Pontifícia Universidade Católica do Rio Grande do Sul
Programa de Pós-Graduação em Zoologia

Análises genômicas da onça-pintada (*Panthera onca*): caracterização do genoma completo e investigação de regiões sob seleção através de comparações interespecíficas e populacionais

Henrique Vieira Figueiró

Tese de Doutorado
Porto Alegre
-2016-

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

Análises genômicas da onça-pintada (*Panthera onca*): caracterização do genoma completo e investigação de regiões sob seleção através de comparações interespecíficas e populacionais

Henrique Vieira Figueiró
Orientador: Dr. Eduardo Eizirik

TESE DE DOUTORADO
PORTO ALEGRE – RS – BRASIL
2016I

SUMÁRIO

"*You, me or nobody is gonna hit as hard as life. But it ain't about how hard you hit. It's about how hard you can get hit and keep moving forward.*"

Rocky Balboa

AGRADECIMENTOS

O doutorado mais que um título, é a soma de todas as influências que recebemos em nossa vida, não apenas como profissionais, mas também como pessoas. É impossível agradecer todos que me influenciaram e tornaram esse caminho possível. Desde sempre contei com o apoio de professores, colegas, amigos e especialmente minha família para as minhas decisões.

Antes de tudo, à minha família, que sempre me apoiou em tudo que foi necessário. Meu pai, que me ensinou a sonhar e especialmente minha mãe, que sempre lutou muito e é meu exemplo de vida.

Aos meus amigos que sempre estiveram lá, algumas vezes longe, mas mesmo assim presentes. Muito obrigado Karen, Luísa, Gustavo, Mari e Guilherme, vocês são demais.

À Helenita, por ser essa pessoa tão especial que tanto me apoiou nesse último ano.

Ao meu orientador, Eduardo Eizirik, pela orientação, por ser um exemplo de pesquisador e ter me ensinado tanto ao longo dos anos, indo além do meio acadêmico. Muito obrigado.

Às minhas duas padawans! Fernanda e Sarah, muito obrigado. Vocês foram essenciais e aprendi muito com as duas, mais do que vocês podem imaginar. Vocês vão longe e tenho muito orgulho de ter acompanhado a formação de vocês.

À Cris, que começou me co-orientando na bancada quando entrei no laboratório e foi uma das pessoas mais importantes para a execução do meu doutorado. Muito obrigado pela parceria e amizade durante esse tempo.

À Maíra, que sem dúvida chegou na hora certa! Muito obrigado por todo o apoio, sugestões e auxílio nesse momento tão conturbado que é o final de um doutorado.

Talita, Manoel, Fernanda Pedone, Ricardo, Flávia, Laura, o nome de vocês não poderia faltar aqui. Muito obrigado pela amizade, conselhos, risadas e momentos do café no Genoma.

Aos genômicos! Todas as gerações que passaram pelo laboratório. Aprendi muito com todos vocês.

À todas as técnicas que passaram pelo laboratório, muito obrigado, sem vocês o Genoma não seria nem de perto o que é hoje. Meu agradecimento especial à Cladi, que mais me acompanhou durante meus momentos de bancada, e até hoje ainda lembro dos ensinamentos dela.

Aos mestres que eu tive durante essa jornada. Guto e Eros, que foram exemplos de professores e uma das maiores influências para eu ter escolhido a biologia. Os professores da Biociências, em especial o Bicca, o Gervásio, o Sandro. E a professora Eliane Santarém, que provavelmente mudou meu rumo como pesquisador, mostrando como biologia molecular é interessante.

Ao professor Rasmus Nielsen, por ter me aceitado no grupo de pesquisa dele. Com certeza, foi um dos momentos mais importantes da minha formação como pesquisador.

Ao pessoal do Nielsen *lab*. Foi muito legal poder conhecer um pouco do resto do mundo através de vocês. Um agradecimento especial ao Tyler, Ke e a Lydia.

À Barbara e ao Andy e seus gatos Bear e Mia. Com certeza, vocês tornaram essa experiência de estar longe de casa muito melhor. Foi uma experiência incrível ter compartilhado esse ano com vocês. Muito obrigado por me receberem na casa de vocês.

Aos colaboradores dos projetos desenvolvidos durante o doutorado. Todos aqueles que contribuíram com amostras ou participaram de alguma etapa do projeto. Em especial, o professo William Murphy no Texas A&M University, o grupo do laboratório de Biotecnologia Animal da ESALQ-USP e os pesquisadores do Centro de Pesquisa René Rachou na FIOCRUZ-MG, além de todos coautores dos trabalhos que compõem a tese.

Ao CNPq e a CAPES pela bolsa de doutorado e doutorado-sanduíche, respectivamente. E todas as agências financiadoras devidamente lembradas nos artigos que compõem essa tese.

À PUCRS pela infraestrutura, e todos seus funcionários por tornarem esse ambiente tão agradável

RESUMO

Nos últimos 10 anos, o sequenciamento genômico de alto desempenho revolucionou a biologia evolutiva. Com os avanços gerados pelo sequenciamento do genoma completo de espécies-modelo, agora é possível aplicar essas técnicas em animais com praticamente nenhum recurso genético disponível. O sequenciamento completo de genomas, bem como o uso de técnicas de representação reduzida, permitem explorar questões evolutivas complexas como, por exemplo, detecção de hibridação e assinaturas de seleção natural em uma escala genômica. Dentre os grupos taxonômicos que podem se beneficiar de tais técnicas está o gênero *Panthera*. O grupo é composto por cinco espécies atuais (*P. onca*, *P. tigris*, *P. leo*, *P. pardus* e *P. uncia*), todas elas apresentando grande porte e atuando como predadores de topo nos ambientes que ocupam. Devido às baixas densidades, alarmante perda de habitat e constantes conflitos com humanos, o nível de ameaça em que essas espécies se encontram é preocupante. Dentre as espécies do grupo, está a onça-pintada (*P. onca*), única integrante do gênero na região Neotropical e o principal foco deste trabalho. Nesse sentido, o presente estudo busca caracterizar pela primeira vez o genoma da onça-pintada, incluindo análises comparativas com as outras quatro espécies do gênero. Além disso, o trabalho tem como objetivo avaliar as populações de onça no Brasil e buscar assinaturas de seleção divergente nos biomas que ela ocupa. Para o sequenciamento do genoma da espécie, foram utilizadas quatro bibliotecas genômicas, com uma cobertura estimada de 84x. A sequência do genoma completo permitiu a anotação de 25.441 genes e a descrição de outros componentes do genoma (p.ex. *ncRNA*, microssatélites, *numts*). Adicionalmente, foi sequenciado o genoma de um leopardo (*P. pardus*) com cobertura estimada de 25x. Com esses dois novos genomas, completou-se um conjunto abrangendo todas as cinco espécies do gênero, permitindo a realização de análises de discordância filogenética para o grupo e detecção de seleção positiva utilizando um conjunto de 13.143 genes ortólogos. Foi possível demonstrar eventos de hibridação durante o processo de especiação das espécies do gênero, bem como sinais de seleção positiva em genes envolvidos em características que se destacam nos grandes felídeos. Entre eles, fenótipos potencialmente afetados por genes sob seleção incluem o crânio e membros robustos da onça-pintada, o comportamento social no leão, adaptação ao frio no leopardo das neves e a presença de listras no tigre. Com o uso de captura de exoma, que tem como objetivo o sequenciamento do conjunto de exons da espécie, foi possível realizar uma nova avaliação das características genéticas de populações de onça-pintada, bem como a detecção de assinaturas de adaptação local. Entre os resultados obtidos está a presença de genes sob seleção relacionados com metabolismo energético em populações da Amazônia, adaptações relacionadas com

desenvolvimento corporal no Pantanal e imunidade na Mata Atlântica. Adicionalmente, foram observados diversos genes de pigmentação com assinaturas de seleção em diferentes biomas. Esses genes, além de afetarem a coloração dos animais, possuem efeitos pleiotrópicos no desenvolvimento e imunidade da espécie. Esses resultados auxiliam no entendimento dos processos evolutivos que moldaram a adaptação das espécies do gênero, e em especial a onça-pintada, aos ambientes que elas ocupam atualmente.

Palavras-chave: genômica comparada; genômica populacional; seleção natural; discordância genealógica; adaptação local; exoma.

ABSTRACT

In the past 10 years, high throughput sequencing has revolutionized evolutionary biology. With the technical advances that emerged with the genome sequencing of model species, it is now possible to apply these techniques to taxonomic groups without any previously available genetic resources. Complete genome sequencing and reduced representation methods have enabled us to explore deeper evolutionary questions, such as detecting ancient hybridization and signatures of selection on a genomic scale. Among the groups that could benefit from these methods is the *Panthera* genus. The group is composed by five species (*P. onca*, *P. tigris*, *P. leo*, *P. pardus* and *P. uncia*), all of which are large felids that exert important ecological role as apex predators in their habitats. Their low densities, alarming rates of habitat loss and chronic conflict with humans, all of them are threatened with extinction in the wild and thus important targets for conservation. One of the species in this group, the jaguar (*P. onca*), is the only member of the genus currently present in the Neotropical region, and the focus of our study. The jaguar has a color pattern similar to that of the leopard, but a much more robust constitution, with massive jaws and shorter limbs. The present study aims to characterize for the first time the jaguar genome, and to perform comparative analyses with the genomes from all other *Panthera* species. In addition, we seek to perform population genomic analyses with Brazilian jaguar populations and search for signatures of divergent selection in different regions. We have sequenced four genomic libraries, with an estimated coverage depth of 84x. The complete genome sequence allowed the annotation of 25,441 genes and the description of other genomic features (e.g. ncRNA, microsatellites, numts). Additionally, we have sequenced the genome of a leopard at low coverage, with an estimated depth of 25x. With the addition of these two genomes, we were able obtain a genomic data set containing all five *Panthera* species, which was used to perform phylogenetic discordance analyses and to detect signatures of selection using a dataset encompassing 13,143 orthologous genes. We were able to demonstrate the presence of hybridization events during the speciation process of the species, as well as signatures of selection in genes potentially involved in important characteristics of these iconic animals. Among them, the jaguar's robust build, the social behavior of lions, cold environment adaptations in the snow leopard and the tiger's stripes. Using an exome capture approach, we performed a population genomics study targeting jaguar populations from different Brazilian biomes. In addition to assessments of genetic diversity and population structure, we detected signals of local adaptation using multiple methods. Among the obtained results is the presence of genes under selection that are related to energetic metabolism in the Amazon, body development in the Pantanal and immunity in the Atlantic Forest. Additionally, we observed several pigmentation-related genes under selection in different biomes. Those genes

affect not only pigmentation, but also have pleiotropic effects in development and immunity routes. Overall, these results help to understand the evolutionary processes that have shaped the adaptation of *Panthera* species, and particularly the jaguar, to the environments where they currently live.

Keywords: comparative genomics; population genomics; natural selection; genealogical discordance, local adaptation; exome.

APRESENTAÇÃO

A presente tese de doutoramento está estruturada na forma de artigos científicos (Capítulos II e III), acompanhados de introdução e discussão gerais (Capítulos I e IV, respectivamente). A Introdução Geral e a Discussão Geral foram redigidas em português, seguindo as normas do periódico *Molecular Ecology* para referências bibliográficas. Os artigos estão em fase final de redação e serão submetidos aos periódicos científicos *Nature* e *Molecular Ecology*, após a incorporação das recomendações recebidas pela banca examinadora.

CAPÍTULO I | **INTRODUÇÃO GERAL**

1. Sequenciamento de Alto Desempenho

1.1 Histórico

Nos últimos 10 anos, com o advento do sequenciamento de alto desempenho, a genética evolutiva passou por mudanças drásticas em termos de volume de dados e métodos de análise. A possibilidade de se sequenciar genomas "completos", a um custo muito inferior ao de métodos tradicionais, permitiu um avanço rápido no nosso entendimento da evolução e funcionamento do genoma, especialmente de espécies-modelo (Shendure & Ji 2008; Quail *et al.* 2012; Ekblom & Wolf 2014; Ellegren 2014). A caracterização de genomas completos se iniciou com espécies-modelo de grande visibilidade, como nós, humanos, ainda através de sequenciamento tradicional, além do chimpanzé, camundongo, cão doméstico e vários outros ao longo dos anos (Ellegren 2014). Essa corrida para descrever e publicar genomas de animais modelo gerou o arcabouço de técnicas de geração de dados, genomas-referência e métodos de análise que puderam ser transferidos para uso em outros organismos (Ellegren 2014). Agora, cerca de 10 anos depois dessa revolução, biólogos evolutivos tem ao seu alcance o uso dessas tecnologias também para o uso em grupos não-modelo.

As primeiras espécies de mamíferos não-modelo a serem sequenciadas foram aquelas de interesse econômico, como gado, ou de alto apelo popular e ameaçadas de extinção, como o panda-gigante (Amadio 2009; Li *et al.* 2010). Tais trabalhos inseriram novas análises, como análise de sintenia, filogenia utilizando os genomas já publicados e estimativas populacionais mais refinadas (Wade *et al.* 2009; Locke *et al.* 2011; Amemiya *et al.* 2013). Foi também nesse período que se começou a incluir mais de um indivíduo, como no caso do orangotango, o que permitiu estimativas demográficas mais robustas e informativas (Locke *et al.* 2011).

Trabalhos mais recentes, envolvendo o esforço de diversos grupos de pesquisa, começaram a explorar técnicas de genômica comparativa para grandes grupos taxonômicos. Um exemplo foi o *Avian Phylogenomics Project*, que em 2014 publicou a primeira análise filogenômica para este grupo, contando com 48 espécies de aves (Zhang *et al.* 2014). Dentre os principais resultados estavam análises de arquitetura genômica que demonstraram como ocorreu a redução do tamanho do genoma do grupo. O genoma das aves é cerca de duas vezes menor do que o genoma de mamíferos. Essa redução de tamanho ocorreu graças à perda de elementos de transposição. Além disso, os autores encontraram diversas rotas sob efeito de seleção, incluindo vocalização, adaptações ao voo e genes relacionados à visão.

As características do sequenciamento genômico de alto desempenho possibilitaram avanços em áreas que anteriormente eram consideradas enormes desafios. A construção de bibliotecas genômicas envolve a quebra do DNA em fragmentos pequenos, p.ex. no caso de bibliotecas *paired-end*, cujos fragmentos tem em torno de 300-400pb. Visto com uma limitação inicialmente, essa abordagem foi um dos fatores que possibilitou o sequenciamento genômico de DNA antigo, como no caso de vários indivíduos de *Homo neanderthalensis*, um hominídeo da caverna Denisova na Rússia, e outros mamíferos como mamute (Haile *et al.* 2008; Green *et al.* 2010; Meyer 2012). Além de espécies já extintas, outra aplicação de técnicas de sequenciamento de nova geração, é o uso de espécimes de museu. Devido ao modo de preservação desse tipo de material, o DNA já se encontra altamente degradado.

Além de material de museu, estas técnicas podem ser aplicadas em amostras de material não invasivo coletadas em campo, como fezes e pelos. Mas para que isso ocorra é necessário que haja os recursos genéticos disponíveis para comparação dos dados. Esse tipo de abordagem é um dos pontos-chave da chamada genômica da conservação, área que apenas agora começa a se desenvolver (Mcmahon *et al.* 2014). Um exemplo de estratégias dessa natureza está na descrição do genoma do cavalo de Przewalski. Essa é uma subespécie de cavalo selvagem tida como a última população natural da espécie. A espécie foi considerada extinta na natureza na década de 1960, e foi mantida em apenas em cativeiro desde então. Utilizando como base o genoma do cavalo doméstico e sequenciando o genoma de indivíduos remanescentes, foi possível descrever o impacto do cativeiro nas populações atuais da subespécie, embasando projetos futuros de reintrodução (Der Sarkissian *et al.* 2015).

1.2 Sequenciamento genômico

Métodos de sequenciamento de alto desempenho, também conhecidos como de nova geração, mudaram a escala de geração de dados nas mais diversas áreas de pesquisa relacionadas com biologia molecular. O método mais utilizado atualmente é o de sequenciamento por síntese. Essa técnica envolve a leitura de milhares ou milhões de agrupamentos de DNA em uma placa, onde vão sendo agregadas novas bases e registradas por meio da geração de imagens fluorescentes (Metzker 2010). Cada adição de uma base pode ter uma intensidade específica que é traduzida com um valor de qualidade. Diferentemente do sequenciamento clássico, onde pode ser gerado no máximo, cerca de 200.000 pares de base por corrida (considerando um sequenciador com 96 capilares), o sequenciamento de alto desempenho permite a geração de mais de 1.8 terabases em uma única corrida (Van Dijk *et al.* 2014).

A repetição do sequenciamento múltiplas vezes para um mesmo sitio é necessária pois diminui drasticamente a taxa de erro associada ao sequenciamento. É recomendada pelo menos uma cobertura de 20x por sitio para identificação de variáveis comuns. Com uma cobertura inferior é recomendado o uso de estimativas de verossimilhança baseadas na qualidade de identificação e mapeamento do sítio (Nielsen *et al.* 2012).

1.3 Métodos de amostragem genômica

Apesar da drástica redução de preços no sequenciamento de alto desempenho, ainda existem barreiras grandes para o sequenciamento de genomas completos para qualquer tipo de estudo, além de tal magnitude de esforço não ser necessária em muitos casos. Um dos principais problemas que surgem ao se trabalhar com genomas completos é o volume de dados e a complexidade que essa quantidade de informação traz (Ekblom & Wolf 2014). Por esse motivo, foram criadas estratégias de preparação de bibliotecas e sequenciamento para reduzir a complexidade através do enfoque em regiões-alvo do genoma. Essas estratégias possibilitam a inclusão de um maior número de indivíduos a um custo reduzido e uma quantidade suficiente de informação para responder a diferentes tipos de pergunta (Jones & Good 2015).

Os dois métodos mais comuns são aqueles baseados em enzimas de restrição ou captura. A diferença entre as técnicas ocorre na forma de seleção das regiões de interesse e corte do genoma. Métodos que utilizam enzimas de restrição cortam regiões que apresentam o sitio da enzima ao longo de todo o genoma, enquanto no método de captura apenas regiões em que sondas previamente desenhadas se ligam são capturadas. Nesse caso, a quebra do DNA é feita de forma mecânica.

As duas técnicas possuem vantagens e desvantagens que devem ser levadas em conta. Métodos que utilizam enzimas de restrição costumam ser mais baratos do que métodos baseado em captura. Isso ocorre pois não existe a necessidade de haver um genoma de referência da espécie ou de uma espécie próxima, nem mesmo a geração de dados preliminares, como é o caso da captura. A desvantagem se dá justamente pela falta dessa informação. Outro ponto negativo é a menor reprodutibilidade do método, pois frequentemente é difícil amostrar as mesmas regiões em diferentes experimentos.

Métodos baseados em captura dependem enormemente dos recursos genômicos disponíveis para a espécie. Essa limitação foi reduzida por técnicas mais recentes, que permitem a utilização de outros tipos de dados como o transcriptoma de um ou mais tecidos (Bi *et al.* 2012). Apesar da vantagem de se trabalhar com regiões-alvos, o que facilitaria responder certas perguntas, isso pode se tornar uma limitação, especialmente no caso de estudos demográficos

em que uma maior quantidade de marcadores não-ligados é benéfica. Ainda assim, o uso de abordagens baseadas em captura de regiões-alvo do genoma é altamente promissor, especialmente em casos em que as regiões possam ser selecionadas de forma eficiente. Dentre estas abordagens, o sequenciamento de exomas (ver abaixo) emerge como uma das técnicas mais poderosas, especialmente quando o foco do estudo é a detecção de regiões codificadoras que estejam sofrendo a influência de seleção natural.

1.4 Exoma

Exoma por definição é o conjunto de éxons de uma determinada espécie. O sequenciamento de exoma surgiu em humanos como um método mais prático de detecção de doenças em uma determinada população (Lohmueller *et al.* 2013). Trabalhos que utilizam o sequenciamento de exoma para fins não médicos ainda são raros. Yi *et al.* (2010) analisaram uma população no Tibete com base no exoma de 50 indivíduos, e conseguiram identificar a ocorrência de seleção natural em um gene relacionado com hipóxia, o *EPAS1*. Li e colaboradores (2010), em uma abordagem semelhante, analisaram o exoma de 200 humanos e identificaram aproximadamente 120.000 *SNPs* (*single nucleotide polymorphisms*), sendo cerca de 53.000 deles codificantes. Com uma cobertura de aproximadamente 12x e uma amostragem grande, eles conseguiram observar alelos raros, com frequências inferiores a 0,5% na população. Devido às características do exoma, mesmo um número amostral pequeno possui a capacidade de diferenciar eventos de seleção de regiões neutras (Tennessen *et al.* 2011).

Recentemente, o uso do sequenciamento de exoma começou a ser explorado com uma alternativa também para espécies não-modelo. Inicialmente, tal técnica era apenas utilizada para espécies com genoma de referência ou espécies próximas (Cosart *et al.* 2011; Bi *et al.* 2012). Apesar de útil, a necessidade de um genoma referência era uma limitação grande para a maioria das espécies e grupos de pesquisa. Por este motivo, algumas iniciativas começaram a surgir, visando a diminuir a necessidade da existência de recursos genômicos pré-existentes (Bi *et al.* 2012). Trabalhos que aplicam esse tipo de técnica começaram há cerca de dois anos. O primeiro deles, realizado como prova de conceito para a técnica desenvolvida pelo seu grupo de pesquisa, utilizou um gênero de esquilos (Bi *et al.* 2013) Neste trabalho, eles comparam a eficiência da captura de exoma com sondas desenhadas a partir do transcriptoma em amostras de tecido frescas e material até cem anos depositado em museu, sugerindo que tais amostras apresentam potencial para este tipo de investigação.

2. Estudo de seleção natural com base em dados genômicos

O estudo da seleção natural é uma importante ferramenta para detecção de regiões com relevância funcional no genoma. Mais especificamente, o foco tende a recair sobre a seleção positiva, que está associada com adaptação e evolução de novas formas e funções. Tais regiões podem ser alvo de desenvolvimento de fármacos, ou no caso de populações naturais, ser uma informação importante para elaboração de estratégias de conservação. Métodos de análise de seleção natural dependem do tipo de marcadores utilizados e do desenho amostral escolhido para o estudo (Nielsen *et al.* 2005). Tais métodos podem ser divididos em duas categorias: técnicas que utilizam os padrões de divergência de sequências entre espécies e métodos que analisam a variação genética em populações (Oleksyk *et al.* 2010).

Nesse sentido, uma das mais conhecidas é o cálculo da razão entre mutações sinônimas e não sinônimas entre sequencias ortólogas de um conjunto de organismos, usualmente de espécies distintas. Mutações que alteram a função de um gene são muito raras. A única chance de uma mutação que altere as sequencias de aminoácidos em uma proteína de se manter na população em longo prazo é caso ela traga algum benefício (Kryazhimskiy & Plotkin 2008). O método assume que mutações sinônimas (medidas pela taxa normalizada 'dS') e mutações não sinônimas (medidas pela taxa normalizada 'dN') ocorrem na mesma frequência e, sob neutralidade, dN/dS = 1. Na presença de seleção negativa, dN/dS < 1 e, na presença de seleção positiva, dN/dS > 1. Devido à dificuldade de uma mutação que altere a função de uma proteína se fixar, a quantidade de mudanças necessárias para que dN/dS > 1 é enorme, o que torna o método extremamente conservador. Uma solução encontrada foi a criação de métodos que levam em conta a distribuição dos valores de dN/dS (Nielsen & Yang, 1998; Yang & Nielsen, 2001). Essa técnica pode ser aplicada tanto para analises sitio-específicas como para analises linhagem-especificas (Yang 2001). Outros métodos comparam a variação intraespecífica com a variação interespecífica para um determinado lócus. Partindo do mesmo princípio do método anterior, caso uma mutação em uma região funcional esteja fixada em uma espécie, este seria o indicativo da presença de seleção (McDonald & Kreitman, 1991). A principal vantagem deste método é sua robustez quando se leva em conta fatores demográficos (Nielsen, 2001).

Outros métodos são mais indicados para estudos populacionais, em que o objetivo é encontrar adaptação local. Tais métodos são baseados em desvios da neutralidade esperada de estatísticas sumárias, como a diversidade nucleotídica (π), D de Tajima e F_{ST} (Tajima 1989; Fu & Li 1993; Akey *et al.* 2002; Beaumont & Balding 2004). Métodos de busca genômica assumem que a seleção atua em regiões gênicas ou regulatórias específicas, enquanto eventos demográficos tem influência ao longo de todo o genoma. Testes que buscam regiões *outliers* estimam inicialmente

uma distribuição nula, de forma que os genes sob seleção (positiva, negativa ou balanceadora) estariam nos limites desta distribuição (Excoffier, 2009). Em regiões sob efeito de seleção positiva, a diversidade nucleotídica seria inferior ao esperado, assim como valores de D de Tajima. Já os valores de diferenciação entre populações ou categorias fenotípicas seriam maiores que a média estimada para o genoma como um todo (Nielsen 2005).

Com a possibilidade de obtenção de genótipos a partir de genomas ou exomas completos, métodos mais exploratórios, como análises do espectro de frequência dos sítios (*site frequency spectrum* [SFS]), também tem sido utilizado para detecção de seleção em estudos genômicos. O SFS é a contagem de quantos alelos derivados em uma amostra de tamanho *n* aparece em 1/n, 2/n, ... (n-1)/n indivíduos (Nielsen, 2005). A partir desse dado, é possível recuperar informações como o número de sítios segregantes e diversidade nucleotídica, bem como sítios sofrendo seleção quando comparando o SFS de duas ou mais populações (chamado de 2DSFS ou 3DSFS) (Schraiber & Akey 2015). Esse teste se baseia no fato de que é esperado que alelos derivados estejam a baixa frequência. Caso seja observado um aumento de frequência desses alelos, este seria um indicativo de seleção positiva. Também é possível observar outros processos, como a presença de um *selective sweep*, com uma redução drástica de diversidade ao redor do sitio sob seleção ou a presença de seleção balanceadora, caso a presença de alelos ancestrais seja maior do que o esperado. Um exemplo foi a descoberta de sinal de seleção do gene *EPAS1* em indivíduos do Tibete, relacionado à adaptação a altitudes elevadas (Yi *et al.* 2010). Assim como em outros testes, efeitos demográficos podem mimetizar o efeito de seleção. Por exemplo, seleção positiva pode ser confundida com o efeito de expansão populacional, e *bottleneck* genético pode ser confundido com a presença de *selective sweep* (Nielsen *et al.* 2005).

3. O gênero *Panthera*

O gênero *Panthera* é um dos mais icônicos entre os mamíferos. Composto por cinco espécies atuais, o gênero faz parte da família Felidae, e encontra-se na linhagem Pantherinae, junto com o leopardo nebuloso (*Neofelis* spp.) (Johnson *et al.* 2006). A linhagem se originou há cerca de seis milhões de anos, e tem como principal característica seu tamanho corporal avantajado em relação as outras espécies da família. Além disso, a capacidade de rugir é uma característica exclusiva do grupo, apesar de o leopardo das neves (*Panthera uncia*) não apresentar tal característica. Acreditava-se que a capacidade de rugir se dava pela ossificação incompleta do hióide, que todos os *Panthera* possuem. Mais recentemente, foi observado que modificações na laringe são as responsáveis (Hast 1989; Weissengruber *et al.* 2002). Outra característica importante é o papel ecológico que estas espécies desempenham. Todas as espécies do grupo são

predadores de topo de cadeia, e regulam o número de presas nos locais onde se encontram. Devido ao papel ecológico que as espécies desempenham, requerem uma extensa área de vida, com alto grau de preservação. Com a atual degradação dos habitats, causada por ação antrópica, bem como conflitos diretos com os humanos, todas as espécies do gênero encontram-se ameaçadas de extinção (Sunquist & Sunquist 2002).

Devido a várias características marcantes do grupo, especialmente a diferença de tamanho em relação aos outros gatos, a monofilia do grupo não é alvo de extenso debate. Por outro lado, apesar de ser um grupo pequeno, com apenas cinco espécies, as relações internas do grupo são alvo de grande discussão, especialmente devido ao grande número de plesiomorfias presentes do grupo (Christiansen 2008). Trabalhos com base em morfometria tem grande dificuldade em estabelecer as relações do grupo, e trabalhos com dados moleculares também apresentam muitas vezes filogenias discrepantes, possivelmente devido ao menos em parte à curadoria errônea dos dados (Christiansen 2008; Davis *et al.* 2010). Existem dois pontos controversos principais: a relação do leopardo das neves com os demais integrantes do gênero, e as relações internas em um sub-clado denominado subgênero *Panthera*, composto pela onça-pintada (*P. onca*), leopardo (*P. pardus*) e o leão (*P. leo*) (Johnson *et al.* 2006; Davis *et al.* 2010; Li *et al.* 2016). Uma hipótese bastante plausível, e formalmente proposta recentemente, é que, por se tratar de um grupo recente, tenham ocorrido eventos de hibridação ao longo da distribuição do gênero, especialmente durante o Pleistoceno e início do Holoceno. Esses processos, somados à recente e rápida radiação do gênero (menos de 4 milhões de anos), ajudariam a explicar a dificuldade na resolução da sua filogenia interpretação dos resultados provenientes de diferentes conjuntos de dados (Li *et al.* 2016).

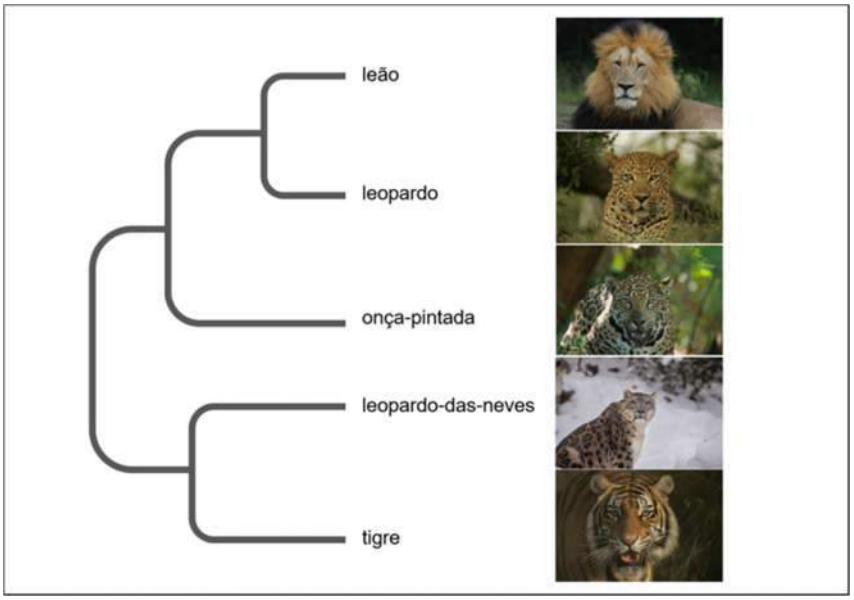


Figura 1 – Relação filogenética entre as espécies do gênero *Panthera*. Com base no trabalho de Davis et al 2010. Fotos: Christian Sperka, Daniel Kantek

Apesar de ser um fenômeno presente em outros grupos, a presença de inserção do DNA mitocondrial no genoma nuclear é um fenômeno bem conhecido e descrito na família Felidae e especialmente no gênero *Panthera* (Lopez *et al.* 1994; Kim *et al.* 2006; Heidtmann 2014). Em *Panthera*, existe uma inserção de 12,5 kb no cromossomo F2 (Kim *et al.* 2006). Essa inserção foi um dos principais problemas para estudos filogenéticos que utilizaram marcadores mitocondriais, devido à dificuldade de se diferenciar sequencias mitocondriais de sequencias nucleares (Davis *et al.*; 2010). Com técnicas de sequenciamento de alto desempenho esse problema praticamente desapareceu. A montagem de genomas mitocondriais pode ser realizada como subproduto de abordagens genômicas, como o sequenciamento de transcriptoma (p. ex., Heidtmann 2014).

A seguir, serão apresentadas as características principais de cada um dos membros atuais do gênero *Panthera*.

3.1 Tigre (*Panthera tigris*)

O tigre (Figura 2) é a maior espécie da família Felidae, com machos adultos podendo medir até 3 metros de comprimento (Mazak 1981). Com coloração marcante, é a única espécie do da família que possui listras no corpo, ao contrário das demais que possuem pintas ou ausência de machas, no caso do leão (*Panthera leo*) (Sunquist & Sunquist 2002). Possuem coloração de fundo laranja em diferentes tons, ventre branco ou amarelado e listras pretas ao longo do dorso, que podem ser utilizadas para identificação individual (Hiby *et al.* 2009). O mecanismo de

formação de listras ainda é desconhecido, mas assim como nas outras espécies, acredita-se que sirva como camuflagem. Pode ocorrer leucismo na espécie, mas esta variação não é encontrada na natureza e todos os indivíduos brancos podem ser rastreados até um único indivíduo nascido em cativeiro (Sunquist & Sunquist 2002).

Figura 2: *Panthera tigris* – A espécie é a única do gênero que possui listras pelo corpo. Foto: Christian Sperka

Sua distribuição histórica envolvia grande parte da Ásia, da Turquia até a Sibéria, China e chegando na península da Malásia (Nowell & Jackson 1996). Atualmente a espécie encontra-se gravemente ameaçada e tem sua distribuição reduzida a apenas pontos com pouquíssimos indivíduos. A espécie possui oito subespécies descritas, sendo que três já se encontram extintas (Mazak 1981). Atualmente, sua classificação quanto ao grau de ameaça geral é ameaçada, mas das cinco remanescentes, três subespécies encontram-se criticamente ameaçadas (*P. tigris altaica*, *P. tigris amoyensis* e *P. tigris sumatrae*) (IUCN 2015).

Existem poucos estudos moleculares com a espécie. Luo e colaboradores (2004), buscaram verificar padrões de estruturação com o uso marcadores mitocondriais e microssatélites. Com amostras de indivíduos de cinco das oito subespécies descritas, eles observaram uma marcada estruturação, confirmando a classificação já existente com a adição de uma subespécie, *P. tigris jacksoni*, ocorrente na península Malaia. Em outro estudo filogeográfico, pesquisadores observaram as relações entre duas subespécies localizadas proximamente. Utilizando DNA antigo, foi observado que o ancestral do tigre de Amur (*P. tigris altaica*) e do tigre cáspio (*Panthera tigris virgata*), subespécie já extinta, colonizou do leste da China até a região da Sibéria há cerca de 10.000 anos (Driscoll *et al.* 2009). Além desses dois, trabalhos mais recentes buscaram avaliar características de determinadas populações com o uso de métodos não invasivos (p.ex. Bhagavatula & Singh 2006; Henry *et al.* 2009).

Em 2013, a espécie teve seu genoma completo publicado, juntamente com análises envolvendo genomas de baixa cobertura do leão e leopardo das neves (Cho *et al.* 2013). Os principais resultados do estudo se encontram na descrição detalhada e anotação do genoma, além de uma primeira análise sobre genes possivelmente sob seleção. Dentre os genes encontrados sob seleção estão alguns relacionados ao metabolismo de proteínas, receptores olfativos e força muscular, características facilmente associadas aos hábitos da espécie, em especial a hipercarnivoria.

3.2 Leopardo das neves (*Panthera uncia*)

Panthera uncia (Schreber, 1775), é o menor dos *Panthera*, com comprimento máximo de 125 cm, com características marcantes que facilmente o diferenciam de outras espécies do gênero (Figura 3). Apresenta rosetas como o leopardo e a onça-pintada, mas geralmente mais esparsas. A pelagem é em tons de cinza com o ventre branco. Durante o verão, a pelagem pode mudar para tons mais amarelados. Por se tratar de uma espécie que habita ambientes frios, possui pelos longos, inclusive nas patas. Tem a cauda mais longa do grupo, e são grandes saltadores, características que ajudariam no deslocamento em locais rochosos e de difícil acesso (Sheldon *et al.* 2009). Além disso, a espécie possui adaptações craniofaciais que auxiliariam na manutenção de calor corporal e respiração em altitudes elevadas (Sunquist & Sunquist 2002).

O leopardo-das-neves tem sua distribuição restrita a regiões montanhosas do centro da Ásia, incluindo Mongólia, Rússia e países contendo a cordilheira do Himalaia. São reconhecidas duas subespécies, *P. u. uncia*, que habita Mongólia e Rússia, e *P. u. uncioides*, que habita os Himalaias e oeste da China (Nowell & Jackson 1996). Recentemente, foi reconhecida uma espécie fóssil muito semelhante a essa espécie, que até o momento é a espécie fóssil mais antiga descrito para o gênero, com cerca de 9 milhões de anos (Tseng *et al.* 2014). Seu status de conservação é ameaçado, segundo a IUCN (IUCN 2015).

Figura 3: *Panthera uncia* – Apesar de outras espécies do gênero habitarem ambiente frios, o Leopardo-das-neves possuis características que o tornam extremamente bem adaptado a esse tipo de clima. O focinho curto e testa em forma de domo ajudariam na respiração em ambientes com temperaturas baixas (Sheldon *et al.* 2009). Foto: Christian Sperka

Além de um trabalho descrevendo o genoma mitocondrial da espécie, existe apenas um trabalho que investiga aspectos genéticos da espécie, através do uso de amostras não invasivas para monitorar populações de leopardo das neves (Janecka *et al.* 2008; Wei *et al.* 2009). A dificuldade em coletar amostras é provavelmente o fator mais importante para isso e abordagens como museômica podem auxiliar nessa questão. Recentemente, a espécie teve seu genoma a baixa cobertura sequenciado, como parte da iniciativa de sequenciamento do genoma do tigre. Dentre os principais resultados, os autores encontraram indícios de alterações funcionais em genes envolvidos na adaptação a altas altitudes (Cho *et al.* 2013).

3.3 Leopardo (*Panthera pardus*)

Atualmente, o leopardo é a espécie do gênero com a distribuição mais ampla. A espécie se encontra em grande parte do continente africano e sul do continente asiático. São descritas nove subespécies (Uphyrkina *et al.* 2001). O leopardo tem grandes semelhanças com a onça-pintada, e é comum a identificação errada em fotos. As duas espécies diferem pelo porte, o leopardo possui um corpo mais grácil e as rosetas não apresentam pintas internas como na onça. Essa semelhança morfológica pode ser explicada pela possível manutenção de características plesiomórficas do gênero, com o leopardo apresentando um alto nível de semelhança com o ancestral dos *Panthera* (Sakamoto & Ruta 2012).

Figura 4: *Panthera pardus* – Devido a sua ampla distribuição, o leopardo possui uma grande variação fenotípica e está dividido em nove subespécies (Uphyrkina *et al.* 2001). Foto: Christian Sperka.

Existem alguns dados moleculares sobre a espécie, especialmente estudos com o objetivo de avaliar a validade taxonômica das subespécies existentes (Miththapala *et al.* 1996; Uphyrkina *et al.* 2001) e sobre genética da conservação (Uphyrkina *et al.* 2002; Busby *et al.* 2009). Assim como a onça-pintada, o leopardo apresenta indivíduos melânicos com frequências variáveis ao longo de sua distribuição, sendo que pode chegar a quase 100% dos indivíduos em algumas regiões da Ásia (Kawanishi *et al.* 2010; Silva 2014). O mecanismo genético do melanismo no leopardo é diferente do observado na onça-pintada. Uma mutação que causaria a perda do quadro de leitura no gene *ASIP*, componente da rota melanogênica, é o responsável por causar melanismo nesta espécie. Indivíduos com essa mutação perderiam completamente a função do gene (Schneider *et al.* 2012).

3.4 Leão (*Panthera leo*)

Dentre as espécies do gênero, o leão é a espécie que historicamente possuía a maior área de distribuição. O registro fóssil e arqueológico indica que a espécie ocorria no Norte da Ásia, América do Norte e África, sendo inclusive, a única espécie a conviver com a onça-pintada no continente americano durante o fim do Pleistoceno (Stuart & Lister 2011). Atualmente a espécie é encontrada apenas na África e Índia e é considerada ameaçada pela IUCN (IUCN 2015).

Dentre os *Panthera*, o leão talvez seja a espécie que apresenta o maior número de traços únicos. Além do tamanho, similar ao tigre, vale ressaltar dois aspectos da espécie. A presença de uma juba nos machos, e o comportamento social, extremamente desenvolvido e único na família Felidae. A juba do leão é um dos exemplos mais bem conhecidos de seleção sexual. Trabalhos de comportamento com a espécie, indicam que machos com jubas maiores e mais escuras são

preferidos pelas fêmeas. Esse atributo apresentou forte relação com os níveis de testosterona dos machos (West & Packer 2002). Os grupos de leões são compostos por um macho "alfa", grupos de caça compostos na sua maioria por fêmeas e machos satélites que podem vir a disputar o controle do grupo (Stander 1992; Funston *et al.* 1998). Grupos demonstram alta correlação genética, e maior proximidade genética com outros grupos mais próximos, inclusive havendo a formação de "supergrupos" em determinadas situações (Spong *et al.* 2002).

Figura 5: *Panthera leo* – O leão é a única espécie de Felidae que possui comportamento social altamente desenvolvido. Foto: Christian Sperka

Leões possuem um extenso registro fóssil, com duas espécies descritas e aceitas atualmente para o grupo (Stuart & Lister 2011). Espécimes fosseis encontrados na Europa recebem o nome especifico *speleae*, enquanto indivíduos encontrados na América do Norte, recebem o nome *atrox*. Essa divisão foi corroborada em trabalhos moleculares que observaram uma forte estruturação do grupo com dois clados e posteriormente em três clados. (Burger *et al.* 2004; Barnett *et al.* 2009). Foi observado que a população existente na América do Norte possivelmente surgiu através do isolamento por distância de populações na Europa próximas ao estreito de Bering. Outro evento observado nos estudos, foi a perda de variabilidade em populações na Europa, aparentemente associadas com a redução do tamanho populacional do bisão, importante fonte de alimento para espécie (Barnett *et al.* 2009). A sua ampla distribuição também indica a possibilidade de hibridação com outras espécies do gênero, com fortes indícios para eventos de hibridação com o leopardo das neves (Li *et al.* 2016).

3.5 Onça-pintada (*Panthera onca*)

A onça-pintada é a maior espécie de felídeo das Américas e o único representante vivo de *Panthera* encontrado no Novo Mundo. Possui características morfológicas semelhantes à do leopardo (*Panthera pardus*) (Figura 6), embora seja maior e mais robusta com uma cauda menor do que o leopardo. A fêmeas são menores, aproximadamente 20%, mas tem estrutura corporal semelhante à dos machos (Seymour 1989) e possuem uma grande variação de tamanho ao longo da distribuição, variação essa atribuída a disponibilidade de alimento (Hoogesteijn & Mondolfi 1996), embora essa associação ainda não esteja totalmente clara. O padrão de pintas entre as duas espécies também difere, enquanto o leopardo possui pontos pretos sólidos em um fundo alaranjado, a onça-pintada possui as pintas formando um padrão de roseta com o fundo alaranjado entre elas e mais escuro no seu interior (Nelson *et al.* 1933; Seymour 1989).

