

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М. В. ЛОМОНОСОВА  
БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

*На правах рукописи*

**Маслакова Айтсана Алексеевна**

**Структурно-функциональный анализ транскриптов гена  
*SERPINA1*: поиск альтернативных продуктов трансляции -  
изоформ и С-концевых пептидов альфа1-антитрипсина  
человека**

1.5.3 «Молекулярная биология»

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Москва-2023

Работа выполнена на кафедре молекулярной биологии Биологического факультета МГУ имени М.В.Ломоносова.

**Научный руководитель:** Рубцов Михаил Александрович, кандидат биологических наук.

**Официальные оппоненты:**

**Лябин Дмитрий Николаевич**, доктор биологических наук, руководитель группы регуляции биосинтеза белка Федерального государственного бюджетного учреждения науки Институт белка Российской академии наук.

**Ельчанинов Андрей Владимирович**, доктор медицинских наук, доцент, заведующий лабораторией роста и развития «Научно-исследовательского института морфологии человека имени академика А.П. Авцына» Федерального государственного бюджетного научного учреждения "Российский научный центр хирургии имени академика Б.В. Петровского".

**Богомазова Александра Никитична** - кандидат биологических наук, заведующая лабораторией клеточной биологии ФНКЦ ФХМ имени академика Ю. М. Лопухина ФМБА России.

Защита диссертации состоится «16» ноября 2023 г. в 16:00 часов на заседании диссертационного совета МГУ.015.4 Московского государственного университета имени М.В.Ломоносова по адресу: 119234, Москва, Ленинские горы, д.1, стр. 12, Биологический факультет, ауд. М1.

E-mail: mgu.03.01.dissovet@gmail.com

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В.Ломоносова (Ломоносовский просп., д. 27) и на сайте ИАС «ИСТИНА»:

<https://istina.msu.ru/dissertations/595879334/>

Автореферат разослан «\_\_» октября 2023 г.

Ученый секретарь  
диссертационного совета,  
доктор биологических наук



Т.В.Комарова

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность темы исследования и степень ее разработанности

Ген *SERPINA1* кодирует альфа1-антитрипсин (ААТ) - главный антипротеолитический агент, поддерживающий в организме баланс протеаза–ингибитор. ААТ является архетипическим представителем надсемейства SERPIN (SERine Protease INhibitor), основная физиологическая функция которого – защита эластических волокон в альвеолах легких от чрезмерного протеолиза эластазой нейтрофилов. К другим функциям ААТ относятся противовоспалительная и иммунорегуляторная, известны его противоапоптотическая, противовирусная и шапероноподобная активности. Концентрация ААТ в крови максимальна в сравнении с другими секреторными серпинами и в норме составляет около 20-30 мкМ (~1,5 г/л). Основными продуцентами ААТ являются гепатоциты, белок секретируется также моноцитами, макрофагами, нейтрофилами, альвеолярными эпителиальными клетками, клетками эпителия роговицы и кишечника. Ограниченный протеолиз полноразмерного ААТ обнажает «спящие» функциональные мотивы в виде его С-концевых пептидов, кодируемых экзоном 5 гена *SERPINA1*, и расширяет многогранные свойства ААТ. Пептиды обладают провоспалительными, митогенными и проопухолевыми свойствами.

Аберрантная экспрессия гена *SERPINA1* ассоциирована с развитием ряда патологических состояний. Дефицит ААТ - наиболее распространенная форма серпинопатии - обусловлен мутациями в гене *SERPINA1* и является наследуемым фактором риска развития заболеваний легких и печени, таких как ранняя хроническая обструктивная болезнь легких, цирроз печени у детей и взрослых. Дефицит ААТ может также являться фактором риска развития сахарного диабета 1 и 2 типов и ряда злокачественных новообразований.

Развитие онкологических заболеваний часто сопровождается активацией (синтез *de novo*) или увеличением экспрессии гена *SERPINA1* и, как следствие, повышением концентрации белка в крови и тканевых жидкостях. Этот феномен выявлен для ряда опухолей, а экспрессия гена *SERPINA1* обнаружена *in vitro* в ряде опухолевых клеточных линий. ААТ может быть вовлечен в универсальные механизмы канцерогенеза, и, возможно, некоторые его функции, а также его изо(протео)формы и их функции, до сих пор остаются неизвестными. Помимо секреторного ААТ, опухолевые клетки, вероятно, производят внутриклеточные протеоформы. Так, клеточная линия опухоли молочной железы MCF7 производит предполагаемый цитоплазматический вариант ААТ, препятствующий аутофагической гибели клеток. ААТ детектирован в ядрах клеток на ранней стадии развития опухоли яичника. К сожалению, структура предполагаемых внутриклеточных протеоформ ААТ и механизм их происхождения неизвестны.

Ген *SERPINA1* транскрибируется тканеспецифически с двух альтернативных промоторов в избыточное количество сплайс-изоформ мРНК (более 11 вариантов), отличающихся структурой 5'-НТО, ввиду чего *SERPINA1* входит в 0,5% генов человека, обладающих наибольшей транскрипционной сложностью. Считается, что с вариантов мРНК

транслируется один и тот же секреторный белок, состоящий из 394 а.о. (после разрезания сигнального пептида в 24 а.о.) и подвергающийся N-гликозилированию, а протеолитический путь - единственно известный механизм образования С-концевых пептидов ААТ. Открытие новых свойств ААТ и его фрагментов указывает на то, что ему отведена более широкая роль, не ограничивающаяся единственной функцией, заложенной в названии, а изучение особенностей экспрессии гена *SERPINA1*, повышающих разнообразие белковых продуктов, является важной научной задачей. Выявление альтернативных изо(протео)форм ААТ и изучение молекулярных механизмов их происхождения представляет интерес не только для фундаментальной науки, но и для решения прикладных задач, таких, как разработка современных терапевтических средств для лечения заболеваний, возможно связанных не только с дефицитом ААТ, но имеющих более сложные механизмы развития и фатальные последствия.

### **Цель и задачи диссертационной работы**

**Цель работы:** идентификация спектра изоформ мРНК гена *SERPINA1*, продуцируемых опухолевыми клеточными линиями человека, и их продуктов трансляции.

Для достижения цели были поставлены следующие **задачи**:

1. Выбор модельной клеточной линии. Скрининг культивируемых клеточных линий человека: определение относительной экспрессии длинных и предполагаемых коротких транскриптов гена *SERPINA1*, содержащих и не содержащих главную ОРС, соответственно, и содержания ААТ в секретах.

2. Идентификация спектра изоформ длинных мРНК *SERPINA1*: детекция в препаратах суммарных РНК и определение структуры их 3'- и 5'-концов.

3. Проверка гипотезы о существовании внутриклеточных изоформ ААТ: трансляция индивидуальных длинных мРНК *SERPINA1* в живых клетках *in vitro*, поиск продуктов трансляции и идентификация альтернативных ОРС.

4. Изучение внутриклеточной локализации химерных продуктов трансляции и предполагаемых эндогенных внутриклеточных изоформ ААТ, транслируемых с длинных транскриптов.

5. Проверка гипотезы об альтернативном механизме образования С-концевых пептидов ААТ - синтезе с коротких транскриптов, содержащих экзон 5 гена *SERPINA1*: детекция коротких транскриптов в препаратах суммарных РНК и определение структуры их 3'- и 5'-концов.

6. Изучение трансляции индивидуальных коротких транскриптов в живых клетках *in vitro*, поиск продуктов трансляции. Изучение внутриклеточной локализации химерных продуктов трансляции и предполагаемых эндогенных протеоформ ААТ.

### **Научная новизна**

Работа посвящена актуальной теме - изучению экспрессии гена *SERPINA1*. Прицельных исследований экспрессии гена на уровне трансляции, а именно возможности синтеза

альтернативных изо(протео)форм ААТ, не зависящих от протеолиза белка-предшественника, и условий, таких как структура транскриптов, при котором этот синтез реализуется, ранее не проводилось.

Мы провели скрининг опухолевых клеточных линий человека методами ПЦР в реальном времени и иммуноблоттинга, а с помощью нозерн-гибридизации детектировали длинные транскрипты *SERPINA1*, содержащие главную ОРС, и впервые - короткие транскрипты, содержащие экзон 5 и не содержащие главную ОРС. На кДНК выбранных клеточных линий DU145 (опухоль эпителия простаты) и HepG2 (опухоль гепатоцеллюлярной карциномы) определили первичную структуру 5'-концов длинных и коротких транскриптов, а также показали, что транскрипты в основном полиаденилированы по проксимальному альтернативному сайту полиаденилирования, приводящему к укорочению 3'-НТО. С помощью подхода *in vitro* мы впервые продемонстрировали, что с длинных транскриптов независимо от структуры 5'-НТО помимо секреторного транслируется также внутриклеточный и не гликозилированный ААТ с альтернативного внутреннего старт-кодона AUG-2. Химерные продукты трансляции локализуются в разных клеточных компартментах: внутриклеточный ААТ-eGFP, в отличие от секреторного, накапливающегося в ЭПР, мигрирует между ядром и цитоплазмой. Мы детектировали предполагаемый эндогенный внутриклеточный ААТ в ядерных спеклах.

Данные FANTOM5 CAGE (Cap Analysis of Gene Expression), извлеченные для *SERPINA1*, позволяют предположить, что короткие транскрипты генерируются необычным путем в результате посттранскрипционного расщепления сплайсированной мРНК, транскрибируемой с гепато-специфического промотора. С помощью подхода *in vitro* мы впервые продемонстрировали, что с коротких гепато-специфических транскриптов транслируется четыре С-концевые протеоформы ААТ, две из которых - с альтернативных изоформных коротких ОРС в экзоне 5. Химеры более удлиненных С-концевых фрагментов, транслированные со стартов в экзоне 3, не секретируются и локализуются в цитоплазме. Химеры С-концевых пептидов ААТ не только секретируются, но также проникают и накапливаются в ядрах клеток, в то время как предполагаемые эндогенные С-концевые пептиды ААТ - в ядрышках.

Экспрессия гена *SERPINA1* сложнее, чем считалось ранее. Наши данные свидетельствуют о повышении биологического разнообразия продуктов гена *SERPINA1* за счет сканирования с утечкой вдоль длинных транскриптов и инициации трансляции с внутреннего альтернативного старт-кодона либо укорочения транскриптов для реализации 3'-проксимальных альтернативных стартовых кодонов.

### **Научная и практическая значимость**

Полученные в работе данные представляют большой интерес для фундаментальной науки и расширяют представления об экспрессии гена *SERPINA1*. Обнаружение ряда альтернативных протеоформ ААТ и механизмов их образования, а в перспективе - их

физиологических функций, открывает новые возможности для более глубокого понимания развития заболеваний: например, изучения взаимосвязи экспрессии *SERPINA1* и канцерогенеза, а также мутаций в гене *SERPINA1* и риска развития патологий. Нельзя исключать и возможности использования полученных данных в качестве отправного пункта для последующих исследований, имеющих прикладное значение, связанное с развитием диагностических приемов и методов.

### **Методология и методы исследования**

В работе использовались современные методы клеточной (выращивание клеточных культур, клонирование клеток для выведения гомозиготных нокаутов по экзону 5, получение клеточных лизатов и секретомов), молекулярной биологии (ПЦР-РВ, нозерн-гибридизация, 3'- и 5'-RACE, методы молекулярного клонирования, геномное редактирование CRISPR/Cas9), иммунобиохимические методы (иммуноблоттинг, ИФА, получение антител с заданной специфичностью, иммунофлуоресценция), методы цито- и иммунофлуоресцентной микроскопии, прижизненного наблюдения с использованием современного оборудования, методы математического и статистического анализа.

