

Выявление породоспецифичности собак методом SNP: что это, за и против.

В последнее время все чаще возникают вопросы у заводчиков и владельцев на тему определения породистости своих собак молекулярно-генетическими методами. Да, действительно, сейчас некоторые лаборатории стали предлагать такую услугу, одной из таких лабораторий и неоспоримым лидером в этом направлении является американская компания Embark, в частности ее ветеринарное подразделение EmbarkVet (<https://embarkvet.com/>). Многие делают такие тесты и получают «удивительные» результаты, которые не совсем понятно, куда и как использовать (если это вообще возможно). Я постараюсь максимально понятным языком описать процедуру, результаты и их применимость в работе с породой.

Итак, **SNP – это полиморфизм одиночных нуклеотидов** (single nucleotide polymorphism). Геном всех живых существ кодируется ДНК (дезоксирибонуклеиновой кислотой), которая представляет из себя последовательно и неразрывно соединенные нуклеотиды, разделяемые только по хромосомам. У собак всего таких нуклеотидов - 2,4 миллиарда в 39 парах хромосом. Всего в ДНК встречается 4 разных нуклеотида: аденин, тимин, гуанин и цитозин, обозначаемых первыми буквами: А, Т, Г, Ц (в английском варианте: А, Т, G, C), в дальнейшем будем называть их просто «буквы». Именно последовательность из различного сочетания этих четырех букв определяет, какие белки (ферменты) и как будут синтезироваться в организме. Незначительные изменения в последовательности букв в большинстве случаев не приводит ни к каким последствиям и только в редких ситуациях их можно наблюдать в виде изменений внешнего вида собаки, или, например, в наследственных заболеваниях, когда работа ферментов нарушается вследствие замен отдельных букв в последовательности ДНК.

Откуда возникают такие замены? Мутагенез – это неотъемлемое свойство всех живых организмов, которое обеспечивает разнообразие и приспособление к окружающей среде. **Мутагенез происходит из-за «ошибок» при копировании ДНК во время деления клетки или под действием мутагенных агентов окружающей среды (сложно сохранить неизменность целых 2,4 млрд букв).** Самое главное – понимать, что это нормальный процесс, который только изредка может приводить к негативным последствиям. Механизм, при котором вследствие мутагенеза происходит изменение свойств организма называется изменчивостью, это длительный процесс, занимающий десятки и сотни поколений животных, поэтому в пределах одного организма ДНК всех клеток одинакова почти на 100%.

Теперь надо напомнить, что одинаковых организмов не бывает, **каждое животное уникально в первую очередь именно из-за постепенного накопления замен в ряду поколений и их комбинирования при оплодотворении и наследовании генетического материала от родителей. Каждый организм имеет отличия в последовательности ДНК.** Именно сочетание этих отличий обеспечивает уникальный внешний вид и особенности.

При выведении породы, человеком отбираются особи, которые максимально полно отвечают требованиям (стандартам), что обеспечивается уникальным сочетанием тех самых изменений в ДНК, при этом **именно те замены, которые нужны для породы, закрепляются, но это не отменяет наличия других замен, не влияющих на внешний вид и характеристики породы.**

Как это можно себе представить? Возьмем небольшой участок ДНК собаки (в данном случае нам не важно какой):

```
ATGTCTGGGCAGGGCCCCAGAGAAGGCTGCTGGGCTCTCTCAATGGCACCTCCCCAGCCACCCCTCACT
TCGAGCTGGCTGCCAACCCAGACCGGGCCCCGGTGCCCTGGAGGTGTCCATTCCCGACGGGCTGTTCCTCAG
CCTGGGGCTGGTGAGCGTFTGTGGAAAATGTGCTGGTGGTGGCCGCCATTGCCAAGAACCAGCAACCTGCAC
TCGCCATGTATTACTTCATCGGTTGCCCTGGCTGTGTCGGACCTGCTGGTGAGCGTGGGCAATGTGCTGG
AGACGGCCGTCATGCTGCTGGTGGCGGCAGGCACCTTGGCTGCTCAGGCTGCTGTGGTGCAGCAGCTGGA
CGACATCATTGACGTGCTCATCTGTGTTCCATGGTATCCAGCCTCTGCTTCCTGGGCGCCATTGCCGTG
```