Figura 6: *Panthera onca* – Única espécie do gênero no continente americano. Assim como o leopardo possui indivíduos melânicos em algumas áreas de sua distribuição. Imagem com individuo focal do Projeto Genoma da Onça-pintada. Foto: Rodrigo Teixeira

Outra característica interessante sobre o padrão de pelagem é a presença de melanismo em indivíduos da espécie, o melanismo ocorre em várias espécies de felídeos, em diferentes genes, embora as forças evolutivas por trás desta característica ainda não estejam claras. O melanismo em onça-pintada tem um modo de herança dominante, com todos os indivíduos melânicos apresentando pelo menos um alelo que contém uma deleção de 15 pares de bases no gene *MC1R*, seguido de duas substituições não sinônimas imediatamente adjacentes a esta deleção (Eizirik *et al.* 2003). Haag e colaboradores (2010) aplicaram técnicas de identificação molecular para diferenciar indivíduos melânicos através das fezes. Este método possibilitou

estimar a frequência de indivíduos melânicos em uma população sem a necessidade de observação direta.

Panthera onca parece ter divergido de um ancestral comum com as outras espécies do gênero há pelo menos 2 milhões de anos e chegou às Américas durante o Pleistoceno (Li *et al.* 2016). A distribuição da espécie na época era mais ampla que a atual ocupando praticamente toda América do Norte e América do Sul, sendo que a considerada distribuição histórica da espécie é apenas um relicto de uma distribuição muito maior. Até o século XIX sua distribuição se estendia do sudoeste dos Estados Unidos até a Patagônia na Argentina. Desde meados de 1900 tem sido extinta em várias regiões, estando hoje presente em cerca de 50% da distribuição original, do norte do México ao norte da Argentina (Província de Misiones) e sul do Brasil (Parque Estadual do Turvo, RS), sob forma de fragmentos populacionais de vários tamanhos (Nowell & Jackson 1996). A espécie já foi dividida em oito subespécies com base em estudos morfológicos (revisado por Seymour, 1989), mas uma revisão de caracteres do crânio não indicou grupos distintos ao analisar estas características (Larson 1997). Corroborando essa informação, dados moleculares (mtDNA e microssatélites) indicaram uma ausência de subdivisões profundas nesta espécie, sugerindo uma estruturação rasa, fruto de uma expansão recente e alta conectividade entre grandes regiões geográficas (Eizirik *et al.* 2001). Apesar disso, há relatos de pesquisadores que existem diferenças no tamanho dos indivíduos em diferentes populações. Especialmente entre populações encontradas em áreas com vegetação mais esparsa, sendo estas maiores (ex. Pantanal e Cerrado) e aquelas encontradas em área de mata fechada com indivíduos menores (ex. Amazônia e Mata Atlântica). Alguns pesquisadores acreditam que esta variação ocorra devido a diferenças na dieta destas populações, com um maior aporte de proteína em populações de área aberta, ou devido a pressões do ambiente (Hoogesteijn & Mondolfi 1996). Ambas as hipóteses ainda não foram testadas, mas com base em genomas já publicados acredita-se que fatores genéticos tem grande influência na determinação do tamanho de indivíduos em diferentes populações em maior ou menor escala (Lindblad-Toh *et al.* 2005).

O registro fóssil da espécie e seus ancestrais é limitado. *Panthera onca georgica* e *Panthera onca gombaszoegensis* (3,2 – 0,126 Ma) são duas subespécies encontradas na Eurásia com datas anteriores à vinda da onça-pintada para o continente americano. O registro de ambas espécies está associado com ambientes similares àqueles onde a onça tem sua distribuição atual. *P. onca gombaszoegensis* tem sua distribuição relacionada com ambientes de floresta úmida e regiões alagadas (Hemmer *et al.* 2010). Hemmer e colaboradores (2001), compararam a espécie com *P. tigris virgata*, que acreditasse ter se dispersado utilizando a margem de grandes rios da região até chegar a Sibéria. Extrapolando para a onça-pintada, a mesma deve ter cruzado o estreito de Bering chegando às Américas, esse evento de migração foi possível com o surgimento

de uma ponte de terra durante o último máximo glacial que conectou a Eurásia e as Américas. Apesar do status taxonômico das espécies fósseis ainda ser discutido, com base em estudos morfológicos e ecológicos, já é claro que estas espécies são proximamente relacionadas com *Panthera onca* como subespécies ou espécies irmãs. Estudos de viabilidade populacional indicam o tamanho restrito dos refúgios na península Itálica durante os eventos de glaciação como a principal causa de extinção para *Panthera onca gombaszoegensis* (O'Regan *et al.*, 2002).

O primeiro registro fóssil da espécie nas Américas data do Pleistoceno (1,8 MA – 11.000 anos) com *Panthera onca augusta*, a subespécie era 15% a 20% maior que a forma atual e encontrava-se distribuída na América do Norte (Hemmer *et al.* 2010). O cenário biogeográfico mais aceito é que *Panthera onca gombaszoegensis* tenha chegado ao continente americano e se dispersado ao longo de continente, com *P. onca augusta* e *P. onca* surgindo através de eventos de especiação alopátrica. Eizirik *et al.* (2001) através de análises de coalescência, estimaram que as linhagens atuais da espécie tenham surgido no norte da América do Sul entre 280.000 e 510.00 anos. Além disso, Christiansen & Harris (2009) analisaram a relação das espécies fósseis do grupo da onça-pintada com *Panthera atrox*, e sugerem que esta espécie também divergiu do ancestral comum do grupo e não daquele que originou o leão moderno. Essas divergências se devem a um grande número de caracteres conservados, o que dificulta o posicionamento de espécies encontradas no registro fóssil dentro da filogenia do grupo. Esta hipótese mostrou-se inválida em trabalhos mais recentes, que utilizaram tantos dados moleculares como morfológicos para estabelecer as relações de parentesco das espécies fósseis do grupo (Tseng *et al.* 2014).

A onça-pintada apresenta um papel importante na estrutura de comunidades, controlando a abundância relativa de espécies, sendo que a remoção deste carnívoro topo de um ecossistema pode provocar um desequilíbrio dentro de uma guilda ou ecossistema (Eisenberg & Redford 1999). A onça-pintada tem sua ocorrência associada fortemente à presença de água, ocorrendo desde habitats de floresta tropical densa até ambientes mais abertos como o cerrado e áreas sazonalmente alagadas como o Pantanal. Os indivíduos que habitam florestas são frequentemente mais escuros e consideravelmente menores em tamanho do que aqueles que habitam áreas mais abertas, diferença que pode ser devida a uma maior abundância de presas em ambientes mais abertos (Hoogesteijn & Mondolfi 1996; Nowell & Jackson 1996). Estudos genéticos sobre essas características podem auxiliar no teste destas hipóteses. As presas naturais mais utilizadas incluem principalmente animais de grande porte como capivaras, veados, antas e porcos-do-mato, e por vezes também répteis como tartarugas e jacarés (Emmons 1989; Weckel *et al.* 2006).

A predação sobre o gado doméstico pode ser considerável em algumas áreas (em geral associada à redução populacional ou extinção antropogênica de suas presas naturais, e à

modificação dos habitats originais), o que gera sérios conflitos com pecuaristas e intensa mortalidade de onças-pintadas (Sunquist & Sunquist 2002). Embora a espécie tenha sido fortemente caçada no passado como um troféu ou para suprir o mercado de peles, no momento a retaliação ou perseguição direta por pecuaristas, em conjunto com a acentuada perda e fragmentação de seus habitats remanescentes, são as principais ameaças à sua sobrevivência (Nowell & Jackson, 1996). A espécie é listada no apêndice I da CITES (Sunquist & Sunquist 2002) e é considerada "quase ameaçada" pela IUCN. No entanto, sabe-se que algumas populações remanescentes são muito pequenas (p.ex. <20 indivíduos) e isoladas, podendo ser consideradas criticamente ameaçadas (Paula *et al.* 2013).

Análises preliminares de viabilidade populacional indicam que alguns destes fragmentos, particularmente os localizados na Mata Atlântica Costeira, têm uma probabilidade de extinção muito alta para as próximas décadas ressaltando a urgência da implementação de estratégias efetivas para sua conservação (Eizirik *et al.* 2001). Trabalhos mais recentes mostram que populações de onça-pintada já apresentam um tamanho efetivo reduzido causado por deriva genética em áreas de Mata Atlântica muito fragmentada e o fluxo gênico entre essas populações encontra-se minimizado devido a heterogeneidade da matriz (Haag *et al.* 2010). Ao contrário do que é observado no Pantanal, onde existe uma alta conectividade e alta diversidade genética (Valdez *et al.* 2015).

Informações sobre os aspectos genéticos das demais populações da onça-pintada no Brasil são inexistentes. Levantamentos populacionais no Cerrado indicam uma população com cerca de 300 animais maduro, com a perda de habitat para produção agrícola sendo o principal fator para o declínio das populações do bioma (Moraes 2012). Populações da Caatinga encontram-se em pior situação, com uma estimativa de apenas 250 indivíduos maduro em todo bioma. Os principais fatores de ameaça são a perda de habitat, com grande parte das populações restritas a apenas parques e reservas. Além disso, é o bioma mais carente de dados. Há pouco mais de dez anos que se começaram estudos mais detalhados da espécie no bioma (Paula *et al.* 2013). Por outro lado, na Amazônia, estima-se que atualmente existam cerca de 10.000 indivíduos em idade reprodutiva. Pesquisadores consideram este bioma como peça fundamental nas estratégias de conservação da espécie a longo prazo (Paula *et al.* 2013). A áreas mais ameaçadas para a espécie estão no que é conhecido como "arco do desmatamento". Essa é uma região na borda do bioma que faz fronteira com as demais regiões do país, portanto de extrema importância para manutenção do fluxo gênico entre os diferentes biomas onde a espécie ocorre.

Características como o melanismo e diferença de tamanho ao longo da distribuição, além do grande nível de ameaça que a espécie sofre em certas áreas, são razões relevantes para a contínua realização de estudos mais aprofundados sobre a onça-pintada. Trabalhos genômicos

com o objetivo de identificar evidências de seleção natural ainda são considerados um desafio e somente agora estudos com espécies não modelo começaram a se tornar realidade. A identificação destes processos em populações naturais além de ser algo pioneiro pode fornecer dados importantes para a conservação da espécie e dos ambientes que ela ocupa. Essa preocupação, aliada com a possibilidade de utilizar novas técnicas de sequenciamento com um poder de geração de dados muito maior, e possíveis descobertas relevantes sobre a biologia da espécie justificam a execução deste tipo de estudo.

3.6 O Projeto Genoma da Onça-Pintada

O presente projeto está ocorrendo em paralelo com a iniciativa para sequenciamento completo do genoma da onça-pintada, liderado por nosso grupo de pesquisa e contando com a participação de várias instituições. O presente trabalho encerra a primeira fase do projeto com o sequenciamento em alta cobertura e anotação inicial do genoma da espécie.

OBJETIVOS

Objetivo Geral

Caracterizar o genoma completo da onça-pintada, *Panthera onca* (Linnaeus, 1758), utilizando metodologias de sequenciamento de alto desempenho, a fim de caracterizar padrões evolutivos e possíveis efeitos da seleção natural sobre variantes fenotípicas observadas nesta espécie.

Objetivos Específicos

- A Caracterizar o genoma de *Panthera onca* e investigar assinaturas de seleção através de análises comparativas com as demais espécies do gênero *Panthera*.

- ACaracterizar a variabilidade genômica e identificar assinaturas de adaptação local em populações naturais de onça-pintada através da análise dos exomas de indivíduos amostrados em diferentes biomas brasileiros.

CAPÍTULO II | **JAGUAR GENOME SHEDS LIGHT ON THE COMPLEX EVOLUTION OF BIG CATS**

Draft, Nature
February 22, 2016.

Jaguar genome sheds light on the complex evolution of the big cats

Henrique V. Figueiró[1], Gang Li[2], Fernanda J. Trindade[1], Juliana Assis[3], Fabiano Pais[3], Gabriel Fernandes[3], Sarah H. D. Santos[1], Graham Hughes[4], Aleksey Komissarov[5], Agostinho Antunes[6], Adauto L. V. Nunes[7], Rodrigo H. F. Teixeira[7], Ronaldo G. Morato[8], Emma Teeling[4], Toni Gabaldón[9], Stephen J. O'Brien[5], Rasmus Nielsen[10], Luiz Lehmann Coutinho[11], Guilherme Correa Oliveira[3], William J. Murphy[2], Eduardo Eizirik[1]

Affiliations:

[1] Faculdade de Biociências, PUCRS. Porto Alegre, RS 90619-900, Brazil
[2] Texas A&M University. College Station, TX, United States
[3] Centro de Pesquisa René Rachou, FIOCRUZ. Belo Horizonte, MG 90190-002, Brazil
[4] Laboratoy of Molecular Evolution and Mammalian Phylogenetics, University College Dublin. Dublin, Ireland
[5] Theodosius Dobzhansky Center for Genome Bioinformatics, Russia
[6] Departamento de Biologia, Faculdade de Ciências da Universidade do Porto. Porto, Portugal
[7] Zoológico Municipal de Sorocaba, Brazil
[8] Instituto Chico Mendes de Conservação da Biodiversidade. Brasília, DF 70670-350, Brazil
[9] Centre for Genomic Regulation, Barcelona. Spain
[10] Institute of Biology, University of California Berkeley. Berkeley, California, United States
[11] Escola Superior de Agricultura Luiz de Queiroz (ESALQ-USP), São Paulo. Brazil

Keywords: genealogical discordance; selection; *Panthera*

Abstract

The great cats of genus *Panthera* are amongst the most iconic animals in the world, capturing historical human fascination and present concern regarding their imperilled status in the wild. The five extant species derive from a recent and rapid radiation, which has posed difficulties to the resolution of their phylogeny, while offering opportunities to understand the adaptive divergence of closely related and widely successful top predators. Here we report the sequence, *de novo* assembly and annotation of the jaguar (*Panthera onca*) genome, and comparative analyses with all other *Panthera* species, including new genomic data from the leopard. We found strong evidence of genealogical discordance across the *Panthera* genomes, caused by both incomplete lineage sorting during speciation and historical interspecific hybridization. We also identified signatures of positive selection in the genomes of *Panthera* species using site-based selection models and genome-wide timetree scans, affecting genes involved in brain, skull and limb development, body size, sensory perception, pigmentation and behaviour. Several of the affected genes undergo imprinting in humans or mice, suggesting a role for this regulatory mechanism in *Panthera* phenotypic evolution. In addition, *Panthera* genomes exhibit strong X-linked selective sweeps that underlie multiple regions harbouring genes implicated in intellectual disability and hearing loss in humans. These findings open up new avenues to understand the genomic basis of phenotypic diversification in the great cats.

Main Text

Recent evolutionary radiations are powerful systems to investigate the interplay of complex processes, such as the origin of ecological divergence and reproductive isolation, and the timing of speciation events and secondary admixture. Genome-wide analyses provide novel opportunities to address these issues, revealing the differential effects of these processes on genes, gene families and other features, and enabling an in-depth assessment of the selective and demographic forces that shape present-day species. The *Panthera* genus is a remarkable radiation to investigate these complex issues, as it comprises five big cat species (Figure 1A) that arose from a recent and rapid diversification process[1,2]. Stemming from an Asian common ancestor ~4 million years ago (MYA), this radiation has successfully colonized a variety of habitats across all continents except Australia, occupying top-predator roles in many of the Earth's terrestrial ecosystems[3]. In this process, they evolved unique features, such as the tiger's stripes, the lion's social behaviour, and the jaguar's stocky build, with a massive head, particularly powerful jaws and relatively shorter forelimbs[2–4]. Understanding the history of these characters depends on resolving the underlying phylogeny of the *Panthera* species, a task that has been notoriously difficult to accomplish[1,5].

Recent analyses have indicated that genealogical discordance caused by incomplete lineage sorting (ILS) and also by post-speciation admixture has contributed to produce such a complex system[2]. Genome-wide analyses of the five extant species should help illuminate this issue, and allow for in-depth investigations of their adaptive divergence, but so far only one high-quality draft genome (tiger) and two low-contiguity genomes (lion and snow leopard) from this group were available[4].

To enable genomic comparisons across the living *Panthera* species, we have sequenced the complete genome of a male jaguar ("Vagalume", *i.e.* "firefly" in Portuguese) from the Brazilian Pantanal region. The genome was sequenced at ~84X coverage using four libraries (see Methods). We performed *de novo* assembly using ALLPATHS[6], which achieved an N50 contig and scaffold length of 28.6 kb and 1.46 Mb, respectively. We annotated the genome with *ab initio* gene prediction and validated protein-coding genes with transcriptome data from six different tissues of the same individual (see Methods). These approaches predicted 25,451 protein-coding genes, of which 24,411 (96%) were supported by our RNA-seq data or other empirical evidence (see Supplementary Material). In addition, we sequenced the genome of a leopard (*P. pardus*) to approximately 25X coverage, thus obtaining a comparative data set comprising all extant *Panthera* species, including previously published tiger, lion and snow leopard whole genome sequence (wgs) data. We used these genome-wide data to perform evolutionary comparisons among the living *Panthera* species, *e.g.* PSMC analyses (Figure 1B) that revealed a common trend for demographic decline in the last 200,000 to 300,000 years (with the leopard showing a more recent episode of population expansion). We also aligned the five genomes (along with that of the domestic cat to be used as an outgroup) and used the resulting alignments in multiple downstream analyses (see online Methods for details). These included a phylome analysis addressing the evolution of gene families in the *Panthera* and several related carnivorans, as well as an in-depth assessment of olfactory receptor (OR) turnover among the great cats (see Supplementary Material).

To investigate genealogical discordance across the genome, we performed maximum likelihood (ML) phylogenetic analyses in fixed-length segments using a sliding-window approach (100-kb window with 100-kb steps). For every species, we also extracted all protein-coding genes and performed separate ML analyses with these gene-based alignments (including exons and introns for each gene). With both window-based and gene-based analyses, the same topology was obtained as the most frequent reconstruction from the autosomes (herein referred to as the 'species tree'; see Figure 1A), which matches a previous Y-chromosome-based phylogeny[5]. However, we observed high levels of genealogical discordance (Figure 1C, Figure 2), beyond what was reported in classical cases such as the human-chimp-gorilla trichotomy[7]. With the gene-based

data set (13,183 different genes shared among the five species, using conservative orthology and coverage criteria), we retrieved all 105 possible tree topologies, with the most frequent (the species tree) being recovered only 23% of the time. Analyses of the 100-kb genomic windows supported the species tree in 64% of the windows (74% when a likelihood ratio test was employed; see Methods). The second and third most frequent topologies varied only in the position of jaguar with respect to lion and leopard, with much greater support for lion+jaguar than leopard+jaguar (Figure 2A).

An analysis of divergence times from these windows revealed consistent point estimates across autosomes when assessing windows that conformed to the species tree, yet much greater variation for those supporting topologies 2 and 3 (Figure 2B). Under a scenario of ILS, theory predicts that the average lion+jaguar and leopard+jaguar divergence times should predate species tree (lion+leopard) divergence times. While this prediction tended to be upheld for the autosomes, divergence estimates for the X chromosome were significantly younger for topology 2 but less so for topology 3 when compared to the species tree. We also observed an increase in X-linked support for topology 2 and 3 relative to the autosomes (Figure 2A). We hypothesize that the reduction in X-linked divergence time is driven by post-divergence admixture between the ancestors of lions and jaguars. Indeed, ABBA-BABA tests showed significant *D*-statistics supporting lion-jaguar admixture, as well as a complex network of ancestral hybridization among several *Panthera* lineages (Figure 2C). Interestingly the lion lineage stands apart as having the most widespread signatures of ancient admixture, which is understandable given the broad historical range of *P. leo* throughout much of the Holarctic region, overlapping with several congeneric lineages. These results also provide expanded confirmation of recent evidence for admixture between the snow leopard and the lion+leopard ancestor leading to mitonuclear discordance[8]. Non-species tree topologies were also enriched on the X chromosome in regions corresponding to large recombination deserts, including two large blocks that support distinctive non-species-tree topologies (12 & 23), in which the leopard lineage is deeply diverged from the remaining *Panthera* species (Figure 2D). This extreme divergence may be caused by either ancient admixture between the leopard and an extinct lineage closely related to *Panthera*, or by strong selection driving accelerated divergence of leopard genes within these low-recombination regions.

To investigate the evolution of adaptive diversity in *Panthera* cats, we employed two different approaches, one based on the 13,183 shared coding genes (Site models and Branch-site models analysed with PAML[9]) and another based on a timetree analysis[10] of the genomic windows described above. The former approach was performed with both the species tree and the locus-specific gene tree to account for genealogical discordance (see Supplementary Material), and

154 revealed > 1,000 genes with significant signatures of positive selection, 155 of which were
155 identified with both the Site and Branch-site models (Figure 3A). The genomic window approach
156 also identified numerous, localized outlier regions across the genome with distorted topologies
157 and significantly aberrant divergence times relative to the genome-wide average. For this
158 method, we focused specifically on the relationships within the lion-leopard-jaguar clade, as these
159 constituted the majority of discordant gene trees (Figure 2A). Following *Z*-transformation of
160 divergence times at each node, we identified 242 windows with dates younger than 2 standard
161 deviations from the mean (sdm) for both terminal and basal nodes (*i.e.* a compressed tree),
162 representing the top 0.85% of all windows. These windows reflect more star-like phylogenies that
163 are a signature of positive selection[10], and contained 161 genes, 17 of which were also detected
164 using the gene-based approach (Figure 3A).
165 We then performed functional enrichment analyses of the three gene sets with signatures
166 of positive selection (Site model, Branch-site model and Outlier window test), using multiple
167 databases (Figure 3B-C; Supplementary Material). A comparison of these results revealed some
168 terms that were consistently detected across the three gene sets (*e.g.* 'Olfactory transduction' and
169 'Protein digestion and absorption' in the Kegg database; and 'Olfactory Signaling Pathway' and
170 three related terms in the Pathway Commons database), while several others were enriched in
171 only one or two of the sets (see Supplementary Material). This initial comparison revealed a
172 striking pattern of positive selection consistently affecting olfactory-related genes in the *Panthera*,
173 highlighting the relevance of this sensory repertoire in the adaptive evolution of this group. This
174 observation was also supported by several instances of species-specific gene duplication events
175 inferred from our whole genome OR annotations, as well as cases that preceded one or more
176 speciation events (Supplementary Material).
177 Detailed analyses of the enrichment results revealed additional insights (Figures 1A, 3D
178 and 4). For example, the significantly overrepresented terms in the Outlier window approach
179 included hyperphagia (*i.e.* abnormally increased appetite for food consumption) and intellectual
180 disability (*e.g.* mental retardation, Prader-Willi and Angelman Syndromes), suggesting an
181 involvement in phenotypes related to body size, metabolism and cognition. Further restricting our
182 analysis to 74 windows with extremely compressed basal nodes (>3 sdm), we observed two
183 general patterns that are consistent with ancient selective sweeps: 1) the windows were often
184 clustered and supported non-species tree relationships, and 2) they showed a precipitous drop in
185 divergence time relative to flanking windows (Figure 4B). These extreme outliers also contained a
186 number of genes previously implicated in intellectual disability (*UBE3A, DACH2, BEX5, MID2, PLP*),
187 morphology and stature (*DOCK3, BMP4, SHROOM4, PJA1*), pigmentation (*EXOC2*), and deafness
188 (*POU3F4, COL4A5/6*).

A very similar set of phenotypes affected by positive selection emerged from detailed analyses of the coding gene data sets, including species-specific signatures of adaptive evolution revealed by the Branch-site tests (Figure 3D). Intriguing examples identified by this approach included body size and social behaviour in the lion, apnea in the snow leopard, melanogenesis in the tiger and craniofacial and limb development in the jaguar. There were also several examples of genes involved in brain development and sensory perception (olfaction, hearing and vision), underscoring the importance of these phenotypes in the history of adaptive divergence among these large predators. Interestingly, several of the positively selected genes identified with this approach have been shown to undergo imprinting in humans or mice, suggesting that this process may play an important role in the evolution of *Panthera* cats. Additional loci that undergo imprinting were also identified with the Outlier genomic window approach (*e.g. UBE3A*; see Figure 4B), corroborating this finding and highlighting the need to further investigate the biological role of genetic imprinting in the phenotypic evolution of non-model organisms.

In several cases, two or more genes associated with the same phenotype were positively selected in the same species (Figure 3D), prompting us to investigate their functional relationships in the context of known biological pathways. An interesting example emerged when we inspected the genes involved in craniofacial and limb development identified in the jaguar. Two of these genes (*ESRP1* and *SSTR4*) have been associated with craniofacial robustness[11,12], which suggests an involvement with the jaguar's distinctively massive head and powerful bite relative to the other living *Panthera* and to the inferred ancestral phenotype for this genus[13] (Figure 4A). These unique features have been hypothesized to represent jaguar adaptations to a diet largely concentrated on heavily armoured reptiles (caiman and freshwater turtles), which may have evolved as a response to the extinction of large mammalian prey at the end of the Pleistocene[14]. For both *ESRP1* and SSTR4, the particular residues bearing signatures of positive selection are likely to have important functional roles (see Figure 4A), laying out a clear hypothesis that can be tested experimentally in model systems.

An additional example of species-specific adaptive evolution involved melanogenesis in the tiger (Figures 3D, 4C). Five different genes belonging to this pathway had significant, tiger-specific signatures of positive selection. These included loci affecting the upstream portion of the pathway (*WNT10B*, *FZD1*), possibly involved in pattern formation, as well as *TYR*, which codes for the main melanogenic enzyme[15]. These findings suggest a potential role for these positively selected changes in the evolution of tiger stripes, relative to the spotted coat inferred for the *Panthera* ancestor[16]. More broadly, we observed a large number of positively selected genes (detected with all three approaches) that belong to the Glypican pathway (Figure 4D), a complex interaction of loci underlying a variety of developmental processes, from bone formation to many

224 complex neurological phenotypes[17]. Remarkably, many of the selected loci were directly
225 connected on the pathway, even though they often bore signatures of selection in different
226 *Panthera* species, or were detected with different approaches (Figure 4D). Dissecting the
227 biological processes mediated by these particular loci and their interactions is thus a promising
228 avenue for further investigating the adaptive evolution of *Panthera* species and likely other
229 mammalian systems.

230 Taken together, the results presented here illustrate the complex interplay among rapid
231 divergence, post-speciation admixture, and natural selection in the context of a highly successful
232 adaptive radiation. The growing genome resources available for the great cats promise to help
233 unravel the relationships among these processes to an unprecedented level, to understand the
234 genetic basis and adaptive relevance of the eco-morphological features that characterize these
235 species, and to provide tools to aid conservation efforts on behalf of these magnificent organisms.
236

References

1. Johnson, W. E. *et al.* The late Miocene radiation of modern Felidae: a genetic assessment. *Science (80-.).* **311,** 73–77 (2006).
2. Mazak. Panthera tigris. *Mammalian Species* **152,** 1–8 (1981).
3. Gonyea, W. J. Adaptative differences in the body proportions of large felids. *Acta Anat. (Basel).* **96,** 81–96 (1976).
4. Mosser, A. & Packer, C. Group territoriality and the benefits of sociality in the African lion, Panthera leo. *Anim. Behav.* **78,** 359–370 (2009).
5. Davis, B. W., Li, G. & Murphy, W. J. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, Panthera (Carnivora: Felidae). *Mol. Phylogenet. Evol.* **56,** 64–76 (2010).
6. Butler, J. *et al.* ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* **18,** 810–20 (2008).
7. Hobolth, A., Dutheil, J. Y., Hawks, J., Schierup, M. H. & Mailund, T. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res.* **21,** 349–56 (2011).
8. Li, G., Davis, B. W., Eizirik, E. & Murphy, W. J. Pervasive signals of ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* 1–11 (2016). doi:10.1101/gr.186668.114.4
9. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).
10. Hunter-Zinck, H. & Clark, A. G. Aberrant time to most recent common ancestor as a signature of natural selection. *Mol. Biol. Evol.* **32,** 2784–2797 (2015).
11. Baujat, G. *et al.* Clinical and molecular overlap in overgrowth syndromes. *Am. J. Med. Genet. - Semin. Med. Genet.* **137 C,** 4–11 (2005).
12. Bebee, T. W. *et al.* The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife* **4,** 1–27 (2015).
13. Sakamoto, M. & Ruta, M. Convergence and divergence in the evolution of cat skulls:

Temporal and spatial patterns of morphological diversity. *PLoS One* **7,** e39752 (2012).

14. Emmons, L. Jaguar predation on chelonians. *J. Herpetol.* **23,** 311–314 (1989).
15. Jackson, I. J. Molecular and developmental genetics of mouse coat color. *Annu. Rev. Genet.* **28,** 189–217 (1994).
16. Ortolani, A. Spots, stripes, tail tips and dark eyes: Predicting the function of carnivore colour patterns using the comparative method. *Biol. J. Linn. Soc.* **67,** 433–476 (1999).
17. Filmus, J., Capurro, M. & Rast, J. Glypicans. *Genome Biol.* **9,** 224 (2008).
18. Cho, Y. S. *et al.* The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat. Commun.* **4,** 2433 (2013).
19. Warzecha, C. C., Sato, T. K., Nabet, B., Hogenesch, J. B. & Carstens, R. P. ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing. *Mol. Cell* **33,** 591–601 (2009).

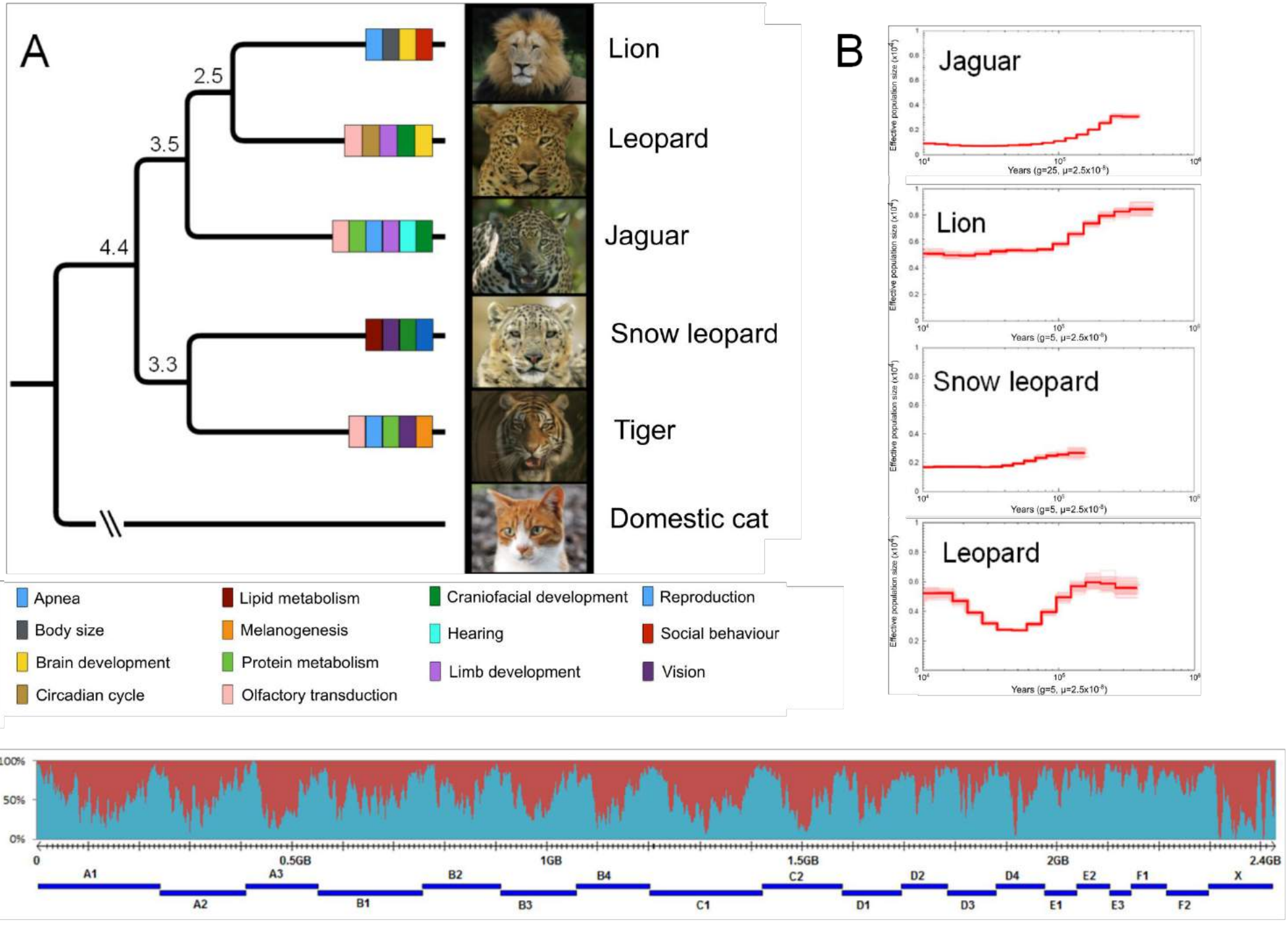
Figure 1. Evolutionary history of the great cats. (A) Species tree of genus *Panthera* estimated from genome-wide data. Numbers next to nodes indicate the age (in 'millions of years ago' [MYA]) of the divergence episodes among present-day lineages, averaged across all autosomal genomic windows that conform to the species tree. Colored rectangles on terminal branches indicate phenotypic categories (defined at the bottom; see Figure 3D and Supplementary Information for more details) affected by species-specific episodes of positive selection. (B) PSMC plots depicting the demographic history of four *Panthera* species inferred from genomic data (see [18] for the tiger PSMC). (C) Remarkable genealogical discordance across the genome of *Panthera* cats, demonstrated by a sliding-window analysis (500kb window size, 100 kb steps) of a full-genome alignment mapped to domestic cat chromosomes (blue lines at the bottom). The Y-axis indicates the percentages of overlapping windows within a given interval that conform to (blue) or reject (red) the species tree. (Photo credits: D. Kantek (jaguar); C. Sperka (others)).

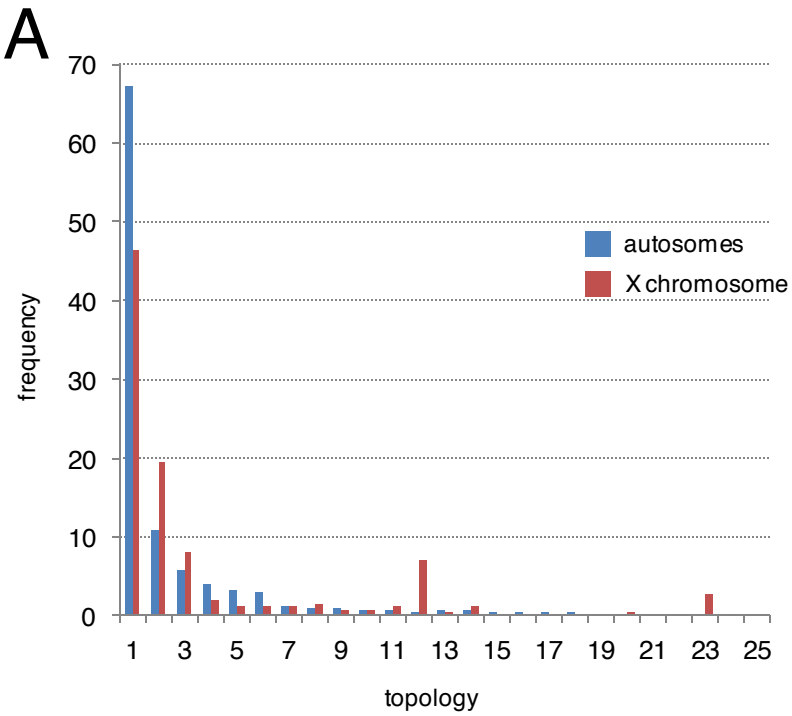
Figure 2. In-depth characterization of phylogenetic discordance in *Panthera*. (**A**) Frequencies of the species tree (topology 1) and alternative topologies on pooled autosomes (blue) and chromosome X (red). Topologies 12 and 23 show the leopard in a basal position relative to the other species (see Supplementary Material). (**B**) Chromosome-wide mean divergence time estimates (in MYA), calculated from 100-kb windows that yielded the three most frequent topologies (shown below the graphs; nodes color-coded). Note the marked drop in divergence time on the X chromosome for non-species-tree topologies. (**C**) Widespread and complex patterns of historical admixture between pairs of *Panthera* lineages, inferred from ABBA-BABA tests. Colored arrows indicate the phylogenetic position of significant *D*-statistics (shown in the table) for different species trios, assuming the domestic cat as an outgroup. (**D**) Patterns of phylogenetic discordance along the X chromosome correlate with the recombination landscape, estimated in 100-kb windows. Regions depicting species-tree relationships (top-1) are shown in white and correspond to higher recombination rates (domestic cat linkage map [Li et al. submitted]) and older divergence times. Recombination deserts are enriched for alternative topologies with reduced divergence times bearing phylogenetic signatures of selection. Two terminally located, lower-recombination regions are enriched for topologies 12 and 23, depicting a basal position for leopard. These regions also harbor tracts of windows with very old divergence times for the leopard from other *Panthera*.

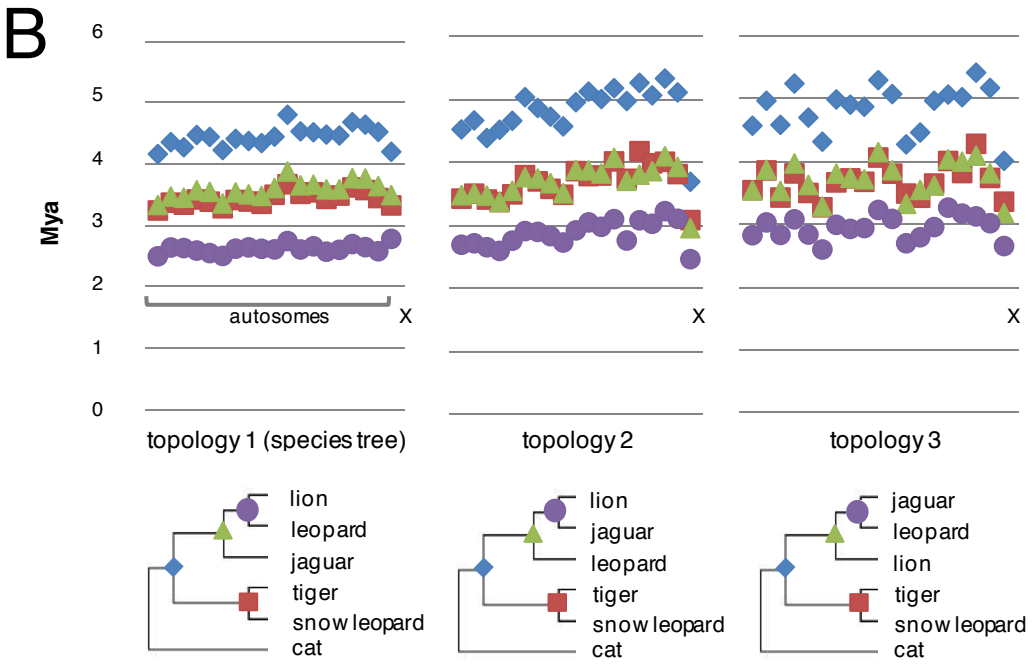
Figure 3. Genomic evidence of natural selection in the *Panthera* genus. (A-C) Venn diagrams depicting shared genes **(A)** and pathways **(B,C)** present it two databases (Kegg **[B]** and Pathway Commons **[C]**) among three approaches that detect signatures of natural selection: PAML Site model and Branch-site model (targeting coding genes) and Outlier window test (based on a genome-wide sliding-windows approach that detects aberrant divergence times induced by selection [Hunter-Zinck & Clark 2015]). **(D)** Detailed results of the Branch-site test, indicating genes bearing significant signatures of positive selection in species-specific lineages; phenotypic categories (color-coded as in Figure 1A) were curated manually based on enrichment results of the Kegg and Pathway Commons databases, as well as additional literature searches on each of the identified genes. Imprinted genes (based on human or mouse data) are shown in bold types.

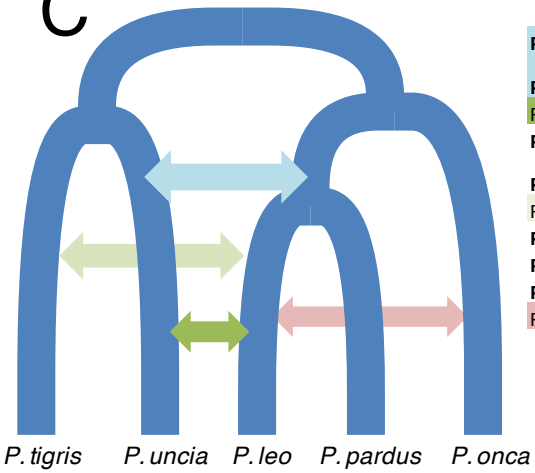
Figure 4. Inferred functional implications of selected genes. (A) Two genes with significant signatures of jaguar-specific positive selection (*ESRP1* and *SSTR4*) affect craniofacial robustness. Silhouettes on the left depict a jaguar's derived robust face relative to the inferred appearance of the *Panthera* ancestor[13]. The positively selected I298Y substitution in the *ESRP1* gene lies in the RRM1 domain, known to bind to the *FGFR2IIIb* gene isoform and affect craniofacial development[19]. The two positively selected substitutions (R39A and R42E) in the *SSTR4* gene imply substantial physicochemical amino acid changes to the affected segment (green: nonpolar; yellow: polar; blue: basic; red: acid). **(B)** Cartoon depicting variation in age estimates for the base of *Panthera* in 100-kb windows across the length of chromosome B3. Insets show age estimates (y-

axis, in MYA) for two nodes per window in two genomic stretches consistently showing significantly lower ages than the genome-wide average (purple: base of *Panthera*; blue/red: internal node [blue: windows conforming to [lion+leopard] topology; red: windows conforming to [lion+jaguar] topology]). These aberrant stretches are characterized by young trees depicting lion-jaguar monophyly, and contain two glypican-pathway loci (*UBE3A* and *BMP4*) that are involved in neuro-muscular and bone development. **(C)** Genes belonging to the KEGG melanogenesis pathway bearing significant signatures of positive selection in the tiger-specific branch-site test. Words in bold are pathway members (genes/gene families), while words underneath each member are the specific genes found to be under selection; numbers on connecting lines indicate additional pathway steps. **(D)** Schematic representation of the Glypican pathway, including only genes that were directly connected to each other, and that exhibited one or more significant signatures of positive selection in the *Panthera*: colored rectangles/cylinders were detected with branch-site tests (green: jaguar; red: lion; orange: tiger; blue: snow leopard; purple: *Panthera vs.* domestic cat); cylinders were detected with the PAML site model; grey rectangles were detected with the outlier genomic window test. Imprinted genes are shown in bold types.