### **Положения, выносимые на защиту**

1. Ген *SERPINA1* экспрессируется в опухолевых клеточных линиях человека в длинные и короткие транскрипты, содержащие и не содержащие главную ОРС, соответственно.

2. С длинных транскриптов *SERPINA1* помимо секреторной синтезируется внутриклеточная изоформа ААТ за счет инициации трансляции с альтернативного старт-кодона.

3. Внутриклеточная изоформа ААТ локализуется в клеточных компартментах, не характерных для секреторной.

4. Короткие транскрипты, содержащие экзон 5 *SERPINA1* - источник усеченных протеоформ ААТ, в том числе С-концевых пептидов ААТ.

5. Продукты трансляции с коротких транскриптов поступают в клеточные компартменты, не характерные для секреторной изоформы.

6. Биологическое разнообразие продуктов гена *SERPINA1* повышается за счет реализации альтернативных сайтов инициации трансляции в составе длинных (5'-проксимальный сайт) либо коротких (3'-проксимальные сайты) транскриптов.

### **Личный вклад соискателя**

Вклад соискателя заключался в выборе предмета и направлений исследования, анализе литературных данных, постановке задач, планировании (соучастие в планировании нозерн-гибридизации, 3'- и 5'-RACE) и проведении экспериментов (за исключением конфокальной микроскопии, клеточного сортирования, конъюгации пептида Е5 с белком-носителем и иммунизации животных при получении антител, дизайна и сборки некоторых плазмидных конструкций и анализа *in silico*), обработке полученных

экспериментальных данных и их интерпретации. Соискатель писал текст всех тезисов и статей, в которых указан первым автором, и занимался их подачей на конференции и в журналы.

### **Степень достоверности результатов и апробация работы**

Результаты работы были получены с использованием современных методик и на современном оборудовании. Результаты статистически достоверны и воспроизводимы. Основные положения и выводы исследования изложены в 5 статьях в рецензируемых научных изданиях, индексируемых в базах данных WoS, Scopus, RSCI и рекомендованных для защиты в диссертационном совете МГУ. Основные результаты работы были представлены на 9 конференциях.

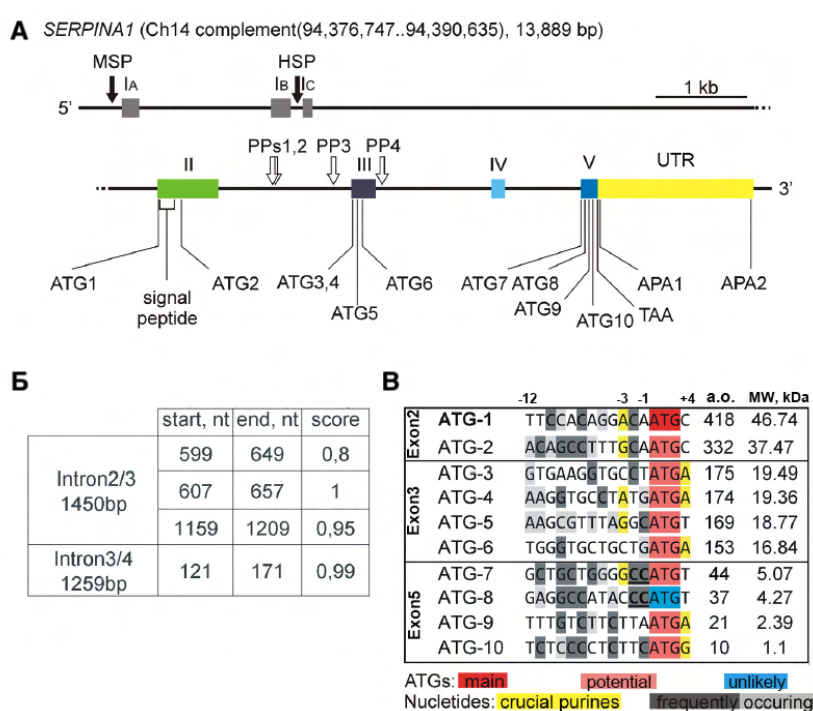
### **Структура и объём диссертации**

Диссертационная работа изложена на 167 страницах, содержит 43 рисунка (8 в обзоре литературы, 19 в результатах, 16 в приложении к результатам) и 6 таблиц (2 в обзоре литературы, 1 в результатах и 3 в приложении к результатам). Диссертация состоит из введения, включающего список публикаций по теме диссертации, обзора литературы, описания материалов и методов, результатов и их обсуждения, заключения, выводов, приложения к результатам, списка цитированной литературы, включающего 457 источников.

## РЕЗУЛЬТАТЫ РАБОТЫ

### Анализ потенциальных альтернативных сайтов инициации трансляции в рамке с главной OPC *SERPINA1*

Поскольку поставленная нами цель в конечном итоге предполагает поиск альтернативных продуктов трансляции, для начала необходимо рассмотреть структуру гена *SERPINA1* на предмет наличия потенциальных альтернативных открытых рамок считывания (альтОПС), находящихся в рамке с *SERPINA1*. Для простоты мы рассмотрим только кодирующую область гена и канонический стартовый кодон AUG.



**Рисунок 1.** Оценка потенциала гена *SERPINA1* для альтернативной экспрессии на уровне трансляции. **А)** Схема гена *SERPINA1*. MSP и HSP - миелоид- и гепатоцит-специфический промотор, PPs - предполагаемые промоторы; APA - альтернативные сайты полиаденилирования (см. раздел “3’-RACE”). **Б)** Интронные позиции и предсказанные вероятности наличия предполагаемых промоторов. **В)** Контекст Козак главного и девяти дополнительных ATG кодонов. Помимо консенсусного остатка +4G, желтой заливкой выделены также остатки +4A, поскольку старт-кодоны с +4A и в отсутствие пуринов в положении -3 обладают потенциалом к инициации трансляции.

Главный старт-кодон ATG-1 (Рисунок 1, А, В) обладает субоптимальным контекстом Козак с остатком аденина в положении -3. Девять дополнительных ATG распределены по всей кодирующей области гена: ATG-2 - ниже ATG-1 во втором экзоне (EX2), четыре (ATG-3-6) - в EX3 и четыре (ATG-7-10) - в EX5 (рисунок 1, А, В). ATG2-6 потенциально могли бы служить альтернативными сайтами инициации трансляции (altTIS), в то время как три из четырех (ATG-7, 9 и 10) в EX5 находятся в подходящем контексте Козак (рисунок 1, В). С учетом общепринятой пороговой длины в 100 кодонов, потенциальные альтиОПС со стартами в EX5 можно рассматривать как *изоформные* короткие OPC (икОПС, *изоформные* по отношению к главной OPC) *SERPINA1*.

На роль altTIS, приводящего к синтезу изоформы ААТ (не являющейся малым белком или пептидом), претендует AUG-2. Поскольку AUG-1 находится в субоптимальном контексте Козак, возможным механизмом трансляции предполагаемой изоформы может быть

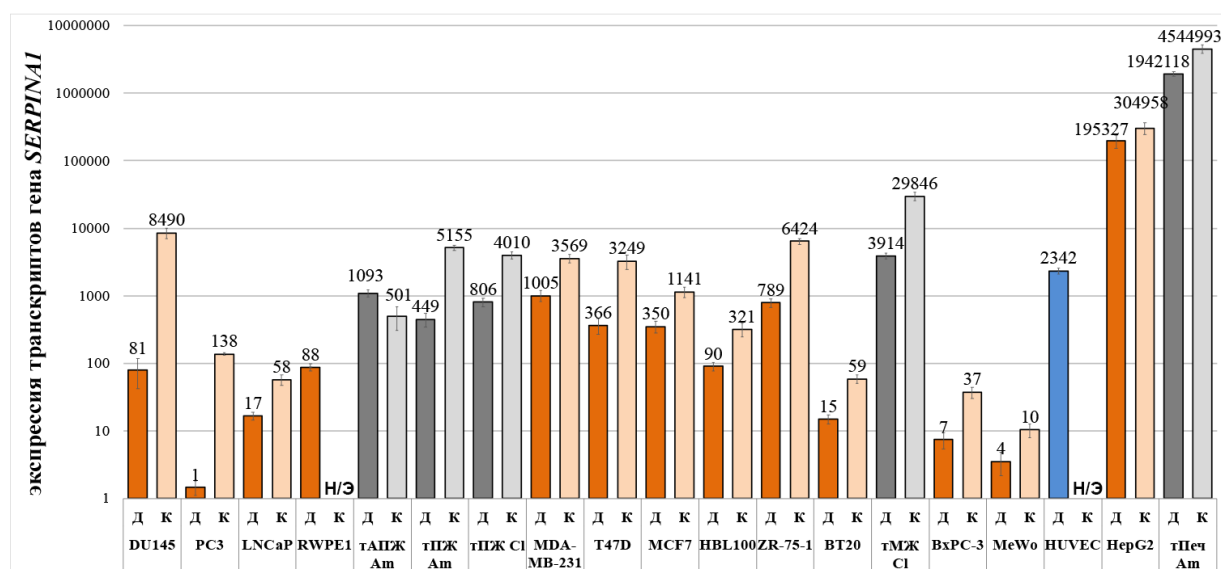


сканирование с утечкой вдоль *длинного* транскрипта, содержащего главную ОРС. Иначе обстоит дело с возможной инициацией трансляции со стартами в EX5. Поскольку им предшествует главный и пять потенциальных стартов, трансляция икОРС за счет сканирования с утечкой вдоль длинного транскрипта не представляется возможной. Это ограничение снимается, если транскрипт усечен с 5'-конца и представляет собой *короткий* транскрипт. Пептиды, точно соответствующие продуктам трансляции предполагаемых икОРС, обнаруживаются *in vivo*; также предсказываются ([https://www.fruitfly.org/seq\\_tools/promoter.html](https://www.fruitfly.org/seq_tools/promoter.html)) внутригенные промоторы с высокими вероятностями в интронах ниже EX2 (рисунок 1, А, Б).

## Выбор модельной клеточной линии

### Определение относительного уровня экспрессии гена *SERPINA1* в клеточных линиях, первичной культуре и тканях человека методом ПЦР в реальном времени

Выбирая клеточную модель, мы руководствовались двумя параметрами: 1) клетки должны обладать существенной экспрессией гена *SERPINA1* и, поскольку С-концевые пептиды кодируются EX5, 2) в этих клетках уровень экспрессии соответствующего участка EX5 должен существенно превышать экспрессию вышестоящих экзонов. Мы определили относительные уровни экспрессии стыка EX2/3, что отражает экспрессию длинных транскриптов, кодирующих полноразмерный белок, и пептид-кодирующей области EX5, что отражает суммарный уровень экспрессии длинных и предполагаемых коротких транскриптов *SERPINA1* в ряде образцов с помощью ПЦР в реальном времени (ПЦР-РВ). Для большинства образцов (16 из 19) экспрессия EX5 выше экспрессии EX2/3, что может свидетельствовать о существовании коротких транскриптов. На диаграмме (рисунок 2) приведены данные, подсчитанные для длинных и предполагаемых коротких транскриптов в отдельности.