GACCGCTACCTCTCCATCTTCTACGCGCTGCGATAACCACAGCATCGTCACACTCCCGCGGGCGTGGCGGG
CCATCTCCGCTATCTGGGTGGCTAGCGTCCCTCTCCAGCACGCTCTTCATTGCCTACTACAATCACACGGC
CGTCCTGCTTTGTCTTGTTCAGCTTCTTTGTAGCCATGCTGGTGCATGGCAGTGTGTACGTCCACATG
CTTGCCCGCGCTGCCAGCACGCCCCGAGGTAATTGCCCGGCTCCATAAGAGGCAGCACTTCATCCCCCAGG
GCTTTGGCCTCAAGGGCGCTGCCACACTCACTATCCTGCTGGCATTTCCTTTCTCTGCTGGGGCCCCCTT
CTTCTTGCACCTCTCACTCATGGTCCCTCTGCCCTCAACACCCCATCTGTGGCTGCGTCTTTTCAGAACTTC
AACCTCTTCCCTACCCCTCATCATCTGCAACTCCATCATTTGACCCCTTCATCTACGCCTTCCGCAGCCAGG
AGCTCTGAAAGACTCTCCAAGAGGTAGTGCTATGTTCCCTGGTGA

Мы видим, что это последовательность из комбинации четырех букв. Если мы возьмем другую собаку, то сможем заметить, что этот же участок может оказаться немного другим:

ATGTCTGGGCAGGGCCCCAGAGAAGGCTGCTGGGCTCTCTCAATGGCACCTCCCCAGCCACCCCTCACT
TCGAGCTGGCTGCCAACCCAGACCGGGCCCCGGTGCCTGGAGGTGTCCATTCCCAGCGGGCTGTTCCCTCAG
CCTGGGGCTGGTGAGCGTTGTGGAAAATGTGCTGGTGGTGGCCGCCATTGCCAAGAACCGAACCTGCAC
TCGCCCATGTATTACTTCATCGGTTGCCCTGGCTGTGTCCGACCTGCTGGTGAGCGTGGGCAATGTGCTGG
AGACGGCCGTCATGCTGCTGGTGGCGGCAGGCACCTTGGCTGCTCAGGCTGCTGTGGTGCAGCAGCTGGA
CGACATCATTTGACGTGCTCATCTGTGTTCCATGGTATCCAGCCTCTGCTTCCCTGGGCGCCATTGCCGTG
GACCGCTACCTCTCCATCTTCTACGCGCTGCGATAACCACAGCATCGTCACACTCCCGCGGGCGTGGCGGG
CCATCTCCGCTATCTGGGTGGCTAGCGTCCCTCTCCAGCACGCTCTTCATTGCCTACTACAATCACACGGC
CGTCCTGCTTTGTCTTGTTCAGCTTCTTTGTAGCCATGCTGGTGCATGGCAGTGTGTACGTCCACATG
CTTGCCCGCGCTGCCAGCACGCCCCGAGGTAATTGCCCGGCTCCATAAGAGGCAGCACTTCATCCCCCAGG
GCTTTGGCCTCAAGGGCGCTGCCACACTCACTATCCTGCTGGCATTTCCTTTCTCTGCTGGGGCCCCCTT
CTTCTTGCACCTCTCACTCATGGTCCCTCTGCCCTCAACACCCCATCTGTGGCTGCGTCTTTTCAGAACTTC
AACCTCTTCCCTACCCCTCATCATCTGCAACTCCATCATTTGACCCCTTCATCTACGCCTTCCGCAGCCAGG
AGCTCTGAAAGACTCTCCAAGAGGTAGTGCTATGTTCCCTGGTGA

Для простоты понимания возьмем только короткие участки ДНК из этих последовательностей, где мы отметили замены и расположим их рядом.

Первая замена:

Собака 1: CGACATCATTTGACGTGCTCATCTGTGTTCCATGGTATCCAGCCTCTGCTTC
Собака 2: CGACATCATTTGACGTGCTCATCTGTGTTCCATGGTATCCAGCCTCTGCTTC

Вторая замена:

Собака 1: GCTGCCACACTCACTATCCTGCTGGCATTTCCTTTCTCTGCTGGGGCCCCCT
Собака 2: GCTGCCACACTCACTATCCTGCTGGCATTTCCTTTCTCTGCTGGGGCCCCCT

Это не привело ни к каким серьезным последствиям, ни для здоровья, ни для внешности собаки, но могло привести к тому, что фермент, кодируемый этим участком ДНК, стал работать чуть лучше или чуть хуже, но в допустимых пределах, а если замена не попала в ген, а содержится в некодирующей области – вообще никак не сказалась. Мы привели всего 954 буквы, а в геноме собаки их в 2,5 миллиона раз больше и такие замены «раскиданы» по всему геному и всего их, на самом деле, десятки тысяч. **Именно это явление отличия животных между собой в пределах одной породы или вида называется SNP или полиморфизм одиночных нуклеотидов.**

А если совсем упростить запись и откинуть вообще все лишнее, т.е. одинаковое, между собаками, то обе замены можно записать так:

Собака 1: GG
Собака 2: TA

Это мы изобразили генетический SNP профиль по двум нуклеотидам, или буквам.