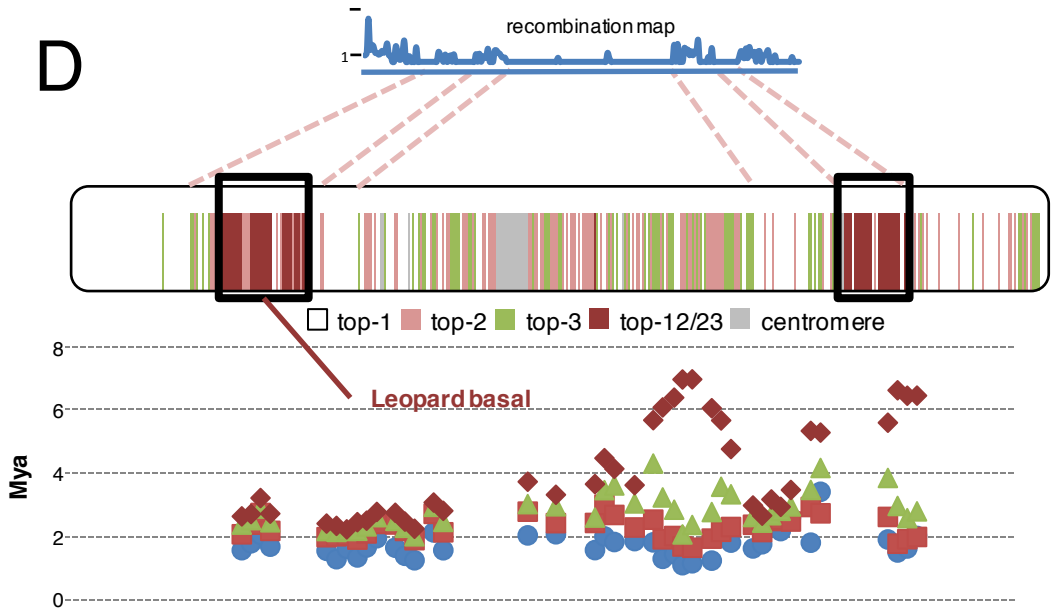

A
2.5
3.5
4.4
3.3
Lion
Leopard
Jaguar
Snow leopard
Tiger
Domestic cat
B
Jaguar
Lion
Snow leopard
Leopard
Effective population size (x10⁴)
Years (g=25, μ=2.5x10⁻⁸)
Years (g=5, μ=2.5x10⁻⁸)
Apnea
Body size
Brain development
Circadian cycle
Lipid metabolism
Melanogenesis
Protein metabolism
Olfactory transduction
Craniofacial development
Hearing
Limb development
Reproduction
Social behaviour
Vision
C
100%
50%
0%
0
0.5GB
1GB
1.5GB
2GB
2.4GB
A1
A2
A3
B1
B2
B3
B4
C1
C2
D1
D2
D3
D4
E1
E2
E3
F1
F2
X



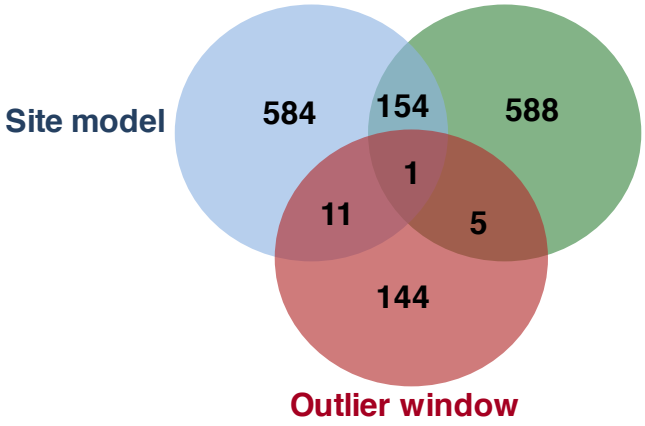




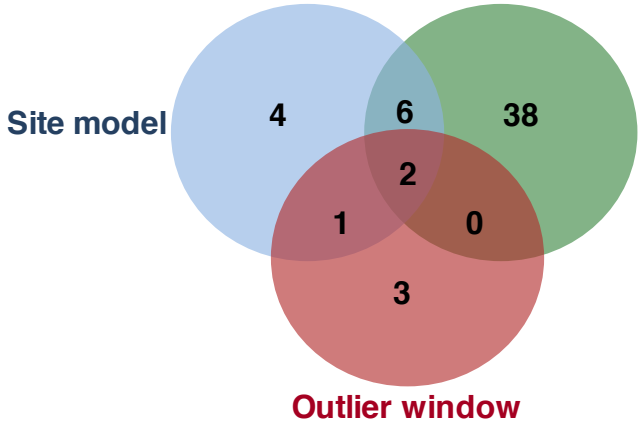
P1	P2	P3	*D*	s.e.	*Z*
PLE	PON	**PUN**	-0.2235	0.0043	-52.5
PPA	PON	**PUN**	-0.1804	0.0042	-43.3
PPA	**PLE**	**PUN**	0.0567	0.0034	16.7
PLE	PON	**PTI**	-0.1739	0.0055	-31.6
PPA	PON	**PTI**	-0.1379	0.0051	-27.2
PPA	**PLE**	**PTI**	0.0482	0.0037	13.1
PTI	PUN	**PON**	-0.1200	0.0040	-30.0
PTI	PUN	**PPA**	-0.0742	0.0028	-26.8
PTI	PUN	**PLE**	-0.0685	0.0026	-26.0
PPA	**PLE**	**PON**	0.0569	0.0048	11.8



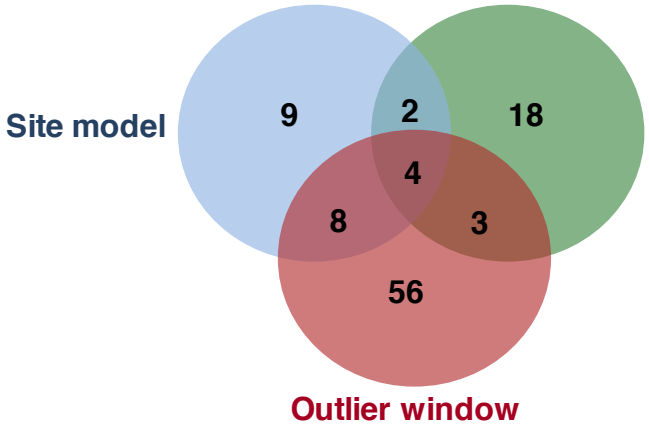
A



B

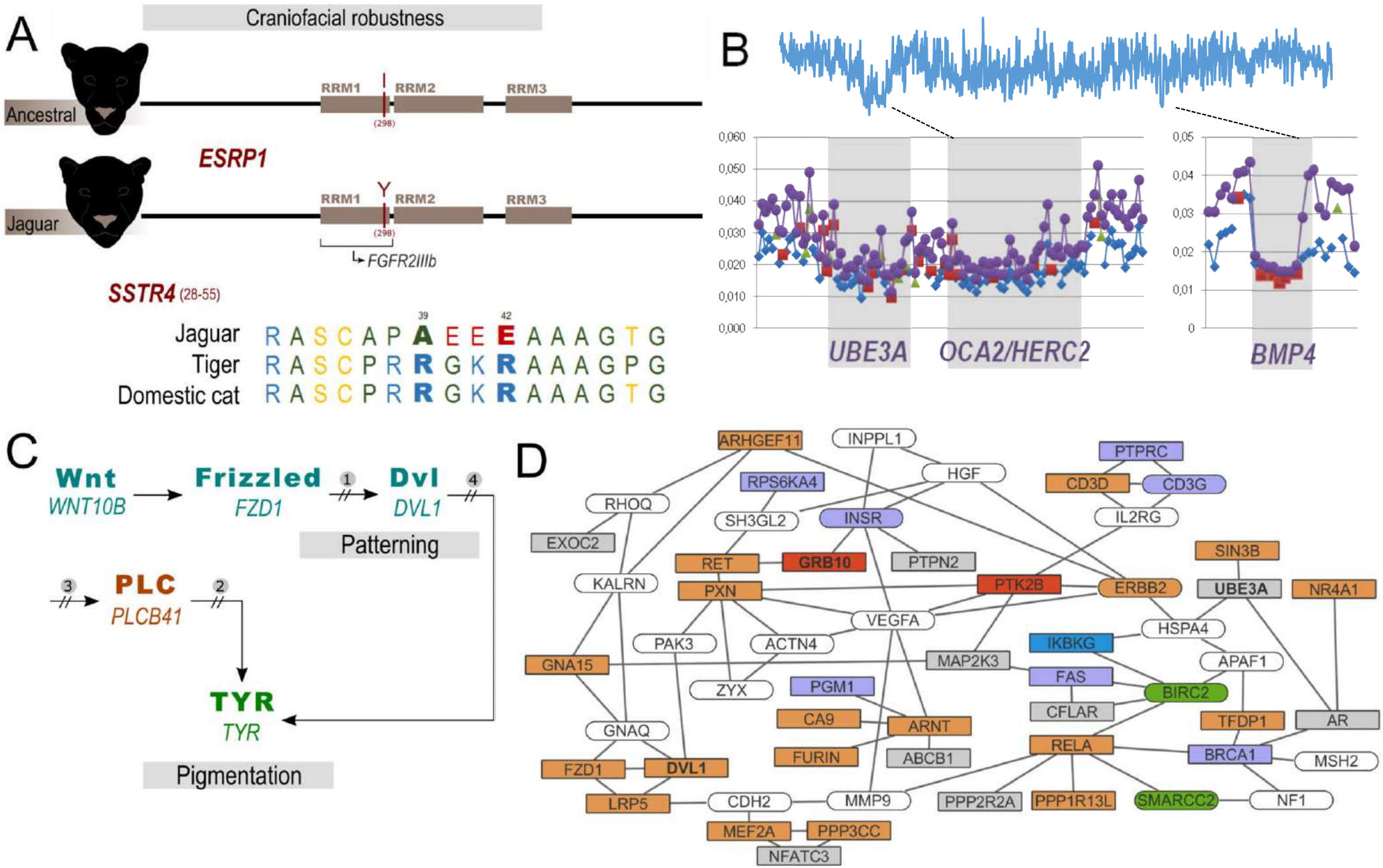


C



D

Phenotype	Species	Genes
Apnea	snow leopard	*TCTN1; TSEN54; UHFR1*
Body Size	lion	*WISP1; **GRB10***
Brain Development	lion	*LHX6; KCNC4*
	leopard	***ATP10A; CCDC86***
Circadian cycle	leopard	*ARNTL; PTPRU*
Craniofacial development	jaguar	*ESRP1; SSTR4; LHX4*
	leopard	*PRDM10; **LDB1***
	snow leopard	*ALX3*
Hearing	jaguar	*SOX10; PCDH15*
Limb development	jaguar	*TPM3; FGD4; SOX10; LIAS*
	leopard	***LDB1**; PODXL*
Lipid Metabolism	snow leopard	*CEL; PNLIPRP2*
Melanogenesis	tiger	*PLCB4; TYR; **DVL1**; WNT10B; FZD1*
Protein Metabolism	jaguar	*SSTR4; SEL1L*
	tiger	*ATP1A1; SLC8A3; COL6A1*
Olfactory transduction	jaguar	*OR10H1; OR3A1; OR13F1; OR2W1; OR2B6; OR4K13; OR9K2*
	leopard	*OR52A5; OR8S1; OR9K2; OR52R1*
	tiger	*OR2T2; OR4L1; OR12D2; OR52D1; OR1D2; OR7C2; OR1F1; OR51I2*
Reproduction	jaguar	*SLC26A3; MSH4*
	lion	*SPATA16, USP26*
	tiger	*CACNA1D; CACNA1F; PDE3A; PIK3CB; PPP2R1B; RPS6KA6; CCNB3; PLCB4; PPP3CC*
Social Behaviour	lion	***GRB10**; EFHC2; ARNTL; PTK2B; PEX13*
Vision	snow leopard	*IKBKG; KIF21A; TCTN1*
	tiger	*CACNA1F; LRP5; GRK1; CLN6; ROM1*


A
Craniofacial robustness
Ancestral
RRM1
RRM2
RRM3
(298)
ESRP1
Jaguar
Y
FGFR2IIIb
SSTR4 (28-55)
39
42
Jaguar R A S C A P A E E E A A A G T G
Tiger R A S C P R R G K R A A A G P G
Domestic cat R A S C P R R G K R A A A G T G
B
0,060
0,050
0,040
0,030
0,020
0,010
0,000
0,05
0,04
0,03
0,02
0,01
0
UBE3A
OCA2/HERC2
BMP4
C
Wnt
WNT10B
Frizzled
FZD1
1
Dvl
DVL1
4
Patterning
3
PLC
PLCB41
2
TYR
TYR
Pigmentation
D
ARHGEF11
INPPL1
PTPRC
RPS6KA4
HGF
CD3D
CD3G
RHOQ
SH3GL2
INSR
IL2RG
EXOC2
SIN3B
RET
GRB10
PTPN2
KALRN
PXN
PTK2B
ERBB2
UBE3A
NR4A1
VEGFA
HSPA4
PAK3
ACTN4
IKBKG
APAF1
GNA15
MAP2K3
FAS
ZYX
PGM1
BIRC2
CFLAR
CA9
ARNT
TFDP1
AR
GNAQ
RELA
FURIN
ABCB1
BRCA1
MSH2
FZD1
DVL1
LRP5
CDH2
MMP9
PPP2R2A
PPP1R13L
SMARCC2
NF1
MEF2A
PPP3CC
NFATC3

Draft, Nature

February 23, 2016.

Supplementary Information

Jaguar genome sheds light on the complex evolution of the big cats

Table of Contents

2.2. Functional annotation of coding genes

2.2.1. Functional annotation with InterProScan5

2.2.2. Functional annotation with CDD

2.2.3. Functional Annotation with Uniprot - Functional Annotation KO

2.2.4. Functional Annotation with RNAseq data

2.2.5. Functional Annotation with OrthoMCL

2.4. Annotation of repetitive regions in the jaguar genome

Supplementary Table S7: **Summary of repeats masking for Jaguar genome assembly**

Supplementary Table S8: **Tandem repeat annotation**

Supplementary Table S9: **Forty largest microsatellites families**.

2.5. Non-coding RNA Annotation

Supplementary Table S10: **Statistics on non-coding RNA annotation**.

2.6. NUMTs in jaguar genome

Supplementary Table S11: **NUMTs identified in jaguar genome.**

3. Genome-wide alignments of the five *Panthera* species

3.1. Genome-wide alignments used to extract coding genes

Supplementary Table S12: **Previously reported genomic resources used in the following analyses**.

Supplementary Table S13: **Consensus genome information**

3.2. Genome-wide alignments used for sliding-window analyses of genealogical discordance

3.3. Identification of orthologous coding genes

Supplementary Figure S1: **Scheme of coding sequences for the *ESRP1* and *SSTR4* genes**

Supplementary Figure S2: **Scheme of coding sequences for the *TPM3*, *FGD4*, *SOX10*, *LIAS* genes**

3.4. Assessment of genealogical discordance in the *Panthera*

4. In-depth analyses of gene families

4.1. Analysis of olfactory receptor genes in the *Panthera*

4.2 Olfactory receptors results

5. Demographic history analyses

5.1. PSMC

5.2. Whole genome introgression (ABBA/BABA) tests

6. Detection of genes under positive selection in the *Panthera* species

6.1 PAML Analyses

Supplementary Table S14: **Parameters used in ctl file by codeml software**

Supplementary Figure S3: **Selection pipeline**

Supplementary Figure S4: **P-value distribution for the branch-site analysis**

Supplementary Figure S5: **P-value distribution for the site model analyses**

6.2. Assessment of the impact of Gene Tree *vs.* Species Tree discordance on the selection analyses

Supplementary Figure S6: **Likelihood ratio tests comparison for the branch-site analysis**.

6.3. Gene enrichment analysis

6.4 Branch-site test results

6.4.1. Cranial and forelimb development in jaguars

Supplementary Table S15: **Branch-site results for the jaguar.**

Supplementary Table S16: **Gene ontology results for the jaguar.**

6.4.2. Social behaviour in Lions

Supplementary Table S17: **Branch-site results for the lion**.

Supplementary Table S18: **Gene ontology results for the lion**

1. Genome sequencing and assembly

1.1. Data collection

1.1.1. Sequencing of the jaguar genome

The target individual was a healthy, wild-caught male (called 'Vagalume'), born in the southern Brazilian Pantanal region and currently housed at the Sorocaba Municipal Zoo, Brazil. Certified veterinarians performed all anesthesia and sampling procedures during routine health check-ups performed by the zoo staff. We collected *ca.* 100 ml of whole blood to use in genomic DNA extractions. These samples were placed in multiple Vacutainer tubes containing the anticoagulant K_2EDTA, and stored at 4ºC for less than 48 hours before DNA extractions were carried out. DNA was extracted using a Qiagen Blood and Tissue Kit (Qiagen).

Vagalume's genome was sequenced at ~84X coverage using four libraries and seven Illumina HiSeq 2500 lanes (two different 180-bp insert paired-end libraries, sequenced in five lanes, and two mate-pair libraries [3-kb and 8-kb insert size] sequenced in one lane each) (Table S1).

Supplementary Table S1: **DNA sequencing information for the reference jaguar individual**. Sequenced libraries with number of reads, insert size, sequence length and GC content information.

Species	Number of reads	Direction	Insert Size	Seq Length	% GC
Jaguar	908,465,441	R1	180bp	101	44
Jaguar	908,465,441	R2	180bp	101	41
Jaguar	211,010,191	R1	3kb	101	41
Jaguar	211,010,191	R2	3kb	101	41
Jaguar	183,752,503	R1	8kb	101	40
Jaguar	183,752,503	R2	8kb	101	40
Total	**2,606,456,270**				

1.1.2. Sequencing of jaguar transcriptomes

Additional blood samples and small biopsies of muscle, testicle, gum and skin were collected from Vagalume during routine veterinarian procedures conducted at the Sorocaba zoo. These samples were immediately stabilized with RNAlater (Invitrogen®), with a 10:1 proportion of this buffer relative to the sample. Total RNA was extracted with a standard Trizol RNA extraction

protocol[1] (Chomczynsky & Sacchi, 1987). We performed RNA quantification and quality assessment (integrity and purity) using a NanoDrop spectrophotometer (Thermo Scientific®), and constructed a cDNA library for each sample using the TruSeq RNA Sample Prep Kit v2 (Illumina®), following the manufacturer's instructions. The cDNA libraries were then used to perform multiplexed/barcoded RNA-seq with an Illumina HiScan sequencer (Table S2).

Supplementary Table S2: **Summary of the jaguar transcriptome sequencing results**. Sequenced tissues used for the transcriptome analysis and annotation. Number of reads, insert size, sequence length and GC content for each library.

Tissue	Library ID	Direction	Number of Reads	Insert Size	Seq Length	% GC
Gum	Gum_1_R1	R1	26672756	180bp	101	49
Gum	Gum_1_R2	R2	26672756	180bp	101	49
Tongue	Tongue_1_R1	R1	20298080	180bp	101	50
Tongue	Tongue_1_R2	R2	20298080	180bp	101	50
Testicle	Testicle_1_R1	R1	26912880	180bp	101	50
Testicle	Testicle_1_R2	R2	26912880	180bp	101	50
Skin	Skin1_R1	R1	24363734	180bp	101	49
Skin	Skin_1_R2	R2	24363734	180bp	101	49
Skin	Skin_2_R1	R1	1473959	180bp	50	52
Skin	Skin_3_R1	R1	4000000	180bp	50	51
Skin	Skin_4_R1	R1	899518	180bp	50	51
Skin	Skin_5_R1	R1	4000000	180bp	50	52
Skin	Skin_6_R1	R1	1942875	180bp	50	52
Skin	Skin_7_R1	R1	2422274	180bp	50	51
Blood	Blood_1_R1	R1	4000000	180bp	50	52
Blood	Blood_2_R1	R1	1805834	180bp	50	52
Muscle	Muscle_1_R1	R1	1807474	180bp	50	51

1.1.3. Sequencing of the leopard genome

To examine the whole genome evolutionary history of genus *Panthera*, we generated whole genome sequence data from a male leopard. DNA was extracted from a fibroblast cell line and used to generate a standard 250-bp insert Illumina library, which was sequenced to ~30x coverage using paired-end 125 bp reads using Illumina on the HiSeq2500 platform (Supplementary Table S3).

Supplementary Table S3: **DNA sequencing information for the leopard genome**. Sequenced libraries with number of reads, direction, insert size, sequenced length and GC content information.

Species	Number of reads	Direction	Insert Size	Seq Length	% GC
Leopard	94016735	R1	180bp	20-125	41
Leopard	94016735	R2	180bp	20-125	41
Total	**188033470**				

1.2. Assembly of the jaguar genome

We applied the standard ALLPATHS-LG *de novo* assembly pipeline to generate a whole genome assembly of Jaguar with genome size of 2.4 GB (contig N50 length of ~29 Kb and scaffold N50 length of 1.5 Mb, see Supplementary Table S4).

Supplementary Table S4**: de novo assembly report of Jaguar genome**

Minimum contig size	1,000
Number of contigs	156,436
Number of contigs per Mb	65.1
Number of scaffolds	7521
Total contig length	2,282,949,125
Total scaffold length, with gaps	2,402,152,336
N50 contig size in kb	28.6
N50 scaffold size in kb, with gaps	1,517
Number of scaffolds per Mb	3.13
Median size of gaps in scaffolds	400
Median dev of gaps in scaffolds	70
% of ambiguous bases	31.68
Ambiguities per 10,000 bases	13.47

1.2.1. Assessment of assembly quality

To assess the quality of the reference jaguar genome assembly, we used the software Quast with eukaryotic settings[2], and for estimation of overall quality of genes assembly we used the software CEGMA[3]. CEGMA reported 93.15% of partially assembled genes including 56.05% of fully assembled genes (Table S5).

Supplementary Table S5: **Assembly quality metrics**.

	Contigs	**Scaffolds**
CEGMA partial	86.29%	93.15%
CEGMA full	49.60%	56.05%
#contigs	158329	7521
#contigs >1kb	156436	7442
Total length	2,284,631,488	2,405,344,986
Total length >1kb	2,282,949,125	2,405,268,288
Largest contig	338,209	8,985,697
GC (%)	41.51	41.51%
N50	28.53 kb	1.52 Mb
L50	23,26	474
%gaps	0%	5%

2. Annotation of the jaguar genome

2.1. De novo prediction of coding genes

We used MAKER2[4], a genome annotation and data management tool designed for second-generation genome projects. *Ab initio* gene predictions were produced by the SNAP program version 2013-02-16[5]. To improve annotation quality, we used three different strategies to evaluate gene models in the MAKER2 pipeline. For the first step, we used tiger proteins and Refseq protein evidence. In the second step, we built the SNAP model (Jaguar.hmm) and used jaguar RNA-seq generated in this study. In the last step, Jaguar.hmm was used to finish the annotation of genes. All predictions were produced in standardized GFF3 format. Evidence-based gene annotations in MAKER2 were produced using default settings. Finally, we aligned the predictions to a TE protein database using BlastP with the E-value set to 1e-10. In total, 25,451 genes were identified (Table S6).

Supplementary Table S6: **Annoted protein-coding genes**

Feature	Number
Base pair	32,384,308
N50	1,827
Longest gene	26,310
Median gene size	927
% genes > 100 b	99.56%
% genes > 200 b	97.27%
% genes > 500 b	73.65%
% genes > 1 kb	46.56%

2.2. Functional annotation of coding genes

2.2.1. Functional annotation with InterProScan5

We used InterProScan5, a tool that combines different protein signature recognition methods into a single resource. This provides an overview of the families that a protein belongs to and the domains and sites it contains, identifying the presence and organization of protein sequence domains as well as critical residues. The genome sequence was submitted in FASTA format. Matches were then calculated against all of the required member database signatures **(**BlastProDom, FPrintScan, HMMPIR, HMMPfam, HMMSmart, HMMTigr, ProfileScan, HAMAP, PatternScan, SuperFamily, SignalPHMM, TMHMM, HMMPanther, Gene3D, Phobius, and Coils) and the results were output in TSV (a simple tab-delimited file) format. The InterProScan5 approach

was able to annotate 22,191 jaguar genes, with only 3,260 predicted genes not showing matches with InterPro's signatures.

2.2.2. Functional annotation with CDD

In this step, we worked with the 3,260 genes that did not show an InterPro signature. For this, we chose CDD: NCBI's conserved domain database, which provides an online tool to annotate protein domains. In total, protein domains were found in 197 of these jaguar genes.

2.2.3. Functional Annotation with Uniprot - Functional Annotation KO

The genes predicted were characterized through a BLAST search against the UniProt Knowledge Base (e-value $1e^{-10}$). Genes were assigned to KEGG Orthotology groups using a UniRef Enriched database (UEKO). In total, 21,279 Jaguar genes (83.6%) showed hits to UniProt and 4,020 genes showed matches with KO.

2.2.4. Functional Annotation with RNAseq data

The RNA-seq dataset generated in this study from the same jaguar individual whose genome was sequenced was used to validate the annotated coding genes. In order to do that, the sequence of each gene was treated as reference and the RNA-seq reads were mapped against them. In total, 16,586 genes (65%) had a RNA-seq coverage of $\geq$ 75 percentage of their predicted sequence.

2.2.5. Functional Annotation with OrthoMCL

We used OrthoMCL, a tool that provides a scalable method for constructing orthologous groups across multiple eukaryotic taxa, to search for orthologs between the jaguar and tiger genomes. This method uses a Markov Cluster algorithm to group putative orthologs and paralogs. In total, 16,680 genes were clustered using this approach.

2.3 Construction of a jaguar gene database

To manage and analyze this large amount of information and to integrate all the annotation data, we built a database named Jaguar_SQL, a Structured Query Language relational database using MySQL as a database management system. To recover potential genes in the database mentioned above, the ID of these genes was used in SQL search and the Final Table was built. Detailed information can be accessed in the file "Final_Table.xls" attached to this report. The

first column of the Final Table corresponds to the identifiers of the 25,451 genes. The remaining columns are: id_jaguar_KO, id_jaguar_InterPro, id_jaguar_CDD, id_jaguar_OrthoMCL, id_jaguar_UniProt respectively. For columns 2-6, we added the values 0 or 1, with 0 being the absence of corresponding evidence for a particular gene and 1 being the presence of evidence for that gene. The 7th column corresponds to the percentage of coverage of the RNAseq and the 8th column corresponds to the gene length. The 9th column represents the sum of the presence of evidence for the gene (sum of the columns 2-6). For validation of genes, the following criteria were used: all genes with sum greater than 1 were considered to have some evidence. In addition, the genes' RNAseq coverage was required to be above 75%.

2.4. Annotation of repetitive regions in the jaguar genome

To estimate the overall genome repeatness, we used WindowMasker software[6]. This masked 29.66% of the genome according to the presence of repeated fragments with exact sequences. For the annotation of known repeats, we used the RepeatMasker software[7] and carnivora-specific library from the Repbase Update library version 20140131[8] . RepeatMasker masked 37.48% of genome including 18.92% of LINEs, 10.40% of SINEs, 5.11% of LTR elements, and 2.86% of DNA elements. To mask simple sequence repeats, we used DUST software[9], and the Jaguar genome contained 11.20%. To detect tandem repeats was used the Tandem Repeats Finder (TRF) software, version 4.07[10] with mismatch and maximum period parameters values set to 5 and 2000; TRF output was processed as published previously in[11]. We found tandem repeats where divided into the following groups with Trevis software: microsatellites, perfect microsatellites, complex tandem repeats, and three groups of large tandem repeats according to array length (Table S8). Additionally, we computed statistics for forty largest families of microsatellites. Two largest families formed by (AG)n and (AC)n microsatellites, 68.72% of (AG)n repeats had imperfect arrays, and 62.1% of (AC)n repeats had perfect microsatellite arrays. The summary of repeat annotation with different tools is shown in Supplementary Table S7.

Supplementary Table S7: **Summary of repeats masking for Jaguar genome assembly**.

Tool	Repeat type	Contigs	Scaffolds
WindowMasker	Any repeated fragments	31.23%	29.66%
TRF	Tandem repeats	2.10%	2.12%
DUST	Simple sequence	6.50%	11.20%
RepeatMasker	Known repeats	38.63%	37.48%
	SINEs	10.65%	10.40%
	LINEs	19.38%	18.92%
	LTR elements	5.41%	5.11%
	DNA elements	3.01%	2.86%
	Other	0.18%	0.19%

Supplementary Table S8: **Tandem repeat annotation**.

Category	Contigs	Scaffolds
All TRs	854704	856590
Microsatellites	344251	343205
Perfect microsatellites	145457	144563
Complex TRs	3097	3091
Large TRs >1 kbp	590	1685
Large TRs >3 kbp	76	341
Large TRs >10 kbp	2	4

Supplementary Table S9: **Forty largest microsatellite families**.

Family	Copy number	Total length (bp)	Imperfect (%)	Perfect (%)
(AG)n	114588	5367594	68.72	31.28
(AC)n	95656	3857060	37.9	62.1
(A)n	27547	1017208	40.77	59.23
(AT)n	18112	1422474	78.57	21.43
(AAAT)n	17070	735046	58.6	41.4
(AAAG)n	14927	1285059	81.18	18.82
(AAGG)n	7867	812581	86.18	13.82
(AAAAC)n	7663	284497	52.72	47.28
(AAAC)n	6572	228002	52.77	47.23
(AAAAT)n	6226	331755	71.86	28.14
(AGAT)n	5991	442948	80.3	19.7
(AAT)n	4971	187591	54.09	45.91
(AATG)n	4706	173181	57.65	42.35
(AAC)n	4445	159310	46.19	53.81
(ATCC)n	3084	411443	83.82	16.18
(C)n	3011	111414	40.42	59.58
(AAAAG)n	2914	226321	86.51	13.49
(AAG)n	2538	174700	60.44	39.56
(ACAT)n	2512	116237	68.39	31.61

(AGG)n	2252	154900	79.97	20.03
(ATC)n	1498	78208	56.81	43.19
(AGGG)n	1385	103320	89.39	10.61
(AAAGG)n	795	80569	82.77	17.23
(ACAG)n	698	26012	84.1	15.9
(AATAT)n	602	38234	80.4	19.6
(AAGGG)n	564	44792	80.14	19.86
(AATT)n	525	16903	38.86	61.14
(AAATT)n	494	22558	61.34	38.66
(AATC)n	486	14553	33.54	66.46
(AGC)n	453	19032	43.05	56.95
(ACC)n	441	24983	51.47	48.53
(AAGAG)n	408	37429	82.35	17.65
(CCG)n	380	14521	70.53	29.47
(AGAGG)n	324	22916	78.4	21.6
(AATAG)n	317	19019	64.67	35.33
(AGGGG)n	315	17641	76.83	23.17
(AACC)n	297	12244	61.62	38.38
(AAATG)n	268	15104	67.54	32.46
(ACTC)n	192	9116	69.79	30.21
(ACCT)n	178	7407	76.97	23.03

2.5. Non-coding RNA Annotation

Non-coding RNA genes were annotated according to the Ensembl recommendations (http://www.ensembl.org/info/genome/genebuild/ncrna.html). All RFAM sequences were aligned against genomic sequences using BLASTN with e=10-5. BLAST hits were clustered and used to seed Infernal searches with the corresponding RFAM covariance models[12]. The resulting BLAST hits were used as supporting evidence for ncRNA genes confirmed by Infernal[13]. The non-coding RNA annotation summary is shown in Supplementary Table S10. For the tRNAs, we excluded tRNA predictions from RFAM results and used results from tRNAscan-SE with default parameters instead[14]. Only predictions with a cove score greater than 20 were considered. tRNA-scan predicted 169,838 regions with possible tRNA genes (supplementary file "genome_jaguar.tRNA.20.html"). The largest part are felidae-specific SINE elements containing SelCys, Arg, Lys, and Gln tRNA fragments (CAN-Sine elements [15,16]). After removing fragmented Arg, Lys and Gln tRNAs with stricter parameters (cove score greater than 60), only 3,299 tRNA genes remained, which is close to the 3,039 tRNA genes reported in the domestic cat genome[17].

Supplementary Table S10: **Statistics on non-coding RNA annotation**.

Type	Subtype	Copy Number	Total length (bp)
miRNA		1,626	130,589
tRNA	All	169,838	4,257,400
	Without repeats	3,299	313,405
rRNA	rRNA	510	94,238
snRNA	snRNA	3454	419,792
	snoRNA	973	106,012
miscRNA		1640	323,969

2.6. NUMTs in jaguar genome

To annotate nuclear mitochondrial translocations (NUMTs) in the jaguar reference genome, we used a BLAST-based approach. Using the complete jaguar mitochondrial genome (generated in this study) as a query, matches were sought with the following search parameters: (i) a hit with at least 16bp; (ii) 10^{-10} e-value threshold; (iii) no DUST filter query; (iv) cost of 0 to open a gap and 2 to extend it; (v) 40 of X-dropoff value, for preliminary gapped extensions; (vi) reward for a match, and penalty for a mismatch, of 1. This search led to the identification of 171 fragments bearing similarity to mitochondrial sequence. These were distributed in 128 scaffolds of the assembled jaguar genome, and ranged in length between 86 and 7361 bp (Table S11).

Supplementary Table S11: **NUMTs identified in jaguar genome.**

Numt ID	Scaffold	Start position	End position	Length (bp)
numt_001	4	3506562	3507835	1273
numt_002	5	2666764	2667031	267
numt_003	5	4159520	4159711	191
numt_004	7	4602010	4599712	2298
numt_005	7	4603732	4602258	1474
numt_006	7	4605573	4604433	1140
numt_007	7	4614835	4611413	3422
numt_008	7	4850669	4851430	761
numt_009	14	1685779	1685214	565
numt_010	14	1685820	1686012	192
numt_011	16	1508880	1509379	499
numt_012	16	1591191	1591301	110
numt_013	31	997313	997411	98
numt_014	35	1417271	1416848	423
numt_015	39	3950127	3949856	271
numt_016	44	2731221	2730916	305
numt_017	44	2732363	2731438	925
numt_018	44	2733145	2732565	580

numt_019	60	587282	586682	600
numt_020	72	3254614	3254006	608
numt_021	81	590452	591085	633
numt_022	86	1389105	1389892	787
numt_023	86	3232539	3232178	361
numt_024	86	3348843	3348180	663
numt_025	90	3125361	3124291	1070
numt_026	99	1975351	1974523	828
numt_027	112	1416604	1417266	662
numt_028	122	1421041	1420335	706
numt_029	133	773043	772957	86
numt_030	135	2208093	2207960	133
numt_031	162	1866763	1866649	114
numt_032	164	47276	46591	685
numt_033	167	80069	80477	408
numt_034	168	2179288	2178960	328
numt_035	187	1794716	1794163	553
numt_036	192	546653	547244	591
numt_037	193	671073	671485	412
numt_038	210	1690114	1690850	736
numt_039	212	1335979	1336336	357
numt_040	226	2064579	2063719	860
numt_041	227	1236236	1234911	1325
numt_042	229	1078776	1079537	761
numt_043	234	688069	687434	635
numt_044	238	1430334	1430132	202
numt_045	259	899640	902782	3142
numt_046	294	840448	842379	1931
numt_047	299	182706	184036	1330
numt_048	305	286737	287264	527
numt_049	321	1678137	1678420	283
numt_050	347	59485	58818	667
numt_051	372	590313	589891	422
numt_052	375	128493	128216	277
numt_053	380	1299674	1299076	598
numt_054	383	1704812	1704522	290
numt_055	389	1586455	1585395	1060
numt_056	395	390104	389307	797
numt_057	405	99744	98990	754
numt_058	411	1256702	1257084	382
numt_059	416	1365893	1365625	268
numt_060	435	829893	829350	543
numt_061	436	1538229	1540236	2007
numt_062	437	1103911	1105560	1649
numt_063	440	40649	39291	1358
numt_064	440	41022	40881	141
numt_065	456	34607	34923	316
numt_066	478	1404685	1404436	249
numt_067	505	812873	813296	423

numt_068	523	669200	669055	145
numt_069	537	1230613	1229983	630
numt_070	538	738211	737403	808
numt_071	554	220269	220624	355
numt_072	565	727980	728219	239
numt_073	565	1277892	1278329	437
numt_074	565	1282248	1282538	290
numt_075	568	333860	333231	629
numt_076	569	322745	322425	320
numt_077	569	324094	322977	1117
numt_078	600	844854	847470	2616
numt_079	600	847488	847778	290
numt_080	600	847953	848972	1019
numt_081	600	850125	851229	1104
numt_082	600	851442	853765	2323
numt_083	615	632282	632848	566
numt_084	618	196197	195056	1141
numt_085	636	667464	667301	163
numt_086	650	526473	526113	360
numt_087	713	994088	993253	835
numt_088	714	665358	664266	1092
numt_089	714	666542	665588	954
numt_090	714	670756	666759	3997
numt_091	776	243443	243901	458
numt_092	838	372763	372304	459
numt_093	871	16353	16089	264
numt_094	871	20722	20284	438
numt_095	873	633110	633476	366
numt_096	945	336402	336536	134
numt_097	958	800600	799980	620
numt_098	962	24438	24238	200
numt_099	991	594374	594785	411
numt_100	1006	439930	439705	225
numt_101	1034	14270	13919	351
numt_102	1050	262882	261869	1013
numt_103	1050	264592	263104	1488
numt_104	1050	266774	264631	2143
numt_105	1088	557282	557905	623
numt_106	1093	227903	227305	598
numt_107	1121	34903	32798	2105
numt_108	1122	560379	558082	2297
numt_109	1143	667050	666845	205
numt_110	1192	13326	12998	328
numt_111	1192	14385	13541	844
numt_112	1229	136605	136386	219
numt_113	1229	137314	136649	665
numt_114	1229	138062	137753	309
numt_115	1229	144412	143443	969
numt_116	1245	578752	579570	818

numt_117	1267	134448	133540	908
numt_118	1267	136771	134670	2101
numt_119	1269	176534	176848	314
numt_120	1273	245861	247127	1266
numt_121	1305	13329	12798	531
numt_122	1373	241091	239997	1094
numt_123	1373	241685	241326	359
numt_124	1373	244231	243634	597
numt_125	1378	388188	388055	133
numt_126	1399	420306	420424	118
numt_127	1422	152687	153216	529
numt_128	1437	344662	344153	509
numt_129	1437	388563	384912	3651
numt_130	1461	490711	490615	96
numt_131	1501	203516	203938	422
numt_132	1607	18994	19307	313
numt_133	1709	248646	248029	617
numt_134	1808	21614	21512	102
numt_135	1868	268903	269537	634
numt_136	1896	236057	236935	878
numt_137	1903	92919	93369	450
numt_138	1979	168245	168465	220
numt_139	2015	229326	227388	1938
numt_140	2063	1697	1	1696
numt_141	2093	135949	135634	315
numt_142	2095	7511	7975	464
numt_143	2108	87699	86870	829
numt_144	2140	26697	26195	502
numt_145	2140	28139	27004	1135
numt_146	2140	30813	28364	2449
numt_147	2140	32725	31032	1693
numt_148	2157	67821	73076	5255
numt_149	2157	73390	74303	913
numt_150	2157	74554	78770	4216
numt_151	2162	183760	183606	154
numt_152	2171	29350	29857	507
numt_153	2171	35239	35919	680
numt_154	2173	131232	130560	672
numt_155	2188	1057	825	232
numt_156	2354	119924	119248	676
numt_157	2354	125259	124750	509
numt_158	2426	39307	39966	659
numt_159	2454	27756	28264	508
numt_160	2454	33095	33779	684
numt_161	2471	97890	99733	1843
numt_162	2588	69606	70227	621
numt_163	2634	45969	49668	3699
numt_164	2814	44303	49526	5223
numt_165	2814	49997	57358	7361

numt_166	3072	7612	8118	506
numt_167	3072	9915	9481	434
numt_168	3125	23119	23810	691
numt_169	3262	15288	15384	96
numt_170	3824	1	5661	5660
numt_171	3824	8200	9604	1404

3. Genome-wide alignments of the five *Panthera* species

To perform comparative analyses across the *Panthera*, we employed two different approaches to construct genome-wide alignments. Approach 3.1 was used to extract orthologous coding genes for all species (see item 3.3 below), while approach 3.2 was employed to perform the sliding-window analyses of genealogical discordance (see item 3.4 below). Each of them will be described in detail below.

3.1. Genome-wide alignments used to extract coding genes

We aligned the genomes of the five *Panthera* species, along with that of the domestic cat (*Felis catus*) (which was used as the outgroup), using the software LAST[18]. This approach enables the alignment of both assembled genomes and unmapped reads to a reference. We used three assembled genomes (jaguar, tiger and domestic cat), and raw genomic reads for the three species with low coverage data (lion, leopard and snow leopard) (Table S12). All genomes were aligned individually using the jaguar genome as reference, and the same exercise was performed separately (for assessment of robustness) using the tiger as reference. By using two different genomes (jaguar and tiger) as reference, we aimed to assess whether any detectable bias would emerge from this process when identifying orthologous coding genes, especially in the low-coverage species (see item 3.3 below).

After mapping the assemblies or reads of each genome towards the reference, we used SAMtools[19] to compute a species-specific consensus, and its coordinates matched exactly those of the reference (Table S12). These alignments could then be used to characterize synteny blocks among jaguar, tiger and domestic cat genomes, as well as to extract coding genes from all species. The latter wereinferred to be orthologous among all species based on their genomic position relative to the reference (see item 3.3).

Supplementary Table S12: **Previously reported genomic resources used in the following analyses**.