**Рисунок 2.** Относительная экспрессия длинных (Д) и коротких (К) транскриптов гена *SERPINA1*. Образцы: опухолевые клеточные линии (предстательной железы DU145, PC3, LNCaP; молочной железы MDA-MB-231, T47D, MCF7, ZR-75-1, BT20; поджелудочной железы VxPC-3; меланомы MeWo; гепатоцеллюлярной карциномы HepG2), иммортализованная линия эпителия нормальной

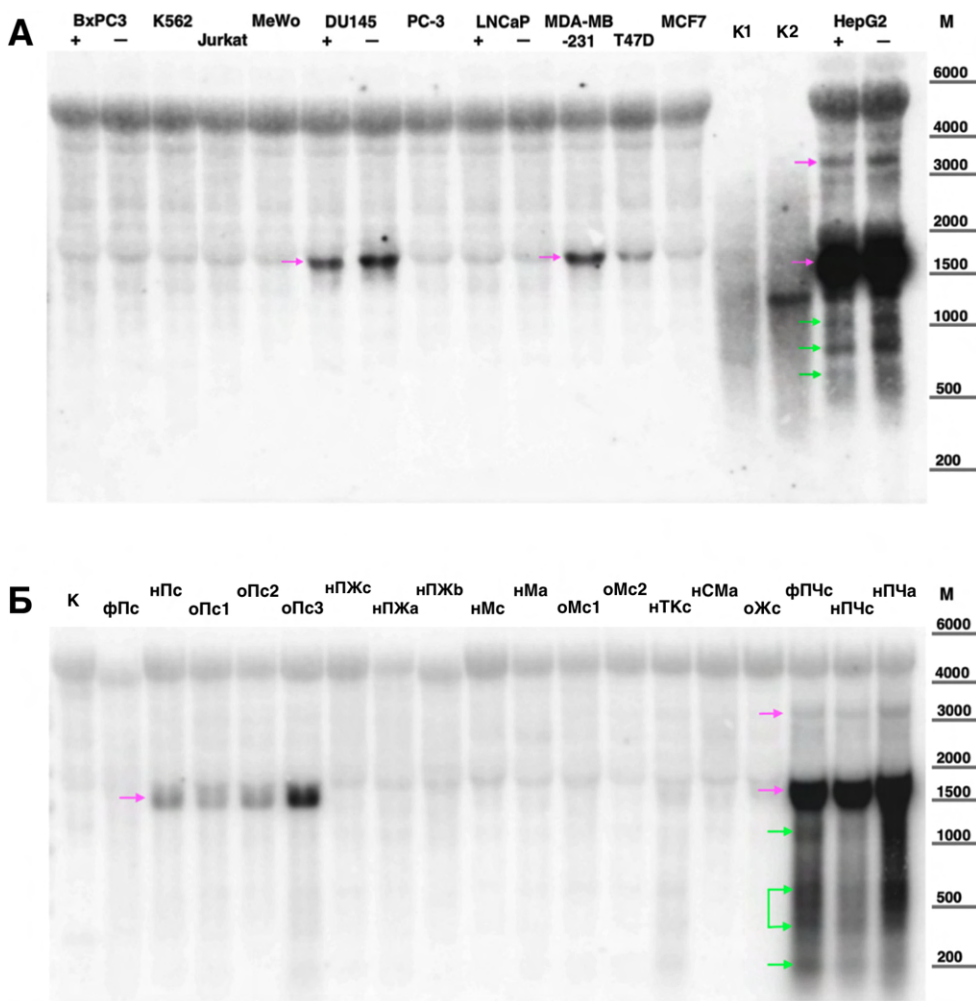
простаты RWPE-1 и спонтанно иммортализованная линия эпителия молочной железы HBL100, первичные клетки HUVEC и ткани (т) человека от производителей Clontech (Cl) и Ambion (Am): АПЖ- аденома простаты, ПЖ- предстательная железа, МЖ- молочная железа, Печ- печень. В качестве гена домашнего хозяйства использовали GAPDH, экспрессия подсчитана относительно PC3 (Д). Данные представлены по трем повторам как среднее±стандартное отклонение.

По полученным результатам в клеточной линии HepG2 максимальна не только экспрессия длинных, но и предполагаемых коротких транскриптов. Из клеток не печеночного происхождения максимальная экспрессия длинных мРНК в линии MDA-MB-231, но по экспрессии предполагаемых коротких транскриптов эта линия уступает клеткам DU145. Наибольшее соотношение экспрессии коротких к длинным наблюдается в гормон-нечувствительных линиях опухоли простаты PC3 и в особенности DU145. Поскольку в клетках PC3 длинные мРНК практически не экспрессируются (среднее значение Ct=35), мы предварительно отобрали линию DU145 (Ct=29,42) наряду с печеночными клетками HepG2.

### **Детекция длинных и коротких транскриптов *SERPINA1* с помощью нозерн-гибридизации**

В качестве зонда для гибридизации использовали радиоактивно меченый кРНК зонд, комплементарный участку мРНК протяженностью 260 н.о., соответствующему EX5 *SERPINA1* (координаты в UCSC Genome Browser: chr14:94,378,432-94,378,692). В качестве положительных контролей - генно-инженерные транскрипты, содержащие участки EX5 *SERPINA1*: короткий (319 н.о.) - в кодирующей части полностью соответствующий зонду и длинный (1187 н.о.) - с координатами кодирующей части chr14:94,378,432-94,379,535 (UCSC Genome Browser). Из клеточных линий наиболее высокий уровень экспрессии гена наблюдается в линии HepG2 (рисунок 3, А). Длинные мРНК детектируются в районе 3300 н.о. (содержат полноразмерный 3'-НТО длиной 1702 н.о.) и 1500 н.о. (содержат укороченный 3'-НТО за счет альтернативного полиаденилирования, см. раздел "3'-RACE"). В области около 1000 н.о. и ниже детектируются короткие транскрипты. Детектируемые варианты коротких транскриптов могут различаться структурой 5'-концов, тогда как диффузность сигналов от коротких транскриптов может быть обусловлена разной длиной поли(А)-хвостов и/или высокой вариабельностью их 5'-концов. Длины коротких транскриптов свидетельствуют о том, что они содержат укороченный 3'-НТО (рисунок 1, А) (см. раздел "3'-RACE").

Из не печеночных максимальная экспрессия длинных транскриптов наблюдается в линиях DU145, MDA-MB-231 и T47D (рисунок 3, А), однако короткие транскрипты практически не детектируются. В тканях максимальная экспрессия гена *SERPINA1* наблюдается в фетальной и нормальной печени взрослого, существенная экспрессия - в ткани нормальной и опухолевой почки взрослого (рисунок 3, Б). Короткие транскрипты в большинстве образцов тканей едва детектируются, кроме ткани печени, части образцов ткани почки и нормальной ткани тонкого кишечника.

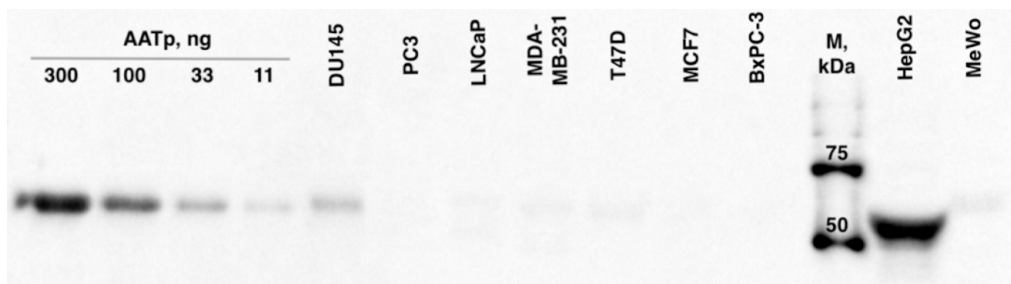


**Рисунок 3.** Нозерн-гибридизация с  $^{32}\text{P}$ -меченым зондом кРНК комплементарным к локусу *SERPINA1*: А) на суммарной РНК из опухолевых клеточных линий. “+” - клетки растили в обычной среде, “-” - среде без фенолового красного с сывороткой после обработки уголь-декстраном. К - положительный контроль. К1: нанесено 1000 амоль длинного смыслового РНК-фрагмента и 10 амоль короткого; К2: 10000 амоль длинного РНК-фрагмента и 100 амоль короткого. Б) на коммерческих препаратах суммарной РНК из тканей человека. К - контроль (10000 амоль длинных и коротких

фрагментов смысловых РНК); ф - фетальная, н - нормальная ткань, о - опухоль; П-почка, ПЖ-простата, М - молочная железа, ТК - тонкий кишечник, СМ - скелетные мышцы, Ж - желудок, ПЧ - печень; с- производитель Clontech, а-Ambion, b-BioChain. М - маркер длин РНК. Малиновыми стрелками отмечены длинные транскрипты, зелеными - короткие.

#### Детекция ААТ в секретах клеточных линий

Мы выбрали несколько клеточных линий, демонстрирующих по результатам ПЦР-РВ и нозерн-гибридизации существенную и низкую экспрессию *SERPINA1* и определили содержание белкового продукта в секретах (Рисунок 4).



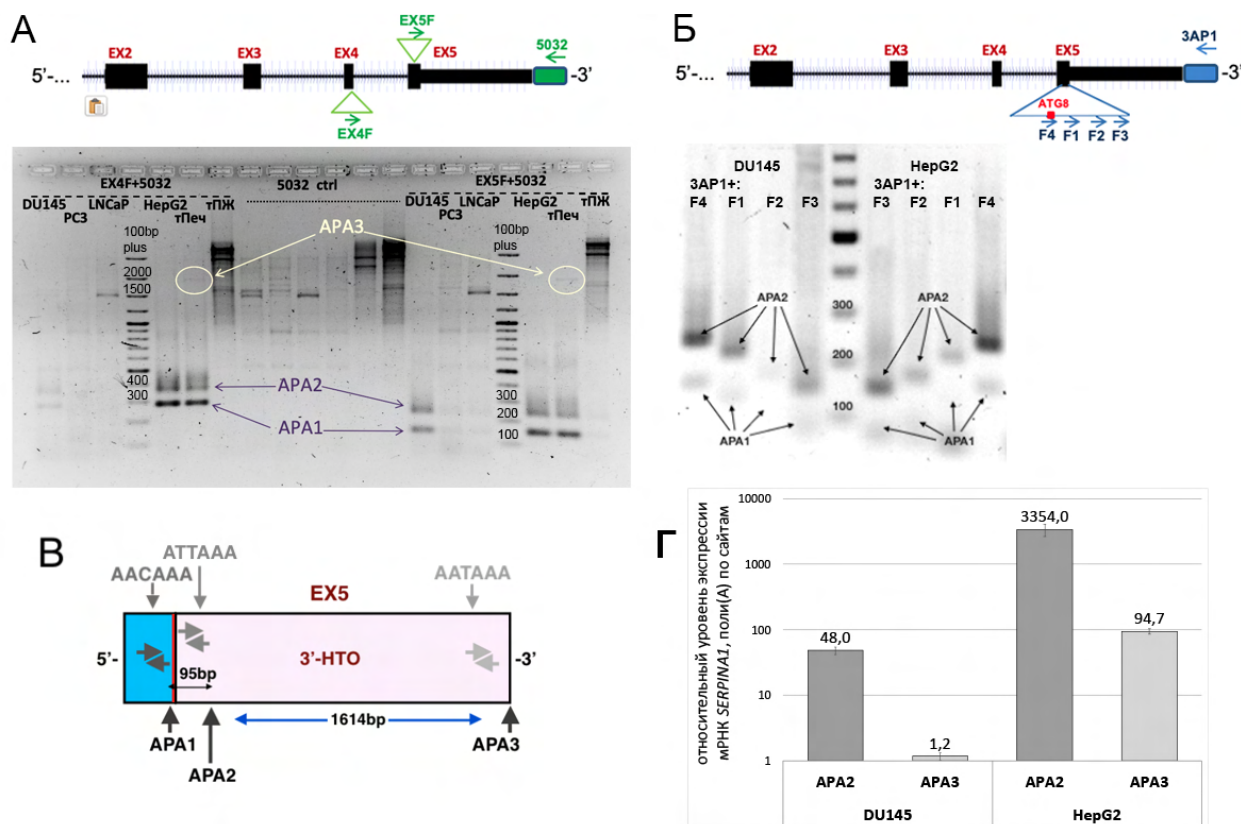
**Рисунок 4.** Детекция ААТ в секретах клеточных линий с помощью иммуноблоттинга с использованием антител против секреторного ААТ. Коммерческий препарат ААТ из плазмы (ААТр) в указанных количествах и секретом HepG2 использовали в качестве положительных контролей.

