	0,19
SSB	shRNA	Liver	HepG2	0,11	0,01	0,10
ID1	siRNA	Lung	A549	9,89	15,92	1,61
ASCL1	siRNA	Lung	DMS79	0,70	2,45	3,50
ID1	siRNA	Lung	H1650	11,20	34,38	3,07
SOX2	siRNA	Lung	LK2	497,55	292,30	0,59
BCL6	siRNA	Lymph	OCI-Ly7 DLBCL	4,14	3,51	0,65
SPDEF	shRNA	Mammary gland	MCF7	54,80	186,93	3,41
TFAP2C	siRNA	Mammary gland	MCF7	51,08	27,18	0,53
ELK1	siRNA	Mammary gland	MDA-MB231	0,00	0,00	0,40
NFYA	siRNA	Mammary gland	MDA-MB231	0,00	0,00	0,46
PTEN	shRNA	Mammary gland	SKBR3	172,25	287,36	1,67
PRRX1	CRISPR	Neuroblastoma	691-MES	2,74	3,39	1,57
SP1	siRNA	Ovary	OVSAYO	17,11	10,59	0,62
HOXB7	siRNA	Pancreas	MIA PaCa-2	2,86	7,33	2,56
CREB1	shRNA	Parotid gland	H3118 MEC	25,96	116,05	4,47
HNF4G	shRNA	Prostate	22RV1	209,69	337,37	1,61
GABPA	shRNA	Prostate	C4-2B	10,26	449,52	43,83
GABPA	shRNA	Prostate	LNCaP	10,89	726,79	66,73
HOXC6	siRNA	Prostate	LNCaP	126,30	196,25	1,55
<b>AR</b>	<b>siRNA</b>	<b>Prostate</b>	<b>LNCaP</b>	<b>26860,30</b>	<b>59764,60</b>	<b>2,23</b>
HTATIP	siRNA	Prostate	LNCaP	26860,30	42162,70	1,57
MEP50	siRNA	Prostate	LNCaP	44820,00	29579,60	0,66
SRF	siRNA	Prostate	LNCaP	6,80	7,41	1,52
STAT3	siRNA	Prostate	LNCaP	26860,30	41283,60	1,54
ETV4	shRNA	Prostate	PC-3	0,00	0,00	2,01
HSF1	shRNA	Skin	A375	9,09	3,60	0,40
PTEN	shRNA	Skin	A431	61,79	37,55	0,61
MITF	shRNA	Skin	COLO829	32,74	77,42	2,36
IRF4	siRNA	Skin	Cutaneous T cell lymphoma HuT-102 cells	12,25	4,42	0,36
MITF	siRNA	Skin	Ma-Mel-15	4,28	5,63	2,55
MITF	siRNA	Skin	Ma-Mel-15	4,29	5,09	1,75
FOXO1	siRNA	Skin	primary human dermal lymphatic endothelial cells	36,73	59,43	1,62
GATA4	siRNA	Stomach	KatoIII	3,48	2,69	0,58
GATA6	siRNA	Stomach	KatoIII	3,48	2,88	0,66
NR2F2	siRNA	Uterus	Primary eutopic stromal cells	65,59	38,51	0,59



**Приложение 6.** Схема определения ингибирования ферментативной активности GCP II.



**Приложение 7.** Схемы плазмид рSBbi-GN (A), рSBtet-Neo (B) и основанных на них плазмид, содержащих последовательности мРНК транскрипционных факторов, для конститутивной экспрессии потенциальных регуляторов транскрипции гена FOLH1.