Common name	Scientific Name	Type of Data	Accession Number	Reference
Tiger	*Panthera tigris*	High-coverage genome	ATCQ01	Cho et al. 2013[20]
Lion	*Panthera leo*	Low-coverage genome (raw reads)	SRX273034	Cho et al. 2013[20]
Snow Leopard	*Panthera uncia*	Low-coverage genome (raw reads)	SRX273036	Cho et al. 2013[20]
Cat	*Felis catus*	High-coverage genome	AANG03	Montague et al. 2014[21]

Supplementary Table S13: **Consensus genome information (Tiger reference)**:

Feature	***Panthera leo***	***Panthera pardus***	***Panthera uncia***
Number of contigs	1,379	1,375	1,379
Number of bases in all contigs	2,390,883,613	2,390,829,645	2,390,861,245
N50	8,860,407	8,860,407	8,860,407
Longest contig	41,607,841	41,607,841	41,607,841
Median contig size	15,815	16,235	16,148
Number of contigs > 100 b	1,376	1,372	1,374
Total of bases into contigs > 100 b	2,390,883,383	2,390,829,405	2,390,860,856
% contigs > 100 b	99	99	99
Number of contigs > 200 b	1,348	1,354	1,351
Total of bases into contigs > 200 b	2,390,878,923	2,390,826,466	2,390,856,977
% contigs > 200 b	97	98	98
Number of contigs > 500 b	1,191	1,187	1,187
Total of bases into contigs > 500 b	2,390,822,375	2,390,766,351	2,390,797,568
% contigs > 500 b	86	86	86
Number of contigs > 1 kb	962	961	962
Total of bases into contigs > 1 kb	2,390,656,905	2,390,603,429	2,390,635,250
% contigs > 1 kb	69	69	69

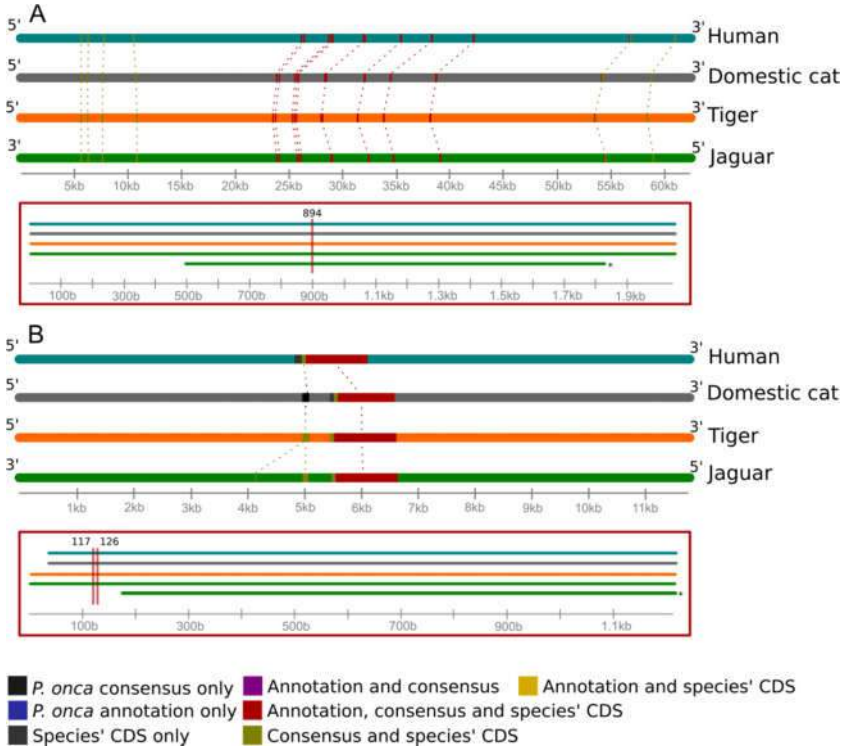
3.2. Genome-wide alignments used for sliding-window analyses of genealogical discordance

We used Trim Galore to trim the raw Illumina reads and screen for potential adapter sequences (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) in data from all five *Panthera* species. Filtered reads were mapped to the whole genome assembly reference of the tiger V1.0[22] with BWA[23]. We reordered all tiger assembly scaffolds relative to the domestic cat genome assembly (version felCat5) to obtain an overview of evolutionary signatures on different chromosomes. Synteny blocks between the tiger and domestic cat genomes were detected on the basis of whole genome alignment results produced with LAST[18] with default parameter settings. Removal of PCR-induced sequence duplicates and calling and filtering raw single nucleotide variants (SNVs) was performed using SAMTOOLS[19]. We applied an analysis pipeline used in a previous study[24] to generate whole genome alignments of the five *Panthera* species and the domestic cat, which was used an outgroup for downstream analyses.

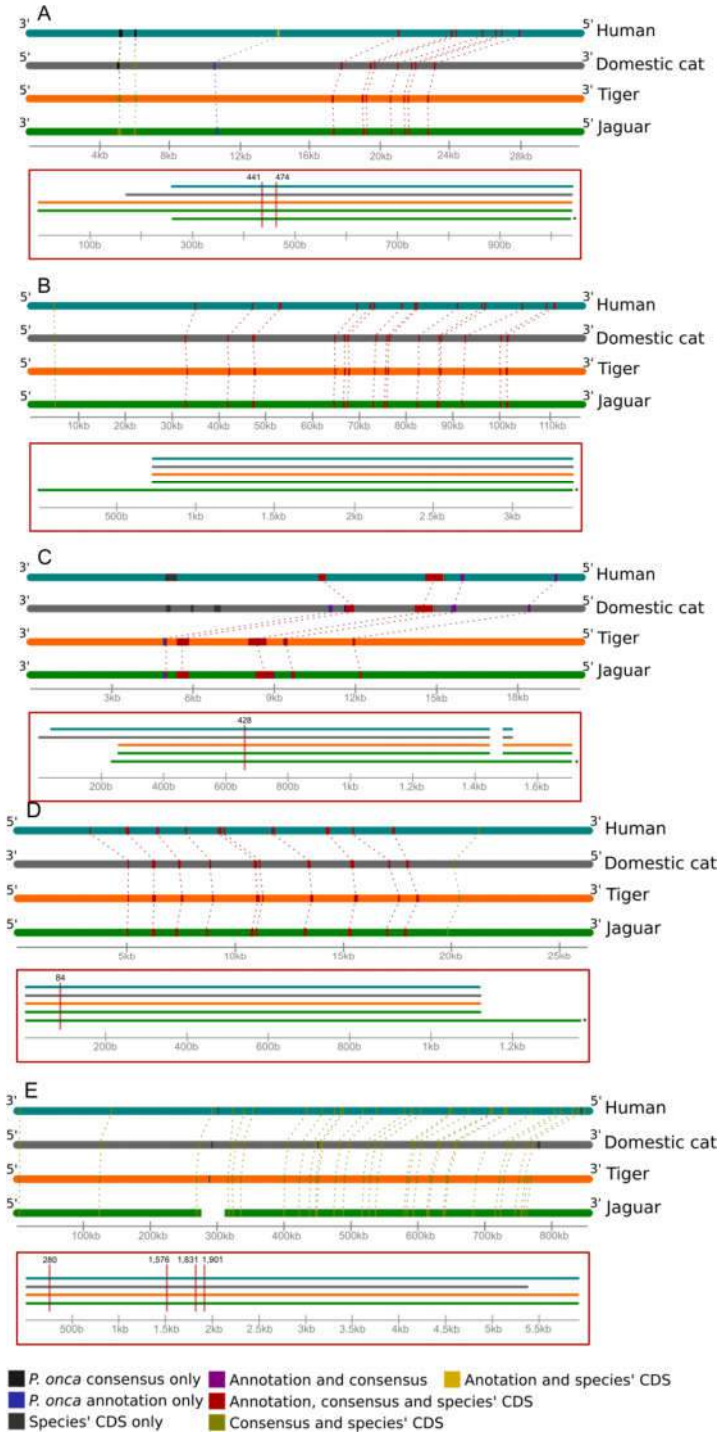
3.3. Identification of orthologous coding genes

We used the genome-wide alignments performed with LAST[18] (and the derived species-specific consensus computed with SAMtools[19]) as a basis to extract orthologous genes from all *Panthera* species and the domestic cat.

Full genes were used to perform genealogical discordance analyses (see item 3.4 below) and to produce gene trees employed in the investigation of shifts in selective regime (especially the occurrence of positive selection) loci under positive selection. The extracted CDS for each gene were aligned and verified (*e.g.* checking for any open reading frame, as well as matching start and stop codon positions, in addition to exon boundaries). All the genes that passed these filtering steps were used in the screen for genes under positive selection in the *Panthera* genus. We also performed manual, in-depth assessments of comparative gene structure between the jaguar annotation and three other genomes (human, domestic cat and tiger) in the case of genes identified as bearing particularly relevant signatures of positive selection (see Figures S1 and S2 for examples).



Supplementary Figure S1: **Scheme of coding sequences of the** *ESRP1* (A) and *SSTR4* (B) genes mapped to genomic region (colored strips) of each species. The dashed lines connect the equivalent portion of the CDS. The box below contains respective coding sequences of each species, being consensus and annotation (star) for the jaguar, and the site under selection at the jaguar consensus. The gray scale indicates the length of sequence (in base pairs). Scaffolds in A: human (NW_004929340.1), domestic cat (NC_018740.2), tiger (NW_006711339.1) and jaguar (scaffold 320). Scaffolds in B: human (NW_004929416.1), domestic cat (NC_018725.2), tiger (NW_006712402.1) and jaguar (scaffold 973).



Supplementary Figure S2: Scheme of coding sequences' fragments *TPM3* (A), *FGD4* (B), *SOX10* (C), *LIAS* (D) and *PCDH15* (E) mapped to genomic region (colored strips) of each species. The dashed lines connect the equivalent portion of the CDS. The respective box below contains coding sequences of each species', being consensus and annotation (star) for the jaguar, and the site under selection at the jaguar consensus. The gray scale indicates the length of sequence (base pair). Scaffolds in A: human (NW_004929293.1), domestic cat (NC_018739.2), tiger (NW_006711962.1) and jaguar (scaffold 563). Scaffolds in B: human (NW_004929383.1), domestic cat (NC_018729.2), tiger (NW_006712031.1) and jaguar (scaffold 172). Scaffolds in C: human (NW_004929430.1), domestic cat (NC_018729.2), tiger (NW_006711815.1) and jaguar (scaffold 941). Scaffolds in D: human (NW_004929317.1), domestic cat (NC_018726.2), tiger (NW_006712726.1) and jaguar (scaffold 623). Scaffolds in E: human (NW_004929376.1), domestic cat (NC_018733.2), tiger (NW_006712591.1) and jaguar (scaffolds 1129 and 866).

3.4. Assessment of genealogical discordance in the *Panthera*

To evaluate and compare the genome-wide differences in phylogenetic signal among the five *Panthera* species, we applied sliding window-based phylogenetic tests on a six-species genome alignment (jaguar, tiger, snow leopard, leopard, lion and domestic cat). We divided the whole genome matrix into non-overlapping 100kb windows and reconstructed a phylogenetic tree for each window. Maximum likelihood (ML) tree searching and bootstrap analysis (200 bootstrap replications) were performed for each window using the software RAxML[25] with a GTR + Γ substitution model. For each window, we used the ML tree and sequence data as input for the program MCMCTree in the software package PAML4[26] to estimate divergence time variation across the genomes of the five *Panthera* species. We employed two soft constraints (1. divergence of Pantherinae and Felinae lineages between 9 and 15 Mya); 2. base of *Panthera* no earlier than 7 Mya[27]. These analyses were performed assuming autocorrelated rates between branches of the tree.

To provide a more accurate estimate of *Panthera* species divergence time that avoids the confounding effects of post-speciation interspecies introgression, we compiled a submatrix of whole-genome sliding windows that conformed to the species tree ((tiger, snow leopard), (jaguar, (lion, leopard))) with strong statistical support (determined using the Approximately Unbiased test). We used this matrix to re-estimate the divergence times for all nodes in genus *Panthera* using MCMCTree.

4. In-depth analyses of gene families

4.1. Analysis of olfactory receptor genes in the *Panthera*

Assembled contigs for each species were scanned for putative olfactory receptor (OR) genes using the 'ORA' Bioperl package[28,29]. ORA utilizes profile hidden Markov models (HMMs), designed using a dataset of mammalian OR genes, to scan a target sequence and identify all possible ORs it contains. Tblastx[30] was also used to locate any ORs that ORA may have failed to locate. Sequences with in-frame stop codons, or with length less than 650bp[28] were considered pseudogenes. ORs that contained unresolved regions designated by one or more 'N' positions were classified as 'unknown'; as their functionality, or lack thereof, could not be determined. Leading or tailing runs of 'N' nucleotides were trimmed while retaining the correct reading frame.

Using blastn[30], the OR repertoire from each individual species was compared against the other 4. This was done to count the number of orthologous OR genes showing a conserved function/ loss of function. Blastn was also used to identify species-specific gene loss events.

To identify species-specific gene duplications, and the subsequent fate of post–duplication ORs, phylogenetic trees were generated for each OR subfamily (13 trees; OR1/3/7, OR2/13, OR4, OR5/8/9, OR6, OR10, OR11, OR12, OR14, OR51, OR52, OR55, OR56;[31]) using amino acid alignments generated by ClustalO[32] and RAxML[25]. The model of protein sequence evolution that best fitted the data was determined using ProtTest3[33]. Each tree was split into all of its possible sub-trees. Sub-trees consisting entirely of OR genes from one species were considered to represent gene duplication events.

4.2 Olfactory receptors results

A total of 5,108 OR genes were found across all 5 species (*P. tigris*: 1,057 ORs, *P. leo*: 1,020 ORs, *P. onca*: 975 ORs, *P. uncia*: 1,026 ORs and *P. pardus*: 1,030 ORs). Due to OR genes with unknown bases, the total number of functional and non–functional could not be fully determined.

An average of 151 ORs were non–functional in all species, with an average of 99 showing conserved stop codons, which indicate loss of functional in the most recent common ancestor. The average number of genes that were functional in all species was 441 ORs. As the loss of function in certain ORs could not be determined ('unknown' ORs), the number of ORs with conserved functionality or loss of function remained conservative. The total number of species-

specific loss of function events were: *P. tigris*: 12 ORs, *P. leo*: 7 ORs, *P. onca*: 11 ORs, *P. uncia:* 20 ORs and *P. pardus*: 13 ORs.

The best model of sequence evolution, for OR families 4 and 6, was determined as JTT+I+ · +F, and JTT+ · +F for the remaining families. RAxML generated trees for each family, and it was determined that there were 5 instances of species specific duplication in *P. tigris*, 1 in *P. leo*, 2 in *P. onca*, 2 in *P. uncia* with no species specific duplications observed in *P. pardus*. This gave a total of 10 duplications, giving rise to 20 OR genes. Of these, 9 have lost their function, only one of the post-duplication genes showed evidence of positive selection for Family 2/13 in *P. onca*.

5. Demographic history analyses

5.1. PSMC

We applied the Pairwise Sequentially Markovian Coalescent analysis (PSMC)[34] to estimate the demographic history of each *Panthera* species. To call diploid sequences, we generated *de novo* assemblies for the lion, leopard and snow leopard using SOAPdenovo2[35] with k-mer set to 31. All quality-trimmed Illumina sequences of each *Panthera* species were mapped to their own *de novo* genome assembly using BWA with default parameter settings. SAMTOOLS was used to estimate average mapping coverage, and to call and filter nucleotide variants. Genome regions with less than half or more than twice the average whole genome mapping depth were excluded from the final diploid sequences. We applied a mutation rate of $1x10^{-8}$ and generation time=5 for all five *Panthera* species. We evaluated the consistency of the PSMC tests by performing 100 bootstrap replicates.

5.2. Whole genome introgression (ABBA/BABA) tests

We applied the ABBA/BABA approach of Green *et al.*[36] to evaluate the imbalanced frequency of alleles present within alternative tree topologies. All trimmed Illumina sequences were mapped to the repeat-masked genome reference of the domestic cat (felCat5) using BWA. The software package ANGSD was used to perform admixture tests, calculate the statistical significance (both *D* and *Z* statistics) and perform the weighted block jackknife tests. We enforced a minimum mapping quality score=40 and a block size of 5Mb of physical genome distance.

6. Detection of genes under positive selection in the *Panthera* species

6.1. PAML analysis

Selection analyses were performed based on the CDS information retrieved from each genome and previously aligned (see section 3.3). Different (nested) models, allowing for neutral evolution and negative selection, or additionally allowing for positive selection, were assessd on the basis of their log likelihoods. For every pair of comparable models, a likelihood ratio test (LRT) was employed to assess the best-fit scenario for a particular gene. To find genes under positive selection, we used two pairs of models available in the software codeml[22]. As input files, we used the CDS alignment and the phylogeny constructed for respective complete gene. In addition, we performed the same analysis considering the species tree, to explore a possible disagreement in the results due to genealogical discordance.

First, we looked for selection at each site of the alignment, through a site-model analysis. For this analysis, we considered normal and beta distributions. Second, we specifically looked for selection on each branch of the phylogeny, but also considering the sites, through a branch-site model analysis. This latter targets episodes of positive selection on each species-specific lineage. The ctl file was constructed with the parameters listed in Supplementary Table S16. For both models, we calculated the LRT as follows: $\Delta \ln \times 2$, being *ln* the log of likelihood values generated for alternative and null hypotheses.

Supplementary Table S14: **Parameters used in ctl file by codeml software**.

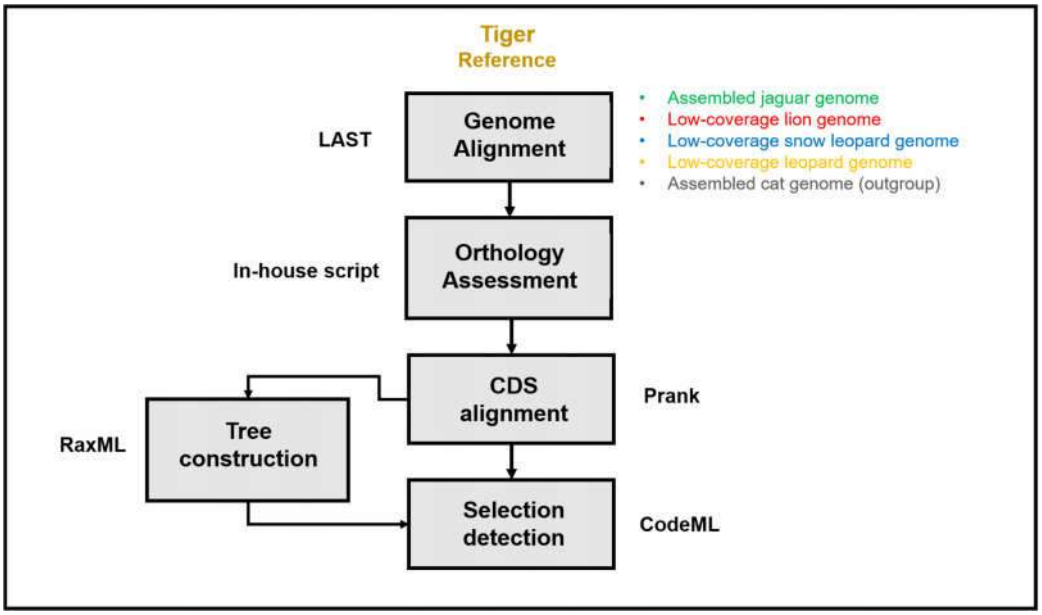
Parameters*	Site-model		Branch-site model	
	Null hypothesis	Alternative hypothesis	Null hypothesis	Alternative hypothesis
noisy	9	9	9	9
verbose	1	1	1	1
runmode	0	0	0	0
CodonFreq	2	2	2	2
clock	0	0	0	0
model	**0**	**0**	**2**	**2**
Nssites	**1,7****	**2,8****	**2**	**2**
icode	0	0	0	0
Mgene	0	0	0	0
fix_kappa	0	0	0	0
kappa	2	2	2	2
fix_omega	**0**	**0**	**1**	**0**
omega	**0.4**	**0.4**	**1**	**0.4**
fix_alpha	1	1	1	1

alpha	0	0	0	0
Malpha	0	0	0	0
ncatG	3	3	3	3
getSE	0	0	0	0
RateAncestor	1	1	1	1
Small_Diff	5.00E-06	5.00E-06	5.00E-06	5.00E-06
cleandata	0	0	0	0
method	0	0	0	0

* As called in ctl file.

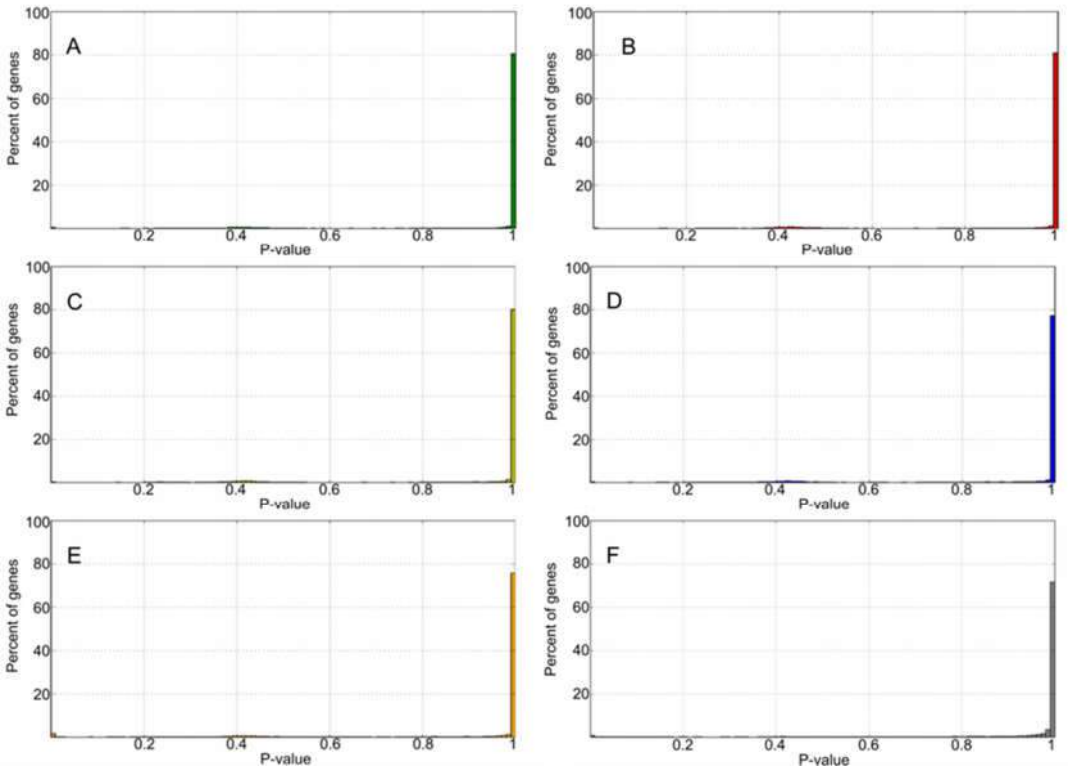
** 1 and 2 for normal distribution; 7 and 8 for beta distribution.

To construct the phylogeny for each gene in our dataset, was used the RaxML software[25]. As cited above, we have also added another layer of analysis by using the species tree, as most studies do. The domestic cat sequences was set as the outgroup; we used the nucleotide substitution model, gamma correction for the rate heterogeneity among sites (GTRGAMMA), and 23,453 as the number of seeds for the parsimony inferences.

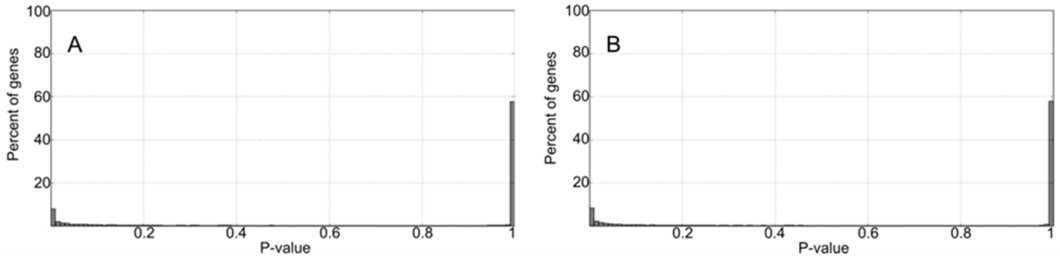


Supplementary figure S3: Selection pipeline showing the analysis used for the position based orthlogy assessment, tree construction and selection detection.

To decide whether we would use FDR (false discovery rate) correction for multiple comparisons, we assessed the p-value distribution for each dataset. When we observed a cluster of loci with p-values near the significance threshold (0.05), we employed the multiple-test correction (Supplementary Figures S4, S5).



Supplementary figure S4: **P-value distribution for the branch-site analysis.** P-value concentrated close to the threshold can indicate the use of FDR correction. (A) jaguar; (B) lion; (C) leopard; (D) snow leopard; (E) tiger; (F) domestic cat.

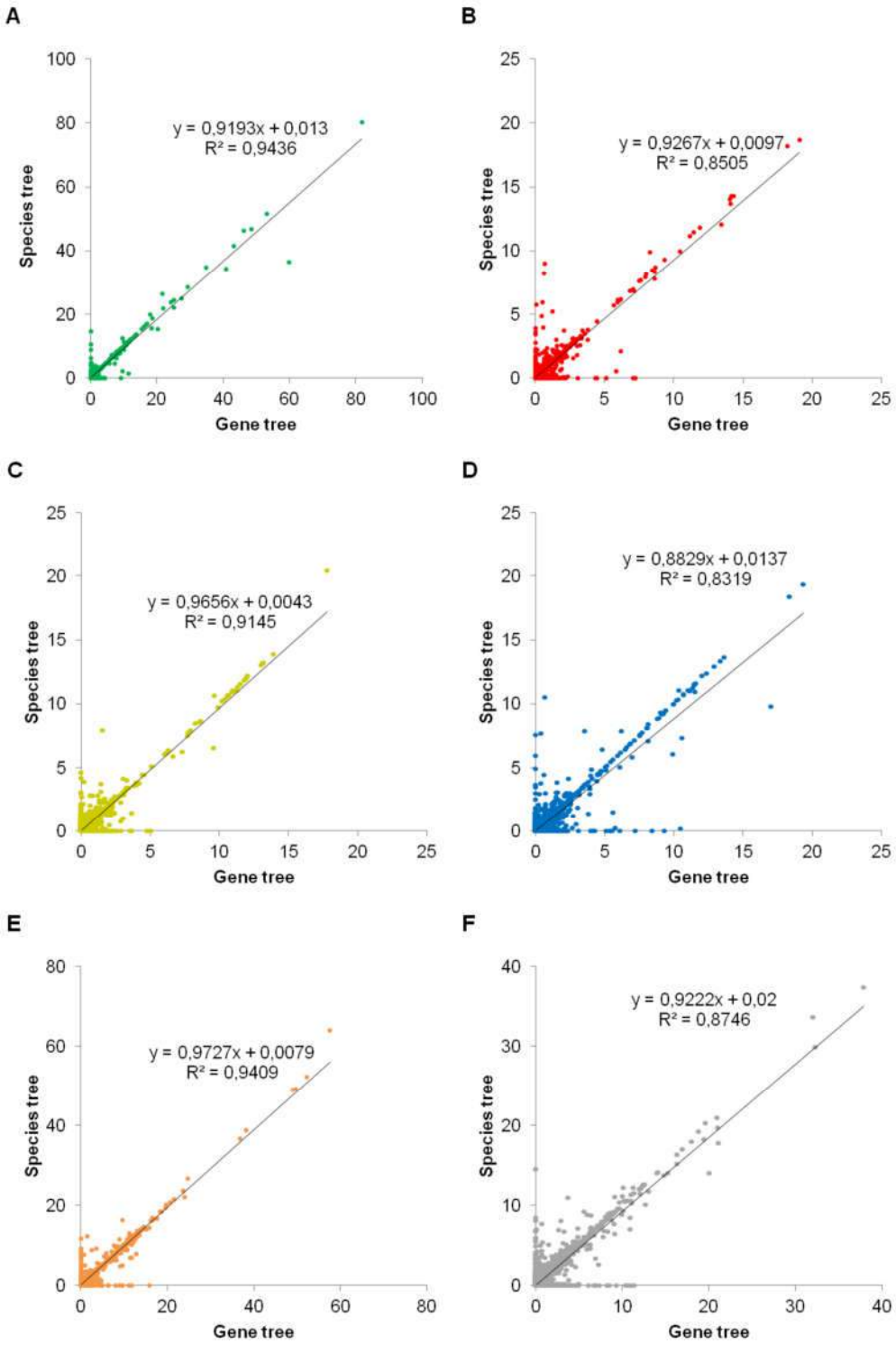


Supplementary figure S5: **P-value distribution for the site model analyses**. P-value concentrated close to the threshold can indicate the use of FDR correction. (A) Site model test using a normal distribution; (B) Site model test using a beta distribution.

6.2. Assessment of the impact of Gene Tree *vs.* Species Tree discordance on the selection analyses

To compare the selection results obtained by the gene tree *vs.* the species tree approaches, we plotted the likelihood values from each analysis (Supplentary Figure S6). This enabled us to asses which genes were identified with each method. Since diferent regions of the genome may have diferent trees, as was strongly demonstrated in this study for the *Panthera*, we

only kept for downstream analyses genes that exhibited significant signatures of positive selection when their locus-specific tree was employed.



Supplementary figure S6: **Likelihood ratio tests comparison for the branch-site analysis**. Graphs depicting the correlation between the calculated values (2*delta) for the branch-site likelihood ratio tests (LRTs) performed with the species tree and the equivalent values estimated with the gene tree (reconstructed for each locus – see text for details). (A) jaguar; (B) lion; (C) leopard; (D) snow leopard; (E) tiger; (F) domestic cat.

6.3. Gene enrichment analysis

We performed a gene enrichment analysis for the genes under selection on each method using WebGestalt[37]. We have used the Human genome as reference for the analysis. The following analyses were performed: Gene ontology[38,39], Pathway Commons[40], KEGG Pathway[41,42], Disease (http://bioinfo.vanderbilt.edu/glad4u) and Phenotypes[43]. All categories had a 0.05 significance threshold and a multiple testing correction (Benjamini–Hochberg) was applied to control for false discovery.

6.4 Branch-site test results

6.4.1. Cranial and forelimb development in jaguars

Craniofacial development

Gene ontology analyses (Table S16) associated two genes, *ESRP1* and *SSTR4*, with acromegaly. Acromegaly is a disease that causes important developmental alterations, with the most apparent ones related to growth and body size proportions [44]. As in humans, acromegaly in cats provokes striking alterations to the musculoskeletal system, such as enlargement of the body, which can include the paws, chin and skull [45]. In model species, both genes are described as associated with important developmental pathways, especially *ESRP1*, which is a splicing regulator acting on several genes during craniofacial and forelimb development. In our selection analysis, the amino acid change in *ESRP1* is located in the RRM domain, which is known to interact with important developmental genes[46]. This gene is a key part of the epithelial-mesenchymal transition (EMT), which is relevant not only for development, but also for wound healing and present in the cancer metastasis mechanism[47].

One of the most important genes affected by *ESRP1* is the *FGFR2IIIb* isoform, responsible for several metabolic routes. It acts during craniofacial development and limb formation, and for this reason is one of the key genes in several lethal syndromes. Mutations in this gene are associated with Crouzon syndrome, Pfeiffer syndrome, Craniosynostosis, Apert syndrome, Jackson-Weiss syndrome, Beare-Stevenson *cutis gyrata* syndrome, Saethre-Chotzen syndrome,

and syndromic craniosynostosis[48–50]. The *SSTR4* gene is known to be active during embryo development but absent in the human adult pituitary [51]. Some studies associate this gene with head and forebrain development and memory[52]. However, there is little information when compared to the other four somatostatin (*SSTR1-3*, *5*) receptors.

Although it is not possible to assess the real impact of these genes in jaguars, based on its distinct phenotype from other Panthera and the important role each gene has during development, it is likely that these mutations have an impact in their morphology and development. The jaguar has been described as a more robust, stocky and strong version of the leopard[53–55], and it is the third largest species in the genus. The species has a highly diverse diet, greatly associated with water bodies. Throughout its distribution, it is known to feed on turtles [56], caiman and capybara[57]. The diet is probably one of the most important drivers for the robust characteristics of the species, with some authors having addressed the diet's importance to help shape cranial features[58,59]. Christensen (2008)[60] described the jaguar skull in comparison with the other Panthera, reporting with significant differences related to the zygomatic arch, palate and other cranial features. All these characteristics are known to be influenced by *ESRP1* and *SSTR4* and affected by acromegaly in domestic cats.

Other genes in our analysis are also related with craniofacial development, but did not show any clear-cut relationship in the gene ontology analysis. This is the case of *SOX10*, which is involved in jaw osteogenesis, neural crest olfactory system development and hearing. Another gene involved in this process is *LHX4*, involved in pituitary development, body weight and length[61].

Limb Development and Muscular strength

Limb development and muscular strength are two important factors for predatory behaviour. Amongst the Panthera, jaguars are described as having shorter limbs and a more robust look[54]. The gene ontology analysis we performed retrieved some genes related to muscular strength, more specifically hindlimb muscular strength. However, we have strong indication that those genes act in other regions of the body and pathways. Amidst these genes, we have found TPM3 with the strongest selection signal; together with *FGD4* and *SOX10*, this locus is related to distal muscle weakness and related myopathies. *TPM3* is also known to be involved in another disease, called nemaline myopathy, which includes muscular weakness in the face, neck and limbs. Other genes related to the detected phenotypes are *PCDH15*, which is also involved in hearing, and *LIAS*. The latter has a homozygous mutation known for causing muscle hypotonia and epilepsy[62].

The phenotypes and diseases observed in the gene ontology analyses and literature searches could indicate the influence of these genes in the characteristic phenotype of the jaguar. In particular, when the overall set of genes found to be under selection in the jaguar is compared to members of pathways associated with craniofacial development, a striking pattern emerges, suggestinf a strong adaptive process influencing the morphological evolution of this species.

Supplementary Table S15: **Branch-site for the jaguar.** Candidate genes under positive selection, p-values for the gene tree and species tree analysis (significance threshold < 0.05) and gene description.

Gene	P-value		Description
	Gene tree	Species tree	
ABCC10	0,995486507	0,012680795	Multidrug resistance-associated protein 7
ABTB2	4,37E-05	4,37E-05	Ankyrin repeat and BTB/POZ domain-containing protein 2
AGER	0,652439781	1,09E-06	Advanced glycosylation end product-specific receptor isoform X2
AKAP13	3,64E-05	3,64E-05	A-kinase anchor protein 13-like
ARNTL	3,72E-02	1,00E+00	Aryl hydrocarbon receptor nuclear translocator-like protein 1 isoform X1
BAHD1	6,59E-04	7,22E-04	Bromo adjacent homology domain-containing 1 protein
BIRC2	8,06E-04	7,92E-04	Baculoviral IAP repeat-containing protein 2
C4BPA	0,02529746	0,034042663	C4b-binding protein alpha chain-like isoform X2
CAPG	6,19E-04	6,19E-04	Macrophage-capping protein isoform X1
CASP8AP2	0,002011208	0,002011235	CASP8-associated protein 2
CNTNAP1	1,59E-05	1,47E-05	Contactin-associated protein 1
DCDC1	0,001353214	0,001353214	Doublecortin domain-containing protein 1
ESRP1	5,31E-07	7,16E-07	Epithelial splicing regulatory protein 1 isoform X1
FCRLB	0,003364714	0,005495659	Fc receptor-like B
FGD4	0,039307091	1	FYVE, rhogef and PH domain-containing protein 4 isoform X1
FOXJ2	0,000493491	0,00049352	Forkhead box protein J2
FSIP1	0,01044599	0,01044599	Fibrous sheath-interacting protein 1
GFI1	0,000314979	0,000315011	Zinc finger protein Gfi-1
GYG1	0,002085254	0,157790319	Glycogenin-1
HJURP	0,011029724	0,011632988	Holliday junction recognition protein
IPCEF1	1	0,031969383	Interactor protein for cytohesin exchange factors 1
KDM3B	0,00020563	0,000226917	Lysine-specific demethylase 3B
KIFC2	0,002279107	0,002279433	Kinesin-like protein KIFC2
KLHL36	0,017964866	0,017964558	Kelch-like protein 36
LHX4	0,035704135	0,035704135	LIM/homeobox protein Lhx4
LIAS	0,0184937	0,032971277	Lipoyl synthase, mitochondrial isoform X1
LOC102948654	2,38E-05	8,36E-06	KH domain-containing protein 3-like
LOC102948925	1,79E-05	7,68E-05	Olfactory receptor 10H1-like
LOC102950007	5,41E-07	2,49E-06	Protein FAM47A-like
LOC102950047	0,003076468	0,003485138	Olfactory receptor 13F1-like
LOC102950787	0,001141224	0,000969606	Keratin-associated protein 4-12-like isoform X3
LOC102952669	1,05E-11	1,05E-11	Repetin-like
LOC102953718	0,031290493	0,031290493	Uncharacterized protein c2orf16 homolog
LOC102953722	0,000239219	0,000241142	Uncharacterized protein LOC102953722

LOC102954036	2,72E-02	2,91E-02	Uncharacterized protein ENSP00000372125-like
LOC102954294	2,04E-03	4,14E-04	Olfactory receptor 1094-like
LOC102954353	0,013534217	0,013534063	Protocadherin alpha-3-like
LOC102955222	7,04E-06	8,59E-05	Uncharacterized protein LOC102955222 isoform X1
LOC102955918	0,991187256	0,034604596	Mucin-2-like
LOC102956118	3,29E-06	2,92E-07	Mucin-16-like
LOC102956679	0,000866456	0,000866455	Olfactory receptor 2B6-like
LOC102957355	3,52E-12	7,62E-12	Olfactory receptor 3A1-like
LOC102957369	3,22E-13	6,78E-13	Uncharacterized protein c1orf167-like
LOC102958958	0,041026354	0,041026306	Uncharacterized F-box/LRR-repeat protein C02F5.7-like isoform X2
LOC102959599	0,998871621	0,000130176	Leukocyte immunoglobulin-like receptor subfamily B member 4-like
LOC102962066	0,005316035	0,01299562	Olfactory receptor 2W1-like
LOC102962664	0,019397835	0,019377421	Nuclear RNA export factor 2-like
LOC102963141	0,001351668	0,001613303	Putative olfactory receptor 10D4-like
LOC102963291	5,30E-04	5,30E-04	Melanoma-associated antigen B4-like isoform X1
LOC102964920	0,000798387	0,000798578	Uncharacterized protein LOC102964920
LOC102965607	0,001040346	0,001350757	Cyclic AMP-dependent transcription factor ATF-5-like
LOC102966661	0,008038109	0,007419355	Olfactory receptor 4K13-like
LOC102968288	0,010116048	0,010348308	Uncharacterized protein LOC102968288
LOC102968863	3,52E-09	3,52E-09	60S ribosomal protein L35-like
LOC102968979	1	0,001159854	Uncharacterized protein LOC102968979
LOC102970804	0,007081901	0,033929997	Serum amyloid A protein-like
LOC102971194	1,65E-07	5,95E-07	Zinc finger protein 135-like
LOC102971284	0,002412155	0,002374378	DNA dc->du-editing enzyme APOBEC3-like
LOC102971659	0,038527383	0,039732947	Lymphocyte antigen 6H-like
LOC102972647	1,44E-03	2,88E-03	Olfactory receptor 9K2-like
LY9	0,558793216	0,036106154	T-lymphocyte surface antigen Ly-9 isoform X1
MAB21L1	1	0,036143359	Protein mab-21-like 1
MRPS5	0,029661194	0,029924819	28S ribosomal protein S5, mitochondrial
MSH4	0,0173564	0,0173564	Muts protein homolog 4
NFE2L3	0,014743074	0,014740465	Nuclear factor erythroid 2-related factor 3
OVCA2	0,009397369	0,009397337	Ovarian cancer-associated gene 2 protein homolog
PCDH15	0,014090319	0,014090319	Protocadherin-15 isoform X1
PDE3A	4,82E-03	6,22E-03	Cgmp-inhibited 3',5'-cyclic phosphodiesterase A
PDS5B	3,53E-02	3,53E-02	Sister chromatid cohesion protein PDS5 homolog B isoform X2
PLA2G4B	0,005457127	0,004430269	Jmjc domain-containing protein 7-like isoform X3
PPP1R15A	0,000179405	0,000179405	Protein phosphatase 1 regulatory subunit 15A isoform X7
PRPF40A	0,036130473	0,994136823	Pre-mrna-processing factor 40 homolog A isoform X4
PRR23A	1,57E-03	7,81E-04	Proline-rich protein 23A
PRSS55	0,006663644	0,006663644	Serine protease 55 isoform X2
PTPN6	1	0,029032454	Tyrosine-protein phosphatase non-receptor type 6
QSOX1	4,49E-03	4,06E-03	Sulfhydryl oxidase 1
RUNDC3A	0,007386977	0,007384038	RUN domain-containing protein 3A
SAPCD2	0,211406246	0,049423751	Suppressor APC domain-containing protein 2
SCUBE2	4,10E-02	1,00E+00	Signal peptide, CUB and EGF-like domain-containing protein 2
SEL1L	0,042979151	1	Protein sel-1 homolog 1

SENP5	0,214833466	0,023342836	Sentrin-specific protease 5 isoform X3
SLC26A3	0,000677633	0,236785196	Chloride anion exchanger
SLC8A3	4,12E-09	4,12E-09	Sodium/calcium exchanger 3 isoform X1
SMARCC2	0,011474809	0,007094612	SWI/SNF complex subunit SMARCC2
SOX10	0,004722239	0,004722239	Transcription factor SOX-10
SSTR4	6,59E-08	8,89E-08	Somatostatin receptor type 4
SYTL2	0,002478955	1	Synaptotagmin-like protein 2 isoform X3
TMX2	0,005888549	0,005888549	Thioredoxin-related transmembrane protein 2
TPM3	8,75E-05	8,75E-05	Tropomyosin alpha-3 chain isoform X6
TRAPPC8	0,002942641	0,005327479	Trafficking protein particle complex subunit 8 isoform X1
TRIP13	0,047870741	0,047460497	Pachytene checkpoint protein 2 homolog
ZFP3	2,98E-06	2,98E-06	Zinc finger protein 3 homolog
ZNF418	0,003538984	0,003538097	Zinc finger protein 418-like
ZNF614	0,048130481	1	Zinc finger protein 614
ZNF664	1	0,002865593	Zinc finger protein 664

Supplementary Table S16: **Gene ontology results for the jaguar.** Analysis refers to the analysis performed in WebGestalt with a significance threshold of < 0.05 and a correction for multiple comparisons. Characteristic refers to the enriched category found. Parameters describes enrichment details. C (Genes presented in the category); O (Genes observed in the dataset); R (Ratio of enrichment); rawP (p-value for the hypergeometric test); adjP (adjusted P value by the multiple correction test).