На основании ПЦР-РВ, нозерн-гибридизации и иммуноблоттинга мы выбрали две клеточные линии для дальнейшего изучения: HepG2 как наиболее перспективный объект для выявления структуры и изучения потенциала к трансляции коротких транскриптов, содержащих EX5, и DU145, в которой наблюдается существенная *de novo* экспрессия гена на уровне мРНК и максимальный выход секреторного продукта среди линий не печеночного происхождения.

## Идентификация структуры длинных и коротких транскриптов

### Идентификация структуры 3'-концов транскриптов методом 3'-RACE

Методом 3'-RACE получили три фрагмента, соответствующих полиаденилированию по трем альтернативным сайтам APA1, APA2 и APA3 (рисунок 5 А, Б). Сигнал к полиаденилированию PAS-APA3 расположен в координатах 1678-1683 н.о. ниже стоп-кодона и представлен консенсусной последовательностью ААТААА. PAS-APA2 представлен вариантом консенсусного PAS - АТТААА - с координатами 56-61 н.о. ниже стоп-кодона. Альтернативное полиаденилирование по APA2 изучалось сравнительно недавно, в т.ч. нами [Maslakova A.A. et al., 2015], и в настоящее время благодаря данным секвенирования транскриптомов аннотируются оба APA. Полиаденилирование по APA1 приводит к образованию мРНК без стоп-кодона (рисунок 5, В), а наиболее подходящий PAS ААСААА расположен в координатах 56-51 выше стоп-кодона. Поскольку такие мРНК подвергаются деградации, функциональным проксимальным сайтом полиаденилирования является APA2.



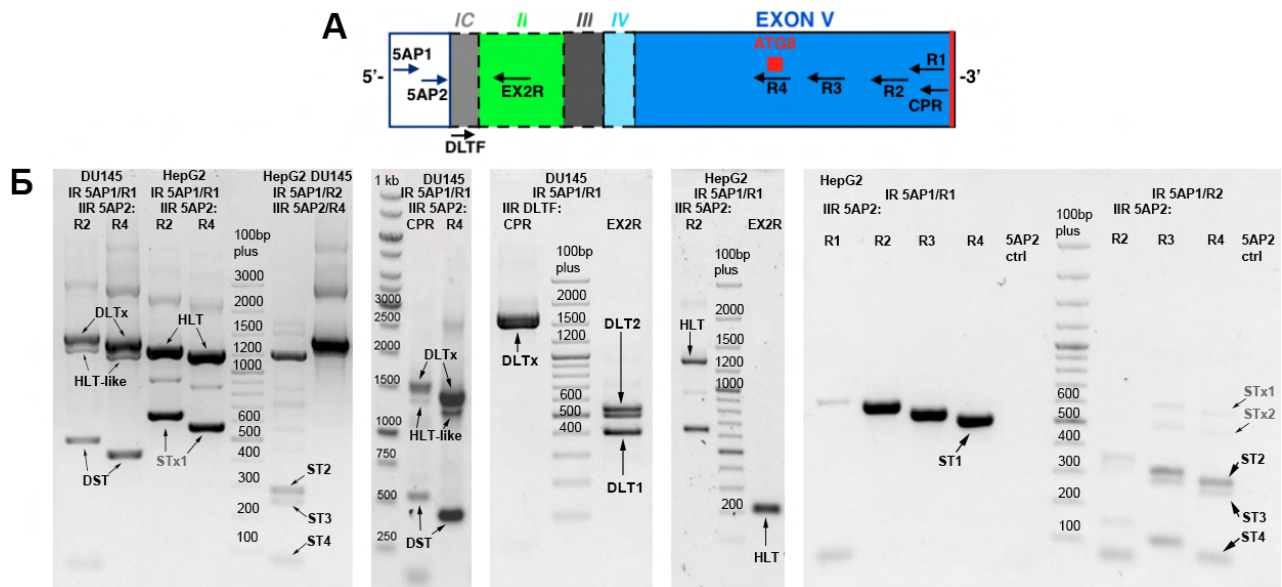
**Рисунок 5.** Анализ структуры 3'-концов транскриптов *SERPINA1* с помощью 3'-RACE. А) 3'-RACE на образцах из клеточных линий и тканей печени (тПеч) и предстательной железы (тПЖ). Сверху -

Схематическая иллюстрация мест отжига праймеров. Снизу - Электрофореграмма продуктов амплификации. **Б)** 3'-RACE на образцах из клеточных линий DU145 и HepG2; в А и Б использованы разные 3'-адапторы. Праймер F4 - 5'-удлиненный вариант праймера EX5F (А). Снизу - Электрофореграмма продуктов амплификации. **В)** Схематическая иллюстрация относительного расположения сайтов APA1-3 в структуре EX5. Красной вертикальной линией отмечен стоп-кодон. **Г)** Относительный уровень экспрессии транскриптов, полиаденилированных по APA2 и APA3, в отдельности. На схеме В) серыми горизонтальными стрелками отмечены пары праймеров, использованные в ПЦР-РВ.

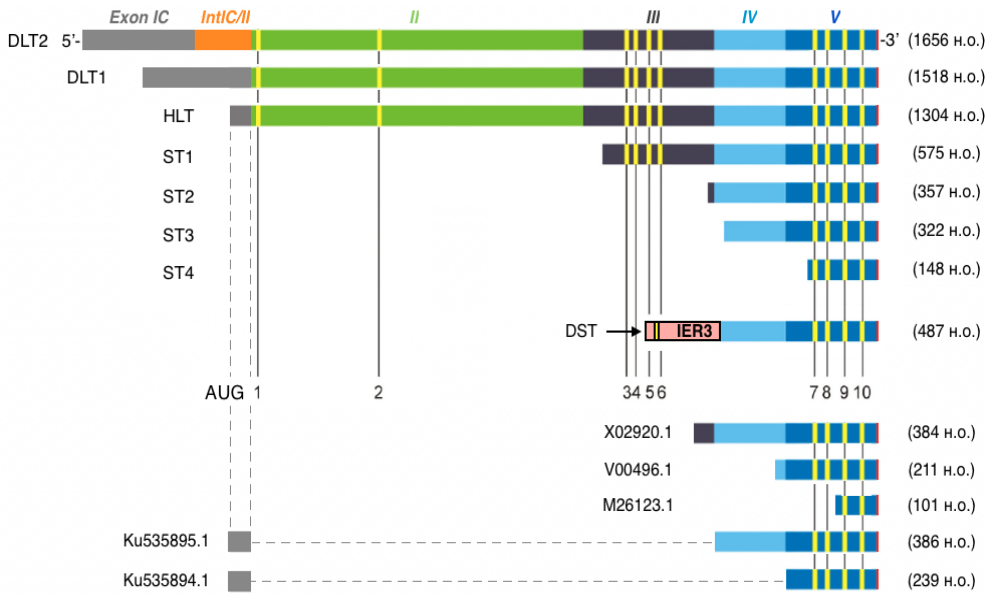
С помощью ПЦР-РВ определили, что в клетках DU145 и HepG2 в составе транскриптов *SERPINA1* преимущественно реализуется сайт APA2 (рисунок 5, Г). Длинный транскрипт, соответствующий APA3, детектируется только в суммарной РНК HepG2 и ткани нормальной печени (рисунок 3, А, Б) (размер около 3300 н.о.), а в 3'-RACE - только в кДНК ткани печени (рисунок 5, А). Мы предполагаем, что в ткани печени доля длинных 3'-НТО выше таковой в HepG2, в которой амплифицируются только короткие продукты.

### Идентификация структуры 5'-концов транскриптов методом 5'-RACE

Методом 5'-RACE на кДНК клеточных линий DU145 и HepG2 получили фрагменты, соответствующие длинным и коротким транскриптам (Рисунок 6). Результаты секвенирования показали, что в клетках HepG2 экспрессируется длинный транскрипт HLT (HepG2 Long Transcript), являющийся классическим гепато-специфическим транскриптом с укороченным вариантом EX1C, соответствующим 5'-концу транскрипта NM\_000295.5.



**Рисунок 6.** Анализ структуры 5'-концов транскриптов *SERPINA1* с помощью 5'-RACE. **А)** Схематическая иллюстрация мест отжига праймеров; белый прямоугольник - 5'-адаптор; красная вертикальная линия - стоп-кодон. **Б)** Сканы гелей с продуктами амплификации во втором раунде вложенной ПЦР. IR и IIR - первый и второй раунд ПЦР. D(L/S)T=DU145 Long/Short Transcript; x - смесь продуктов DLT1 и DLT2; HLT=HepG2 Long Transcript; HLT-like - транскрипт, идентичный HLT; ST1-4=Short Transcript 1-4 из HepG2. STx - минорные короткие продукты, полученные в режиме амплификации кДНК коротких транскриптов, исключенные из дальнейшего анализа.



**Рисунок 7.** Схематические структуры 5'-концов транскриптов *SERPINA1* в клеточных линиях DU145 и HepG2. D(L/S)T=DU145 Long/Short Transcript; HLT=HepG2 Long Transcript; ST1-4=Short Transcript 1-4 из HepG2. Снизу представлены короткие транскрипты из других работ. AUG кодоны отмечены

желтыми вертикальными линиями и стоп-кодон красной вертикальной линией. Справа - длины транскриптов до стоп-кодона включительно.

Основной длинный транскрипт DU145, DLT1 (DU145 Long Transcript), содержит 5'-удлиненный вариант EX1C (с охватом всего интрона IB/IC) и соответствует 5'-НТО транскрипта NM\_000295.4 (с заменами T34>C, A39>G) (рисунок 6, Б; рисунок 7). При амплификации с прямым праймером на самый 5'-конец транскрипта DLT1 и обратным из EX5 и EX2 обнаружили, что этот продукт представлен смесью продуктов (рисунок 6, Б, третий скан слева), из которых выделили и секвенировали два - DLT1 и DLT2. Последний не аннотирован и отличается частичным удержанием интронных 138 н.о. смежных с 5'-концом EX2. Сравнение интронного локуса с тегами экспрессируемых последовательностей (EST, expressed sequence tag) предполагает его экспрессию в фетальной печени (GenBank ID BX506843.1) и нормальной печеночной ткани (GenBank ID AV658834.1). Границы EX1C в DLT2 также соответствуют NM\_000295.4 без замен. DLT1 и DLT2 могут происходить из разных аллелей *SERPINA1*. В DU145 обнаружены и транскрипты, идентичные HLT, но в меньших количествах (рисунок 6, Б).