Как же происходит анализ таких букв? Конечно, все десятки тысяч нуклеотидов никто не исследует для рутинного анализа при коммерческом тестировании – это избыточно и дорого, по большому счету просто не нужно. **Для исследования выбирается, как правило, несколько сотен таких точек во всем геноме и по ним проводится анализ.** Но мы помним, что букв всего четыре – это ограничение всех возможностей изменений в одной конкретной точке.

Для наглядности возьмем 50 букв, построим из них SNP профиль и попробуем сравнить случайных собак между собой (помним, что, на самом деле, между каждой из букв может стоять сотни тысяч других, не взятых в анализ):

```

Собака 1: TACGGACAGGACGTC AAGTGCCTAACTGCAACGAGTAAGTCGTTCCGGTT
Собака 2: TACAGACAGGACGTCAGTGCCTAACTGCACGAGTAAGACGTTCCGGTT
Собака 3: TACGGACAGGACGTCAGAGCCTAACTGCAACGGGTAAGTCGTTCCGGTT
Собака 4: TACAGACTGGACGTCAGTGCCTAACTGCAACGAGTAAGTCGTTCCGGTT
Собака 5: TACGGACAGGACGTCAGTGCCTAACTGCAACGAGTACGTCGTTGCGGTT
Собака 6: TACGGACTGCACGTCAAGTGCCTAACGGCAACGGGTAAGTCGTTGCGGTT
Собака 7: TACAGACTGCACGTCAAGTGCCTAACGGCACGGGTAAGTCGTTGCGGTT
Собака 8: TACGGACTGCACGTCAAGAGCCTAACGGCAACGGGTAAGTCGTTGCGGTT
Собака 9: TACAGACTGGACGTC AAGTGCCTAACGGCAACGGGTAAGACGTTGCGGTT
Собака 10: TACGGACTGCACGTCAAGTGCCTAACGGCAACGGGTAAGTCGTTGCGGTT

```

Мы видим, что каждая собака оказывается уникальной, но наша задача понять, как из этого можно определить породу. **Самое важное ДО этого иметь достаточную базу данных уже проанализированных собак, для которых известна порода!** Представим себе, что первые пять (1-5) наших собак принадлежат породе чихуахуа, а вторые пять (6-10) – породе бигль:

Чихуахуа :

	1	2	3	4	5	6	7	8
Собака 1:	TACGGAC	AGGACGTC	AAGTGCCTAACTGCAACG	AGTAAGTCGTT	CCGGTT			
Собака 2:	TACAGAC	AGGACGTC	AGTGCCTAACTGCACGAGTAAG	ACGTT	CCGGTT			
Собака 3:	TACGGAC	AGGACGTC	AGAGCCTAACTGCAACGGGTAAGTCGTT	CCGGTT				
Собака 4:	TACAGACT	GGACGTC	AGTGCCTAACTGCAACGAGTAAGTCGTT	CCGGTT				
Собака 5:	TACGGAC	AGGACGTC	AGTGCCTAACTGCAACGAGT	ACGTCGTT	GCGGTT			

Бигль :

Собака 6:	TACGGACTGC	CACGTCAAGTGCCTAACGGCAACGGGTAAGTCGTTGCGGTT
Собака 7:	TACAGACTGC	CACGTCAAGTGCCTAACGGCACGGGTAAGTCGTTGCGGTT
Собака 8:	TACGGACTGC	CACGTCAAGAGCCTAACGGCAACGGGTAAGTCGTTGCGGTT
Собака 9:	TACAGACTGGACGTC	AAGTGCCTAACGGCAACGGGTAAGACGTTGCGGTT
Собака 10:	TACGGACTGC	CACGTCAAGTGCCTAACGGCAACGGGTAAGTCGTTGCGGTT

Можно заметить, что на фоне отличий всех собак между собой отмечается, что по некоторым точкам отличия наблюдаются большей частью только между породами – выделены рамками и пронумерованы. Именно по таким точкам, которые встречаются наиболее часто у собак одной породы формируется SNP профиль породы. Обратите внимание, что не обязательно у всех собак одной породы должны быть одни и те же буквы в одной точке – их встречаемость должна быть не менее 80% среди собак одной породы, чтобы точка была включена в профиль породы. Но, конечно, надо учитывать все точки для всех пород: если еще больше сократить профили до пород, убрав разнообразие и не отличающиеся точки между собаками то мы получим:

```

12345678
чихуахуа: AGTTTATC
бигль: TCAAGGGG