Phenotype	Parameters	Genes
Distal lower limb muscle weakness	C=7;O=2;E=0.02;R=81.56;rawP=0.0002;adjP=0.0216	*FGD4; TPM3*
Distal lower limb amyotrophy	C=8;O=2;E=0.03;R=71.36;rawP=0.0003;adjP=0.0216	*FGD4; TPM3*
Lower limb amyotrophy	C=10;O=2;E=0.04;R=57.09;rawP=0.0005;adjP=0.0240	*FGD4; TPM3*
Distal muscle weakness	C=68;O=3;E=0.24;R=12.59;rawP=0.0013;adjP=0.0312	*FGD4; TPM3; SOX10*
Distal amyotrophy	C=66;O=3;E=0.23;R=12.97;rawP=0.0012;adjP=0.0312	*FGD4; TPM3; SOX10*
Motor delay	C=157;O=4;E=0.55;R=7.27;rawP=0.0013;adjP=0.0312	*FGD4; TPM3; LIAS; PCDH15*

PheWas	Parameters	Genes
Erectile dysfunction	C=37;O=2;E=0.11;R=18.79;rawP=0.0042;adjP=0.0336	*PDE3A; PDS5B*
Insomnia	C=32;O=2;E=0.09;R=21.72;rawP=0.0032;adjP=0.0336	*SLC8A3; PDE3A*
Pruritus and related conditions	C=46;O=2;E=0.13;R=15.11;rawP=0.0065;adjP=0.0347	*SLC8A3; PDE3A*
Psychogenic disorder	C=57;O=2;E=0.16;R=12.20;rawP=0.0099;adjP=0.0352	*SLC8A3; PDE3A*
Acid-base balance disorder	C=60;O=2;E=0.17;R=11.59;rawP=0.0110;adjP=0.0352	*SLC8A3; PDE3A*

KEGG Pathway	Parameters	Genes
Pancreatic secretion	C=101;O=2;E=0.12;R=16.42;rawP=0.0067;adjP=0.0280	*SLC26A3; PLA2G4B*
Toxoplasmosis	C=132;O=2;E=0.16;R=12.57;rawP=0.0112;adjP=0.0280	*PLA2G4B; BIRC2*
Protein processing in endoplasmic reticulum	C=165;O=2;E=0.20;R=10.05;rawP=0.0170;adjP=0.0283	*PPP1R15A; SEL1L*

Olfactory transduction	C=388;O=7;E=0.64;R=10.96;rawP=3.66e-06;adjP=2.20e-05	*LOC102948925; LOC102957355; LOC102950047; LOC102962066; LOC102956679; LOC102966661; LOC102972647*

Diseases	Parameters	*Genes*
Accelerated phase chronic myeloid leukemia	C=16;O=2;E=0.02;R=103.67;rawP=0.0002;adjP=0.0063	*GFI1; CASP8AP2*
Acromegaly	**C=20;O=2;E=0.02;R=82.94;rawP=0.0003;adjP=0.0063**	***SSTR4; ESRP1***
Hyperpituitarism	C=26;O=2;E=0.03;R=63.80;rawP=0.0005;adjP=0.0070	*SSTR4; ESRP1*
Adenoma	C=157;O=3;E=0.19;R=15.85;rawP=0.0009;adjP=0.0095	*SLC26A3; SSTR4; ESRP1*
Blast Crisis	C=46;O=2;E=0.06;R=36.06;rawP=0.0014;adjP=0.0118	*AKAP13; GFI1*
Brenner tumour of ovary	C=222;O=3;E=0.27;R=11.21;rawP=0.0025;adjP=0.0175	*CNTNAP1; C4BPA; FOXJ2*
Muscle Hypotonia	**C=86;O=2;E=0.10;R=19.29;rawP=0.0049;adjP=0.0257**	***LIAS; TPM3***
Pituitary Diseases	C=81;O=2;E=0.10;R=20.48;rawP=0.0043;adjP=0.0257	*SSTR4; LHX4*
Congenital Abnormalities	C=643;O=4;E=0.78;R=5.16;rawP=0.0075;adjP=0.0350	*FGD4; SOX10; LHX4; PCDH15*
Puberty, Delayed	C=119;O=2;E=0.14;R=13.94;rawP=0.0091;adjP=0.0382	*SSTR4; LHX4*
Endocrine System Diseases	C=429;O=3;E=0.52;R=5.80;rawP=0.0151;adjP=0.0407	*SSTR4; LHX4; SEL1L*
Hearing Loss, Sensorineural	C=140;O=2;E=0.17;R=11.85;rawP=0.0125;adjP=0.0407	*SOX10; PCDH15*
Peripheral Nervous System Diseases	C=157;O=2;E=0.19;R=10.57;rawP=0.0155;adjP=0.0407	*FGD4; SOX10*
Endocrine disorder NOS	C=429;O=3;E=0.52;R=5.80;rawP=0.0151;adjP=0.0407	*SSTR4; LHX4; SEL1L*
Pancreatic Diseases	C=147;O=2;E=0.18;R=11.28;rawP=0.0137;adjP=0.0407	*SSTR4; SEL1L*
Endocrine disturbance NOS	C=429;O=3;E=0.52;R=5.80;rawP=0.0151;adjP=0.0407	*SSTR4; LHX4; SEL1L*
Infertility, Male	C=188;O=2;E=0.23;R=8.82;rawP=0.0217;adjP=0.0449	*SLC26A3; MSH4*
Deafness	**C=206;O=2;E=0.25;R=8.05;rawP=0.0257;adjP=0.0449**	***SOX10; PCDH15***
Demyelinating Diseases	C=178;O=2;E=0.21;R=9.32;rawP=0.0196;adjP=0.0449	*FGD4; SOX10*
Neuroendocrine Tumors	C=210;O=2;E=0.25;R=7.90;rawP=0.0267;adjP=0.0449	*SOX10; SSTR4*
Leukemia, Myeloid, Acute	C=208;O=2;E=0.25;R=7.97;rawP=0.0262;adjP=0.0449	*KDM3B; GFI1*
Pancreatic Neoplasms	C=197;O=2;E=0.24;R=8.42;rawP=0.0237;adjP=0.0449	*SSTR4; SEL1L*
Infertility	C=209;O=2;E=0.25;R=7.94;rawP=0.0264;adjP=0.0449	*SLC26A3; MSH4*
Muscle Weakness	**C=174;O=2;E=0.21;R=9.53;rawP=0.0188;adjP=0.0449**	***FGD4; TPM3***
Brain Neoplasms	C=198;O=2;E=0.24;R=8.38;rawP=0.0239;adjP=0.0449	*SOX10; OVCA12*
Nervous System Malformations	C=221;O=2;E=0.27;R=7.51;rawP=0.0293;adjP=0.0473	*FGD4; LHX4*

Pathway Commons	Parameters	Genes
Olfactory Signaling Pathway	C=370;O=6;E=0.61;R=9.85;rawP=3.43e-05;adjP=0.0010	*LOC102948925; LOC102957355; LOC102962066; LOC102956679; LOC102966661; LOC102972647*
GPCR downstream signaling	C=553;O=7;E=0.91;R=7.69;rawP=3.59e-05;adjP=0.0010	*LOC102948925; LOC102957355; LOC102962066; LOC102956679;*

		LOC102966661; LOC102972647; PDE3A
Signaling by GPCR	C=823;O=8;E=1.35;R=5.90;rawP=6.27e-05;adjP=0.0012	*LOC102948925; LOC102957355; LOC102972647; LOC102956679; LOC102962066; LOC102966661; PDE3A; SSTR4*
Signal Transduction	C=1231;O=8;E=2.03;R=3.95;rawP=0.0009;adjP=0.0126	*LOC102948925; LOC102957355; LOC102972647; LOC102956679; LOC102962066; LOC102966661; PDE3A; SSTR4*
TNF alpha/NF-kB	C=171;O=3;E=0.28;R=10.66;rawP=0.0029;adjP=0.0325	*SMARCC2*; *BIRC2*; *CASP8AP2*
Endogenous TLR Signaling	C=57;O=2;E=0.09;R=21.31;rawP=0.0040;adjP=0.0336	*LOC102970804; BIRC2*
CD40/CD40L Signaling	C=58;O=2;E=0.10;R=20.95;rawP=0.0042;adjP=0.0336	*C4BPA*; *BIRC2*
Transport of Inorganic Cations/anions and Amino Acids/oligopeptides	C=94;O=2;E=0.15;R=12.92;rawP=0.0106;adjP=0.0742	*SLC26A3*; *SLC8A3*
mTOR Signaling Pathway	C=1288;O=4;E=2.12;R=1.89;rawP=0.1625;adjP=0.1931	*SMARCC2*; *GFI1*; *PPP1R15A*; *BIRC2*
Alpha9 Beta1 Integrin Signaling Events	C=1305;O=4;E=2.15;R=1.86;rawP=0.1679;adjP=0.1931	*SMARCC2; GFI1*; *PPP1R15A*; *BIRC2*

6.4.2. Social behaviour in Lions

Lions are known to be the only species in the Felidae to present a highly structured social behaviour [63]. They have a complex network of relationship among and within groups[64]. Their interaction is mediated by a set of different behaviours, which include roaring, scent marking and direct aggression towards member of other groups. They are also described as having a high number of different facial expressions[65]. In the long term, these behaviours aid lions in their adaptation to their environments and life history. Recent studies showed that group cohesiveness and size are strongly related with reproductive success for females[64]. As of now, there are no studies that address such behaviour from a genetic perspective.

We have found several genes under selection that are related with different forms of behaviour, some of them directly related to psychological disorders in model species.

GRB10 is an imprinted gene in humans with different roles in different tissues and during development as well. Like other imprinted genes, *GRB10* has its differential expression tied to the allele that is being used. In humans and mice, the *GRB10* allele expressed in the placenta is of maternal origin, and is involved in growth and body size at birth [66]. The paternal allele is expressed in the brain, and recent studies show its relationship with social dominance behaviour[67].

Another gene that presented signatures of selection was *LHX6*. This gene plays an important role in the cell differentiation of medial ganglionic eminence and cell migration to the

the cortex[68,69]. Also is involved in odontogenesis together with *LHX7*[70], and palate formation with *LHX8*[71]. It is a developmental gene that acts in the forebrain and is related with innate reproductive behaviour, more specifically activated by reproductive olfactory stimuli[72].

The *EFHC2* gene is expressed in the amygdala and has been associated with fear recognition in Turner syndrome. According to Weiss *et al* (2007)[73], it is possible that *EFHC2* has a diverse neuronal function related to normal social cognitive competence.

In our genome-wide selection scan, we have found another imprinted gene, *UBE3A*. This gene is associated with a well-known syndrome, Angelman's syndrome, one of the key examples of imprinted gene diseases. As is the case of *GRB10*, it can affect growth and behaviour depending on the tissue and parental allele expressed.

Other genes under selection include: *ARNTL*, which is involved in the seasonal affective disorder[74], *PTK2B* associated with Alzheimer's disease[75] and sleep duration in certain populations[76], and *PEX13*, a gene that acts in the peroxisomes and whose mutations can cause diseases such as Zellweger syndrome[77].

This is the first time that the genes potentially related to social behaviour in lions are described. Further studies are needed to build a more complete scenario. Such an avenue of research can aid in future actions regarding *in-situ* and especially *ex-situ* conservation programmes, focusing on specific characteristics presented by lions regarding their social behaviour.

Supplementary Table S17: **Branch-site for the lion**. Candidate genes under positive selection, p-values for the gene tree and species tree analysis (significance threshold < 0.05) and gene description.

Gene	P-value		Description
	Gene tree	Species tree	
AFM	0,003401644	0,005208019	Afamin
ARNTL	0,037184051	1	Aryl hydrocarbon receptor nuclear translocator-like protein 1 isoform X1
ART3	1,96E-05	1,00E+00	Ecto-ADP-ribosyltransferase 3 isoform X2
CROCC	3,30E-03	3,22E-03	Rootletin
DBF4B	0,023744048	0,987486183	Protein DBF4 homolog B
DIS3L	0,003652386	0,003652386	DIS3-like exonuclease 1 isoform X1
DLG1	1	0,038423493	Disks large homolog 1 isoform X1
DPCR1	3,33E-03	0,003823866	Diffuse panbronchiolitis critical region protein 1
EFHC2	0,007948052	1	EF-hand domain-containing family member C2
ESRP1	0,013007217	0,151364069	Epithelial splicing regulatory protein 1 isoform X1
FLT4	0,509421115	0,042084763	Vascular endothelial growth factor receptor 3 isoform X1
GRAMD1B	0,006186105	0,005786768	GRAM domain-containing protein 1B isoform X4
GRB10	0,000154843	0,000154863	Growth factor receptor-bound protein 10
KCNC4	0,000178952	0,000178955	Potassium voltage-gated channel subfamily C member 4
KLRB1	0,015757309	0,46177866	Killer cell lectin-like receptor subfamily B member 1

KRT40	1,26E-05	1,55E-05	Keratin, type I cytoskeletal 40
KRT76	2,00E-05	2,00E-05	Keratin, type II cytoskeletal 2 oral isoform X2
LHX6	0,002296397	0,002296397	LIM/homeobox protein Lhx6 isoform X2
LOC102949218	0,000248934	0,000506201	Olfactory receptor 10H4-like
LOC102951325	3,36E-07	0,998871621	Complement C3-like
LOC102952612	0,001258053	0,0015839	Protocadherin beta-15-like
LOC102952669	0,01503389	0,01503389	Repetin-like
LOC102954617	0,004039725	0,001664406	Uncharacterized protein c7orf72 homolog isoform X1
LOC102955878	5,08E-01	0,027973855	Zinc finger protein 33B-like
LOC102958876	0,009055135	0,009055135	Protein monoglycylase TTLL8-like
LOC102960268	0,012920067	0,012920067	Olfactory receptor 4K14-like
LOC102961273	0,00798679	0,00837811	Mucin-2-like isoform X10
LOC102966647	0,42011697	0,002686881	Orexin-like
LOC102969904	0,430811043	0,004083925	Olfactory receptor 1L3-like
LOC102970180	0,005771271	0,005673381	Zinc finger and SCAN domain-containing protein 31-like
LOC102971471	0,457278448	0,047033715	Krueppel-like factor 15-like
MUC20	0,004903428	0,004903428	Mucin-20
NLRP14	0,014820599	0,013558196	NACHT, LRR and PYD domains-containing protein 14
PCDHB6	0,839967045	0,016259124	Protocadherin beta-6
PEX13	0,000577842	0,000577842	Peroxisomal membrane protein PEX13
PODXL	0,266261529	0,022053946	Podocalyxin
PTK2B	0,007223747	1	Protein-tyrosine kinase 2-beta isoform X1
RPL27A	1	0,048790749	
RPS6KL1	0,004772861	0,004352516	Ribosomal protein S6 kinase-like 1
SCUBE2	0,035530613	1	Signal peptide, CUB and EGF-like domain-containing protein 2
SLC24A1	0,007647888	0,008900162	Sodium/potassium/calcium exchanger 1
SLC37A3	0,000169878	0,000156196	Sugar phosphate exchanger 3
SPATA16	0,000836741	0,000836737	Spermatogenesis-associated protein 16 isoform X1
THADA	0,480970494	0,014961448	Thyroid adenoma-associated protein homolog
USP26	0,013507972	0,013534599	Ubiquitin carboxyl-terminal hydrolase 26
WISP1	0,017175278	0,017175278	WNT1-inducible-signaling pathway protein 1
YBX1	1	0,000268396	Nuclease-sensitive element-binding protein 1
YIPF2	0,000173238	0,000218707	Protein YIPF2
ZFP69B	0,000717504	0,000717504	Zinc finger protein ZFP69B
ZSCAN12	0,034837788	0,034837502	Zinc finger and SCAN domain-containing protein 12

Supplementary Table S18: **Gene ontology results for the lion**. Analysis refers to the analysis performed in WebGestalt with a significance threshold < 0.05 and multiple correct testing (FDR). Characteristic refers to the enriched category found. Parameters describes enrichment details. C (Genes presented in the category); O (Genes observed in the dataset); R (Ratio of enrichment); rawP (p-value for the hypergeometric test); adjP (adjusted P value by the multiple correction test).

Characteristic	Parameters	Gene
Oligospermia	C=98;O=2;E=0.07;R=30.35;rawP=0.0020;adjP=0.0120	*SPATA16; USP26*
Growth Disorders	C=183;O=2;E=0.12;R=16.25;rawP=0.0067;adjP=0.0131	*WISP1; GRB10*
Infertility	C=209;O=2;E=0.14;R=14.23;rawP=0.0087;adjP=0.0131	*SPATA16; USP26*
Infertility, Male	C=188;O=2;E=0.13;R=15.82;rawP=0.0071;adjP=0.0131	*SPATA16; USP26*

Male Urogenital Diseases	C=423;O=2;E=0.28;R=7.03;rawP=0.0327;adjP=0.0327	*SPATA16; USP26*
Autoimmune Diseases	C=414;O=2;E=0.28;R=7.18;rawP=0.0315;adjP=0.0327	*MUC20; KLRB1*

6.4.3. Craniofacial, limb and brain development in leopard

Craniofacial and limb development

Some authors believe that the *Panthera* ancestor was similar to the extant leopard[78]. Compared to the other species, the leopard is a medium-sized animal with some similarities with the jaguar regarding the color pattern[79]. The gene ontology analyses showed diverse pathways under selection, which do not correlate clearly with obvious characteristics of this species. Still, some patterns have emerged while searching for gene information in the literature and will be discussed here, but further analysis will be required.

LDB1 is an imprinted gene in humans [80], highly conserved and acting in different tissues during development[81]. This gene acts as a co-factor for important development genes, such as the LIM-homeodomain family. Knockout mice for *LDB1* have severe and lethal development alterations, such as posterior axis duplications, absence of heart formation and loss of foregut indentation[82]. Specifically in the brain, mutations in *LDB1* can alter the expression of LIM-homeodomain proteins, affecting the formation of Purkinje cells in the cerebellum. Like jaguars, leopards also present genes involved in craniofacial development under selection. Being involved in several developmental routes, *LDB1* is also involved in secondary palate formation. Almaidhan *et al.* (2014)[83], using mouse knockouts, discovered that mutants presented cleft secondary palate due to low expression of other genes (*WNT5a*, *OSR2* and *PAX9*) regulated by *LDB1*. *PRDM10*, another gene under selection in the leopard, is also expressed during craniofacial development; its expression was observed in the mandible region, reticular dermis layer of differentiated head tissues, and the adrenal medulla[84]. *LDB1* is also associated with limb development and growth working as a co-factor for *LHX2* and *LHX9*[85]. Finally, its association with *PODXL* can cause knee problems[86].

Brain development and circadian cycle

PODXL was found to be highly expressed in laminated regions of the brain, such as the the olfactory bulb, cerebral cortex, hippocampus and cerebellum and is required for correct synaptogenesis. Knockout mice for this gene show important alterations in the central nervous system and neuromuscular system[86].

ATP10A is another imprinted gene in humans with a signature of positive selection. It is located on chromosome 15 and is associated with the Angelman syndrome[87]. Angelman

syndrome is characterized by gait ataxia, microcephaly, severe mental retardation and speech limitations.

CCDC86 is an imprinted gene associated with schizophrenia and is known to be expressed in the hippocampus[88] and the placenta. It was primarily described as being expressed in lymphocytes in mice, where it is responsible for the regulation of T-cells. Other genes act on specific regions of the brain, more specifically in the cerebellum and hippocampus. *ELAVL2* and PAPLN are associated with behaviour[89,90]. *ARNTL* is involved in the circadian cycle, and is known to cause seasonal mood disorders in humans and affect fertility[91]. *PTPRU* is another gene related to the circadian cycle, more specifically to sleep duration[76].

Supplementary Table S19: **Branch-site for the leopard**. Candidate genes under positive selection, p-values for the gene tree and species tree analysis (significance threshold < 0.05) and gene description.

Gene	P-value		Description
	Gene tree	Species tree	
ALDH1A3	1	0,040479036	Aldehyde dehydrogenase family 1 member A3
ARNTL	0,037116721	1	Aryl hydrocarbon receptor nuclear translocator-like protein 1 isoform X4
ATP10A	0,000529299	0,000480505	Probable phospholipid-transporting atpase VA
BROX	0,001292934	0,001292934	BRO1 domain-containing protein BROX isoform X1
BRSK1	0,003369073	0,003369092	Serine/threonine-protein kinase BRSK1
CCDC162P	0,068204974	0,04593714	Coiled-coil domain-containing protein 162-like
CCDC86	0,044425555	0,055695646	Coiled-coil domain-containing protein 86
CLNK	0,045696154	0,049727309	Cytokine-dependent hematopoietic cell linker
DIRAS3	0,001139122	0,001139121	GTP-binding protein Di-Ras3
ELAVL2	0,016016133	1	ELAV-like protein 2 isoform X1
FBXL8	0,011855004	0,011855004	F-box/LRR-repeat protein 8 isoform X1
GCC1	1	1,63E-05	GRIP and coiled-coil domain-containing protein 1
GHRL	0,207793948	0,00490667	Appetite-regulating hormone isoform X2
HDAC2	1	0,001094072	Histone deacetylase 2
HDHD3	0,00406934	0,003579217	Haloacid dehalogenase-like hydrolase domain-containing protein 3 isoform X1
HJURP	0,000929296	0,000901781	Holliday junction recognition protein
KIAA1211	0,000596637	0,000592702	Uncharacterized protein KIAA1211 homolog
LDB1	0,013695501	0,013695501	LIM domain-binding protein 1 isoform X3
LOC102948481	0,004901975	0,004874934	Uncharacterized protein c10orf68 homolog
LOC102948891	2,44E-05	6,14E-06	Uncharacterized protein LOC102948891
LOC102949390	0,005354625	0,005354625	Transmembrane 9 superfamily member 2-like
LOC102951695	0,0006805	0,000678123	Olfactory receptor-like protein OLF4-like
LOC102954321	0,652463793	0,048522508	Olfactory receptor 147-like
LOC102955930	0,039311628	0,05185713	Zinc finger protein 416-like isoform X1
LOC102958153	0,000686815	0,000686815	Uncharacterized protein c16orf78 homolog isoform X1
LOC102958356	0,000193121	0,000193121	Uncharacterized protein LOC102958356
LOC102958954	0,028544346	1	Olfactory receptor 52A5-like
LOC102959355	0,009897846	0,015331645	Probable inactive protein kinase-like protein sgk071

LOC102960461	0,000937984	0,000937974	Zinc finger protein 416-like
LOC102960856	0,001918639	0,001131641	Olfactory receptor 8S1-like
LOC102961943	0,000308041	0,000308043	Olfactory receptor 9K2-like
LOC102962222	0,079421911	0,042784111	Transmembrane and coiled-coil domain-containing protein 5B-like
LOC102963918	0,00194525	0,010575879	Antileukoproteinase-like isoform X1
LOC102964575	0,001381643	0,001381651	Cytochrome b reductase 1-like
LOC102965923	7,91E-04	8,97E-04	Uncharacterized protein LOC102965923
LOC102967712	0,000933547	0,000933692	Probable methyltransferase BTM2 homolog
LOC102969593	0,00681252	0,01263762	Olfactory receptor 52R1-like
LOC102971342	0,001612294	0,001797108	Uncharacterized protein LOC102971342
LOC102972668	0,001072962	0,001072962	Uncharacterized protein LOC102972668 isoform X1
NINL	0,050051999	0,000306554	Ninein-like protein
NR1I2	0,024843127	1	Nuclear receptor subfamily 1 group I member 2
PAPLN	0,012785761	0,012785761	Papilin
PODXL	0,035292942	0,035303282	Podocalyxin
PRDM10	0,013202823	0,013202823	PR domain zinc finger protein 10
PTPRU	0,000278305	0,000278305	Receptor-type tyrosine-protein phosphatase U
RNF133	0,000772235	0,000771867	E3 ubiquitin-protein ligase RNF133
STXBP4	0,000562058	0,00055113	Syntaxin-binding protein 4
TBC1D31	0,005461893	0,006093299	TBC1 domain family member 31 isoform X1
THAP4	0,00029125	0,000291245	THAP domain-containing protein 4
TSPAN18	1	0,031947424	Tetraspanin-18 isoform X1
TTC37	0,032369838	0,034848723	Tetratricopeptide repeat protein 37
TTC8	0,000522816	0,000490953	Tetratricopeptide repeat protein 8 isoform X4
ZBTB7B	0,001131092	0,001131688	Zinc finger and BTB domain-containing protein 7B
ZC3HC1	0,000658543	0,318670322	Nuclear-interacting partner of ALK
ZNF620	0,003236307	0,003236307	Zinc finger protein 620-like
ZNF710	0,023724077	0,023804203	Zinc finger protein 710
ZNF836	0,035421634	0,231618593	Zinc finger protein 836-like isoform X2

Supplementary Table S20: **Gene ontology results for the leopard.** Analysis refers to the analysis performed in WebGestalt with a significance threshhold < 0.05 and multiple correct testing (FDR). Characteristic refers to the enriched category found. Parameters describes enrichment details. C (Genes presented in the category); O (Genes observed in the dataset); R (Ratio of enrichment); rawP (p-value for the hypergeometric test); adjP (adjusted P value by the multiple correction test).

Disease	Parameters	Genes
Subarachnoid Hemorrhage	C=104;O=2;E=0.07;R=28.60;rawP=0.0022;adjP=0.0268	*ATP10A; CCDC86*
Alexander Disease	C=147;O=2;E=0.10;R=20.23;rawP=0.0044;adjP=0.0268	*PRDM10; THAP4*
Myocardial Infarction	C=242;O=2;E=0.16;R=12.29;rawP=0.0115;adjP=0.0268	*ATP10A; CCDC86*
Carcinoma, Hepatocellular	C=208;O=2;E=0.14;R=14.30;rawP=0.0086;adjP=0.0268	*DIRAS3;NR1I2*
Liver Neoplasms	C=242;O=2;E=0.16;R=12.29;rawP=0.0115;adjP=0.0268	*DIRAS3;NR1I2*
Nelson Syndrome	C=673;O=3;E=0.45;R=6.63;rawP=0.0102;adjP=0.0268	*THAP4; ZC3HC1; RNF133*
Obesity	C=323;O=2;E=0.22;R=9.21;rawP=0.0199;adjP=0.0398	*ATP10A; TTC8*
Endocrine System Diseases	C=429;O=2;E=0.29;R=6.93;rawP=0.0336;adjP=0.0433	*DIRAS3; PODXL*

Endocrine Disorder NOS	C=429;O=2;E=0.29;R=6.93;rawP=0.0336;adjP=0.0433	*DIRAS3; PODXL*
Urogenital Neoplasms	C=432;O=2;E=0.29;R=6.88;rawP=0.0340;adjP=0.0433	*DIRAS3; PODXL*
Endocrine Disturbance NOS	C=429;O=2;E=0.29;R=6.93;rawP=0.0336;adjP=0.0433	*DIRAS3; PODXL*
Williams Syndrome	C=461;O=2;E=0.31;R=6.45;rawP=0.0383;adjP=0.0447	*THAP4; RNF133*

Phenotype	**Parameters**	**Genes**
Hepatic Fibrosis	C=56;O=2;E=0.04;R=45.88;rawP=0.0005;adjP=0.0335	*TTC8; TTC37*
Abnormality of the Biliary System	C=137;O=2;E=0.11;R=18.75;rawP=0.0028;adjP=0.0375	*TTC8; TTC37*
Decreased Liver Function	C=132;O=2;E=0.10;R=19.46;rawP=0.0026;adjP=0.0375	*TTC8; TTC38*
Hepatic Failure	C=116;O=2;E=0.09;R=22.15;rawP=0.0020;adjP=0.0375	*TTC8; TTC39*
Renal Cysts	C=126;O=2;E=0.10;R=20.39;rawP=0.0024;adjP=0.0375	*TTC8; TTC40*
Downslanted Palpebral Fissures	C=152;O=2;E=0.12;R=16.90;rawP=0.0035;adjP=0.0391	*TTC8; TTC41*

KEGG Pathway	**Parameters**	**Genes**
Olfactory transduction	C=388;O=4;E=0.32;R=12.35;rawP=0.0003;adjP=0.0003	*LOC102958954; LOC102960856; LOC102961943; LOC102969593*

Pathway commons	**Parameters**	**Genes**
Olfactory Signaling Pathway	C=370;O=4;E=0.31;R=12.95;rawP=0.0003; adjP=0.0018	*LOC102958954; LOC102960856; LOC102961943; LOC102969593*
GPCR Downstream Signaling	C=553;O=4;E=0.46;R=8.67;rawP=0.0011; adjP=0.0033	*LOC102958954; LOC102960856; LOC102961943; LOC102969593*
Signal Transduction	C=1231;O=5;E=1.03;R=4.87;rawP=0.0034; adjP=0.0068	*PTPRU; LOC102958954; LOC102960856; LOC102961943; LOC102969593*
Signaling by GPCR	C=823;O=4;E=0.69;R=5.82;rawP=0.0048; adjP=0.0072	*LOC102958954; LOC102960856; LOC102961943; LOC102969593*
Transmembrane Transport of Small Molecules	C=379;O=2;E=0.32;R=6.32;rawP=0.0399; adjP=0.0479	*ATP10A; CYBRD1*
LKB1 Signaling Events	C=1308;O=2;E=1.09;R=1.83;rawP=0.2985; adjP=0.2985	*BRSK1; STXBP4*

6.4.4. High Altitude adaptation and Lipid Metabolism in Snow Leopard

High altitude and cold environment adaptation

Snow leopards inhabit the Tibetan plateau, and have a limited range when compared to the other big cats. The species is described as having a series of morphological adaptations to live in cold climate and high altitudes[92,93]. Our gene ontology analyses presented some clues about those adaptations that were further investigated through literature searches.

TCTN1 and *TSEN54* mutations are associated with several ciliary syndromes. Those syndromes can affect the lungs, reproductive system, particularly spermatozoa, brain and other

systems. In the gene ontology analysis, both genes were identified as part of an enriched set related to apnea.

Regarding this characteristic, both genes are expressed during the cerebellum development[94]. *TCTN1* is known to be associated with Joubert syndrome, which affects the cerebellum and has symptoms such as ataxia and hyperpnea[95]. *TSEN54,* the other gene associated with apnea, causes Pontocerebellar ataxia, also affecting the cerebellum and causing facial alterations[96]. Another gene found to be under positive selection in the snow leopard was ubiquitin-like, containing PHD and RING finger domains 1 (*UHRF1*). This gene is highly involved in the methylation process and a recent study associated it with hypoxia. The gene was not listed as associated with apnea or hypoxia terms in gene ontology, but two studies concluded its involvement with hypoxia[97]. More precisely, Ban *et al.* (2015) observed a higher level of expression of the gene in the lungs of Tibetan pigs, when compared to closely related breeds from low altitudes[97]. They discussed the possibility that the increased expressed of this gene could be related to the downregulation of several others, not necessarily in high altitude environments.

Finally, a ALX homeobox 3 transcription factor (*ALX3*) mutation is associated with frontonasal malformations in humans but not in mice[98]. The snow leopard has a short muzzle with enlarged nasal cavities. This is thought to be an adaptation to high altitudes and cold climates[99].

Lipid Metabolism

Carboxyl ester lipase (*CEL*) and pancreatic lipase-related protein 2 (*PNLIPRP2*) are positively selected genes for *Panthera uncia,* both of which are components of the lipid digestion metabolism. They are expressed in the pancreas and mammary glands, the latter being of great importance to the offspring. This is because, early in their lives, they have an insufficient pancreatic function[100]. Thus, with the help of the enzymes coded by these genes and present in the milk, newborn cubs are able to digest the lipids, thus avoiding nutritional issues. In addition, another feature of the *PNLIPRP2* gene is that it can hydrolyze the galactolipid substrate present in the membrane of plants. This may be important for the snow leopard since this species also eats grass, a unique feature among the *Panthera*[101].

Supplementary Table S21: **Branch-site for the snow leopard**. Candidate genes under positive selection, p-values for the gene tree and species tree analysis (significance threshold < 0.05) and gene description. Parameters describes enrichment details. C (Genes presented in the category); O (Genes observed in the dataset); R (Ratio of enrichment); rawP (p-value for the hypergeometric test); adjP (adjusted P value by the multiple correction test).

Gene	P-value		Description
	Gene tree	Species tree	
ADRA2A	0,007531778	0,007532324	Alpha-2A Adrenergic Receptor

ALX3	0,029319517	0,029319517	Homeobox Protein Aristaless-like 3
ANKLE2	0,030488298	0,033736114	Ankyrin Repeat and LEM Domain-containing Protein 2
ART3	0,991043956	6,26E-05	Ecto-ADP-ribosyltransferase 3 isoform X3
ATP2A3	0,002362773	0,002362771	Sarcoplasmic/endoplasmic reticulum calcium atpase 3
ATP5SL	0,017761634	0,219005465	ATP synthase subunit s-like protein isoform X2
CASC5	0,001396348	0,00137624	Protein CASC5
CCDC136	0,016112162	0,635897365	Coiled-coil domain-containing protein 136 isoform X1
CCS	0,017765141	0,017765141	Copper chaperone for superoxide dismutase
CEL	0,010790811	0,011439879	Bile salt-activated lipase
CELF6	0,007219297	1	CUGBP Elav-like family member 6
CENPQ	0,003008286	0,003008286	Centromere protein Q isoform X1
CEP63	0,000257752	0,000257752	Centrosomal protein of 63 kda isoform X1
CHRNA1	0,001128338	0,0068033	Acetylcholine receptor subunit alpha
CNOT10	4,69E-02	0,066270398	CCR4-NOT transcription complex subunit 10 isoform X1
CPSF7	0,003716085	1	Cleavage and polyadenylation specificity factor subunit 7 isoform X1
DBF4B	0,02374413	1	Protein DBF4 homolog B
DIAPH1	0,000666738	0,000657464	Protein diaphanous homolog 1 isoform X1
DIRAS3	0,000695884	0,000695885	GTP-binding protein Di-Ras3
DIS3	4,38E-03	4,38E-03	Exosome complex exonuclease RRP44
DUSP27	0,045323966	0,036447796	Inactive dual specificity phosphatase 27
EMC1	0,038863917	1	ER membrane protein complex subunit 1
FAM205A	0,002278794	1	Protein FAM205A
FAM227A	0,010121344	0,010121344	Putative uncharacterized protein LOC388900
GDPD4	0,00157708	0,00157708	Glycerophosphodiester phosphodiesterase domain-containing protein 4
HELT	0,006039383	0,00603941	Hairy and enhancer of split-related protein HELT
HS6ST1	1	1,38E-06	Heparan-sulfate 6-O-sulfotransferase 1
IKBKG	1,09E-05	1,09E-05	NF-kappa-B essential modulator
INO80C	1	0,014927589	INO80 complex subunit C isoform X1
KIAA2026	0,005489185	0,005494454	Uncharacterized protein KIAA2026 homolog
KIF21A	0,0090353	0,009035289	Kinesin-like protein KIF21A
LIMD1	0,005441252	0,005441252	LIM domain-containing protein 1 isoform X2
LOC102948543	0,00084155	0,000913108	Protein polybromo-1-like
LOC102949721	0,038682062	1	Keratin-associated protein 16-1-like
LOC102950484	0,034214037	0,034214037	Uncharacterized protein LOC102950484
LOC102950787	0,001491778	1	Keratin-associated protein 4-12-like isoform X1
LOC102951271	0,001268541	0,000879844	Uncharacterized protein c12orf56 homolog
LOC102951869	0,003594232	0,003444859	Uncharacterized protein LOC102951869 isoform X1
LOC102953554	0,000440492	0,000440472	Long-chain-fatty-acid--coa ligase ACSBG2-like
LOC102953684	0,022969662	0,023012226	Olfactory receptor 10C1-like
LOC102955108	0,543024982	0,005544886	Olfactory receptor 13A1-like
LOC102955930	0,03353892	0,034933103	Zinc finger protein 416-like isoform X1
LOC102957202	0,000703247	0,000703248	Uncharacterized protein LOC102957202
LOC102958227	0,023762886	0,028327148	Protein piccolo-like
LOC102958616	1,83E-05	1,83E-05	Uncharacterized protein LOC102958616
LOC102960188	0,059152521	0,005120811	Vomeronasal type-1 receptor 1-like

LOC102961248	0,03175765	0,033249151	DLA class II histocompatibility antigen, DR-1 beta chain-like
LOC102963746	0,043918821	0,028036493	Zinc finger protein with KRAB and SCAN domains 7-like
LOC102964060	0,437533737	0,03560232	Mucin-5AC-like
LOC102964471	0,005983682	0,006130344	Uncharacterized protein LOC102964471
LOC102964775	0,002103911	0,002103911	Uncharacterized protein LOC102964775 isoform X1
LOC102965668	0,000901548	0,000901548	Transcriptional repressor NF-X1 homolog
LOC102965808	0,002842153	0,002842451	Carcinoembryonic antigen-related cell adhesion molecule 1-like
LOC102966492	0,002653338	0,002293125	Immunoglobulin iota chain-like
LOC102966691	0,014747106	0,014747106	Uncharacterized protein c6orf229 homolog
LOC102968979	0,419530514	0,001188267	Uncharacterized protein LOC102968979
LOC102970787	0,000325316	0,000326722	Uncharacterized protein c17orf50 homolog
LOC102970896	1,63E-03	1,40E-02	Olfactory receptor 2AP1-like
LOC102972872	0,020247342	1	Olfactory receptor 9A4-like
LRCH1	0,012846446	0,005079732	Leucine-rich repeat and calponin homology domain-containing protein 1
LRRC8A	0,046098049	1	Leucine-rich repeat-containing protein 8A
MASP2	0,024709142	0,024709284	Mannan-binding lectin serine protease 2
MDC1	0,211854201	0,049803684	Mediator of DNA damage checkpoint protein 1 isoform X2
MUC13	0,004254596	0,007804076	Mucin-13
NEK10	0,002306498	0,002306498	Serine/threonine-protein kinase Nek10 isoform X1
NGFR	0,027578421	0,011417219	Tumor necrosis factor receptor superfamily member 16
NUDT13	0,008127438	0,015654239	Nucleoside diphosphate-linked moiety X motif 13
PDZRN4	0,019792908	0,019792908	PDZ domain-containing RING finger protein 4
PLCB2	1	0,005975241	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-2
PLOD1	0,018275561	1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1
PNLIPRP2	0,000749457	0,00080727	Pancreatic lipase-related protein 2
PPAPDC3	0,04880342	0,060230652	Probable lipid phosphate phosphatase PPAPDC3
PRG3	0,013641665	0,025002659	Proteoglycan 3
PRIMPOL	1	0,026399939	DNA-directed primase/polymerase protein
REP15	0,121611565	0,042423141	Rab15 effector protein
RIOK1	0,002343834	0,002404378	Serine/threonine-protein kinase RIO1 isoform X1
RRP36	0,004271241	0,003697166	Ribosomal RNA processing protein 36 homolog isoform X1
RSPH4A	0,050114311	0,247145453	Radial spoke head protein 4 homolog A
SCUBE2	0,041021706	1	Signal peptide, CUB and EGF-like domain-containing protein 2
SEMA3G	0,013489945	1	Semaphorin-3G isoform X1
SIX4	0,001074407	0,001074407	Homeobox protein SIX4
SYCP2	0,034836442	0,046068008	Synaptonemal complex protein 2
SYNPO	0,000688578	0,000955531	Synaptopodin
TCTN1	0,001314726	0,001314726	Tectonic-1
TERF1	0,050343894	0,048686199	Telomeric repeat-binding factor 1
THAP4	0,001057447	0,001061391	THAP domain-containing protein 4
TIMM44	0,002405056	0,002405014	Mitochondrial import inner membrane translocase subunit TIM44
TLE2	0,002448251	0,002448251	Transducin-like enhancer protein 2
TNIP2	0,000221917	0,000221916	TNFAIP3-interacting protein 2 isoform X1
TSEN54	0,000515221	0,00048088	Trna-splicing endonuclease subunit Sen54
TSKS	0,001591796	0,001591812	Testis-specific serine kinase substrate
UHRF1	3,65E-05	0,001758798	E3 ubiquitin-protein ligase UHRF1

USP40	0,001205208	0,679958991	Ubiquitin carboxyl-terminal hydrolase 40 isoform X1
ZKSCAN7	0,042538161	0,042011421	Zinc finger protein with KRAB and SCAN domains 7 isoform X1
ZNF37A	0,013149178	0,013149178	Zinc finger protein 37A-like

Supplementary Table S22: **Gene ontology results for the snow leopard**. Analysis refers to the analysis performed in WebGestalt with significance threshold < 0.05 and multiple correct testing (FDR). Characteristic refers to the enriched category found. Parameters describes enrichment details. C (Genes presented in the category); O (Genes observed in the dataset); R (Ratio of enrichment); rawP (p-value for the hypergeometric test); adjP (adjusted P value by the multiple correction test).