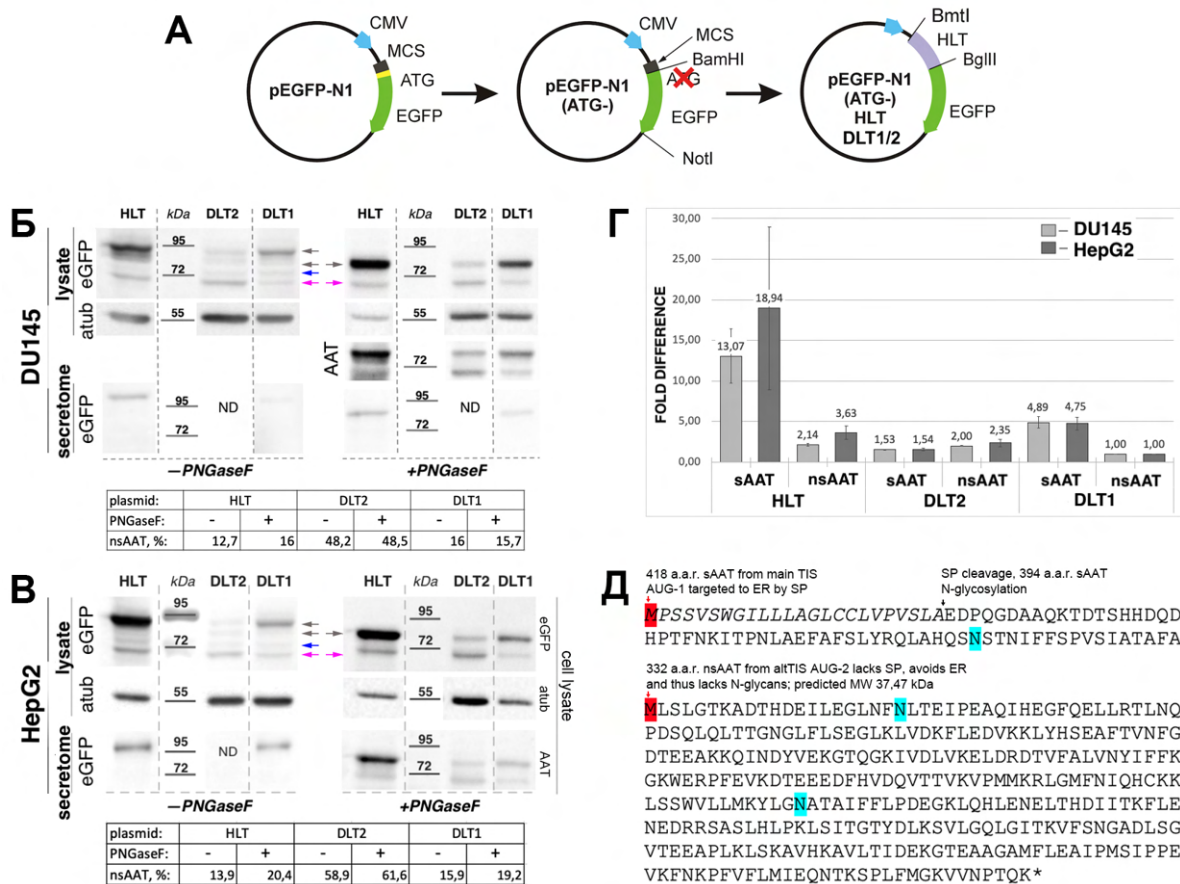
Продукт 5'-RACE, соответствующий короткому транскрипту DU145 (DST, DU145 Short Transcript), является химерой с IER3 (Immediate Early Response 3) без сдвига рамки считывания (рисунок 7 и 8). Поскольку короткий транскрипт не детектирован в нозерн-гибридизации в клетках DU145 в явном виде и не выявлялся при амплификации на других препаратах кДНК DU145, мы исключили его из дальнейшего анализа. В HepG2 наблюдается несколько коротких продуктов амплификации (рисунок 7, ST1-4, Short Transcript): транскрипт ST1 состоит из EX3 (с 37го н.о. экзона)-EX4-EX5; ST2 - EX3 (с 255го н.о.)-EX4-EX5; ST3 - EX4 (с 19го н.о.)-EX5; ST4 - EX5 (с 45го н.о.). Ввиду особенностей методики 5'-RACE сделали вывод о том, что короткие транскрипты кэпированы и полиаденилированы по APA2, как и короткие печеночные транскрипты,

идентифицированные в библиотеках кДНК печени ранее: V00496.1, X02920.1 и M26123.1 (рисунок 7). На основании данных FANTOM5 CAGE и ENCODE Chip-Seq, извлеченных для гена *SERPINA1*, мы предположили, что основным источником усеченных внутригенных транскриптов может являться посттранскрипционное расщепление сплайсированных длинных транскриптов *SERPINA1*, синтез которых направлен гепатоцит-специфическим промотором (данные представлены в полном тексте диссертации).

## Изучение трансляции индивидуальных изоформ длиной мРНК *SERPINA1* в живых клетках *in vitro*

### С длинных мРНК синтезируется предполагаемая несекреторная изоформа ААТ

Мы клонировали кДНК *SERPINA1*, соответствующие DLT1, DLT2 и HLT, без сдвига рамки считывания в мутированный нами вектор pEGFP-N1(ATG-) (Рисунок 8, А). Для каждой конструкции в клеточных лизатах обнаружили несколько eGFP+ полос в районе 69-88 кДа, а в секрете HLT и DLT1 детектирована единственная полоса 99 кДа, соответствующая зрелому ААТ-eGFP (Рисунок 8, Б).



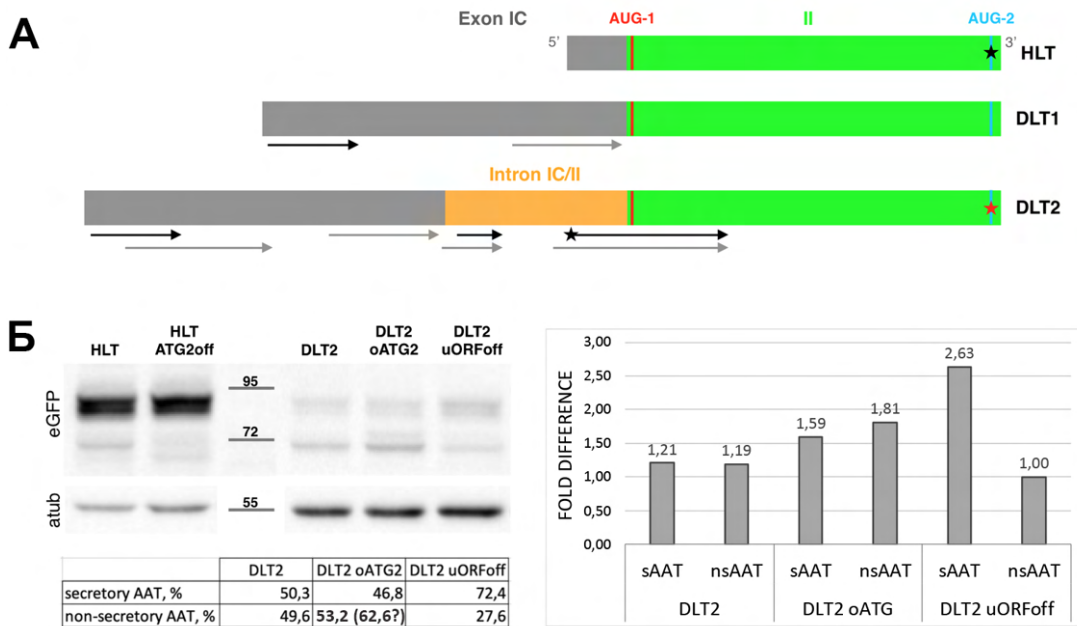
**Рисунок 8.** *In vitro* трансляция длинных мРНК *SERPINA1*. А) Схема клонирования кДНК. Б) Иммуноблоттинг лизатов и секретомов трансфицированных клеток DU145 (Б) и HepG2 (В) с использованием антител против GFP. Серые стрелки - секреторный sAAT, малиновая - предполагаемая несекреторная изоформа nsAAT. Мембрану с образцами DU145 +PNGaseF перепроверили антителами против ААТ из плазмы. Г) Денситометрический анализ экспрессии sAAT и nsAAT с конструкций; сигналы нормировали к интенсивности Ponceau S, затем к полосе с минимальной экспрессией (nsAAT с DLT1 для каждой клеточной линии); данные представлены как

среднее  $\pm$  стандартное отклонение по двум независимым экспериментам. Д) Последовательность ААТ (UniProtKB ID P01009). Красная заливка указывает на остатки метионина, соответствующие s/nsААТ; SP - сигнальный пептид; ER - эндоплазматический ретикулум; GA - аппарат Гольджи; MW - молекулярная масса; голубая заливка - остатки аспарагина, подвергающиеся гликозилированию в sААТ.

После обработки лизатов ферментом PNGaseF eGFP+ продукты 81-88 кДа объединились в 81 кДа продукт (Рисунок 8, Б), что указывает на различия их полисахаридного состава, а не первичной структуры. Полоса 69 кДа, соответствующая главному внутриклеточному продукту (Рисунок 8, Б, малиновая стрелка), оставалась неизменной под воздействием фермента (Рисунок 8, Б, таблица под сканом). Результаты, наблюдаемые в клетках DU145, воспроизводились также и в HepG2 (Рисунок 8, В, Г). На основании двухфакторного дисперсионного анализа мы пришли к выводу о том, что содержание белков и соотношение секреторного к несекреторному ААТ определяется самими транскриптами ( $p=0.0003$  и  $p=0.0026$ ) и не зависит от типа клеток ( $p=0.44$  и  $p=0.54$ , соответственно).

### Сайт-направленный мутагенез плазмидных конструкций ранжирует внутриклеточную протеоформу в изоформу ААТ

Источником 69 кДа слитого белка может являться трансляция с внутреннего altTIS *SERPINA1*. *SERPINA1* AUG-2 является наиболее вероятным кандидатом, так как он канонический, встроен в подходящий контекст Козак и дает продукт 332 а.о. с расчетной молекулярной массой 37,47 кДа (Рисунок 8, Д), практически идентичный 38,7 кДа ААТ-компоненту в химерном белке. Мы мутировали предполагаемый альтернативный стартовый кодон ATG-2>AAG в конструкции HLT. Кроме того, мы оптимизировали его контекст Козак GCAATGC>GCAATGG в DLT2.



**Рисунок 9.** Идентификация альтОПС длинных мРНК *SERPINA1*. **А)** Схематическая иллюстрация расположения 5'-ОПС. Черные стрелки - 5'-ОПС со стартом AUG, серые - со стартом CUG. AUG-2 - предполагаемый старт трансляции. Черные звезды - мутации AUG>AAG, красные - оптимизирующие



контекст Козак. **Б) Слева** - иммуноблоттинг лизатов DU145 после трансфекции конструкциями: ATG2off - мутация ATG-2>AAG, uORFoff - ATG>AAG предполагаемого 5'-ОРС, перекрывающегося с главной, oATG2 - ATG-2 с оптимизацией контекста Козак; **Справа** - денситометрический анализ экспрессии s/nsAAT.