```

Следует еще раз отметить, что для формирования таких SNP профилей породы необходима достаточно большая выборка животных, желательно не меньше 100 собак, с установленной породой. Возникает первый вопрос: кто заявляет, что это именно та порода, что в ней не было подмеса и какова ее история? На данный момент, при формировании коммерческих баз данных лаборатории руководствуются данными, заявленными владельцами – это может внести некоторый сумбур – теоретически можно предположить такую ситуацию, когда сдается 100-200 собак породы ньюфаундленд, но заявляются как таксы – лаборатория же не видит собак – это может внести сумбур во всю структуру базы, но, это, скорее, теоретическая ситуация, на практике,

конечно, не реализуемая. Самое главное понимать, что то, что заявляет заказчик анализа, то и примет лаборатория для базы данных.

Теперь попробуем взять одиннадцатую случайную собаку, породу которой мы не знаем, сначала ее SNP профиль:

```
          1 2          3          4 5          6          7 8
Собака 11: TGC GCACTGGACCTCAAGAGCCAAACGGCAAGGGGTTAGTCGTTCCGGCT
```

Получим из нее по пронумерованным точкам профиль породы и подставим уже известные нам профили чихуахуа и бигля:

```
          12345678
Собака 11: TGAAGGTG
чихуахуа: AСТТТАТС
бигль: TCAAGGGG
```

И так, мы получили, что наша собака имеет по 6 точкам из 8 (а помним, что на самом деле их несколько сотен) сходство с чихуахуа и по двум – с биглем. Да, такое возможно, это же изменчивость и, если предки собаки даже той же породы разошлись много поколений назад – такое вполне может быть. Т.е. 75% точек собаки 11 совпадает с чихуахуа и 25% совпадает с биглем.

А теперь результат, который мы получим от лаборатории:

Собака 11 имеет сходство с чихуахуа на 75% и на 25% - с биглем. Получается, что у нас к чихуахуа подмешан бигль? НЕТ, ни в коем случае! Просто часть точек породного профиля не совпали! Т.е. мы можем с уверенностью говорить, что наша собака – чихуахуа, но никак не можем утверждать, что туда подмешан бигль. Названная ситуация работает, когда обнаруживается сходство с какой-то породой более 70%, конечно, **если мы видим, что самый похожий профиль совпадает лишь на 50% и меньше - это говорит о том, что перед нами метис или представитель породы, которой нет в базе данных.**

Последнее утверждение очень важно понимать в отношении редких и эндемичных пород, достаточной выборки которых может не оказаться у лаборатории – она выдаст сходство с теми профилями, которые есть. И **если у лаборатории нет готового достоверного профиля восточноевропейской овчарки, то она не сможет сказать точно, к какой породе принадлежит собака.**

А если породного профиля нет, а надо удостовериться, что собаки принадлежат одной породе и она не похожа ни на одну другую? Для этого: **1. Необходимо собрать достаточно собак одной породы и сдать их биоматериал на анализ. 2. Сравнить полные генетические профили собак между собой**, как мы это делали в самом начале для профилей из 50 букв и удостовериться, что мы можем вывести комплекс точек, по которым нет расхождения между 80% собак. Если это не получается, то нельзя говорить, что у нас собаки принадлежат вообще одной породе. **3. Получить результат, в котором схожесть с другими породами не будет превышать 60%** (иначе можно будет сказать, что это просто отдаленная помесь уже известной породы). **Соблюдение условий по второму и третьему пунктам позволит нам получить SNP профиль породы и говорить о том, что перед нами новая порода, значительно отличающаяся от других известных.**

В заключение можно сказать, что, исходя из принципов, лежащих в основе получения SNP профиля этот **метод применим** в случаях: **1. установления родства** (сейчас может использоваться параллельно с STR анализом) – в этом случае сравниваются полные SNP профили собак, а не породные; **2. анализа метисов:** от кого же пошло это чудо (это, скорее, вопрос интереса) или для доказательства или опровержения принадлежности собаки той или иной породе; **3. установления генетической обособленности породы.** И метод **не применим** для выявления так называемой

«породистости» - процент схожести больше 70% будет в любом случае указывать на принадлежность определенной породе, но оставшиеся меньше 30% не докажут подмес, а покажут генетическую отдаленность вашей собаки от общего генетического профиля, имеющегося в лаборатории. Тут следует учитывать и географическое происхождение собак: конечно, в американскую лабораторию попадает больше собак американского и канадского разведения, которые зачастую могут иметь очень большую отдаленность в поколениях от собак Старого света, Азии и России, а, соответственно, и породные SNP профили пород могут несколько отличаться.