Disease	Parameters	Genes
Strabismus	**C=74;O=3;E=0.10;R=29.14;rawP=0.0002;adjP=0.0096**	***IKBKG; KIF21A; TCTN1***
Encephalocele	C=31;O=2;E=0.04;R=46.37;rawP=0.0009;adjP=0.0192	*TCTN1; ALX3*
Alexander Disease	C=147;O=3;E=0.20;R=14.67;rawP=0.0012;adjP=0.0192	*EMC1; NEK10; THAP4*
Skin Diseases	C=417;O=4;E=0.58;R=6.89;rawP=0.0027;adjP=0.0259	*DIRAS3; PLOD1; IKBKG; NEK10*
Ciliary Motility Disorders	C=54;O=2;E=0.08;R=26.62;rawP=0.0026;adjP=0.0259	*RSPH4A; TCTN1*
Apnea	**C=66;O=2;E=0.09;R=21.78;rawP=0.0039;adjP=0.0276**	***TCTN1; TSEN54***
Hernia	C=70;O=2;E=0.10;R=20.54;rawP=0.0043;adjP=0.0276	*TCTN1; ALX3*
Skin and Connective Tissue Diseases	C=481;O=4;E=0.67;R=5.98;rawP=0.0046;adjP=0.0276	*DIRAS3; PLOD1; IKBKG; NEK10*
Neuromuscular Diseases	C=276;O=3;E=0.38;R=7.81;rawP=0.0068;adjP=0.0363	*KIF21A; CHRNA1; CCS*
Depressive Disorder, Major	C=98;O=2;E=0.14;R=14.67;rawP=0.0083;adjP=0.0398	*NGFR; ADRA2A*
Depressive Disorder	C=112;O=2;E=0.16;R=12.84;rawP=0.0107;adjP=0.0428	*NGFR; ADRA2A*
Necrosis	C=324;O=3;E=0.45;R=6.66;rawP=0.0105;adjP=0.0428	*IKBKG; TNIP2; NGFR*
Breast Diseases	C=350;O=3;E=0.49;R=6.16;rawP=0.0129;adjP=0.0442	*DIRAS3; NEK10; UHRF1*
Congenital Abnormalities	C=643;O=4;E=0.89;R=4.47;rawP=0.0124;adjP=0.0442	*PLOD1; IKBKG; RSPH4A; ALX3*
Ataxia Telangiectasia	C=129;O=2;E=0.18;R=11.14;rawP=0.0140;adjP=0.0448	*IKBKG; TERF1*
Breast Neoplasms	C=377;O=3;E=0.52;R=5.72;rawP=0.0157;adjP=0.0471	*DIRAS3; NEK10; UHRF1*

Molecular Function	Parameters	Genes
Acylglycerol Lipase Activity	**C=8;O=2;E=0.02;R=83.07;rawP=0.0002;adjP=0.0184**	***CEL; PNLIPRP2***
Retinyl-palmitate Esterase Activity	**C=15;O=2;E=0.05;R=44.30;rawP=0.0009;adjP=0.0368**	***CEL; PNLIPRP2***
Triglyceride Lipase Activity	**C=17;O=2;E=0.05;R=39.09;rawP=0.0012;adjP=0.0368**	***CEL; PNLIPRP2***

KEGG Pathway	Parameters	Genes
Pancreatic Secretion	C=101;O=3;E=0.14;R=21.35;rawP=0.0004;adjP=0.0028	*CEL; PNLIPRP2; ATP2A3*
Fat Digestion and Absorption	**C=46;O=2;E=0.06;R=31.25;rawP=0.0019;adjP=0.0051**	***CEL; PNLIPRP2***
Glycerolipid Metabolism	C=50;O=2;E=0.07;R=28.75;rawP=0.0022;adjP=0.0051	*CEL; PNLIPRP2*
Shigellosis	C=61;O=2;E=0.08;R=23.57;rawP=0.0033;adjP=0.0058	*IKBKG; DIAPH1*

RNA Degradation	C=71;O=2;E=0.10;R=20.25;rawP=0.0044;adjP=0.0062	*DIS3; CNOT10*
Adipocytokine Signaling Pathway	C=68;O=2;E=0.11;R=17.87;rawP=0.0057; adjP=0.0093	*IKBKG; ACSBG2*

Pathway Commons	Parameters	Genes
Digestion of Dietary Lipid	**C=4;O=2;E=0.01;R=359.39;rawP=1.14e-05;adjP=0.0008**	***CEL; PNLIPRP2***
Chromosome Maintenance	C=69;O=3;E=0.10;R=31.25;rawP=0.0001;adjP=0.0037	*CENPQ; CASC5; TERF1*
Triacylglycerol Degradation	**C=14;O=2;E=0.02;R=102.68;rawP=0.0002;adjP=0.0049**	***CEL; PNLIPRP2***
Nucleosome Assembly	C=35;O=2;E=0.05;R=41.07;rawP=0.0011;adjP=0.0146	*CENPQ; CASC5*
M Phase	C=158;O=3;E=0.22;R=13.65;rawP=0.0014;adjP=0.0146	*CENPQ; CEP63; CASC5*
Deposition of New CENPA-Containing Nucleosomes at the Centromere	C=35;O=2;E=0.05;R=41.07;rawP=0.0011;adjP=0.0146	*CENPQ; CASC5*
Deadenylation-dependent mRNA Decay	C=46;O=2;E=0.06;R=31.25;rawP=0.0019;adjP=0.0146	*DIS3; CNOT10*
Meiotic Synapsis	C=47;O=2;E=0.07;R=30.59;rawP=0.0020;adjP=0.0146	*SYCP2; TERF1*
Interleukin-1 Signaling	C=39;O=2;E=0.05;R=36.86;rawP=0.0014;adjP=0.0146	*IKBKG; TNIP2*
Lipid Digestion, Mobilization, and transport	**C=47;O=2;E=0.07;R=30.59;rawP=0.0020;adjP=0.0146**	***CEL; PNLIPRP2***
Mitotic M-M/G1 Phases	C=242;O=3;E=0.34;R=8.91;rawP=0.0047;adjP=0.0286	*CENPQ; CEP63; CASC5*
Meiosis	C=72;O=2;E=0.10;R=19.97;rawP=0.0046;adjP=0.0286	*SYCP2; TERF1*
DNA Replication	C=261;O=3;E=0.36;R=8.26;rawP=0.0058;adjP=0.0326	*CENPQ; CEP63; CASC5*
Signaling by Interleukins	C=92;O=2;E=0.13;R=15.63;rawP=0.0074;adjP=0.0386	*IKBKG; TNIP2*
Mitotic Prometaphase	C=99;O=2;E=0.14;R=14.52;rawP=0.0085;adjP=0.0414	*CENPQ; CASC5*
Cell Cycle, Mitotic	C=318;O=3;E=0.44;R=6.78;rawP=0.0100;adjP=0.0456	*CENPQ; CEP63; CASC5*

Phewas	Parameters	Genes
Diseases of the Salivary Glands	C=110;O=2;E=0.25;R=7.90;rawP=0.0219;adjP=0.0438	*USP40; PDZRN4*

6.4.5. Melanogenesis in the Tiger

The tiger is the only felid species with exclusively black stripes as its coat pattern[99] . The background is reddish orange, whereas the belly and some other areas are white[102]. The stripes are vertical on the flank, with different arrangements between the sides; ringed at the distal tail; and symmetric on the face. This coat probably was selected due to camouflage in natural environments, attenuating the body line and merging the animal to its surroundings[99]. The pattern establishment may occur by two processes in the embryonic stages of development. It begins with the spatial orientation of the pattern, by epithelial cell differentiation; and then for differential pigmentation, according to this pattern[103]. The pathway Wnt/beta-catenin is a candidate for these

processes, since it has a role at epithelial development and cell differentiation from the neural crest[104,105]; in addition to being part of the melanogenesis pathway [KEGG map04916].

The Wnt signaling pathway is related to different embrionary and postnatal development processes. Among these, there are the development and morphogenesis of hair follicle associated to epithelial differentiation, including melanocytes[106,107]. In the canonical pathway, which involves the stabilization of beta-catenin, the Wnt family members bind to membrane receptors frizzled and *LRP6*, promoting *Dvl* recruitment. This latter one is related to *LRP6* phosphorylation and, by interaction with G protein, recruits Axin. The beta-catenin stabilization occurs downstream, when *Dvl* interacts with it, inhibiting its degradation. Then beta-catenin moves to the nucleus and activates the transcription of target genes[108]. This pathway can be related to coat pattern due to the activation of MITF[109]; or by reaction-diffusion model, through the interaction with *Dkk* and modification of hair follicle density[110,111].

The Wnt family has 19 members among which is *WNT10B*. For this gene, there is no knowledge about its frizzled receptor[112], but it is known to be functional both at embrionary and postnatal stages[113]. It is related to *in vitro* hair follicle development and *in vivo* hair growth[114]; epithelial cell differentiation in adult tissue and hair shaft elongation[113]; promotion of melanocyte maturation (melanogenesis from immature cells) and increased pigmentation (tyrosinase activity) in the hair bulb[107]. Furthermore, this gene is also related to hair follicle disposition according to its interaction with the inhibitor *Fgf20*[115]. Given these characteristics and linking it to the tiger's striped phenotype, we hypothesize that this pathway is under selection due to its influence on the coat pattern. In the striped hair follicles, there may occur increased differentiation of stem cells into melanocytes, maturation of these and higher melanin production by increased tyrosinase expression and/or activity (since this gene itself is also under selection in the tiger).

Supplementary Table S23: **Branch-site for the tiger**. Candidate genes under positive selection, p-values for the gene tree and species tree analysis (significance threshold < 0.05) and gene description.

Gene	P-value		Description
	Gene tree	Species tree	
ABHD15	0,034838522	0,034838033	Abhydrolase domain-containing protein 15
ACAP3	0,018407898	0,021899174	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 3
ACTN3	0,005859068	0,005859003	Alpha-actinin-3
ADAMTSL3	0,027542091	0,027542091	ADAMTS-like protein 3
ADRA2A	0,001166873	0,001166866	Alpha-2A adrenergic receptor
AGER	0,018345923	0,032886925	Advanced glycosylation end product-specific receptor isoform X2
AHCY	0,02372941	0,02372941	Adenosylhomocysteinase isoform X1
AK7	0,0106417	0,0106417	Adenylate kinase 7
ALS2CL	0,00850549	0,00850549	ALS2 C-terminal-like protein
ANGPTL2	0,000532401	0,000532405	Angiopoietin-related protein 2

ANXA11	3,04E-03	3,04E-03	Annexin A11
AP2A2	0,001349453	0,001349466	AP-2 complex subunit alpha-2
ARHGEF11	0,001562951	0,001562951	Rho guanine nucleotide exchange factor 11
ARHGEF18	0,005757252	0,009083506	Rho guanine nucleotide exchange factor 18
ARHGEF2	0,009397748	0,014741217	Rho guanine nucleotide exchange factor 2
ARMC9	3,28E-03	2,80E-03	Lish domain-containing protein ARMC9 isoform X6
ARNT	0,000492	0,000473629	Aryl hydrocarbon receptor nuclear translocator isoform X2
ATG16L2	1	0,031851015	Autophagy-related protein 16-2
ATP1A1	0,00305356	0,003260638	Sodium/potassium-transporting atpase subunit alpha-1
ATP2B1	0,001083389	0,001445705	Plasma membrane calcium-transporting atpase 1 isoform X1
ATP2C2	0,002422659	0,002211014	Calcium-transporting atpase type 2C member 2
ATP5S	0,00158684	0,00158684	ATP synthase subunit s, mitochondrial isoform X1
ATPAF2	0,003207762	0,003207762	ATP synthase mitochondrial F1 complex assembly factor 2
ATXN1L	0,000301745	0,000228504	Ataxin-1-like isoform X1
BAIAP2L2	0,000245587	0,00039056	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 2
BICD2	1,03E-05	1,03E-05	Protein bicaudal D homolog 2
C3AR1	0,001476521	0,001996886	C3a anaphylatoxin chemotactic receptor
CA9	0,000434486	0,000434486	Carbonic anhydrase 9 isoform X1
CACNA1D	0,000868468	0,001146946	Voltage-dependent L-type calcium channel subunit alpha-1D isoform X1
CACNA1F	9,42E-07	2,57E-06	Voltage-dependent L-type calcium channel subunit alpha-1F
CALCOCO1	0,00750401	0,00750401	Calcium-binding and coiled-coil domain-containing protein 1 isoform X3
CARKD	0,003290489	0,004802795	ATP-dependent (S)-NAD(P)H-hydrate dehydratase
CCDC8	0,004421215	0,004719495	Coiled-coil domain-containing protein 8
CCER1	0,004560721	0,006473842	Coiled-coil domain-containing glutamate-rich protein 1
CCM2L	0,000281475	0,000281475	Cerebral cavernous malformations 2 protein-like
CCNB3	4,16E-03	4,43E-03	G2/mitotic-specific cyclin-B3
CD3D	0,000369211	0,000735621	T-cell surface glycoprotein CD3 delta chain isoform X1
CDK13	1,76E-12	2,26E-12	Cyclin-dependent kinase 13
CHAF1A	0,010495398	0,009340344	Chromatin assembly factor 1 subunit A
CHDC2	0,057176516	0,041838489	Calponin homology domain-containing protein 2
CHRNA4	0,002477105	0,002477105	Neuronal acetylcholine receptor subunit alpha-4
CLCN4	0,032785295	0,03278541	H(+)/Cl(-) exchange transporter 4
CLDN7	0,000258495	0,000258495	Claudin-7
CLIC1	0,00634244	0,010930965	Chloride intracellular channel protein 1
CLN6	0,000838331	0,000442617	Ceroid-lipofuscinosis neuronal protein 6 homolog
COL6A1	5,92E-05	5,97E-05	Collagen alpha-1(VI) chain
CPEB4	0,000513142	0,000513142	Cytoplasmic polyadenylation element-binding protein 4 isoform X1
CPNE6	0,000207937	0,000207937	Copine-6 isoform X2
CTDSP1	8,81E-06	7,05E-06	Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 1
CTGF	1	0,010887928	Connective tissue growth factor
DBF4B	1	0,0237484	Protein DBF4 homolog B
DBX1	0,002090779	0,002009679	Homeobox protein DBX1
DCAF11	0,00064362	0,000510131	DDB1- and CUL4-associated factor 11 isoform X2
DCAF15	0,036022314	0,037903442	DDB1- and CUL4-associated factor 15

DDX18	0,000906596	0,001045743	ATP-dependent RNA helicase DDX18
DDX26B	0,000851728	0,000309431	Protein DDX26B-like
DHCR24	0,03432062	0,993923551	Delta(24)-sterol reductase
DLGAP3	0,000219055	0,000219055	Disks large-associated protein 3
DLL4	0,000617182	0,001397263	Delta-like protein 4
DNMT1	0,006707238	0,006707238	DNA (cytosine-5)-methyltransferase 1
DPCR1	0,222638963	0,000472395	Diffuse panbronchiolitis critical region protein 1
DPF1	0,007383075	0,007383075	Zinc finger protein neuro-d4
DRAXIN	0,021353088	0,006987171	Draxin
DUSP2	0,026554464	1	Dual specificity protein phosphatase 2
DVL1	7,64E-03	3,73E-03	Segment polarity protein dishevelled homolog DVL-1
EBNA1BP2	0,00569227	0,001558008	Probable rrna-processing protein EBP2
EFHC2	0,030522713	0,03052321	EF-hand domain-containing family member C2
ELF4	0,001858945	0,001858959	ETS-related transcription factor Elf-4
EME1	0,013039684	0,899644011	Crossover junction endonuclease EME1
EMILIN1	0,000912663	0,001101892	EMILIN-1
EML2	0,000500643	0,000500643	Echinoderm microtubule-associated protein-like 2
ENTPD7	0,000810868	0,000810868	Ectonucleoside triphosphate diphosphohydrolase 7
EPS8L2	0,008682516	0,008682438	Epidermal growth factor receptor kinase substrate 8-like protein 2
ERBB2	0,786599743	0,00578863	Receptor tyrosine-protein kinase erbb-2
ESAM	0,117899284	0,002934417	Endothelial cell-selective adhesion molecule
ESRP1	1	0,005266162	Epithelial splicing regulatory protein 1 isoform X2
ESYT1	0,001339882	0,002397886	Extended synaptotagmin-1
FAM111A	0,001221766	0,001313305	Protein FAM111A
FAM160A2	0,047844319	0,048013194	FTS and Hook-interacting protein isoform X3
FAM73B	0,004295283	0,004306686	Protein FAM73B isoform X1
FANCD2	0,032297758	0,242576549	Fanconi anemia group D2 protein
FAXC	0,000538561	0,000826115	Failed axon connections homolog
FGD4	0,039198641	0,998871621	FYVE, rhogef and PH domain-containing protein 4 isoform X1
FIGN	0,011920616	1	Fidgetin
FMNL3	0,000526207	0,000526207	Formin-like protein 3 isoform X2
FN1	0,953216101	0,003038348	Fibronectin isoform X8
FN3KRP	0,040906163	0,040905728	Ketosamine-3-kinase isoform X1
FOXP3	0,001906605	5,18E-05	Forkhead box protein P3 isoform X2
FURIN	0,001981548	0,001981548	Furin
FZD1	3,08E-03	3,08E-03	Frizzled-1
GABRQ	0,001979884	0,001979727	Gamma-aminobutyric acid receptor subunit theta
GALNT12	0,061359047	0,002360655	Polypeptide N-acetylgalactosaminyltransferase 12
GCM2	0,001504624	0,001576352	Chorion-specific transcription factor gcmb
GDPD2	9,35E-05	9,35E-05	Glycerophosphoinositol inositolphosphodiesterase GDPD2 isoform X1
GDPD2	0,000137299	0,000137299	Glycerophosphoinositol inositolphosphodiesterase GDPD2 isoform X2
GLYAT	0,000760667	0,00903783	Glycine N-acyltransferase isoform X1
GLYCTK	0,004883197	0,004746571	Glycerate kinase isoform X1
GPR108	0,001300586	0,001300586	Protein GPR108 isoform X2
GPR124	0,001453728	0,003165567	G-protein coupled receptor 124

GPR137	0,000226486	0,000225543	Integral membrane protein GPR137
GRIP2	0,00274248	1	Glutamate receptor-interacting protein 2
GRK1	0,023269304	0,022398901	Rhodopsin kinase
HID1	7,27E-05	1,51E-04	Protein HID1
HIF3A	0,014847853	0,014847853	Hypoxia-inducible factor 3-alpha
HOOK2	0,034538062	0,034538062	Protein Hook homolog 2
HS6ST1	0,001750144	0,001750165	Heparan-sulfate 6-O-sulfotransferase 1
IGSF9B	0,005571118	0,011625753	Protein turtle homolog B
ILF3	0,001246958	0,001275128	Interleukin enhancer-binding factor 3 isoform X3
INHBE	0,045827495	0,068874243	Inhibin beta E chain
IQCD	0,005472783	0,005472783	IQ domain-containing protein D
IREB2	0,004592524	0,00191533	Iron-responsive element-binding protein 2
IRF9	9,08E-07	9,09E-07	Interferon regulatory factor 9 isoform X2
ISLR2	0,014502349	0,014339245	Immunoglobulin superfamily containing leucine-rich repeat protein 2
KATNBL1	0,001351258	0,001351254	KATNB1-like protein 1
KCND1	0,00070975	0,000709748	Potassium voltage-gated channel subfamily D member 1
KIAA0232	0,002105849	0,00210595	Uncharacterized protein KIAA0232 homolog
KIAA0355	0,002975916	0,00071179	Uncharacterized protein KIAA0355 homolog isoform X1
LAD1	0,003281146	0,0032812	Ladinin-1
LBR	1,98E-02	2,44E-02	Lamin-B receptor
LHX6	0,00185701	0,00185701	LIM/homeobox protein Lhx6 isoform X2
LOC102948479	1	0,008204085	Olfactory receptor 2W1-like
LOC102948964	4,50E-06	5,45E-06	Zinc finger protein 705F-like isoform X1
LOC102949469	0,009128827	0,009127763	Uncharacterized protein LOC102949469
LOC102950007	0,003421619	0,003298619	Protein FAM47A-like
LOC102950097	0,012773588	0,012773588	Uncharacterized protein c6orf106 homolog
LOC102950159	0,032191942	0,032191942	BCL2/adenovirus E1B 19 kda protein-interacting protein 3-like
LOC102950214	0,044899457	0,03667613	Keratin, type II cuticular Hb5-like
LOC102950368	0,003940968	0,003943638	Olfactory receptor 6C74-like
LOC102950423	6,84E-05	1	Lysine-specific demethylase 4B-like
LOC102950516	0,000987202	0,000987151	Olfactory receptor 52D1-like
LOC102950619	0,025181808	0,025182128	Lymphocyte-specific protein 1-like isoform X1
LOC102950787	1	0,015895204	Keratin-associated protein 4-12-like isoform X3
LOC102950901	0,001600964	0,001661637	Cytochrome b-245 heavy chain-like
LOC102951123	0,006694409	0,006694409	Uncharacterized protein R102.4-like
LOC102951158	0,000290276	0,000380939	Modulator of retrovirus infection homolog
LOC102951432	0,001027743	0,001267481	Olfactory receptor 51I2-like
LOC102951477	8,57E-04	0,000857609	UPF0609 protein c4orf27 homolog
LOC102952097	0,002131481	0,002132393	AP-4 complex subunit sigma-1-like
LOC102952385	1	0,000659911	Olfactory receptor-like protein OLF3-like
LOC102952556	0,002598175	0,002598175	Olfactory receptor 7C2-like
LOC102952684	0,019816861	0,019816861	Uncharacterized protein c11orf35 homolog
LOC102953350	0,006313409	0,011551635	Hepatoma-derived growth factor-related protein 2-like
LOC102953622	1,40E-03	1,38E-03	Trophoblast glycoprotein-like
LOC102953718	0,02544355	0,02544355	Uncharacterized protein c2orf16 homolog
LOC102954169	1	2,62E-11	Olfactory receptor 6C4-like

LOC102954373	0,007971838	0,007971802	Tyrosine-protein phosphatase non-receptor type 23-like isoform X2
LOC102955052	4,73E-07	4,73E-07	Zinc finger protein 548-like isoform X1
LOC102955918	0,403698492	0,005892143	Mucin-2-like
LOC102956551	0,008571499	0,011032431	TATA box-binding protein-associated factor, RNA polymerase I, subunit C-like
LOC102957246	3,66E-02	2,62E-02	Ly-6/neurotoxin-like protein 1-like
LOC102957369	0,012161344	0,012368717	Uncharacterized protein c1orf167-like
LOC102957797	0,047210799	0,226210407	Olfactory receptor 12D2-like
LOC102958279	0,004696944	0,012551624	Class I histocompatibility antigen, Gogo-B*0101 alpha chain-like isoform X3
LOC102958291	0,007070056	0,007070056	Pepsin A-like
LOC102958677	0,012777962	0,01702951	Leucine zipper putative tumor suppressor 3-like isoform X2
LOC102959335	0,01304809	0,01022713	UDP-N-acetylglucosamine transferase subunit ALG14 homolog
LOC102959459	0,022812454	0,022812664	Specifically androgen-regulated gene protein
LOC102959599	0,032701134	0,32666539	Leukocyte immunoglobulin-like receptor subfamily B member 4-like
LOC102960209	0,000376441	0,000364153	Uncharacterized protein LOC102960209
LOC102960468	0,004851202	0,005727286	Uncharacterized protein LOC102960468
LOC102960583	0,004029698	0,012142585	Olfactory receptor 2T2-like
LOC102960673	0,001521337	0,002373546	Histone H4-like
LOC102961627	3,53E-06	3,53E-06	Guanine nucleotide-binding protein subunit alpha-15-like
LOC102961810	0,00033324	0,000333244	Histone H1.5-like
LOC102962684	4,99E-13	4,99E-13	Putative uncharacterized zinc finger protein 814-like
LOC102962945	2,27E-02	2,27E-02	Homeobox protein Hox-A2-like
LOC102963013	0,009420711	0,009420711	Melanoma-associated antigen B3-like
LOC102963221	0,00390694	0,003907172	Carbonyl reductase [NADPH] 1-like
LOC102964892	0,005947601	0,005947601	AF4/FMR2 family member 3-like
LOC102965074	0,016976071	0,007207001	Uncharacterized protein c14orf80 homolog
LOC102965402	0,000618128	1	Uncharacterized protein LOC102965402
LOC102965574	4,86E-05	4,09E-05	Inactive serine/threonine-protein kinase TEX14-like
LOC102965607	0,00107874	0,001350757	Cyclic AMP-dependent transcription factor ATF-5-like
LOC102966729	2,73E-03	3,06E-02	Steroid 21-hydroxylase-like
LOC102966841	0,000587527	0,008305433	Eukaryotic translation initiation factor 2 subunit 3, Y-linked-like isoform X1
LOC102966952	0,002832315	0,002832101	Gtpase IMAP family member 7-like
LOC102967234	0,005709254	0,005709273	Olfactory receptor 4L1-like
LOC102967258	0,041542377	0,041539629	Histone H1.4-like
LOC102967512	7,66E-04	7,66E-04	Collagen alpha-1(I) chain-like
LOC102967952	0,000939282	0,000784133	Uncharacterized protein c9orf142 homolog
LOC102968009	0,03636974	0,03528707	Olfactory receptor 24-like
LOC102968393	0,008236797	0,012553212	Keratin-associated protein 9-4-like
LOC102968875	0,048305483	0,379251733	Basic proline-rich protein-like
LOC102970611	0,000863662	0,000981767	Uncharacterized protein LOC102970611
LOC102970658	0,004838591	0,008237275	Neutrophil gelatinase-associated lipocalin-like isoform X1
LOC102970816	0,000412065	0,000375931	Olfactory receptor 1D2-like
LOC102971239	9,84E-05	0,000190066	Uncharacterized protein LOC102971239
LOC102971360	0,002710813	0,003266921	Olfactory receptor 1F1-like
LOC102971471	0,001561439	0,002063884	Krueppel-like factor 15-like

LOC102971582	0,020619611	0,018502683	Uncharacterized protein LOC102971582
LOC102972563	0,003096448	0,003096448	Cytochrome P450 2U1-like
LONRF2	0,000736883	0,000715637	LON peptidase N-terminal domain and RING finger protein 2
LPXN	0,027285791	0,027285791	Leupaxin
LRP5	0,000613049	0,000809707	Low-density lipoprotein receptor-related protein 5
LRRC61	1,11E-03	7,21E-04	Leucine-rich repeat-containing protein 61
LTBP4	0,00049136	0,000555416	Latent-transforming growth factor beta-binding protein 4
LZTS2	0,014734428	0,014734428	Leucine zipper putative tumor suppressor 2
MAPRE3	0,006279921	0,006279921	Microtubule-associated protein RP/EB family member 3
MARK3	0,00072029	0,000852371	MAP/microtubule affinity-regulating kinase 3 isoform X2
MATN4	0,000196545	0,000196545	Matrilin-4
MDGA1	0,006707044	0,006707044	MAM domain-containing glycosylphosphatidylinositol anchor protein 1
MEF2A	0,040941321	0,040941321	Myocyte-specific enhancer factor 2A isoform X1
MFSD10	1,87E-05	1,73E-05	Major facilitator superfamily domain-containing protein 10
MKL1	0,000977733	0,000978042	MKL/myocardin-like protein 1
MORC4	0,001076291	0,001076291	MORC family CW-type zinc finger protein 4
MROH2B	2,75E-05	4,28E-05	Maestro heat-like repeat-containing protein family member 2B isoform X1
MROH5	0,000820705	0,000820705	Maestro heat-like repeat family member 5
MTR	0,000312765	0,000312768	Methionine synthase
MYH1	1,51E-05	1,86E-05	Myosin-1
NAALAD2	0,000282802	0,00045337	N-acetylated-alpha-linked acidic dipeptidase 2 isoform X1
NCDN	6,55E-10	4,65E-10	Neurochondrin
NFE2	0,010789245	0,009047704	Transcription factor NF-E2 45 kda subunit isoform X1
NHSL1	0,001479928	0,001251411	NHS-like protein 1
NNT	4,43E-05	4,82E-05	NAD(P) transhydrogenase, mitochondrial isoform X1
NOB1	0,011267086	0,011267086	RNA-binding protein NOB1
NOL11	0,027436454	0,027436454	Nucleolar protein 11
NONO	0,002231615	0,002230884	Non-POU domain-containing octamer-binding protein
NOTCH1	1,17E-06	1,17E-06	Neurogenic locus notch homolog protein 1
NOTCH3	0,014654258	0,004786915	Neurogenic locus notch homolog protein 3
NOTUM	0,001797618	0,001508176	Protein notum homolog
NR4A1	0,003906248	0,003906248	Nuclear receptor subfamily 4 group A member 1 isoform X1
NR5A1	6,45E-04	6,45E-04	Steroidogenic factor 1
NUAK2	1	0,003683317	NUAK family SNF1-like kinase 2
NUP133	0,040994264	0,194594979	Nuclear pore complex protein Nup133
NUP210	0,000985413	0,00066949	Nuclear pore membrane glycoprotein 210-like
OVCH1	0,052352572	0,019564116	Ovochymase-1
PARL	0,000666487	0,00066603	Presenilins-associated rhomboid-like protein, mitochondrial isoform X1
PARVG	0,00104474	0,000966351	Gamma-parvin isoform X3
PBX1	0,013773943	0,013773943	Pre-B-cell leukemia transcription factor 1-like isoform X3
PC	1,38E-09	1,34E-09	Pyruvate carboxylase, mitochondrial
PCDH8	0,001895808	0,00185485	Protocadherin-8
PCDHA6	0,993923551	0,002356024	Protocadherin alpha-6
PCDHB4	0,016467496	0,016468452	Protocadherin beta-4

PCGF2	0,000207572	0,000207572	Polycomb group RING finger protein 2 isoform X3
PCP2	0,005191095	0,0051911	Purkinje cell protein 2 homolog
PDCL	0,004824302	1	Phosducin-like protein isoform X1
PDE3A	0,001248607	0,001199405	Cgmp-inhibited 3',5'-cyclic phosphodiesterase A
PEF1	0,000732425	0,000732426	Peflin isoform X1
PGAP1	0,000204905	0,000204905	GPI inositol-deacylase
PHKA2	0,014935656	0,010103684	Phosphorylase b kinase regulatory subunit alpha, liver isoform
PIANP	0,000361599	0,005376819	PILR alpha-associated neural protein
PICK1	0,039117349	0,02152393	PRKCA-binding protein
PIK3CB	0,001207944	0,001208143	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform
PKDREJ	0,009911732	0,009911743	Polycystic kidney disease and receptor for egg jelly-related protein
PLA2G15	0,001472987	0,001472979	Group XV phospholipase A2 isoform X1
PLCB4	0,000468559	0,000468559	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-4
PLD3	0,000402284	0,000402279	Phospholipase D3 isoform X1
POU5F1	0,004496792	0,004496812	POU domain, class 5, transcription factor 1
PPP1R13L	0,025971609	0,025971609	Rela-associated inhibitor
PPP2R1B	0,008759964	0,010537958	Serine/threonine-protein phosphatase 2A 65 kda regulatory subunit A beta isoform
PPP3CC	0,027236161	0,027236161	Serine/threonine-protein phosphatase 2B catalytic subunit gamma isoform
PROS1	0,00201528	1	Vitamin K-dependent protein S
PRPF40B	5,61E-04	0,000561435	Pre-mrna-processing factor 40 homolog B isoform X2
PSMC3IP	0,002191897	0,0021919	Homologous-pairing protein 2 homolog
PTGDR2	0,003898742	0,004304122	Prostaglandin D2 receptor 2
PTGDS	0,002869152	0,00286918	Prostaglandin-H2 D-isomerase
PXN	0,002957513	0,002957513	Paxillin-like isoform X1
RAB11FIP1	0,001387575	0,001336503	Rab11 family-interacting protein 1
RABGAP1	0,00182756	0,00182756	Rab gtpase-activating protein 1
RAD23B	2,95E-03	2,95E-03	UV excision repair protein RAD23 homolog B
RASAL1	0,00837414	0,00837414	Rasgap-activating-like protein 1
RASGEF1C	0,003660074	0,003075962	Ras-GEF domain-containing family member 1C
RBM15	0,011418582	0,011418582	Putative RNA-binding protein 15
RDH14	0,009613053	0,009612686	Retinol dehydrogenase 14
RELA	0,005034489	0,005034489	Transcription factor p65
REPIN1	2,31E-14	2,67E-15	Replication initiator 1 isoform X3
RET	0,001705078	0,001705078	Proto-oncogene tyrosine-protein kinase receptor Ret
RHPN1	0,00247415	0,002475146	Rhophilin-1
RNF112	0,013130143	0,009246825	RING finger protein 112
RNF157	0,35740014	0,025424502	RING finger protein 157
RNMTL1	2,49E-12	2,49E-12	RNA methyltransferase-like protein 1
ROM1	0,000813878	0,000851014	Rod outer segment membrane protein 1
RPS6KA6	0,003755993	0,003756064	Ribosomal protein S6 kinase alpha-6
RPS6KL1	0,045554613	0,063805019	Ribosomal protein S6 kinase-like 1
RRM1	0,025329722	0,042438241	Ribonucleoside-diphosphate reductase large subunit
RRP7A	0,002473864	0,002477741	Ribosomal RNA-processing protein 7 homolog A
RRS1	0,000546162	0,000509652	Ribosome biogenesis regulatory protein homolog

SASH3	0,007200203	0,010820507	SAM and SH3 domain-containing protein 3
SCAP	0,000753881	0,000753886	Sterol regulatory element-binding protein cleavage-activating protein
SCARF1	1,45E-03	1,45E-03	Scavenger receptor class F member 1
SEL1L2	0,000214047	0,000214047	Protein sel-1 homolog 2
SEMA5B	0,012619902	0,015373816	Semaphorin-5B
SHISA4	0,001358844	0,001358844	Protein shisa-4
SIN3B	0,000167343	0,000167343	Paired amphipathic helix protein Sin3b
SLC16A5	0,000642864	0,000642865	Monocarboxylate transporter 6 isoform X4
SLC22A13	0,001181209	0,002679013	Solute carrier family 22 member 13
SLC29A3	0,005296829	0,005297076	Equilibrative nucleoside transporter 3
SLC35F1	0,997743245	0,008879034	Solute carrier family 35 member F1
SLC8A3	0,003963303	0,005530073	Sodium/calcium exchanger 3 isoform X1
SMOX	0,00245463	0,002110644	Spermine oxidase
SNAPC1	0,000318842	0,000318859	Snrna-activating protein complex subunit 1
SNX15	0,004785307	0,004688173	Sorting nexin-15
SOX13	0,035122908	0,035122908	Transcription factor SOX-13 isoform X3
SPPL2A	0,00196383	1	Signal peptide peptidase-like 2A
SREBF2	1,13E-03	1,13E-03	Sterol regulatory element-binding protein 2
STXBP5	0,000131775	0,000131775	Syntaxin-binding protein 5 isoform X1
TBC1D21	0,000451383	0,000449287	TBC1 domain family member 21 isoform X2
TEP1	0,001312896	0,001312896	Telomerase protein component 1 isoform X3
TERF2IP	0,011731317	0,011731317	Telomeric repeat-binding factor 2-interacting protein 1
TFDP1	0,000355085	0,000355087	Transcription factor Dp-1
THAP4	0,000620766	0,000620766	THAP domain-containing protein 4
THOC2	0,995212721	0,047588075	THO complex subunit 2
TKTL1	0,002265836	0,004483486	Transketolase-like protein 1
TMEM129	0,000642532	0,000642532	Transmembrane protein 129
TMEM132A	0,000687067	0,000522469	Transmembrane protein 132A
TMEM205	2,69E-02	2,69E-02	Transmembrane protein 205 isoform X1
TMEM67	3,65E-03	1,88E-03	Meckelin isoform X1
TNFSF13	0,014652064	0,014636664	Tumor necrosis factor ligand superfamily member 13
TNIK	0,000133747	0,000133747	TRAF2 and NCK-interacting protein kinase
TNKS2	6,52E-07	2,34E-07	Tankyrase-2
TRPV4	0,006340586	0,006340586	Transient receptor potential cation channel subfamily V member 4
TSPAN32	0,000502943	0,001077014	Tetraspanin-32
TST	0,01105106	0,009884158	Thiosulfate sulfurtransferase
TTC24	0,880964658	0,00361269	Tetratricopeptide repeat protein 24
TTC7B	0,000418692	4,33E-04	Tetratricopeptide repeat protein 7B isoform X4
TXNDC16	5,82E-06	5,48E-06	Thioredoxin domain-containing protein 16
TYR	0,000145694	0,000180153	Tyrosinase
UBE2O	0,000576474	0,000596994	Ubiquitin-conjugating enzyme E2 O
UNC13D	0,001179001	0,001391734	Protein unc-13 homolog D
UTP3	0,001412955	0,001412958	Something about silencing protein 10
VBP1	0,004157002	0,004157002	Prefoldin subunit 3
VPS51	0,021393216	0,02313833	Vacuolar protein sorting-associated protein 51 homolog
WDR44	1,56E-03	1,56E-03	WD repeat-containing protein 44 isoform X1

WFIKKN2	1,97E-03	0,025577031	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2
WNT10B	5,84E-05	5,74E-05	Protein Wnt-10b
XRCC3	0,000778833	0,000778833	DNA repair protein XRCC3
YIPF4	0,001743874	0,001743874	Protein YIPF4
ZBED1	0,005747754	0,005747766	Zinc finger BED domain-containing protein 1
ZBTB46	0,00353891	0,003538844	Zinc finger and BTB domain-containing protein 46
ZCCHC12	0,016468565	0,016468565	Zinc finger CCHC domain-containing protein 12
ZDHHC14	0,011613717	0,011613717	Probable palmitoyltransferase ZDHHC14
ZFP36	0,004818915	0,004818915	Tristetraprolin
ZMYM3	0,000520854	0,000415752	Zinc finger MYM-type protein 3
ZMYND12	8,41E-03	8,41E-03	Zinc finger MYND domain-containing protein 12 isoform X1
ZNF333	0,026972851	0,026972851	Zinc finger protein 333
ZNF37A	0,001865626	0,001865626	Zinc finger protein 37A-like
ZNF395	0,023462897	0,043787425	Zinc finger protein 395
ZNF699	0,004406284	0,004406284	Zinc finger protein 699-like
ZNF836	1,12E-09	0,998871621	Zinc finger protein 836-like isoform X1
ZSCAN2	1,49E-14	1,04E-13	Zinc finger and SCAN domain-containing protein 2 isoform X1
ZSWIM8	0,000131701	0,000129008	Zinc finger SWIM domain-containing protein 8 isoform X1

Supplementary Table S24: **Gene ontology results for the tiger.** Analysis refers to the analysis performed in WebGestalt.com with a significance threshold < 0.05 and multiple correct testing (FDR). Characteristic refers to the enriched category found. Parameters describes enrichment details. C (Genes presented in the category); O (Genes observed in the dataset); R (Ratio of enrichment); rawP (p-value for the hypergeometric test); adjP (adjusted P value by the multiple correction test).