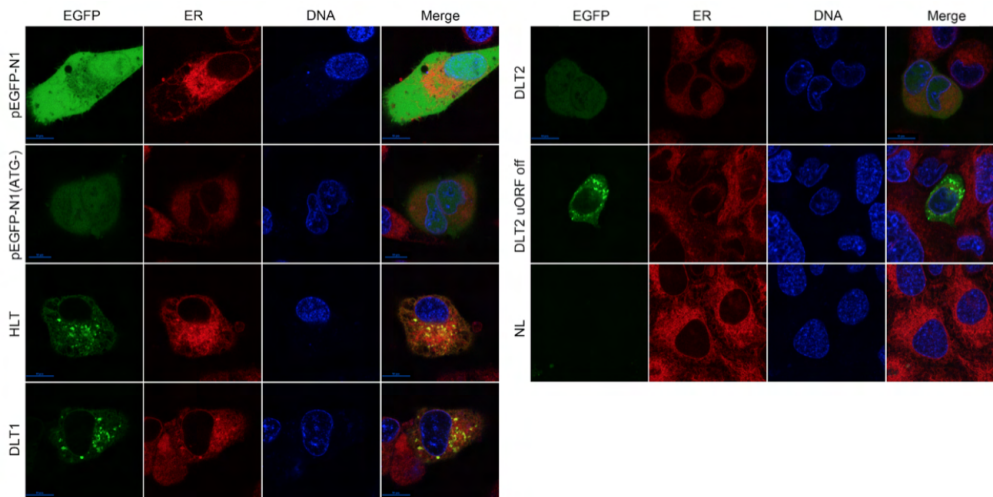
AUG-2 служит старт-кодоном для трансляции несекреторной изоформы ААТ, что подтверждается исчезновением соответствующей полосы 69 кДа (Рисунок 9, Б). Оптимизация контекста Козак у AUG-2 привела к увеличению доли внутриклеточного ААТ на несколько процентов (Рисунок 9, Б). Если минорная полоса 75 кДа (рисунок 8, Б, синяя стрелка), ставшая более отчетливой после оптимизации (Рисунок 9, Б), - внутриклеточный ААТ-eGFP, подвергшийся посттрансляционной модификации, то увеличение доли составляет  $\approx 13$  %. Синтез внутриклеточной изоформы с AUG-2, по-видимому, происходит за счет сканирования с утечкой, поскольку основной старт-кодон AUG-1 находится в субоптимальном контексте Козак (рисунок 1, В).

С транскриптов с более длинной 5'-НТО (DLT2>DLT1>HLT) транслируется меньше секреторного белка ААТ-eGFP (Рисунок 8, Б-Г) предположительно ввиду ингибирующего влияния 5'-ОРС (Рисунок 9, А). Удержание части интрона в DLT2 приводит к появлению двух дополнительных пар котерминальных 5'-ОРС (Рисунок 9, А). Вторая пара перекрывает главную ОРС и мы мутировали предполагаемый канонический старт 5'-ОРС (ATG>AAG) в DLT2. В результате доля секреторного ААТ увеличилась на 22 % по сравнению с не мутантным DLT2 (Рисунок 9, Б). 5'-ОРС DLT1/2 не являются триггерами трансляции с AUG-2, а регулируют частоту инициации трансляции с главного AUG-1 и практически не влияют на инициацию с AUG-2, поскольку мы наблюдали трансляцию с последнего в HLT, 5'-НТО которого не содержит 5'-ОРС (Рисунок 8, В; 9, А). Мы сделали вывод о том, что клеточный аппарат регулирует баланс секреторной и внутриклеточной изоформ ААТ путем транскрипционного контроля на уровне альтернативного сплайсинга пре-мРНК *SERPINA1*.

## **Изучение внутриклеточной локализации химерных продуктов трансляции с длинных транскриптов и предполагаемой эндогенной внутриклеточной изоформы ААТ**

### **Слитые белки изоформ ААТ-eGFP локализуются в разных компартментах клетки**

Мы провели микроскопию на живых клетках, используя сигналы от eGFP и ЭПР-трекера. HLT и DLT1 показали цитоплазматическую локализацию слитого белка с ожидаемой солокализацией с ЭПР-трекером (Рисунок 10). DLT2 утратил обычный паттерн распределения ААТ в ЭПР, слитые белки распределены более диффузно и мигрировали между цитоплазмой и ядром (Рисунок 10). Те же паттерны распределения ААТ-eGFP наблюдали в клеточной линии НерG2 (Рисунок 5П, приложение к диссертации).

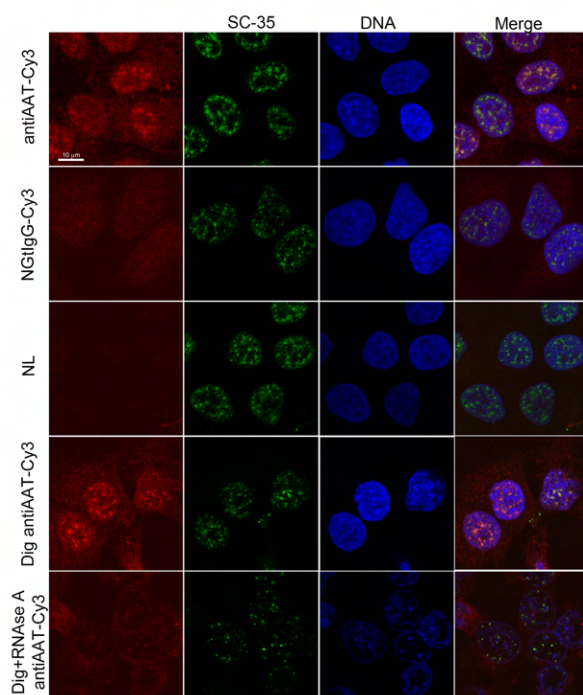


**Рисунок 10.** Изучение внутриклеточной локализации слитых белков ААТ-eGFP в клетках DU145 с помощью конфокальной микроскопии на живых клетках. Клетки трансфицировали конструкциями HLT, DLT1 и DLT2. DLT2 uORF off -

мутация ATG>AAG 5'-ОПС, перекрывающейся с главной ОПС *SERPINA1* в конструкции DLT2, NL - отрицательный контроль без добавления плазмидной ДНК.

### Предполагаемая эндогенная внутриклеточная изоформа ААТ накапливается в ядерных спеклах

Мы окрасили фиксированные клетки с помощью Су3-меченых козьих поликлональных антител, выращенных против ААТ из плазмы человека. Мы предварительно блокировали фиксированные клетки немечеными нормальными козьими антителами (NGtIgG) и окрашивали их специфическими IgG-Су3 в смеси с избытком немеченых NGtIgG. В этих условиях мы детектировали предполагаемую эндогенную изоформу ААТ в ядре и цитоплазме (Рисунок 11). Совместная окраска с анти-Sc-35 мышинными антителами показала, что предполагаемая изоформа ААТ накапливается в ядерных спеклах - субъядерном компартменте, обогащенном малыми РНП частицами и различными факторами сплайсинга (Рисунок 11). *In situ* обработка клеток РНКазой А предполагает, что изоформа ААТ - компонент РНП-комплекса (Рисунок 11).



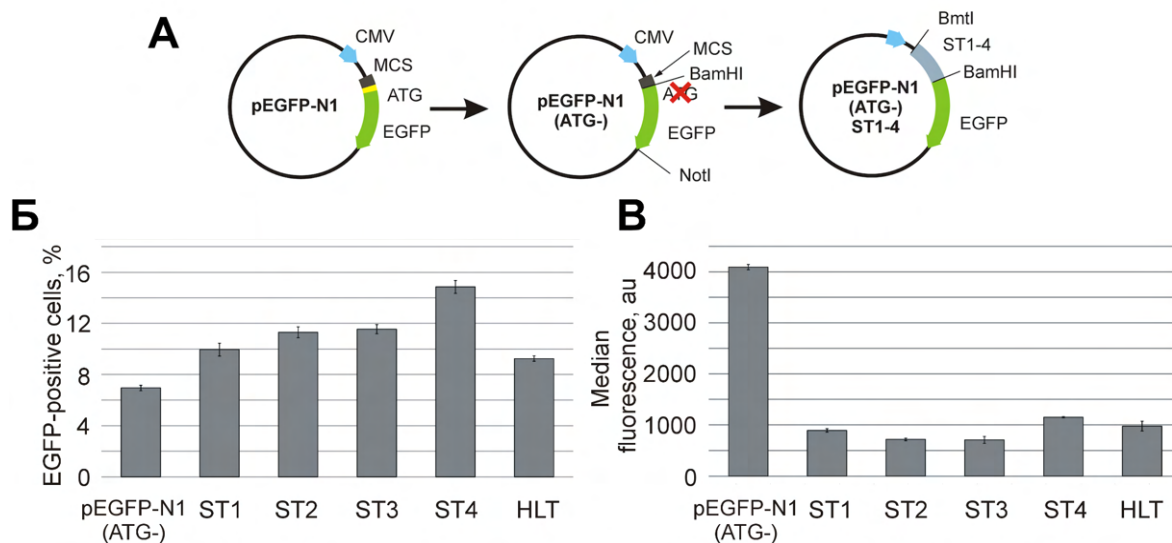
**Рисунок 11.** Выявление предполагаемой эндогенной изоформы ААТ в клетках DU145 методом иммунофлуоресцентной конфокальной микроскопии. Фиксированные клетки предварительно блокировали нормальными козьими IgG (NGtIgG), затем окрашивали с помощью анти-ААТ IgG-Cy3 и анти-Sc35 в смеси с немеченым NGtIgG. В качестве отрицательного контроля использовали NGtIgG-Cy3 в эквивалентной концентрации. Нижние панели - клетки обрабатывали или не обрабатывали RNase А после *in situ* пермеабиллизации дигитонином, фиксировали, блокировали и окрашивали анти-ААТ IgG-Cy3 и анти-Sc35 (в паре со вторичными IgG-Alexa 488) в конкурентных условиях.

Нам не удалось детектировать эндогенный внутриклеточный ААТ с помощью иммуноблоттинга в клеточных экстрактах с использованием тех же специфических антител, в то время как мы consistently детектировали ААТ в секретах. Это можно объяснить низкой представленностью изоформы ААТ и не оптимальным распознаванием используемыми антителами. Детекцию ААТ в иммунофлуоресценции можно объяснить сохранением некоторых конформационных эпитопов нативного ААТ. Для идентификации эндогенной изоформы ААТ необходимо получить антитела против усеченной и негликозилированной молекулы ААТ для увеличения шансов на ее специфическое распознавание. Это помогло бы "поймать" и охарактеризовать предсказанную *in vitro* внутриклеточную изоформу из опухолевых и/или нормальных тканей.

## Изучение трансляции индивидуальных коротких транскриптов *SERPINA1* в живых клетках *in vitro*

### Короткие транскрипты *SERPINA1* транслируются в живых клетках

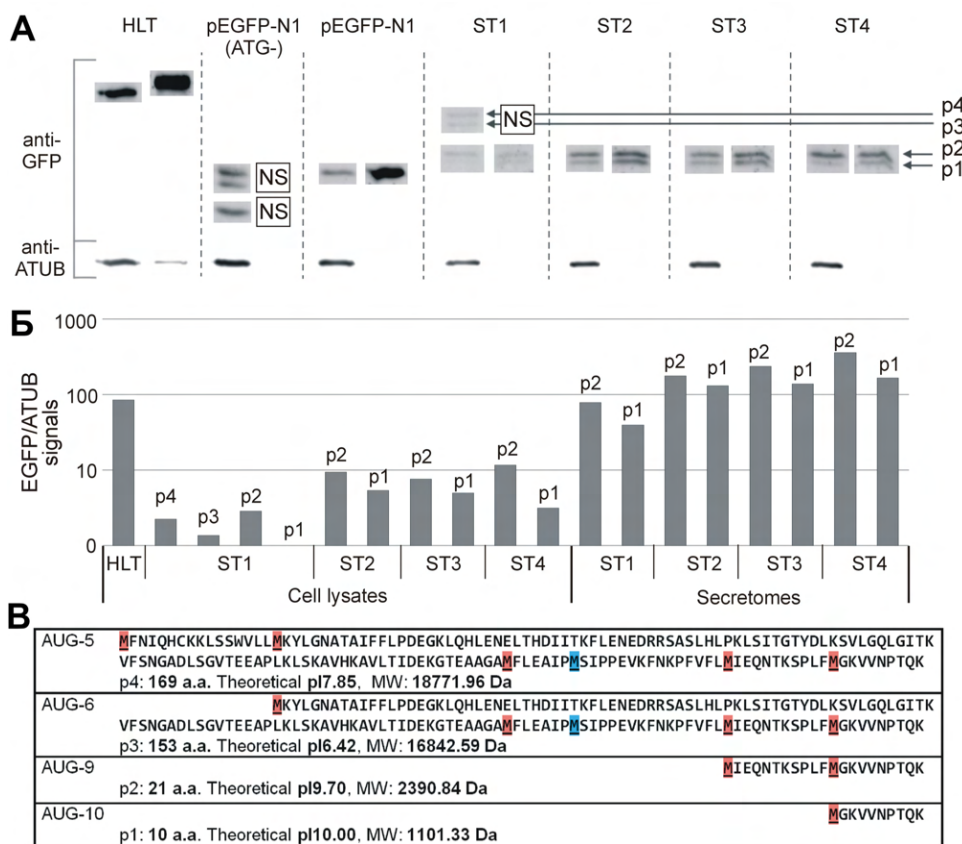
Короткие транскрипты содержат предполагаемые альтернативные сайты начала трансляции (АЛТ) в рамке *SERPINA1*, начинающиеся в EX3 и EX5 со стартовыми кодонами (кроме AUG-8) в подходящем контексте Козак (Рисунок 1, В). Мы клонировали кДНК, соответствующие коротким транскриптам ST1-4, в тот же мутантный вектор pEGFP-N1(ATG-) (рисунок 12, А).