KEGG Pathway	Parameters	Gene
Calcium Signaling Pathway	C=177; O=7;E=1.01;R=6.93;rawP=7.59e-05;adjP=0.0017	*PLCB4; PPP3CC; CACNA1F; CACNA1D; SLC8A3; PHKA2; ATP2B1*
Wnt Signaling Pathway	C=150; O=7;E=0.86;R=8.18;rawP=2.65e-05;adjP=0.0017	*PLCB4; PPP3CC; LRP5; DVL1; WNT10B; PPP2R1B; FZD1*
Cysteine and Methionine Metabolism	C=36; O=4;E=0.21;R=19.48;rawP=5.27e-05;adjP=0.0017	*AHCY; MTR; DNMT1; TST*
Melanogenesis	**C=101; O=5;E=0.58;R=8.68;rawP=0.0003;adjP=0.0034**	***PLCB4; TYR; DVL1; WNT10B; FZD1***
Chagas Disease (*American trypanosomiasis*)	C=104; O=5;E=0.59;R=8.43;rawP=0.0003;adjP=0.0034	*PLCB4; CD3D; RELA; PIK3CB; PP2R1B*
Notch Signaling Pathway	C=47; O=4;E=0.27;R=14.92;rawP=0.0002;adjP=0.0034	*NOTCH1; DVL1; NOTCH3; DLL4*
MAPK Signaling Pathway	C=268; O=7;E=1.53;R=4.58;rawP=0.0009;adjP=0.0089	*DUSP2; PPP3CC; CACNA1F; CACNA1D; RELA; NR4A1; RPS6KA6*
Olfactory Transduction	C=388; O=9;E=2.51;R=3.59;rawP=0.0010; adjP=0.0072	*LOC102960583; LOC102967234; LOC102957797; LOC102950516; LOC102970816; LOC102952556; LOC102971360; LOC102951432*
Hepatitis C	C=134; O=5;E=0.76;R=6.54;rawP=0.0011;adjP=0.0095	*RELA; PIK3CB; CLDN7; PPP2R1B; IRF9*
Progesterone-mediated Oocyte Maturation	C=86; O=4;E=0.49;R=8.15;rawP=0.0015;adjP=0.0115	*CCNB3; PIK3CB; PDE3A; RPS6KA6*
Carbohydrate Digestion and Absorption	C=44; O=3;E=0.25;R=11.95;rawP=0.0020;adjP=0.0138	*CACNA1D; ATP1A1; PIK3CB*
Metabolic Pathways	C=1130; O=15;E=6.45;R=2.33;rawP=0.0023;adjP=0.0144	*PC; AK7; PTGDS; DHCR24; NNT; RRM1; PLCB4; TYR; AHCY; MTR;*

		PGAP1; DNMT1; TKTL1; GLYCTK; TST
Pathways in Cancer	C=326; O=7;E=1.86;R=3.76;rawP=0.0028;adjP=0.0161	*DVL1; WNT10B; RELA; PIK3CB; ARNT; RET; FZD1*
T cell Receptor Signaling Pathway	C=108; O=4;E=0.62;R=6.49;rawP=0.0035;adjP=0.0173	*CD3D; PPP3CC; RELA; PIK3CB*
Amoebiasis	C=106; O=4;E=0.60;R=6.62;rawP=0.0033;adjP=0.0173	*PLCB4; RELA; ACTN3; PIK3CB*
Basal Cell Carcinoma	C=55; O=3;E=0.31;R=9.56;rawP=0.0039;adjP=0.0179	*DVL1; WNT10B; FZD1*
Vascular Smooth Muscle Contraction	**C=116; O=4;E=0.66;R=6.05;rawP=0.0045;adjP=0.0180**	***PLCB4; CACNA1F; CACNA1D; ARHGEF11***
Leukocyte Transendothelial Migration	C=116; O=4;E=0.66;R=6.05;rawP=0.0045;adjP=0.0180	*PXN; ACTN3; PIK3CB; CLDN7*
Chemokine Signaling Pathway	C=189; O=5;E=1.08;R=4.64;rawP=0.0047;adjP=0.0180	*PLCB4; PXN; GRK1; RELA; PIK3CB*
Endocytosis	C=201; O=5;E=1.15;R=4.36;rawP=0.0061;adjP=0.0207	*AP2A2; GRK1; ACAP3; RET; RAB11FIP1*
Osteoclast Differentiation	C=128; O=4;E=0.73;R=5.48;rawP=0.0063;adjP=0.0207	*PPP3CC; RELA; PIK3CB; IRF9*
Focal Adhesion	C=200; O=5;E=1.14;R=4.38;rawP=0.0060;adjP=0.0207	*PXN; PARVG; ACTN3; PIK3CB; COL6A1*
Tight Junction	C=132; O=4;E=0.75;R=5.31;rawP=0.0071;adjP=0.0223	*ACTN3; MYH1; CLDN7; PPP2R1B*
Long-term Potentiation	**C=70; O=3;E=0.40;R=7.51;rawP=0.0076;adjP=0.0228**	***PLCB4 ; PPP3CC; RPS6KA6***
Dorso-ventral Axis Formation	C=24; O=2;E=0.14;R=14.61;rawP=0.0082;adjP=0.0236	*NOTCH1; NOTCH3*
VEGF Signaling Pathway	C=76; O=3;E=0.43;R=6.92;rawP=0.0095;adjP=0.0242	*PPP3CC; PXN; PIK3CB*
Arrhythmogenic Right Ventricular Cardiomyopathy	C=74; O=3;E=0.42;R=7.11;rawP=0.0088;adjP=0.0242	*CACNA1F; CACNA1D; ACTN3*
B cell Receptor Signaling Pathway	C=75; O=3;E=0.43;R=7.01;rawP=0.0092;adjP=0.0242	*PPP3CC; RELA; PIK3CB*
Cardiac Muscle Contraction	**C=77; O=3;E=0.44;R=6.83;rawP=0.0098;adjP=0.0242**	***CACNA1F; CACNA1D; ATP1A1***
Protein Digestion and Absorption	**C=81; O=3;E=0.46;R=6.49;rawP=0.0113;adjP=0.0260**	***ATP1A1; SLC8A3; COL6A1***
Homologous Recombination	C=28; O=2;E=0.16;R=12.52;rawP=0.0111;adjP=0.0260	*XRCC3; EME1*
TGF-beta Signaling Pathway	C=84; O=3;E=0.48;R=6.26;rawP=0.0124;adjP=0.0276	*TFDP1; PPP2R1B; INHBE*
Apoptosis	C=87; O=3;E=0.50;R=6.05;rawP=0.0137;adjP=0.0295	*PPP3CC; RELA; PIK3CB*
Salivary Secretion	C=89; O=3;E=0.51;R=5.91;rawP=0.0145;adjP=0.0303	*PLCB4; ATP1A1; ATP2B1*
Alzheimer's Disease	**C=167; O=4;E=0.95;R=4.20;rawP=0.0157;adjP=0.0319**	***PLCB4; PPP3CC; CACNA1F; CACNA1D***
GnRH Signaling Pathway	**C=101; O=3;E=0.58;R=5.21;rawP=0.0203;adjP=0.0389**	***PLCB4; CACNA1F; CACNA1D***
Pancreatic Secretion	C=101; O=3;E=0.58;R=5.21;rawP=0.0203;adjP=0.0389	*PLCB4; ATP1A1; ATP2B1*
Aldosterone-regulated Sodium Reabsorption	C=42; O=2;E=0.24;R=8.35;rawP=0.0240;adjP=0.0448	*ATP1A1; PIK3CB*
Oocyte Meiosis	C=112; O=3;E=0.64;R=4.70;rawP=0.0265;adjP=0.0481	*PPP3CC; PPP2R1B; RPS6KA6*

Disease	Parameters	Gene
Von Hippel-Lindau Disease	C=38; O=6;E=0.22;R=27.68;rawP=7.68e-08;adjP=3.30e-05	*VBP1; NR4A1; ARNT; RET; CA9; HIF3A*
Urinary Bladder Neoplasms	C=138; O=7;E=0.79;R=8.89;rawP=1.55e-05;adjP=0.0033	*TMEM129; POU5F1; MTR; DNMT1; RAD23B; CA9*

Cancer or Viral Infections	C=951; O=16;E=5.42;R=2.95;rawP=0.0001;adjP=0.0107	*NOTCH1 ; DLL4; RAD23B ; RET; TEP1; RRM1; FANCD2; XRCC3; PPP1R13L; OU5F1 ; RELA ; DNMT1; MTR; TKTL1; CLDN7; CA9*
Gonadal Dysgenesis	C=45; O=4;E=0.26;R=15.58;rawP=0.0001;adjP=0.0107	*POU5F1; PSMC3IP; EFHC2; NR5A1*
Drug Interaction with Drug	C=349; O=9;E=1.99;R=4.52;rawP=0.0002;adjP=0.0143	*SIN3B; PXN; CHRNA4; GPR124; FANCD2; RELA; ARNT; PIK3CB; PICK1*
Xeroderma Pigmentosum	C=96; O=5;E=0.55;R=9.13;rawP=0.0002;adjP=0.0143	*TERF2IP; XRCC3; PPP1R13L; EME1; RAD23B*
Neoplasms, Basal Cell	C=103; O=5;E=0.59;R=8.51;rawP=0.0003;adjP=0.0161	*NOTCH1; XRCC3; RELA; RET; CA9*
Adenocarcinoma	C=362; O=9;E=2.06;R=4.36;rawP=0.0003;adjP=0.0161	*NOTCH1; RET; RRM1; XRCC3; POU5F1; DNMT1; CLDN7; TKTL1; CA9*
Tic Disorders	C=33; O=3;E=0.19;R=15.94;rawP=0.0009;adjP=0.0206	*DLGAP3; LHX6; ADRA2A*
Vascular Diseases	C=357; O=8;E=2.04;R=3.93;rawP=0.0011;adjP=0.0206	*TNFSF13; MEF2A; PROS1; NOTCH3; SREBF2; AGER; MTR; ATP2B1*
Adenomatous Polyposis Coli	C=78; O=4;E=0.44;R=8.99;rawP=0.0011;adjP=0.0206	*DVL1; MAPRE3; MYH1; FZD1*
Carcinoma, Renal Cell	C=121; O=5;E=0.69;R=7.24;rawP=0.0007;adjP=0.0206	*VBP1; CLDN7; CA9; RPS6KA6; HIF3A*
Translocation, Genetic	C=431; O=9;E=2.46;R=3.66;rawP=0.0009;adjP=0.0206	*SLC8A3; SLC29A3; MKL1; PBX1; TYR; RBM15; RELA; ELF4; ADRA2A*
Granulosa Cell Tumor	C=31; O=3;E=0.18;R=16.97;rawP=0.0007;adjP=0.0206	*POU5F1; NR4A1; NR5A1*
Disorder of Uterus NOS	C=135; O=5;E=0.77;R=6.49;rawP=0.0011;adjP=0.0206	*NOTCH1; LHX6; TKTL1; CA9; INHBE*
Neoplasm of Unspecified Nature of Digestive System	C=445; O=9;E=2.54;R=3.55;rawP=0.0011;adjP=0.0206	*NOTCH1; RRM1; XRCC3; DNMT1; MTR; MYH1; CLDN7; TKTL1; CA9*
Urologic Neoplasms	C=185; O=6;E=1.06;R=5.69;rawP=0.0007;adjP=0.0206	*XRCC3; TMEM129; POU5F1; MTR; CLDN7; CA9*
Vascular Resistance	C=75; O=4;E=0.43;R=9.35;rawP=0.0009;adjP=0.0206	*DLL4; EMILIN1; AGER; CA9*
Congenital, Hereditary, and Neonatal Diseases and Abnormalities	C=78; O=4;E=0.44;R=8.99;rawP=0.0011;adjP=0.0206	*CACNA1F; AK7; RET; TRPV4*
Carcinoma	C=522; O=10;E=2.98;R=3.36;rawP=0.0009;adjP=0.0206	*NOTCH1; RET; RRM1; XRCC3; PPP1R13L; POU5F1; DNMT1; CLDN7; TKTL1; CA9*
Gastrointestinal Diseases	C=413; O=9;E=2.36;R=3.82;rawP=0.0007;adjP=0.0206	*FOXP3; RET; XRCC3; DNMT1; MTR; MYH1; CLDN7; TKTL1; CA9*
Intestinal Diseases	C=331; O=8;E=1.89;R=4.24;rawP=0.0007;adjP=0.0206	*FOXP3; RET; XRCC3; DNMT; MTR; MYH1; CLDN7; CA9*
Uterine Neoplasms	C=136; O=5;E=0.78;R=6.45;rawP=0.0011;adjP=0.0206	*NOTCH1; LHX6; TKTL1; CA9; INHBE*
Carcinoma, Embryonal	C=87; O=4;E=0.50;R=8.06;rawP=0.0016;adjP=0.0215	*TSPAN32; POU5F1; TFDP1; NR5A1*
Lymphatic Diseases	C=363; O=8;E=2.07;R=3.86;rawP=0.0012;adjP=0.0215	*UNC13D; PBX1; TNFSF13; FOXP3; NOTCH1; MTR; SLC29A3; ANXA11*
Disease Progression	C=220; O=6;E=1.25;R=4.78;rawP=0.0017;adjP=0.0215	*FOXP3; LTBP4; AGER; TKTL1; CA9; RRM1*
Adhesion	C=647; O=11;E=3.69;R=2.98;rawP=0.0013;adjP=0.0215	*TERF2IP; PXN; PCDH8; EMILIN1; GPR124; MDGA1;*

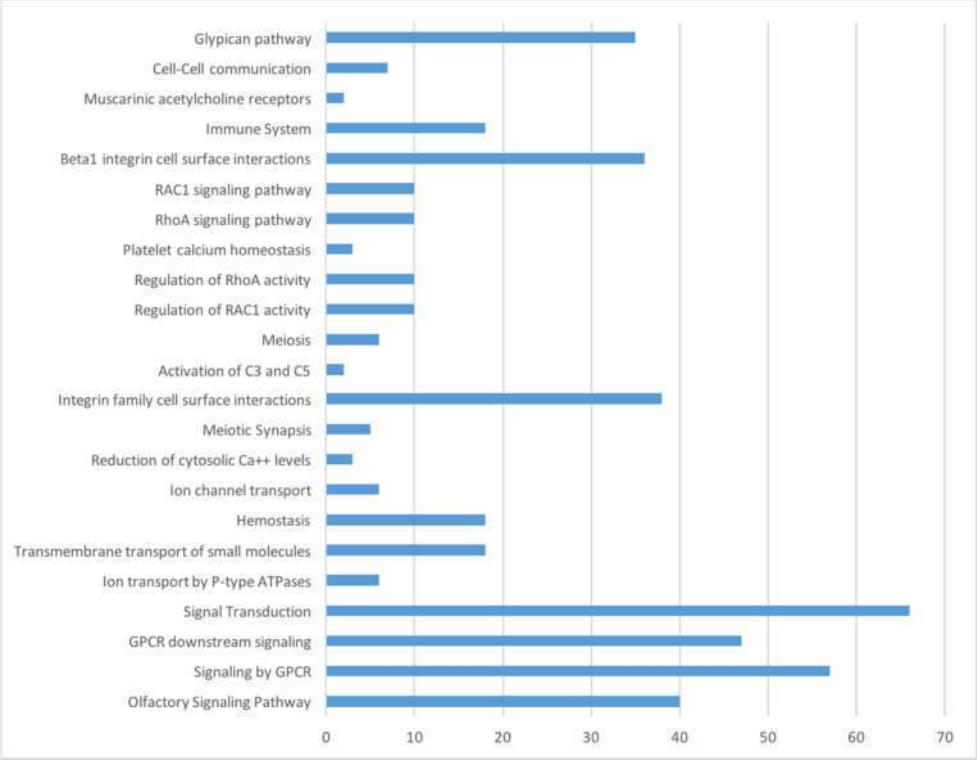
		SCARF1; CLDN7; PCDHB4; CA9; LPXN
Respiratory Tract Diseases	C=373; O=8;E=2.13;R=3.76;rawP=0.0015;adjP=0.0215	*FOXP3; IREB2; TRPV4; RRM1; XRCC3; PTGDR2; PPP1R13L; CA9*
Tourette Syndrome	C=40; O=3;E=0.23;R=13.15;rawP=0.0015;adjP=0.0215	*DLGAP3; LHX6; ADRA2A*
Retinoblastoma	C=221; O=6;E=1.26;R=4.76;rawP=0.0018;adjP=0.0215	*SIN3B; PXN; ARNT; DNMT1; TFDP1; RPS6KA6*
Megakaryocytic Leukaemia	C=41; O=3;E=0.23;R=12.83;rawP=0.0017;adjP=0.0215	*NFE2; RBM15; MKL1*
Arterial Occlusive Diseases	C=219; O=6;E=1.25;R=4.80;rawP=0.0017;adjP=0.0215	*MEF2A; SREBF2; AGER; SCAP; NR4A1; MTR*
Wounds and Injuries	C=90; O=4;E=0.51;R=7.79;rawP=0.0018;adjP=0.0215	*XRCC3; LRP5; PTGDS; RET*
Albinism, Ocular	C=86; O=4;E=0.49;R=8.15;rawP=0.0015;adjP=0.0215	*TYR; CLCN4; CACNA1F; ROM1*
Neoplasms	C=854; O=13;E=4.87;R=2.67;rawP=0.0014;adjP=0.0215	*TNFSF13; NOTCH1; RAD23B; RET; RRM1; XRCC3; PPP1R13L; POU5F1; DNMT1; MTR; CLDN7; TKTL1; CA9*
Patent Foramen Ovale	C=11; O=2;E=0.06;R=31.88;rawP=0.0017;adjP=0.0215	*NOTCH3; LTBP4*
Hypertension	C=227; O=6;E=1.29;R=4.63;rawP=0.0020;adjP=0.0232	*ATP1A1; NOTCH3; EMILIN1; PTGDS; ATP2B1; ADRA2A*
Adenoma	C=157; O=5;E=0.90;R=5.58;rawP=0.0021;adjP=0.0238	*MYH1; GCM2; CLDN7; NR5A1; RET*
Myoclonic Epilepsy, Juvenile	C=45; O=3;E=0.26;R=11.69;rawP=0.0022;adjP=0.0243	*CHRNA4; CLN6; EFHC2*
Neurodegenerative Diseases	C=404; O=8;E=2.30;R=3.47;rawP=0.0024;adjP=0.0256	*PRPF40B; CHRNA4; DHCR24; TRPV4; FGD4; PARL; CLN6; AGER*
Hypospadias	C=47; O=3;E=0.27;R=11.19;rawP=0.0025;adjP=0.0256	*TMEM67; CCDC8; NR5A1*
Neoplasms, Squamous Cell	C=236;O=6;E=1.35;R=4.46;rawP=0.0025;adjP=0.0256	*NOTCH1; XRCC3; TKTL1; CLDN7; CA9; RRM1*
Cranial Nerve Diseases	C=101;O=4;E=0.58;R=6.94;rawP=0.0027;adjP=0.0268	*PARL; TYR; CACNA1F; TRPV4*
Carcinoma, Squamous Cell	C=242;O=6;E=1.38;R=4.35;rawP=0.0028;adjP=0.0268	*NOTCH1; XRCC3; TKTL1; CLDN7; CA9; RRM1*
Recurrence	C=167;O=5;E=0.95;R=5.25;rawP=0.0028;adjP=0.0268	*NOTCH3; TKTL1; CLDN7; RET; CA9*
Dementia	C=247;O=6;E=1.41;R=4.26;rawP=0.0031;adjP=0.0290	*PLCB4 ; PRPF40B ; NOTCH3 ; CHRNA4; AGER; DHCR24*
Blindness	C=172;O=5;E=0.98;R=5.10;rawP=0.0032;adjP=0.0293	*CACNA1F; LRP5; GRK1; CLN6; ROM1*
Carcinoma, Large Cell	C=110;O=4;E=0.63;R=6.38;rawP=0.0037;adjP=0.0294	*CLDN7; RET; CA9; RRM1*
Pigmentation Disorders	C=110;O=4;E=0.63;R=6.38;rawP=0.0037;adjP=0.0294	*TYR; CLCN4; SLC29A3; RAD23B*
Migraine without Aura	C=56;O=3;E=0.32;R=9.39;rawP=0.0041;adjP=0.0294	*NOTCH3 ; GABRQ ; ATP2C2*
Inflammation	C=435;O=8;E=2.48;R=3.22;rawP=0.0037;adjP=0.0294	*TNFSF13; FOXP3; ANGPTL2; ZFP36; PTGDR2; RELA; AGER; C3AR1*
Carcinoma, Non-Small-Cell Lung	C=177;O=5;E=1.01;R=4.95;rawP=0.0036;adjP=0.0294	*XRCC3; PPP1R13L; TKTL1; CA9; RRM1*
Kidney Diseases, Cystic	C=56;O=3;E=0.32;R=9.39;rawP=0.0041;adjP=0.0294	*TMEM67; PKDREJ; CA9*
Head and Neck Neoplasms	C=262;O=6;E=1.49;R=4.01;rawP=0.0041;adjP=0.0294	*NOTCH1; XRCC3; GCM2; CLDN7; RET; CA9*

Colorectal Neoplasms	C=260;O=6;E=1.48;R=4.05;rawP=0.0039;adjP=0.0294	*XRCC3; MYH1; MTR; DNMT1; TKTL1; CA9*
Carcinoma, Neuroendocrine	C=56;O=3;E=0.32;R=9.39;rawP=0.0041;adjP=0.0294	*NOTCH1; CLDN7; RET*
Oat Cell Carcinoma of Lung	C=181;O=5;E=1.03;R=4.84;rawP=0.0039;adjP=0.0294	*XRCC3; PPP1R13L; TKTL1; CA9; RRM1*
Metabolism, Inborn Errors	C=346;O=7;E=1.97;R=3.55;rawP=0.0039;adjP=0.0294	*TYR; CLCN4; PC; PHKA2; CLN6; MTR; GLYCTK*
Migraine with Aura	C=56;O=3;E=0.32;R=9.39;rawP=0.0041;adjP=0.0294	*NOTCH3; GABRQ; ATP2C2*
Brain Stem Neoplasms	C=53;O=3;E=0.30;R=9.92;rawP=0.0035;adjP=0.0294	*NOTCH1; POU5F1; RET*
Pulmonary Edema	C=17;O=2;E=0.10;R=20.63;rawP=0.0042;adjP=0.0296	*ATP1A1; AGER*
Gastrointestinal Neoplasms	C=354;O=7;E=2.02;R=3.47;rawP=0.0044;adjP=0.0305	*XRCC3; MYH1; MTR; DNMT1; TKTL1; CLDN7; CA9*
Quadriplegia	C=18;O=2;E=0.10;R=19.48;rawP=0.0047;adjP=0.0311	*UNC13D; MYH1*
Intestinal Neoplasms	C=268;O=6;E=1.53;R=3.92;rawP=0.0046;adjP=0.0311	*XRCC3; MYH1; MTR; DNMT1; CLDN7; CA9*
Hemiplegia	C=18;O=2;E=0.10;R=19.48;rawP=0.0047;adjP=0.0311	*UNC13D; MYH1*
Thrombosis	C=120;O=4;E=0.68;R=5.84;rawP=0.0051;adjP=0.0324	*PROS1; MTR; PDE3A; STXBP5*
Uterine Cervical Neoplasms	C=121;O=4;E=0.69;R=5.80;rawP=0.0052;adjP=0.0324	*NOTCH1; LHX6; TKTL1; CA9*
Preleukemia	C=19;O=2;E=0.11;R=18.45;rawP=0.0052;adjP=0.0324	*PBX1; NOTCH1*
Hip Injuries	C=19;O=2;E=0.11;R=18.45;rawP=0.0052;adjP=0.0324	*LRP5; WNT10B*
Carcinoma, Small Cell	C=278;O=6;E=1.59;R=3.78;rawP=0.0054;adjP=0.0327	*NOTCH1; PPP1R13L; POU5F1; TKTL1; CA9; RRM1*
Mental Disorders	C=564;O=9;E=3.22;R=2.80;rawP=0.0054;adjP=0.0327	*DLGAP3; PPP3CC; NOTCH3; CHRNA4; MDGA1; PTGDS; DHCR24; PICK1; ADRA2A*
Colonic Diseases	C=279;O=6;E=1.59;R=3.77;rawP=0.0055;adjP=0.0328	*XRCC3; MYH1; MTR; DNMT1; CLDN7; RET*
Thrombocytosis	C=21;O=2;E=0.12;R=16.70;rawP=0.0063;adjP=0.0356	*NFE2; PROS1*
Carcinoma in Situ	C=128;O=4;E=0.73;R=5.48;rawP=0.0063;adjP=0.0356	*DLL4;LHX6; CLDN7; RET*
Arthritis, Juvenile Rheumatoid	C=65;O=3;E=0.37;R=8.09;rawP=0.0062;adjP=0.0356	*UNC13D; FOXP3; AGER*
Histiocytosis	C=21;O=2;E=0.12;R=16.70;rawP=0.0063;adjP=0.0356	*UNC13D; SLC29A3*
Sinusitis	C=67;O=3;E=0.38;R=7.85;rawP=0.0067;adjP=0.0369	*FOXP3; PTGDR2; AK7*
Night Blindness	C=67;O=3;E=0.38;R=7.85;rawP=0.0067;adjP=0.0369	*CACNA1F; GRK1; ROM1*
Epilepsies, Myoclonic	C=68;O=3;E=0.39;R=7.73;rawP=0.0070;adjP=0.0376	*CHRNA4; CLN6; EFHC2*
Anoxia	C=68;O=3;E=0.39;R=7.73;rawP=0.0070;adjP=0.0376	*ARNT; CA9; HIF3A*
Hypogonadism	C=69;O=3;E=0.39;R=7.62;rawP=0.0073;adjP=0.0388	*HS6ST1; SLC29A3; NR5A1*
Hermaphroditism	C=70;O=3;E=0.40;R=7.51;rawP=0.0076;adjP=0.0399	*POU5F1; EFHC2; NR5A1*
Disorder of Sexual Differentiation	C=71;O=3;E=0.40;R=7.41;rawP=0.0079;adjP=0.0404	*POU5F1; EFHC2; NR5A1*
Arteriosclerosis	C=214;O=5;E=1.22;R=4.10;rawP=0.0079;adjP=0.0404	*MEF2A; SREBF2; AGER; SCAP; MTR*

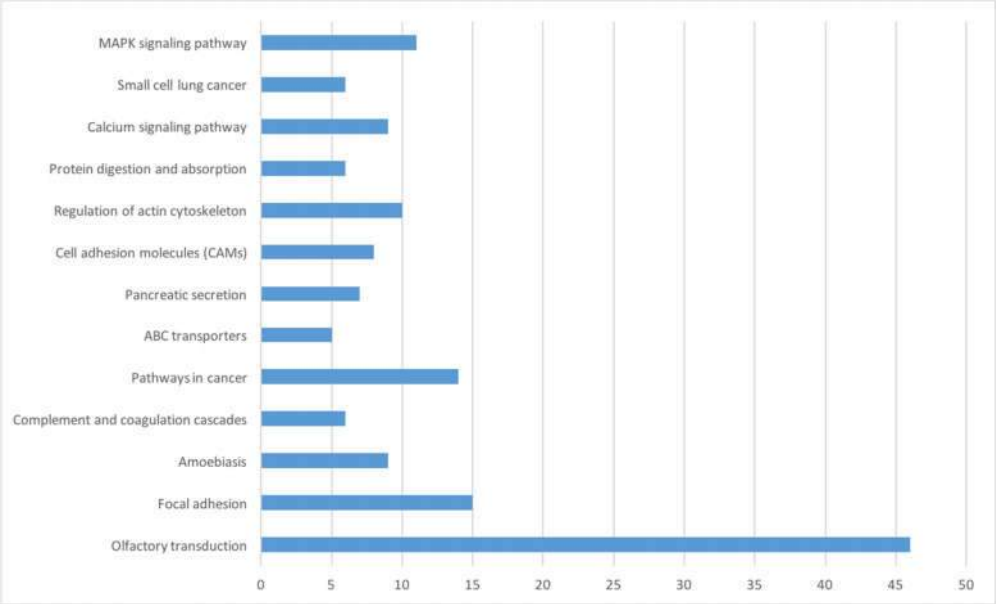
Precursor Cell Lymphoblastic Leukemia-Lymphoma	C=138;O=4;E=0.79;R=5.08;rawP=0.0082;adjP=0.0405	*PBX1; NOTCH1; NOTCH3; MTR*
Pseudohermaphroditism	C=72;O=3;E=0.41;R=7.30;rawP=0.0082;adjP=0.0405	*POU5F1; EFHC2; NR5A1*
Leukemia-Lymphoma, Adult T-Cell	C=72;O=3;E=0.41;R=7.30;rawP=0.0082;adjP=0.0405	*FOXP3; NOTCH1; NOTCH3*
Food Hypersensitivity	C=25;O=2;E=0.14;R=14.03;rawP=0.0089;adjP=0.0407	*FOXP3; PTGDR2*
Nystagmus, Pathologic	C=25;O=2;E=0.14;R=14.03;rawP=0.0089;adjP=0.0407	*TYR; CACNA1F*
Neural Tube Defects	C=73;O=3;E=0.42;R=7.20;rawP=0.0085;adjP=0.0407	*AHCY; TMEM67; MTR*
Lung Neoplasms	C=306;O=6;E=1.75;R=3.44;rawP=0.0085;adjP=0.0407	*XRCC3; PPP1R13L; POU5F1; PPP2R1B; CA9; RRM1*
Strabismus	C=74;O=3;E=0.42;R=7.11;rawP=0.0088;adjP=0.0407	*TYR; CACNA1F; TMEM67*
Sleep Initiation and Maintenance Disorders	C=25;O=2;E=0.14;R=14.03;rawP=0.0089;adjP=0.0407	*AHCY; SLC29A3*
Leukemia, Megakaryoblastic, Acute	C=25;O=2;E=0.14;R=14.03;rawP=0.0089;adjP=0.0407	*RBM15; MKL1*
Nervous System Malformations	C=221;O=5;E=1.26;R=3.97;rawP=0.0090;adjP=0.0407	*FGD4; TMEM67; MTR; SNX15; TRPV4*
Skin Neoplasms	C=145;O=4;E=0.83;R=4.84;rawP=0.0098;adjP=0.0430	*TYR ; FOXP3; NOTCH1; XRCC3*
Ovarian Failure, Premature	C=77;O=3;E=0.44;R=6.83;rawP=0.0098;adjP=0.0430	*PSMC3IP; NR5A1; INHBE*
Multiple Endocrine Neoplasia Type 1	C=26;O=2;E=0.15;R=13.49;rawP=0.0096;adjP=0.0430	*VPS51; SNX15*
Hyperhomocysteinemia	C=27;O=2;E=0.15;R=12.99;rawP=0.0104;adjP=0.0443	*AHCY; MTR*
Homocysteinemia	C=27;O=2;E=0.15;R=12.99;rawP=0.0104;adjP=0.0443	*AHCY; MTR*
Paresis	C=27;O=2;E=0.15;R=12.99;rawP=0.0104;adjP=0.0443	*MYH1; TRPV4*
Carcinoma, Basal Cell	C=80;O=3;E=0.46;R=6.57;rawP=0.0109;adjP=0.0450	*NOTCH1; XRCC3; CA9*
Glycogen Storage Disease	C=28;O=2;E=0.16;R=12.52;rawP=0.0111;adjP=0.0450	*PHKA2; ACTN3*
Protein Deficiency Disease	C=234;O=5;E=1.33;R=3.75;rawP=0.0113;adjP=0.0450	*FOXP3; PROS1; PC; PHKA2; ACTN3*
Male Urogenital Diseases	C=423;O=7;E=2.41;R=2.90;rawP=0.0112;adjP=0.0450	*XRCC3; POU5F1; AGER; PKDREJ; PSMC3IP; NR5A1; CA9*
Ischemic Attack, Transient	C=28;O=2;E=0.16;R=12.52;rawP=0.0111;adjP=0.0450	*NOTCH3; PLA2G15*
Color Vision Defects	C=28;O=2;E=0.16;R=12.52;rawP=0.0111;adjP=0.0450	*CACNA1F; PHKA2*
Dementia, Vascular	C=28;O=2;E=0.16;R=12.52;rawP=0.0111;adjP=0.0450	*NOTCH3; SREBF2*
Hyperostosis	C=29;O=2;E=0.17;R=12.09;rawP=0.0119;adjP=0.0469	*LRP5; COL6A1*
Myoglobinuria	C=30;O=2;E=0.17;R=11.69;rawP=0.0127;adjP=0.0487	*PHKA2; TST*
Adrenocortical Carcinoma	C=30;O=2;E=0.17;R=11.69;rawP=0.0127;adjP=0.0487	*TSPAN32; NR5A1*
Precancerous Conditions	C=85;O=3;E=0.48;R=6.19;rawP=0.0128;adjP=0.0487	*NOTCH1; XRCC3; RAD23B*
Synovitis	C=30;O=2;E=0.17;R=11.69;rawP=0.0127;adjP=0.0487	*TNFSF13; ANGPTL2*

6.5. *Site model results*

To assess the presence of overall signatures of positive selection in the *Panthera*, we used a site model test without branch constraint. We have found 779 candidate genes under selection for the gene tree dataset and 754 candidate genes for the species tree dataset. The enrichment tests for those genes showed some interesting patterns regarding body growth, olfactory perception, protein metabolism according to our gene enrichment analysis (Figure S8, S9).



Supplementary Figure S7 - **Pathway Commons Enrichment**: Results for the site model test using the gene tree dataset.



Supplementary Figure S8 – **KEGG enrichment analysis**: Results for the site model test using the gene tree dataset.

6.5.1. Body size, olfactory receptors and protein metabolism in *Panthera*

Increased body size is one of the most striking characteristics of the *Panthera,* since they comprise the largest species amongst the living Felidae[99]. Felidae is divided between two major lineages. Felinae, the small cats (the only exceptions being the Cheetah and the Cougar), and Pantherinae, which is comprised solely by the Panthera lineage (*Panthera* and *Neofelis*)[27]. The site model analysis retrieved some interesting patterns regarding growth and one of the enriched pathways was the glypican pathway. This route is highly conserved in mammals and is responsible for interacting with very important developmental pathways such as *WNT*, *SSH* and *BMP*[116]. Also, it interacts with fibroblast growth factors[117]. Considering their broad scope of functions and interactions, it may not be surprising that this pathway has several genes bearing signals of selection for one of the species analyzed in this study.

In addition, we have found several olfactory receptors to be under selection using the site model approach. Smell is one of the most important senses for felids, for both social interactions among individuals that share the same area and for hunting[99].

Another important characteristic found to be under selection was hypercarnivory. Species from *Panthera* are hypercarnivores, feeding almost exclusively on meat, the only exception being the snow leopard (see section 6.5.4 for more information). For this reason, is no surprise that they

present genes under selection for this trait. This pattern was already observed in previous studies[20,21] and the enrichment analyses here corroborate those results.

Supplementary Table S25: **Gene ontology results for site model analysis.** Analysis refers to the analysis performed in WebGestalt with a significance threshold < 0.05 and multiple correct testing (FDR). Characteristic refers to the enriched category found. Parameters describes enrichment details. C (Genes presented in the category); O (Genes observed in the dataset); R (Ratio of enrichment); rawP (p-value for the hypergeometric test); adjP (adjusted P value by the multiple correction test).

Disease	Parameters	Genes
Myocardial Infarction	C=242;O=20;E=3.85;R=5.19;rawP=2.73e-09;adjP=8.63e-07	*MMP9; F13B; MIA3; GYLTL1B; ATP10A; UGGT2; CD14; WRN; MEGF11; PECAM1; MYO7B; NOS1AP; MARCH10; CCDC86; HGF; TNNT2; SPTBN5; FGA; CD34; AGER*
Infarction	C=236;O=20;E=3.76;R=5.32;rawP=1.77e-09;adjP=8.63e-07	*MMP9; MIA3; GYLTL1B; UGGT2; CD14; WRN; MEGF11; PECAM1; MYO7B; NOS1AP; MARCH10; CCDC86; MMP8; HGF; COL4A1; TNNT2; SPTBN5; FGA; CD34; AGER*
Adhesion	C=647;O=34;E=10.31;R=3.30;rawP=1.91e-09;adjP=8.63e-07	*MMP9; EMR2; LASP1; PTPN23; PECAM1; VNN3; HMCN1; IGFBP7; ZYX; RAC2; ESAM; ICAM5; COL4A1; IPCEF1; CEACAM1; CD48; PODXL; PTPRM; FCGBP; TNS1; MAG; DIAPH1; CNTNAP1; TROAP; MUC1; IGSF11; GPR124; CLCA2; CDH17; ACTN4; DSG1; HGF; CD34; CDH2;*
Gastrointestinal Neoplasms	C=354;O=24;E=5.64;R=4.26;rawP=3.86e-09;adjP=9.15e-07	MMP9; KRT20; GALNT12; ERBB2; WISP1; MUC2; ABCC2; VEGFA; APC2; TKTL1; MUC5AC; IGFBP7; MDK; DMBT1; MUC1; MSH2; CDH17; CDH2; GATA5; CEACAM1; DNMT1; CBX7; MUC13; GLI1
Gastrointestinal Diseases	C=413;O=25;E=6.58;R=3.80;rawP=1.77e-08;adjP=2.89e-06	*MMP9; FOXP3; KRT20; GALNT12; ERBB2; WISP1; MUC2; ABCC2; CD14; VEGFA; IL22; TKTL1; MUC5AC; IGFBP7; CCR6; DMBT1; SLC26A3; MUC1; MSH2; CDH17; GATA5; CEACAM1; DNMT1; MUC13; GLI1*
Colonic Neoplasms	C=171;O=16;E=2.72;R=5.87;rawP=1.83e-08;adjP=2.89e-06	*KRT20; NEU3; GALNT12; ATP2A3; MUC2; VEGFA; PIGR; APC2; IGFBP7; SLC26A3; MUC1; MSH2; CDH2; CEACAM1; DNMT1; GLI1*
Adenocarcinoma	C=362;O=23;E=5.77;R=3.99;rawP=2.71e-08;adjP=3.67e-06	*MMP9; KRT20; CHD1L; ERBB2; MUC2; VEGFA; MCM2; TKTL1; MUC5AC; IGFBP7; DIRAS3; MDK; MUC1; MUC16; MSH2; CDH17; HGF; CDH2; GATA5; RNF139; CEACAM1; DNMT1 ; GLI1*
Colorectal Neoplasms	C=260;O=19;E=4.14;R=4.59;rawP=4.93e-08;adjP=5.84e-06	*MMP9; KRT20; GALNT12; WISP1; MUC2; VEGFA; APC2; TKTL1; MUC5AC; DFNA5; IGFBP7; DMBT1; SLC26A3; MUC1; MSH2; CDH17; CEACAM1; DNMT1; GLI1*
Respiratory Tract Infections	C=244;O=18;E=3.89;R=4.63;rawP=9.59e-08;adjP=9.09e-06	*MMP9; C3; FOXP3; PHF11; IL2RG; C5; MUC2; CD14; PECAM1; IL22; IKBKAP; IL18R1; MUC5AC; MMP8; CCDC40; PTGDR2; FCER1A; DOCK8*
Skin Neoplasms	C=145;O=14;E=2.31;R=6.06;rawP=9.53e-08;adjP=9.09e-06	*FOXP3; KRT20; NF1; PTPRD; FOXN1; MSH2; PTCH1; WISP1; GNAQ; APAF1; CD34; GLI1; PRPH; IGFBP7*
Intestinal Diseases	C=331;O=21;E=5.27;R=3.98;rawP=1.11e-07;adjP=9.57e-06	*FOXP3; KRT20; GALNT12; WISP1; MUC2; CD14; VEGFA; IL22; APC2; MUC5AC; IGFBP7; CCR6; DMBT1; SLC26A3; MUC1; SLC9A3; MSH2; CDH17; DNMT1; MUC13; GLI1*
Intracranial Hemorrhages	C=55;O=9;E=0.88;R=10.27;rawP=2.07e-07;adjP=1.64e-05	*MMP9; F13B; MARCH10; SERPINA3COL4A1; SPTBN5; MEGF11; VEGFA; MYO7B*
Stroke NOS	C=235;O=17;E=3.74;R=4.54;rawP=2.86e-07;adjP=1.83e-05	*MMP9; F13B; GYLTL1B; ARHGEF10; UGGT2; C5; CD14; MEGF11; MYO7B; MARCH10; CCDC86; SERPINA3; COL4A1; TNNT2; SPTBN5; FGA; SLC1A2*
Stroke	C=235;O=17;E=3.74;R=4.54;rawP=2.86e-07;adjP=1.83e-05	*MMP9; F13B; GYLTL1B; ARHGEF10; UGGT2; C5; CD14; MEGF11; MYO7B; MARCH10; CCDC86; SERPINA3; COL4A1; TNNT2; SPTBN5; FGA; SLC1A2*
Stomach Neoplasms	C=183;O=15;E=2.92;R=5.15;rawP=2.89e-07;adjP=1.83e-05	*MMP9; KRT20; MUC1; IGSF11; CDH17; MSH2; ERBB2; MUC2; VEGFA; GATA5; DNMT1; CBX7; MUC13; SAPCD2; MUC5AC;*

Common Cold	C=246;O=17;E=3.92;R=4.34;rawP=5.44e-07;adjP=3.22e-05	*MMP9; C3; FOXP3; PHF11; IL2RG; C5; MUC2; CD14; IL22; IKBKAP; IL18R1; MUC5AC; MMP8; CCDC40; PTGDR2; FCER1A; DOCK8*
Nervous System Neoplasms	C=194;O=15;E=3.09;R=4.85;rawP=6.12e-07;adjP=3.41e-05	*MMP9; DMBT1; RAC2; NF2; NF1; PTPRM; AQP1; PTCH1; ERBB2; CD34; VEGFA; PEG3; KIAA1549GLI1; PRPH*
Cancer or Viral Infections	C=951;O=37;E=15.15;R=2.44;rawP=8.09e-07;adjP=4.26e-05	*MMP9; NF2; NF1; TRIM5; ERBB2; MDC1; TKTL1; IGFBP7; DIRAS3; DMBT1; MUC16; MSH2; CEACAM1; DNMT1; GLI1; KRT20; WISP1; PTCH1; ABCC2; MUC2; WRN; MCM2; VEGFA; MDK; TET2; MAGEA10; MUC1; CLCA2; CDH17; BST2; HGF; APAF1; CD34; CDH2; DND1; TERF1; CBX7*
Subarachnoid Hemorrhage	C=104;O=11;E=1.66;R=6.64;rawP=9.04e-07;adjP=4.51e-05	*MMP9; MARCH10; LTBP4; CCDC86; SERPINA3GYLTL1B; UGGT2; ATP10A; SPTBN5; MEGF11; MYO7B*
Cerebral Hemorrhage	C=49;O=8;E=0.78;R=10.25;rawP=1.01e-06;adjP=4.74e-05	*MMP9; SPTBN5; COL4A1; MEGF11; MARCH10; MYO7B; SERPINA3; GYLTL1B*
Neoplasm of unspecified nature of digestive system	C=445;O=23;E=7.09;R=3.24;rawP=1.05e-06;adjP=4.74e-05	*MMP9; KRT20; CHD1L; ERBB2; WISP1; MUC2; ABCC2; VEGFA; TKTL1; MUC5AC; IGFBP7; DIRAS3; MDK; MUC1; MSH2; CDH17; HGFCDH2; GATA5; CEACAM1; DNMT1; MUC13; GLI1*
Skin Diseases	C=417;O=22;E=6.64;R=3.31;rawP=1.28e-06;adjP=5.52e-05	*MMP9; FOXP3; LASP1; CYP2D6; ABCC11; ATP2C1; PTCH1; ERBB2; KRT10; VEGFA; IL22; TRPS1; IGFBP7; DIRAS3; DST; MUC1; AGPAT2; CTSC; TACC2; FOXN1; DSG1; SCGB1D2*
Glioma	C=207;O=15;E=3.30;R=4.55;rawP=1.39e-06;adjP=5.65e-05	*MARK4; MMP9; DMBT1; RAC2; NF2; NF1; PTPRM; AQP1; PTCH1; VEGFA; PEG3; ATF5; SPOCK3; KIAA1549; GLI1*
Carcinoma	C=522;O=25;E=8.32;R=3.01;rawP=1.43e-06;adjP=5.65e-05	*MMP9; KRT20; PODXL; CHD1L; PTCH1; ERBB2; MUC2; ABCC2; VEGFA; MCM2; TKTL1; MUC5AC; IGFBP7; MDK; MUC1; MUC16; MSH2; CDH17; HGF; CDH2; GATA5; CEACAM1; DNMT1; CBX7; GLI1*
Cerebral Infarction	C=236;O=16;E=3.76;R=4.26;rawP=1.50e-06;adjP=5.69e-05	*MMP9; F13B; MARCH10; CCDC86; SERPINA3; GYLTL1B; ARHGEF10; UGGT2; SPTBN5; TNNT2; COL4A1; CD14; FGA; MEGF11; MYO7B; SLC1A2*
Bronchitis	C=237;O=16;E=3.78;R=4.24;rawP=1.58e-06;adjP=5.76e-05	*C3; FOXP3; PHF11; IL2RG; C5; MUC2; CD14; PECAM1; IL22; IKBKAP; IL18R1; MUC5AC; DPCR1; MUC1; CCDC40; FCER1A*
Intestinal Neoplasms	C=268;O=17;E=4.27;R=3.98;rawP=1.77e-06;adjP=6.21e-05	*KRT20; GALNT12; ERBB2; WISP1; MUC2; VEGFA; APC2; MUC5AC; IGFBP7; SLC26A3; MUC1; MSH2; CDH17; CEACAM1; DNMT1; MUC13; GLI1*
Bronchiolitis	C=217;O=15;E=3.46;R=4.34;rawP=2.49e-06;adjP=8.43e-05	*C3; DPCR1; FOXP3; MUC1; PHF11; IL2RG; C5; MUC2; CD14; FCER1A; PECAM1; IL22; IKBKAP; IL18R1; MUC5AC*
Bronchial Diseases	C=307;O=18;E=4.89;R=3.68;rawP=2.71e-06;adjP=8.86e-05	*MMP9; C3; FOXP3; PHF11; C5; MUC2; CD14; IL22; IKBKAP; IL18R1; MUC5AC; DCBLD2; DPCR1; MUC1; MMP8; CCDC40; PTGDR2; FCER1A*
Central Nervous System Neoplasms	C=173;O=13;E=2.76;R=4.72;rawP=4.66e-06;adjP=0.0001	*MMP9; DMBT1; RAC2; NF2; NF1; PTPRM; AQP1; PTCH1; ERBB2; VEGFA; PEG3; KIAA1549; GLI1*
Ischemia	C=146;O=12;E=2.33;R=5.16;rawP=4.29e-06;adjP=0.0001	*MMP9; C3; F13B; C5; HGF; TNNT2; CD14; FGA; VEGFA; CD34; AGER; SLC1A2*
epithelial cancers	C=316;O=18;E=5.03;R=3.58;rawP=4.06e-06;adjP=0.0001	*MMP9; KRT20; ESRP1; PTCH1; ERBB2; MUC2; MUC5AC; DIRAS3; DMBT1; MUC1; MUC16; KLF17; FOXN1; HGF; CDH2; CEACAM1; IFI16; GLI1*
Vascular Diseases	C=357;O=19;E=5.69;R=3.34;rawP=5.88e-06;adjP=0.0002	*MMP9; F13B; MIA3; C5; CD14; PECAM1; VEGFA; ULK4; SREBF2; MMP8; SERPINA3; HGF; COL4A1; TNNT2; SPTBN5; FGA; ECE1; CD34; AGER*
Adenoma	C=157;O=12;E=2.50;R=4.80;rawP=9.05e-06;adjP=0.0003	*KRT20; MUC1; SLC26A3; NF2; PCSK1; ESRP1; MSH2; MUC2; MCM2; APC2; MUC5AC; PDE11A*
Neuromuscular Diseases	C=276;O=16;E=4.40;R=3.64;rawP=1.11e-05;adjP=0.0003	*FLNC; TTN; NF2; NF1; MAG; IKBKAP; CA3; SPG7; ETFDH; MUSK; FGD4; KIF21A; MTMR1; SLC1A2; SETX; PRPH*
Colonic Diseases	C=279;O=16;E=4.44;R=3.60;rawP=1.26e-05;adjP=0.0003	*KRT20; GALNT12; WISP1; MUC2; VEGFA; APC2; MUC5AC; IGFBP7; SLC26A3; MUC1; MSH2; CDH17; CEACAM1; DNMT1; MUC13; GLI1*