**Рисунок 12.** Трансляция гепато-специфических коротких транскриптов *SERPINA1* *in vitro*. **А**) Схема клонирования кДНК. **Б**) Проточная цитофлуориметрия клеток HepG2 после трансфекции плазмидными конструкциями. Эффективность трансфекции и **В**) медианы флуоресценции eGFP+ продуктов трансляции. Данные представлены как среднее ± стандартное отклонение по двум независимым экспериментам.

Мы проанализировали клетки методом проточной цитофлуориметрии и сравнили эффективность трансфекции и медианы флуоресценции от разных конструкций (Рисунок 12, Б, В). eGFP+ клетки, выявленные после трансфекции “пустым” вектором pEGFP-N1(ATG-), отражают количество клеток, способных транслировать транскрипты с неканонических старт-кодонов, присутствующих в 5'-проксимальной области, кодирующей eGFP. Плазмиды со вставками *SERPINA1* давали больше eGFP+ клеток, чем мутантная плазида без вставки

(Рисунок 12, Б), что, возможно, отражает трансляцию коротких транскриптов. Мы отсортировали eGFP<sup>+</sup> клетки и сравнили eGFP-продукты от конструкций и плазмид без вставки методом иммуноблоттинг на клеточных экстрактах и секретомах (Рисунок 13, А; Рисунок 8П). Серпиновая компонента в слитом белке подтверждена детекцией сигнала от секретома HLT антителами против ААТ из плазмы, где, наряду с эндогенным секретируемым ААТ, мы наблюдали продукт массой 100 кДа, - как и в случае использования анти-eGFP антител в этом образце, но отсутствующий в секретома от pEGFP-N1 (Рисунок 8П). Так как для клонирования кДНК во всех конструкциях мы использовали обратный праймер на одну и ту же кодирующую область и подтвердили положение вставок секвенированием, мы продемонстрировали *SERPINA1* природу N-концов химерных продуктов.



**Рисунок 13.** *In vitro* трансляция коротких транскриптов *SERPINA1*. А)

Иммуноблоттинг лизатов eGFP<sup>+</sup> клеток и секретомов HepG2, трансфицированных плазмидами со вставками кДНК *SERPINA1*: HLT - длинный транскрипт, ST1-4 - короткие транскрипты; pEGFP-N1(ATG-) - мутантная плазида без вставки. Слева лизаты клеток, справа секретомы. NS - не секретируется.

Сигналы проявлены антителами против GFP. Б)

Интенсивность сигнала eGFP нормирована на альфа-тубулин и минимальный сигнал (ST1-p1) и отражает

относительную эффективность соответствующих предполагаемых альтОПС. Значения представлены в логарифмической шкале для лизатов клеток и секретомов. Значение для секретома HLT отсутствует из-за быстрого насыщения соответствующего сигнала. В) Последовательности предполагаемых продуктов трансляции, расчетные изоэлектрические точки (pI) и молекулярные массы продуктов p1-4. Расчеты проведены с использованием онлайн-программы ([https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)).

Трансляция в случае трансфекции мутантной плазмиды без вставки происходит с неканоническими стартов ввиду отсутствия главного AUG eGFP. В отличие от плазмид без вставки, ST1-4 дают продукты с большим молекулярным весом, предполагая трансляцию со стартов, расположенных в последовательности вставок. Детектированы четыре продукта ААТ, синтезируемых с коротких транскриптов: p1, p2, p3 и p4, которые имеют большую

электрофоретическую подвижность, чем eGFP в отдельности, на 0,7, 2,2, 15,6 и 18,7 кДа, соответственно (Рисунок 13, А; 8П). Отсутствие продуктов, соответствующих продуктам трансляции с неканонических CUG старт-кодонов в EX4 (в подходящем контексте) и нижележащих неканонических стартов мутантного вектора позволили сделать вывод о том, что короткие транскрипты транслируются с канонических AUG-кодонов. Разница в электрофоретической подвижности с eGFP указывает на то, что в ST1 имеются две активные альтОРС со стартами в EX3 (ATG-5 и 6, Рисунок 1, А, В; 13, В) и две икОРС со стартами в EX5 (ATG-9 и 10, Рисунок 1, А, В; 13, В). В ST2-4 активны те же две икОРС. Расчетные длины продуктов p1-p4 составляют 10, 21, 153 и 169 а.о., соответственно (Рисунок 13, В).

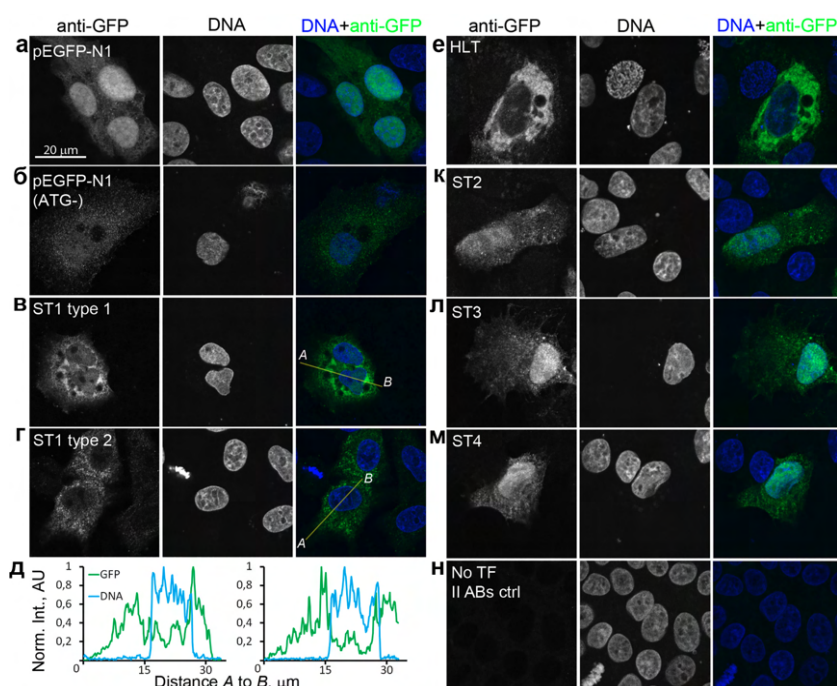
С каждого варианта усеченного транскрипта транслируются как минимум две альтОРС, вероятно, посредством сканирования с утечкой из-за неоптимального контекста Козак или, возможно, с отдельных по-разному свернутых молекул мРНК. Мы оценили относительную эффективность трансляции альтОРС (Рисунок 13, Б). икОРС в ST2-4, которые не содержат старт-кодоны из EX3, проявляют более высокую эффективность трансляции, чем ST1, а продукты p1/p2, по-видимому, являются основными продуктами с коротких транскриптов, вероятно, благодаря оптимальной структуре соответствующих 5'-НТО. Во всех коротких транскриптах икОРС-p2 активнее, чем икОРС-p1 (Рисунок 8; Рисунок 16, Б).

Пептидные продукты, происходящие из EX5, секретируются. Поскольку мутантный eGFP не обнаруживается в секретоме, мы предполагаем, что секреция p1(p2)-eGFP отражает естественное поведение С-концевых пептидов ААТ. Более длинные (p3/4) усеченные протеоформы ААТ не секретируются, возможно, поскольку последовательности, кодируемые EX3/EX4, влияют на биохимические свойства пептидов, происходящих из EX5, компенсируя их общий положительный заряд (рисунок 13, В). Известно, что положительный заряд на N-конце белка повышает эффективность секреции через SRP-независимый путь.

## **Изучение внутриклеточной локализации химерных продуктов трансляции с коротких транскриптов и предполагаемых эндогенных С-концевых пептидов ААТ**

### **С-концевые пептиды ААТ-eGFP накапливаются в ядрах**

Мы проанализировали трансфицированные клетки HepG2 после окрашивания их антителами против eGFP (Рисунок 14). Не мутантный eGFP накапливается в ядрах, в то время как мутантный равномерно распределен между цитоплазмой и ядром (Рисунок 18, а, б). Продукты с конструкций ST2-4 (p1/p2) обогащают ядра в отличие от продуктов с родительской мутантной плазмиды (Рисунок 14, б, к-м). Продукты ST1 дают два типа внутриклеточного распределения: с ядерным окрашиванием (Рисунок 14, в, д (слева)) и без него (Рисунок 14, г, д (справа)). Мы предполагаем, что в ядре наблюдаются только продукты ST1-p1/p2, в то время как более крупные p3/p4 в ядра не попадают. Как и ожидалось, продукт НЛТ локализуется в цитоплазме и предназначен для секреции (Рисунок 14, е).

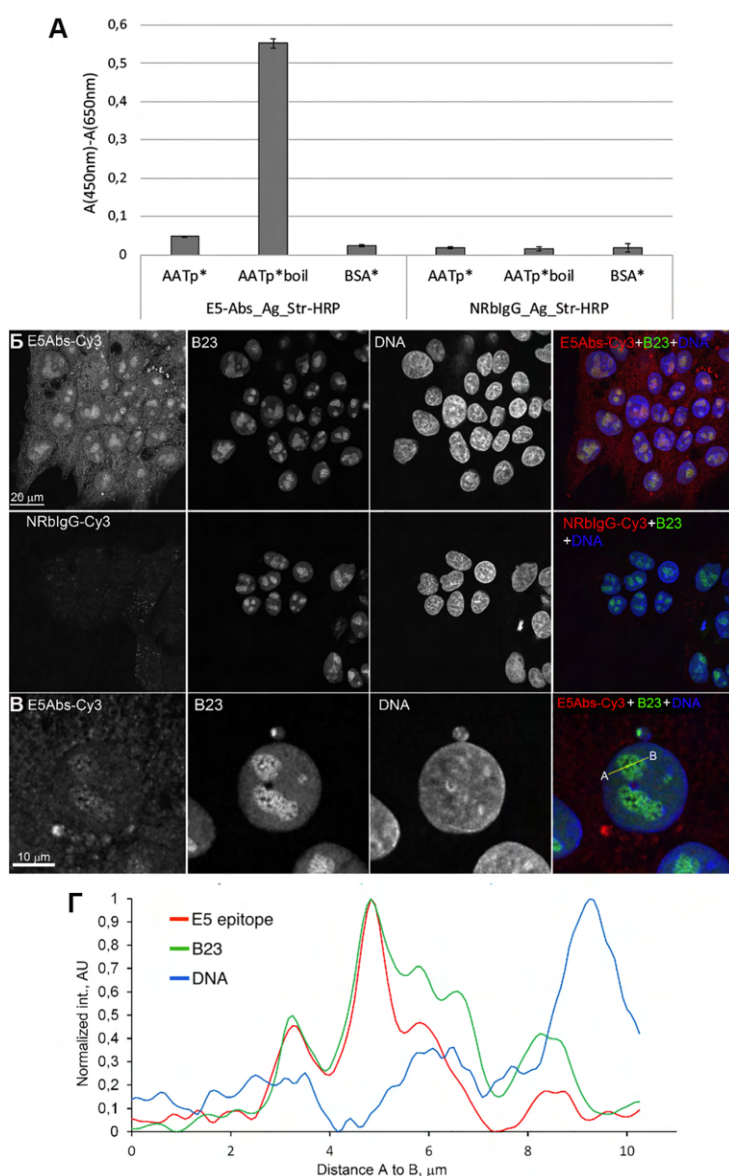


**Рисунок 14.** Иммунофлуоресцентный анализ продуктов трансляции после трансфекции клеток HepG2 плазмидами с и без вставок: pEGFP-N1 (а), pEGFP-N1 (ATG-) (б), ST1 (в, г) и соответствующие графики нормализованной интенсивности флуоресценции вдоль профиля линии (д, слева - тип 1, справа - тип 2), HLT (е), ST2-4 (ж-м, соответственно). н - отрицательный контроль.

### Поиск эндогенных С-концевых пептидов ААТ, синтезируемых опухолевыми клетками

Транслируются ли короткие транскрипты *SERPINA1 in vivo*? Мы получили кроличьи поликлональные антитела против последних 9 а.о. ААТ (Е5-эпитоп, кодируемый EX5) (Рисунок 10П). Мы провели ИФА по "вылавливанию" секреторного ААТ из раствора и показали, что эпитоп Е5 спрятан внутри белка и недоступен для распознавания без термически индуцированной денатурации антигена (Рисунок 15, А), - таким образом, антитела против Е5-эпитопа способны отличать усеченные протеоформы ААТ от полноразмерного нативного белка.

В отличие от отрицательного контроля, антитела против Е5-эпитопа равномерно окрашивают цитоплазму и ядро, в особенности ядрышки (Рисунок 15, Б-Г). Эпитопы Е5 солокализуются с белком ядрышка В23 (Рисунок 15, Г) и, по-видимому, локализуются в гранулярном компоненте ядрышка. Мы предполагаем, что именно пептиды р1 (С10) и/или р2 (С21) проникают в ядра и обогащают ядрышки. Стоит отметить, что предполагаемые эндогенные С-концевые пептиды ААТ, транслируемые с коротких транскриптов *SERPINA1*, и предполагаемая внутриклеточная изоформа ААТ, транслируемая с длинных транскриптов, схожи в их проникновении в ядро, но внутри ядра локализуются в разных субъядерных компартментах (Рисунок 12П).



**Рисунок 15.** Выявление предполагаемых эндогенных С-концевых пептидов ААТ. **А)** Тестирование распознавания Е5-эпитопа анти-Е5 антителами (Е5Abs) в структуре нативного и термически денатурированного ААТ из плазмы, меченого биотином (ААТр\*), методом ИФА. Отрицательные контроли: BSA\* и NRbIgG. Ag=антиген, Str-HRP= конъюгат стрептавидина с пероксидазой хрена. Данные представлены как среднее±стандартное отклонение по трем повторам. **Б)** Прямой иммунофлуоресцентный анализ с использованием Е5Abs-Cy3 и NRbIgG-Cy3 (отрицательный контроль). Фиксированные клетки блокировали немеченой неиммунной сывороткой кролика. Анти-B23 антитела использовали для окрашивания ядрышек. **В)** Изображение ядра с большим увеличением. **Г)** График нормализованных интенсивностей флуоресценции вдоль профиля линии на рисунке В.

Учитывая, что наша система предполагает сверхэкспрессию транскриптов с CMV промотора, ожидается, что эндогенные продукты трансляции с коротких транскриптов представлены в гораздо меньших количествах. Малый размер основных продуктов трансляции (p1/p2) и их низкая представленность существенно затрудняют их обнаружение, что, возможно, объясняет, почему этот феномен, а именно экспрессия *SERPINA1* с икОРС, не изучался ранее. Эндогенные С-концевые пептиды ААТ не детектируются с помощью иммуноблоттинга, по-видимому, ввиду невозможности разделения коротких пептидов в полиакриламидном геле. Так, синтетический пептид С36 consistently разделяется в геле, в то время как пептид С10, по-видимому, не связывается с додецилсульфатом натрия и не проникает в гель вовсе (рисунок 13П). В перспективе мы планируем провести

иммунопреципитацию E5-эпитоп-содержащих продуктов, совмещенную с их идентификацией методами масс-спектрометрии.

Мы также пробовали протестировать детектируемые предполагаемые С-концевые пептиды ААТ на их происхождение из EX5 гена *SERPINA1* с помощью геномного редактирования CRISPR/Cas9. Мы отобрали две сублинии HepG2 с гомозиготным нокаутом практически всей кодирующей области EX5 (Рисунок 14П). На ранних пассажах в нокаутных клонах практически отсутствовал сигнал от E5-эпитопа в ядрышках (Рисунок 15П), однако на более поздних сигнал восстановился (Рисунок 16П). Полученные результаты можно объяснить срабатыванием компенсаторных механизмов, обеспечивающих генетическую целостность. Недавние исследования продемонстрировали, что механизмы генетической компенсации путем транскрипционной адаптации срабатывают в случае, если не нарушена транскрипция мутированного гена, и обеспечиваются гомологией фрагментов деградированной аберрантной мРНК. В нашем случае функции предполагаемых транскрибированных протеоформ ААТ могли быть замещены продуктами генов-паралогов *SERPINA1* (Рисунок 14П, Г). Для верификации происхождения внутриклеточного E5-эпитопа необходимо провести эксперимент по нокауту всей области гена *SERPINA1*.

## Заключение

Изученные нами особенности экспрессии гена *SERPINA1* выводят на новый уровень понимания его функционирования и свидетельствуют о том, что, по крайней мере в опухолевых клетках человека, задействуются “резервы” гена в виде альтернативных сайтов инициации трансляции. 5'-проксимальный альтернативный старт-кодон AUG-2 реализуется в составе длинных транскриптов, а 3'-проксимальные старты (AUG-5, 6, 9 и 10) - в отсутствие ограничений главной ОПС в составе коротких транскриптов, содержащих экзон 5 и образующихся, по-видимому, путем посттранскрипционного процессинга гепато-специфического сплайсированного мРНК-предшественника. Данные свидетельствуют в пользу повышения биологического разнообразия продуктов гена *SERPINA1* за счет реализации альтернативных сайтов инициации трансляции, поскольку химерные продукты, синтезированные с главного и альтернативных стартовых кодонов, и соответствующие им предполагаемые эндогенные продукты трансляции локализуются в разных клеточных компартментах. Дальнейшие исследования этого явления могут раскрыть новые функциональные роли укороченных изо(протео)форм ААТ и их участие в различных биологических процессах и/или развитии патологических состояний.



## ВЫВОДЫ

1. Экспрессия длинных и предполагаемых коротких транскриптов, содержащих и не содержащих главную ОРС, соответственно, и содержание ААТ в секретах максимально в линии гепатоцеллюлярной карциномы HepG2, из линий не печеночного происхождения - в линии опухоли эпителия простаты DU145.

2. В препаратах суммарной РНК клеточных линий опухоли печени (HepG2), предстательной (DU145) и молочной железы (MDA-MB-231 и T47D) и тканей печени и почки детектируются длинные транскрипты. В HepG2 идентифицирован один, а в DU145 три варианта длинных мРНК, отличающихся структурой 5'-НТО. Длинные транскрипты преимущественно содержат укороченный вариант 3'-НТО.

3. С длинных транскриптов *SERPINA1* помимо секреторной транслируется внутриклеточная изоформа ААТ с внутреннего старт-кодона AUG-2 независимо от структуры 5'-НТО. 5'-НТО регулирует частоту инициации трансляции с главного AUG-1 и определяет соотношение обеих изоформ.

4. Химерный продукт трансляции с внутриклеточной изоформой ААТ мигрирует между ядром и цитоплазмой, в отличие от такового с секреторной, локализуящегося в ЭПР. Предполагаемая эндогенная внутриклеточная изоформа ААТ накапливается в ядерных спеклах.

5. В суммарной РНК HepG2 и тканей печени детектируются короткие транскрипты *SERPINA1*, содержащие экзон 5, генерируемые предположительно путем посттранскрипционного процессинга мРНК-предшественника. В HepG2 идентифицирована структура 5'-концов четырех вариантов коротких транскриптов, содержащих укороченную 3'-НТО.

6. Короткие транскрипты *SERPINA1* служат источником четырех альтернативных протеоформ ААТ, в том числе С-концевых пептидов, с предполагаемых стартов AUG-5, 6, 9 и 10. Химерные продукты, соответствующие пептидам ААТ, накапливаются в ядрах, предполагаемые эндогенные пептиды - в ядрышках.

## Список опубликованных работ по теме диссертации

### Публикации в журналах

1. **Maslakova A.A.**; Golyshev S.A.; Potashnikova D.M.; Moisenovich A.M.; Orlovsky I.V.; Smirnova O.V.; Rubtsov M.A. *SERPINA1* long transcripts produce non-secretory alpha1-antitrypsin isoform: *In vitro* translation in living cells // *Int J Biol Macromol.* 2023. Vol. 241. P. 124433. IF Web of Science = 8,2
2. **Maslakova A.A.**; Didych D.A.; Golyshev S.A.; Katrukha I.A.; Viushkov V.S.; Zamalutdinov A.V.; Potashnikova D.M.; Rubtsov M.A.; Smirnova O.V.; Orlovsky I.V. Towards unveiling the nature of short *SERPINA1* transcripts: Avoiding the main ORF control to translate alpha1-antitrypsin C-terminal peptides // *Int J Biol Macromol.* 2022. Vol. 203. P. 703-717. IF Web of Science = 8,2
3. **Maslakova A.A.**; Telkov M.V.; Orlovsky I.V.; Sokolova O.S. Comparative analysis of *SERPINA1* gene expression in tumor cell lines // *Moscow University biological sciences bulletin.* 2015. Vol. 70, № 3. P. 127–131. IF Scopus = 0,9. Статья доступна на русском языке в журнале Вестник Московского университета. Серия 16: Биология. 2015. Т. 70, № 3. С. 26–31.
4. **Докрунова А.А. (Маслакова А.А.)**; Соколова О.С. Содержание  $\alpha 1$ -антитрипсина в сыворотке при заболеваниях предстательной железы // *Урология.* 2012. Т. 5. С. 77-80. IF Scopus = 0,293; IF РИНЦ Web of Science = 0,94
5. **Докрунова А.А. (Маслакова А.А.)**; Уласова Н.Ю.; Крамина Т.Е.; Альбицкий И.А.; Хасигов П.З.; Соколова О.С. Полиморфизм альфа1-антитрипсина при раке и гиперплазии предстательной железы // *Молекулярная медицина.* 2010, Т. 6. С. 43-49. IF РИНЦ Web of Science = 0,5