Brain Neoplasms	C=198;O=13;E=3.15;R=4.12;rawP=1.99e-05;adjP=0.0005	*MMP9; DMBT1; RAC2; NF1; PTPRM; MSH2; AQP1; PTCH1; ERBB2; VEGFA; PEG3; KIAA1549; GLI1*
Metaplasia	C=170;O=12;E=2.71;R=4.43;rawP=2.02e-05;adjP=0.0005	*MMP9; KRT20; CDH17; AQP1; HGF; MUC2; PECAM1; VEGFA; CD34; ANTXR1; CEACAM1; MUC5AC*
Nevus	C=53;O=7;E=0.84;R=8.29;rawP=2.06e-05;adjP=0.0005	*PTCH1; GNAQ; APAF1; CD34; KRT10; IGFBP7; PRPH*
Carcinoma, Pancreatic Ductal	C=75;O=8;E=1.19;R=6.70;rawP=2.63e-05;adjP=0.0006	*KRT20; MUC1; MUC16; CDH17; MUC2; DNMT1; MUC5AC; GLI1*
Eye Diseases	C=368;O=18;E=5.86;R=3.07;rawP=3.15e-05;adjP=0.0007	*C3; CACNA1F; RP1L1; IMPG2; PCDH15; LRP5; VEGFA; HMCN1; BFSP1; WFS1; CHML; GPR98; COL4A1; GNAQ; TG; KIF21A; AGER; SETX*
Infertility	C=209;O=13;E=3.33;R=3.90;rawP=3.51e-05;adjP=0.0008	*SLC26A3; SPEF2; SPAG16; WIPF3; SYCP1; CCDC40; AURKC; MSH4; FSHR; FMR1; HSPA2; USP26; PCSK5*
Immune System Diseases	C=680;O=26;E=10.83;R=2.40;rawP=4.60e-05;adjP=0.0010	*C3; FOXP3; TRIM5; PHF11; IL2RG; CD3G; CR2; C5; CD14; SAA1; IKBKAP; IL22; IL18R1; CD5; CCR6; DST; MUSK; FOXN1; PHTF1; BST2; PTGDR2; DSG1; TG; FCER1A; AGER; DOCK8*
Neoplasm Metastasis	C=315;O=16;E=5.02;R=3.19;rawP=5.49e-05;adjP=0.0012	*MMP9; KRT20; LASP1; PODXL; ERBB2; VEGFA; TKTL1; CCR6; MUC1; MUC16; CLCA2; CDH17; HGF; CDH2; CEACAM1; GLI1*
Endocrine disorder NOS	C=429;O=19;E=6.83;R=2.78;rawP=7.29e-05;adjP=0.0014	*FOXP3; PODXL; FSHR; VEGFA; BCL11A; PDE11A; DIRAS3; MUC16; INSR; IYD; PCSK1; WFS1; PHTF1; TG; GPR119; AVP; FMR1; AGER; DND1*
Brenner tumour of ovary	C=222;O=13;E=3.54;R=3.68;rawP=6.52e-05;adjP=0.0014	*C3; CCR6; CNTNAP1; RAC2; CLCA2; ICAM5; FCGRT; C5; PTGDR2; HGF; FCER1A; DEF6; C8B*
Endocrine disturbance NOS	C=429;O=19;E=6.83;R=2.78;rawP=7.29e-05;adjP=0.0014	*FOXP3; PODXL; FSHR; VEGFA; BCL11A; PDE11A; DIRAS3; MUC16; INSR; IYD; PCSK1; WFS1; PHTF1; TG; GPR119; AVP; FMR1; AGER; DND1*
Endocrine System Diseases	C=429;O=19;E=6.83;R=2.78;rawP=7.29e-05;adjP=0.0014	*FOXP3; PODXL; FSHR; VEGFA; BCL11A; PDE11A; DIRAS3; MUC16; INSR; IYD; PCSK1; WFS1; PHTF1; TG; GPR119; AVP; FMR1; AGER; DND1*
Inflammation	C=435;O=19;E=6.93;R=2.74;rawP=8.74e-05;adjP=0.0015	*MMP9; IL36B; C3; FOXP3; C5; CD14; PECAM1; SAA1; VEGFA; IL22; NLRP10; IL18R1; MUC5AC; CCR6; MMP8; OSMR; PTGDR2; FCER1A; AGER*
H Syndrome	C=394;O=18;E=6.28;R=2.87;rawP=7.58e-05;adjP=0.0015	*C3; NF1; PCDH15; PTCH1; FXR2; WRN; NOS1AP; TRPS1; PDE11A; INPP5E; INSR; AGPAT2; CTSC; WFS1; MSH2; GPR98; FMR1; DOCK8*
Urinary Bladder Neoplasms	C=138;O=10;E=2.20;R=4.55;rawP=7.84e-05;adjP=0.0015	*KRT20; CYP2D6; TMEM129; PARP4; MSH2; ERBB2; APAF1; VEGFA; CDH2; DNMT1*
Eczema	C=45;O=6;E=0.72;R=8.37;rawP=7.68e-05;adjP=0.0015	*FOXP3; CROCC; CD14; FCER1A; PHF11; DOCK8*
Recurrence	C=167;O=11;E=2.66;R=4.13;rawP=8.29e-05;adjP=0.0015	*MMP9; KRT20; MUC1; CYP2D6; ERBB2; TG; VEGFA; CDH2; TKTL1; GLI1*
Cerebrovascular Disorders	C=171;O=11;E=2.72;R=4.04;rawP=0.0001;adjP=0.0015	*MMP9; F13B; SREBF2; SERPINA3; ARHGEF10; SPTBN5; COL4A1; CD14; FGA; CD34; AGER*
Drowning	C=7;O=3;E=0.11;R=26.90;rawP=0.0001;adjP=0.0015	*NF2; FMR1; AQP1*
Neoplasms	C=854;O=29;E=13.60;R=2.13;rawP=0.0001;adjP=0.0015	*MMP9; KRT20; NF2; NF1; PTCH1; ERBB2; WISP1; MUC2; VEGFA; TKTL1; IGFBP7; DIRAS3; MDK; MUC1; MAGEA10; DMBT1; TET2; MUC16; CLCA2; MSH2; CDH17; HGF; APAF1; CD34; CDH2; CBX7; CEACAM1; DNMT1; GLI1*
Adenocarcinoma, Mucinous	C=50;O=6;E=0.80;R=7.53;rawP=0.0001;adjP=0.0015	*ERBB2; MUC1; MUC2; MUC16; MUC5AC; MSH2*
Hypersensitivity	C=238;O=13;E=3.79;R=3.43;rawP=0.0001;adjP=0.0015	*MMP9; C3; FANCB; FOXP3; PHF11; C5; PTGDR2; FCER1A; CD14; IL22; FANCM; IL18R1; MUC5AC*
Neurodegenerative Diseases	C=404;O=18;E=6.44;R=2.80;rawP=0.0001;adjP=0.0015	*NF2; NF1; SIGMAR1; LRRK2; CALB1; IKBKAP; DLST; SPG7; HSPA4; ATP7B; WFS1; SERPINA3FGD4; ECE1; AGER; SETX; PRPH; SLC1A2*
Blister	C=32;O=5;E=0.51;R=9.81;rawP=0.0001;adjP=0.0015	*MMP9; DST; DSG1; KRT10; KLRC2*
Sinusitis	C=67;O=7;E=1.07;R=6.56;rawP=9.59e-05;adjP=0.0015	*MMP9; CCDC40; FOXP3; PTGDR2; SPAG16; MMP8; MUC5AC*
Polyps	C=66;O=7;E=1.05;R=6.66;rawP=8.71e-05;adjP=0.0015	*MMP9; PTGDR2; SLC26A3; VEGFA; CEACAM1; MUC5AC; MSH2*

Neuroectodermal Tumors	C=313;O=15;E=4.99;R=3.01;rawP=0.0002;adjP=0.0027	*MARK4; MAGEA10; DMBT1; NF2; NF1; PTPRD; TCH1; GNAQ; APAF1; VEGFA; CDH2; KIAA1549; GLI1; PRPH; IGFBP7*
Cardiovascular Diseases	C=425;O=18;E=6.77;R=2.66;rawP=0.0002;adjP=0.0027	*MMP9; F13B; TTN; MIA3; CD14; PECAM1; SAA1; VEGFA; NOS1AP; MARCH10; SREBF2; COL4A1; TNNT2; HGF; FGA; ECE1; AVP; AGER*
Nervous System Diseases	C=694;O=25;E=11.06;R=2.26;rawP=0.0002;adjP=0.0027	*NF2; NF1; PCDH15; LRRK2; ATP2B2; FXR2; IKBKAP; DFNA5; SPG7; ETFDH; PAK3; PC; MUSK; ATP7B; WFS1; SERPINA3; GPR98; COL4A1; KIF21A; FMR1; MCPH1; MTMR1; PRPH; SETX; SLC1A2*
Urologic Diseases	C=285;O=14;E=4.54;R=3.08;rawP=0.0002;adjP=0.0027	*C3; KRT20; PKD2L2; TMEM129; WFS1; PEX16; LRP2; ACTN4; TNNT2; PKHD1; AVP; VEGFA; AGER; RNF139*
Disease Progression	C=220;O=12;E=3.50;R=3.42;rawP=0.0002;adjP=0.0027	*MMP9; FOXP3; MUC1; TET2; LTBP4; MMP8; ERBB2; APAF1; IL22; VEGFA; AGER; TKTL1*
Arteriosclerosis	C=214;O=12;E=3.41;R=3.52;rawP=0.0002;adjP=0.0027	*MMP9; SREBF2; MIA3; C5; TNNT2; CD14; ECE1; FGA; PECAM1; SAA1; AGER*
Diabetes Insipidus	C=8;O=3;E=0.13;R=23.54;rawP=0.0002;adjP=0.0027	*AVP; WFS1; AQP1*
Infertility, Male	C=188;O=11;E=2.99;R=3.67;rawP=0.0002;adjP=0.0027	*SLC26A3; SPEF2; SPAG16; WIPF3; SYCP1; CCDC40; AURKC; MSH4; FSHR; HSPA2; USP26*
Cystic Fibrosis	C=155;O=10;E=2.47;R=4.05;rawP=0.0002;adjP=0.0027	*ASZ1; MUC1; SLC26A3; ABCC12; HSPA4; ABCC11; CLCA2; SLC9A3; MUC2; MUC5AC*
Raynaud Disease	C=36;O=5;E=0.57;R=8.72;rawP=0.0003;adjP=0.0040	*TET2; COL4A1; GRIPAP1; BRWD3; NIN*
Retinal Degeneration	C=200;O=11;E=3.19;R=3.45;rawP=0.0004;adjP=0.0049	*C3; CACNA1F; RP1L1; NMNAT1; CHML; IMPG2; PCDH15; GPR98; VEGFA; HMCN1; TARM1*
Respiratory Tract Diseases	C=373;O=16;E=5.94;R=2.69;rawP=0.0004;adjP=0.0049	*MMP9; DPCR1; FOXP3; MUC1; MMP8; PHF11; SERPINA3; ERBB2; PTGDR2; CD14; FCER1A; IL22; VEGFA; IL18R1; MUC5AC; DCBLD2*
Essential Tremor	C=23;O=4;E=0.37;R=10.92;rawP=0.0004;adjP=0.0049	*LINGO4; USP14; FMR1; LRRK2*
Cystadenocarcinoma, Mucinous	C=23;O=4;E=0.37;R=10.92;rawP=0.0004;adjP=0.0049	*MUC16; FSHR; OVGP1; PEG3*
Pulmonary Heart Disease	C=85;O=7;E=1.35;R=5.17;rawP=0.0004;adjP=0.0049	*MMP9; CCDC40; TNNT2; TNS1; MUC16; OR12D2; AGER*
Hearing Loss, Conductive	C=83;O=7;E=1.32;R=5.29;rawP=0.0004;adjP=0.0049	*MYO1A; DIAPH1; PCDH15; ATP2B2; MUC5AC; DFNA5; TRIOBP*
Meningioma	C=86;O=7;E=1.37;R=5.11;rawP=0.0005;adjP=0.0058	*MMP9; ERBB2; RAC2; NF2; VEGFA; IGFBP7; MGEA5*
Asthma	C=275;O=13;E=4.38;R=2.97;rawP=0.0005;adjP=0.0058	*MMP9; C3; CSF2RB; FOXP3; PTPRD; PHF11; PTGDR2; FCER1A; CD14; IL22; IKBKAP; IL18R1; MUC5AC*
Bacterial Infections	C=175;O=10;E=2.79;R=3.59;rawP=0.0005;adjP=0.0058	*MMP9; C3; MUC1; C5; ATP4A; CD14; PIGR; IL22; KRT10; CEACAM1*
Barrett Esophagus	C=42;O=5;E=0.67;R=7.47;rawP=0.0005;adjP=0.0058	*ERBB2; KRT20; MAGEA10; MUC2; MUC5AC*
Appendiceal Neoplasms	C=11;O=3;E=0.18;R=17.12;rawP=0.0006;adjP=0.0066	*PTCH1; MUC2; CDH2*
Pancreatic Diseases	C=147;O=9;E=2.34;R=3.84;rawP=0.0006;adjP=0.0066	*KRT20; MUC1; MUC16; CDH17; MSH2; MUC2; DNMT1; GLI1; MUC5AC*
Dermatitis	C=118;O=8;E=1.88;R=4.26;rawP=0.0006;adjP=0.0066	*OSMR; FOXP3; PTGDR2; CD14; FCER1A; IL22; PHF11; DOCK8*
Lung Diseases	C=354;O=15;E=5.64;R=2.66;rawP=0.0006;adjP=0.0066	*MMP9; MUC1; CLCA2; PHF11; SERPINA3; ERBB2; PTGDR2; CD14; FCER1A; PIGR; IL22; VEGFA; AGER; IL18R1; MUC5AC*
Retinal Diseases	C=247;O=12;E=3.93;R=3.05;rawP=0.0007;adjP=0.0075	*C3; CACNA1F; RP1L1; NMNAT1; IMPG2; PCDH15; GPR98; LRP5; VEGFA; AGER; HMCN1; TARM1*
Infection	C=516;O=19;E=8.22;R=2.31;rawP=0.0007;adjP=0.0075	*C3; FOXP3; TRIM5; CD3G; CR2; C5; CD14; SAA1; IFIT3; IL22; ISG15; CCR6; DMBT1; SP100; BST2; IFIT1; CEACAM1; SIGLEC1; DOCK8*
Acute-Phase Reaction	C=46;O=5;E=0.73;R=6.82;rawP=0.0008;adjP=0.0082	*MMP9; OSMR; CD14; SAA1; SERPINA3*
Urologic Neoplasms	C=185;O=10;E=2.95;R=3.39;rawP=0.0008;adjP=0.0082	*KANK1; KRT20; MUC1; TMEM129; MSH2; ERBB2; APAF1; VEGFA; CDH2; RNF139*
Arterial Occlusive Diseases	C=219;O=11;E=3.49;R=3.15;rawP=0.0008;adjP=0.0082	*MMP9; SREBF2; MMP8; MIA3; TNNT2; CD14; FGA; ECE1; PECAM1; SAA1; AGER*

Pathologic Processes	C=561;O=20;E=8.94;R=2.24;rawP=0.0008;adjP=0.0082	*MMP9; FOXP3; KRT20; NLRP13; DISC1; PHF11; LRRK2; ERBB2; WISP1; C5; CD14; VEGFA; IL18R1; NOS1AP; MMP8; MSH2; ERAP1; KIF21A; CDH2; AGER*
Skin and Connective Tissue Diseases	C=481;O=18;E=7.66;R=2.35;rawP=0.0008;adjP=0.0082	*MMP9; FOXP3; ABCC11; ATP2C1; PTCH1; ERBB2; C5; SAA1; IL22; VEGFA; KRT10; DIRAS3; DST; MUC1; AGPAT2; CTSC; DSG1; SCGB1D2*
Heart Failure	C=125;O=8;E=1.99;R=4.02;rawP=0.0009;adjP=0.0091	*HRC; FANCB; TTN; MUC16; GNAQ; TNNT2; NFATC4; AVP*
Dermatitis, Atopic	C=98;O=7;E=1.56;R=4.48;rawP=0.0010;adjP=0.0095	*FOXP3; PTGDR2; CD14; FCER1A; IL22; PHF11; DOCK8*
glomerular disease	C=127;O=8;E=2.02;R=3.95;rawP=0.0010;adjP=0.0095	*ACTN4; C3; C5; PKHD1; PECAM1; AGER; NAT8B; LRP2*
Cerebral Palsy	C=48;O=5;E=0.76;R=6.54;rawP=0.0010;adjP=0.0095	*SPG7; KANK1; DOCK8; SLC1A2*
Gastroesophageal Reflux	C=72;O=6;E=1.15;R=5.23;rawP=0.0010;adjP=0.0095	*MUC2; ATP4A; OVGP1; CCDC88C; GATA5; MUC13*
Unspecified keratoconjunctivitis	C=13;O=3;E=0.21;R=14.49;rawP=0.0010;adjP=0.0095	*MUC16; IL22; MUC5AC*
Metabolic Diseases	C=612;O=21;E=9.75;R=2.15;rawP=0.0010;adjP=0.0095	*MEPE; IL2RG; WRN; SAA1; FANCM; BCL11A; FANCB; ETFDH; INSR; PC; AGPAT2; HLCS; PCSK1; ATP7B; WFS1; PEX16; MSH2; GPR119; AGER; SETX; PRPH*
Neoplastic Processes	C=411;O=16;E=6.55;R=2.44;rawP=0.0011;adjP=0.0099	*MMP9; KRT20; LASP1; ERBB2; WISP1; MUC2; VEGFA; TKTL1; MDK; MUC1; MUC16; CDH17; HGF; CDH2; CEACAM1; GLI1*
Autoimmune Diseases	C=414;O=16;E=6.59;R=2.43;rawP=0.0011;adjP=0.0099	*C3; FOXP3; CR2; C5; MAG; SAA1; IL22; IKBKAP; CD5; CCR6; DST; MUSK; PHTF1; TG; DSG1; AGER*
Ataxia Telangiectasia	C=129;O=8;E=2.05;R=3.89;rawP=0.0011;adjP=0.0099	*RNF168; WRN; MDC1; MCM2; MCPH1; TERF1; RIF1; SETX*
Laryngeal Neoplasms	C=73;O=6;E=1.16;R=5.16;rawP=0.0011;adjP=0.0099	*MMP9; ERBB2; CCR6; VEGFA; MSH2; AQP1*
Urticaria	C=29;O=4;E=0.46;R=8.66;rawP=0.0011;adjP=0.0099	*FCER1A; FOXN1; NCDN; ZNF521*
Arthritis	C=302;O=13;E=4.81;R=2.70;rawP=0.0012;adjP=0.0106	*IL36B; CCR6; FOXP3; FRZB; ACAN; C5; ERAP1; CR2; LRP5; SAA1; IL22; AGER; KLRC2*
Carcinoma, Papillary	C=101;O=7;E=1.61;R=4.35;rawP=0.0012;adjP=0.0106	*DUOX1; KRT20; MUC1; MUC2; TG; PECAM1; CBX7*
Joint Diseases	C=266;O=12;E=4.24;R=2.83;rawP=0.0013;adjP=0.0111	*IL36B; CCR6; FOXP3; FRZB; ACAN; C5; ERAP1; CD14; SAA1; IL22; AGER; KLRC2*
Eosinophilia	C=51;O=5;E=0.81;R=6.15;rawP=0.0013;adjP=0.0111	*CSF2RB; PTGDR2; FCER1A; DOCK8; MUC5AC*
Pancreatic Neoplasms	C=197;O=10;E=3.14;R=3.19;rawP=0.0013;adjP=0.0111	*KRT20; MUC1; PODXL; MUC16; CDH17; MSH2; MUC2; DNMT1; GLI1; MUC5AC*
Telangiectasis	C=30;O=4;E=0.48;R=8.37;rawP=0.0013;adjP=0.0111	*ANKRD49; WRN; TERF1; SETX*
Chronic Disease	C=307;O=13;E=4.89;R=2.66;rawP=0.0014;adjP=0.0118	*MMP9; C3; FOXP3; TET2; MMP8; SERPINA3; TNNT2; CD14; IL22; VEGFA; AGER; MUC5AC; CD5*
liver transplantation	C=77;O=6;E=1.23;R=4.89;rawP=0.0015;adjP=0.0123	*FOXP3; HGF; PKHD1; ABCC2; IL22; ATP7B*
Disorder of uterus NOS	C=135;O=8;E=2.15;R=3.72;rawP=0.0015;adjP=0.0123	*MMP9; MUC16; LHX6; INHBE; MSH2; ERBB2; VEGFA; TKTL1*
Gastroenteritis	C=200;O=10;E=3.19;R=3.14;rawP=0.0015;adjP=0.0123	*CCR6; FOXP3; SLC26A3; DMBT1; SLC2A4RG; AMIGO3; MUC2; CD14; IL22; HSPA2*
Neoplasm, Residual	C=53;O=5;E=0.84;R=5.92;rawP=0.0015;adjP=0.0123	*ERBB2; HGF; TG; MUC16; CD34*
Amyotrophic Lateral Sclerosis	C=78;O=6;E=1.24;R=4.83;rawP=0.0016;adjP=0.0127	*VAPA; VEGFA; SLC1A2; SETX; NOP56; PRPH*
Kidney Diseases	C=274;O=12;E=4.36;R=2.75;rawP=0.0016;adjP=0.0127	*C3; PKD2L2; WFS1; PEX16; LRP2; ACTN4; TNNT2; PKHD1; AVP; VEGFA; AGER; RNF139*
Uterine Neoplasms	C=136;O=8;E=2.17;R=3.69;rawP=0.0016;adjP=0.0127	*MMP9; MUC16; LHX6; INHBE; MSH2; ERBB2; VEGFA; TKTL1*
Nasal Polyps	C=54;O=5;E=0.86;R=5.81;rawP=0.0017;adjP=0.0133	*MMP9; PTGDR2; VEGFA; MMP8; MUC5AC*
Skin Diseases, Vascular	C=79;O=6;E=1.26;R=4.77;rawP=0.0017;adjP=0.0133	*MMP9; ERBB2; MUC1; COL4A1; VEGFA; COL15A1*

Fibroma	C=33;O=4;E=0.53;R=7.61;rawP=0.0018;adjP=0.0134	*PTCH1; CREB3L2; CD34; OVGP1*
Embryo Loss	C=33;O=4;E=0.53;R=7.61;rawP=0.0018;adjP=0.0134	*MUC1; KHDC3L; APAF1; IKBKAP*
aspirin-induced asthma	C=33;O=4;E=0.53;R=7.61;rawP=0.0018;adjP=0.0134	*DPCR1; PTGDR2; FCER1A; DCBLD2*
Cicatrix	C=33;O=4;E=0.53;R=7.61;rawP=0.0018;adjP=0.0134	*MMP9; MYH4; RSPO2; SEMA6C*
Deglutition Disorders	C=33;O=4;E=0.53;R=7.61;rawP=0.0018;adjP=0.0134	*MUC2; MUC16; OVGP1; MUC13*
Anthrax	C=33;O=4;E=0.53;R=7.61;rawP=0.0018;adjP=0.0134	*C3; FANCB; DUOX1; ANTXR1*
Disease Susceptibility	C=825;O=25;E=13.14;R=1.90;rawP=0.0019;adjP=0.0136	*C3; NLRP13; DISC1; CYP2D6; PTPRD; PHF11; LRRK2; C5; LRP5; CD14; FSHR; TCERG1L; PCLO; IL18R1; NOS1AP; DPCR1; FRZB; WFS1; MMP8; SERPINA3; MSH2; ERAP1; TG; FGA; FCER1A*
Brain Ischemia	C=109;O=7;E=1.74;R=4.03;rawP=0.0019;adjP=0.0136	*MMP9; F13B; C5; HGF; FGA; AGER; SLC1A2*
Hearing Loss, Sensorineural	C=140;O=8;E=2.23;R=3.59;rawP=0.0019;adjP=0.0136	*MYO1A; WFS1; LBX2; PCDH15; ATP2B2; DFNA5; TRIOBP; GPR98*
Heat Stroke	C=16;O=3;E=0.25;R=11.77;rawP=0.0019;adjP=0.0136	*HSPA4; VDAC1; CD34*
Carotid Artery Diseases	C=81;O=6;E=1.29;R=4.65;rawP=0.0019;adjP=0.0136	*MMP9; CD14; CD34; AGER; MMP8; MIA3*
Endometriosis	C=82;O=6;E=1.31;R=4.59;rawP=0.0020;adjP=0.0140	*MMP9; HGF; AFM; MUC16; FSHR; VEGFA*
Carcinoma, Hepatocellular	C=208;O=10;E=3.31;R=3.02;rawP=0.0020;adjP=0.0140	*MDK; DIRAS3; CHD1L; CDH17; HGF; SALL3; IL22; VEGFA; CDH2; DNMT1*
Rectal Neoplasms	C=82;O=6;E=1.31;R=4.59;rawP=0.0020;adjP=0.0140	*WISP1; KRT20; MUC2; VEGFA; MUC5AC; MSH2*
Schizophrenia	C=360;O=14;E=5.73;R=2.44;rawP=0.0021;adjP=0.0145	*CSF2RB; DISC1; CMYA5; CYP2D6; EHHADH; SIGMAR1; CHRM1; KALRN; MAG; HTR5A; ERMN; ATF5; NOS1AP; SLC1A2*
Angina, Unstable	C=57;O=5;E=0.91;R=5.51;rawP=0.0021;adjP=0.0145	*MMP9; TNNT2; CD14; ARHGAP29; MMP8*
Nerve Degeneration	C=112;O=7;E=1.78;R=3.92;rawP=0.0022;adjP=0.0150	*CALB1; KALRN; MAG; NMNAT1; LRRK2; SLC1A2; PRPH*
Macular Degeneration	C=112;O=7;E=1.78;R=3.92;rawP=0.0022;adjP=0.0150	*C3; C5; RP1L1; NMNAT1; VEGFA; HMCN1; IMPG2*
Gallbladder Neoplasms	C=35;O=4;E=0.56;R=7.17;rawP=0.0023;adjP=0.0152	*ERBB2; MUC1; MUC5AC; MSH2*
Obstetric Labor Complications	C=212;O=10;E=3.38;R=2.96;rawP=0.0023;adjP=0.0152	*MMP9; F13B; HSPA4; COL4A2; MMP8; COL4A1; CD14; PECAM1; VEGFA; AGER*
Gastritis	C=58;O=5;E=0.92;R=5.41;rawP=0.0023;adjP=0.0152	*MMP9; FOXP3; MUC1; ATP4A; CD14*
Pharyngitis	C=17;O=3;E=0.27;R=11.08;rawP=0.0023;adjP=0.0152	*AHCY; DMBT1; SIGLEC1*
Precancerous Conditions	C=85;O=6;E=1.35;R=4.43;rawP=0.0024;adjP=0.0157	*ERBB2; MUC2; CEACAM1; MUC5AC; MSH2*
Liver Diseases	C=250;O=11;E=3.98;R=2.76;rawP=0.0024;adjP=0.0157	*FOXP3; NUP210; ATP7B; PEX16; SERPINA3; CHD1L; CDH17; HGF; ABCC2; PKHD1; IL22*
Protein-Losing Enteropathies	C=5;O=2;E=0.08;R=25.11;rawP=0.0025;adjP=0.0158	*FOXP3; FCGRT*
Korsakoff Syndrome	C=5;O=2;E=0.08;R=25.11;rawP=0.0025;adjP=0.0158	*DLST; TKTL1*
Spermatocele	C=5;O=2;E=0.08;R=25.11;rawP=0.0025;adjP=0.0158	*SLC26A3; SLC9A3*
Pyruvate carboxylase deficiency	C=5;O=2;E=0.08;R=25.11;rawP=0.0025;adjP=0.0158	*HLCS; PC*
Pain	C=146;O=8;E=2.33;R=3.44;rawP=0.0025;adjP=0.0158	*MMP9; TNNT2; CD14; MUC16; CYP2D6; MMP8; SERPINA6; ACAN*
Sore on skin	C=18;O=3;E=0.29;R=10.46;rawP=0.0027;adjP=0.0161	*MMP9; HGF; MMP8*
Brain Edema	C=18;O=3;E=0.29;R=10.46;rawP=0.0027;adjP=0.0161	*MMP9; VEGFA; AQP1*

Neovascularization, Pathologic	C=148;O=8;E=2.36;R=3.39;rawP=0.0027;adjP=0.0161	*MMP9; AQP1; HGF; VEGFA; CD34; PECAM1; ANTXR1; CEACAM1*
Rupture	C=148;O=8;E=2.36;R=3.39;rawP=0.0027;adjP=0.0161	*MMP9; F13B; COL4A2; MMP8; OOEP; KHDC3L; CD14; VEGFA*
Uterine Hemorrhage	C=18;O=3;E=0.29;R=10.46;rawP=0.0027;adjP=0.0161	*MMP9; FGA; VEGFA*
Dyslexia	C=60;O=5;E=0.96;R=5.23;rawP=0.0027;adjP=0.0161	*DOCK4; RP1L1; NCAPD3; CTSO; SLC47A2*
Ear Diseases	C=217;O=10;E=3.46;R=2.89;rawP=0.0027;adjP=0.0161	*MYO1A; NF2; WFS1; CHML; PCDH15; ATP2B2; GPR98; DIAPH1; DFNA5; MUC5AC*
Esophageal Diseases	C=148;O=8;E=2.36;R=3.39;rawP=0.0027;adjP=0.0161	*KRT20; MUC1; ERBB2; MUC2; VEGFA; GATA5; MUC5AC; GLI1*
Thiamine Deficiency	C=18;O=3;E=0.29;R=10.46;rawP=0.0027;adjP=0.0161	*SLC25A19; DLST; TKTL1*
Fibrosis	C=183;O=9;E=2.92;R=3.09;rawP=0.0028;adjP=0.0165	*MMP9; ASZ1; SLC26A3; ABCC12; PKHD1; HGF; KIF21A; IL22; MUC5AC*
Leukemia, Myelomonocytic, Chronic	C=37;O=4;E=0.59;R=6.79;rawP=0.0028;adjP=0.0165	*TET1; TET2; NF1; BCL11A*
Depression	C=184;O=9;E=2.93;R=3.07;rawP=0.0029;adjP=0.0169	*DISC1; CYP2D6; E195; SIGMAR1; WFS1; AVP; HTR5A; FMR1; PCLO; PDE11A*
Rhinitis	C=88;O=6;E=1.40;R=4.28;rawP=0.0029;adjP=0.0169	*FOXP3; PTGDR2; CD14; FCER1A; PHF11; MUC5AC*
Shock, Surgical	C=38;O=4;E=0.61;R=6.61;rawP=0.0031;adjP=0.0179	*ABRA; CD14; HSPA4; SREBF2*
Diabetes Mellitus	C=336;O=13;E=5.35;R=2.43;rawP=0.0032;adjP=0.0183	*FOXP3; VPS13C; INSR; AGPAT2; PCSK1; WFS1; PHTF1; VEGFA; GPR119; AGER; NOS1AP; BCL11A; PRKAG3*
Dysautonomia, Familial	C=19;O=3;E=0.30;R=9.91;rawP=0.0032;adjP=0.0183	*DST; ELP4; IKBKAP*
Colitis	C=153;O=8;E=2.44;R=3.28;rawP=0.0033;adjP=0.0186	*FOXP3; KRT20; SLC26A3; DMBT1; FKBP15; MUC2; CD14; IL22*
Astrocytoma	C=121;O=7;E=1.93;R=3.63;rawP=0.0033;adjP=0.0186	*MMP9; DMBT1; VEGFA; NF1; KIAA1549; PTPRM; GLI1*
Myocardial Ischemia	C=261;O=11;E=4.16;R=2.65;rawP=0.0034;adjP=0.0188	*MMP9; SREBF2; MIA3; TNNT2; CD14; FGA; PECAM1; VEGFA; CD34; AGER; NOS1AP*
Bronchiectasis	C=39;O=4;E=0.62;R=6.44;rawP=0.0034;adjP=0.0188	*CCDC40; SPAG16; SAA1; MMP8*
Musculoskeletal Diseases	C=462;O=16;E=7.36;R=2.17;rawP=0.0034;adjP=0.0188	*FLNC; TTN; ETFDH; MEPE; FRZB; CDK5RAP2; ACAN; PTCH1; C5; ERAP1; LRP5; TG; SAA1; MCPH1; MTMR1; TRPS1*
Fragile X-associated Tremor/Ataxia Syndrome	C=6;O=2;E=0.10;R=20.93;rawP=0.0036;adjP=0.0193	*FMR1; DIP2A*
Impetigo NOS	C=6;O=2;E=0.10;R=20.93;rawP=0.0036;adjP=0.0193	*DSG1; AHCY*
Aneurysm	C=64;O=5;E=1.02;R=4.90;rawP=0.0036;adjP=0.0193	*MMP9; COL4A1; LTBP4; MMP8; SERPINA3*
Male Urogenital Diseases	C=423;O=15;E=6.74;R=2.23;rawP=0.0036;adjP=0.0193	*PODXL; PKD2L2; WFS1; PEX16; LRP2; ACTN4; PKHD1; FSHR; VEGFA; AVP; AGER; DND1; RNF139; USP26; IGFBP7*
Diabetes Insipidus, Nephrogenic	C=6;O=2;E=0.10;R=20.93;rawP=0.0036;adjP=0.0193	*AVP; AQP1*
Radiculopathy	C=6;O=2;E=0.10;R=20.93;rawP=0.0036;adjP=0.0193	*ACAN; SETX*
Familial aplasia of the vermis	C=20;O=3;E=0.32;R=9.42;rawP=0.0037;adjP=0.0195	*INPP5E; PAK3; NOP56*
Peritoneal Diseases	C=40;O=4;E=0.64;R=6.28;rawP=0.0037;adjP=0.0195	*KRT20; MUC2; MUC16; VEGFA*
Blepharoptosis	C=20;O=3;E=0.32;R=9.42;rawP=0.0037;adjP=0.0195	*MRPS22; KANK1; KIF21A*
Arthritis, Juvenile Rheumatoid	C=65;O=5;E=1.04;R=4.83;rawP=0.0038;adjP=0.0199	*ERAP1; FOXP3; C5; AGER; PHTF1*
Smallpox	C=125;O=7;E=1.99;R=3.52;rawP=0.0040;adjP=0.0207	*ATP10A; DAGLA; NLRP13; IL18R1; SPOCK3; PCDH15; GOLGA3*

Infertility, Female	C=94;O=6;E=1.50;R=4.01;rawP=0.0040;adjP=0.0207	*MUC1; MSH4; FSHR; VEGFA; FMR1; PCSK5*
Pleural Diseases	C=41;O=4;E=0.65;R=6.12;rawP=0.0041;adjP=0.0211	*MMP9; SAA1; VEGFA; IL22*
Wiskott-Aldrich Syndrome	C=95;O=6;E=1.51;R=3.96;rawP=0.0042;adjP=0.0214	*KIAA1033; RHOQ; FKBP15; WIPF3; DIAPH1; NYAP2*
Ataxia	C=269;O=11;E=4.29;R=2.57;rawP=0.0042;adjP=0.0214	*INPP5E; SLC9A6; RNF168; WRN; MCM2; MDC1; FMR1; MCPH1; TERF1; NOP56; SETX*
Hydronephrosis	C=21;O=3;E=0.33;R=8.97;rawP=0.0043;adjP=0.0217	*TSHZ2; LTBP4; TSHZ3*
Retinal Drusen	C=21;O=3;E=0.33;R=8.97;rawP=0.0043;adjP=0.0217	*C3; C5; HMCN1*
Skin Diseases, Genetic	C=310;O=12;E=4.94;R=2.43;rawP=0.0044;adjP=0.0220	*MUC1; CYP2D6; AGPAT2; ABCC11; CTSC; ATP2C1; ERBB2; PTCH1; DSG1; FCER1A; KRT10; SCGB1D2*
Breast Diseases	C=350;O=13;E=5.58;R=2.33;rawP=0.0044;adjP=0.0220	*MMP9; DIRAS3; LASP1; MUC1; CYP2D6; ABCC11; TACC2; CLCA2; ERBB2; VEGFA; SCGB1D2; UHRF1; IGFBP7*
Mental Disorders	C=564;O=18;E=8.98;R=2.00;rawP=0.0046;adjP=0.0227	*DISC1; CYP2D6; SIGMAR1; PTPRD; CHRM1; ASMT; KALRN; CHRM4; HTR5A; DLST; PCLO; NOS1AP; PAK3; WFS1; SERPINA3; AVP; FMR1; SLC1A2*
Fibrosarcoma	C=68;O=5;E=1.08;R=4.62;rawP=0.0046;adjP=0.0227	*MMP9; CREB3L2; AKAP12; IFIT1; CD34*
Syndrome	C=654;O=20;E=10.42;R=1.92;rawP=0.0047;adjP=0.0231	*NF1; PCDH15; PTCH1; FXR2; WRN; FSHR; NOS1AP; TRPS1; INPP5E; FKBP15; INSR; AGPAT2; WIPF3; CTSC; WFS1; PEX16; MSH2; GPR98; FMR1; DOCK8*
Pleural Effusion, Malignant	C=22;O=3;E=0.35;R=8.56;rawP=0.0049;adjP=0.0235	*MMP9; MUC1; VEGFA*
Oligospermia	C=98;O=6;E=1.56;R=3.84;rawP=0.0049;adjP=0.0235	*SYCP2; MSH4; FSHR; SYCP1; HSPA2; USP26*
Depressive Disorder, Major	C=98;O=6;E=1.56;R=3.84;rawP=0.0049;adjP=0.0235	*DISC1; CYP2D6; AVP; PCLO; WFS1; PDE11A*
Amyloidosis	C=43;O=4;E=0.68;R=5.84;rawP=0.0048;adjP=0.0235	*OSMR; FGA; SAA1; AGER*
Learning Disorders	C=69;O=5;E=1.10;R=4.55;rawP=0.0049;adjP=0.0235	*DISC1; FMR1; NF1; SAMD3; SLC1A2*
Unspecified corneal ulcer	C=7;O=2;E=0.11;R=17.94;rawP=0.0050;adjP=0.0238	*MMP9; MUC5AC*
Endometrial Neoplasms	C=99;O=6;E=1.58;R=3.80;rawP=0.0051;adjP=0.0241	*ERBB2; MUC1; MUC16; VEGFA; PEG3; MSH2*
Liver Cirrhosis	C=99;O=6;E=1.58;R=3.80;rawP=0.0051;adjP=0.0241	*HGF; DMBT1; CD14; NUP210; ATP7B; AQP1*
Osteoarthritis, Knee	C=70;O=5;E=1.12;R=4.48;rawP=0.0052;adjP=0.0242	*LRP5; RSPO2; FRZB; IL18R1; ACAN*
Neoplastic Cells, Circulating	C=44;O=4;E=0.70;R=5.71;rawP=0.0052;adjP=0.0242	*ERBB2; KRT20; MUC1; VEGFA*
Intracranial Aneurysm	C=44;O=4;E=0.70;R=5.71;rawP=0.0052;adjP=0.0242	*MMP9; COL4A1; LTBP4; SERPINA3*
Glioblastoma	C=132;O=7;E=2.10;R=3.33;rawP=0.0054;adjP=0.0250	*MARK4; MMP9; DMBT1; VEGFA; ELK1; PTPRM; GLI1*
Streptococcal Infections	C=45;O=4;E=0.72;R=5.58;rawP=0.0057;adjP=0.0253	*C5; CD14; PIGR; MUC5AC*
Cranial Nerve Diseases	C=101;O=6;E=1.61;R=3.73;rawP=0.0057;adjP=0.0253	*CACNA1F; KIF21A; NF2; NF1; WFS1; SETX*
Vaccinia	C=101;O=6;E=1.61;R=3.73;rawP=0.0057;adjP=0.0253	*DAGLA; NLRP13; GANAB; WIPF3; PCSK1; PCSK5*
Ascites	C=45;O=4;E=0.72;R=5.58;rawP=0.0057;adjP=0.0253	*MUC2; MUC16; VEGFA; AQP1*
Pemphigus	C=23;O=3;E=0.37;R=8.19;rawP=0.0056;adjP=0.0253	*DST; DSG1; ATP2C1*
Otitis	C=45;O=4;E=0.72;R=5.58;rawP=0.0057;adjP=0.0253	*CCDC40; C3; DOCK8; MUC5AC*
Hepatitis, Toxic	C=45;O=4;E=0.72;R=5.58;rawP=0.0057;adjP=0.0253	*OR5H2; ABCC2; CYP2D6; IL22*
Crohn Disease	C=167;O=8;E=2.66;R=3.01;rawP=0.0055;adjP=0.0253	*CCR6; DMBT1; CD14; IL22; HSPA2; SLC2A4RG; AMIGO3; LRRK2*

Deaf-Blind Disorders	C=23;O=3;E=0.37;R=8.19;rawP=0.0056;adjP=0.0253	*WFS1; PCDH15; GPR98*
Liver Neoplasms	C=242;O=10;E=3.85;R=2.59;rawP=0.0058;adjP=0.0256	*MMP9; MDK; DIRAS3; CHD1L; CDH17; HGF; ABCC2; VEGFA; DNMT1; IGFBP7*
Genetic Predisposition to Disease	C=808;O=23;E=12.87;R=1.79;rawP=0.0059;adjP=0.0259	*NLRP13; DISC1; CYP2D6; SIGMAR1; PTPRD; PHF11; LRRK2; C5; LRP5; CD14; FSHR; TCERG1L; PCLO; IL18R1; NOS1AP; DPCR1; FRZB; WFS1; MMP8; SERPINA3; MSH2; ERAP1; FCER1A*
Cystadenocarcinoma	C=46;O=4;E=0.73;R=5.46;rawP=0.0061;adjP=0.0264	*KRT20; MUC16; VEGFA; OVGP1*
Kidney Neoplasms	C=135;O=7;E=2.15;R=3.26;rawP=0.0061;adjP=0.0264	*KANK1; MUC1; APAF1; VEGFA; RNF139; PDE11A; MSH2*
Arthritis, Reactive	C=46;O=4;E=0.73;R=5.46;rawP=0.0061;adjP=0.0264	*ERAP1; CD14; SAA1; OR6C4*
Bone Cysts	C=24;O=3;E=0.38;R=7.85;rawP=0.0063;adjP=0.0269	*PTCH1; NF1; GLI1*
Neurilemmoma	C=24;O=3;E=0.38;R=7.85;rawP=0.0063;adjP=0.0269	*NF2; CD34; NF1*
Common Variable Immunodeficiency	C=24;O=3;E=0.38;R=7.85;rawP=0.0063;adjP=0.0269	*CR2; FOXP3; DOCK8*
Pre-Eclampsia	C=104;O=6;E=1.66;R=3.62;rawP=0.0065;adjP=0.0276	*MMP9; HSPA4; VEGFA; AGER; KLRC2; PAPPA2*
Infiltrating duct carcinoma of breast	C=74;O=5;E=1.18;R=4.24;rawP=0.0066;adjP=0.0278	*ERBB2; DIRAS3; MUC1; VEGFA; LTBP4*
Blindness	C=172;O=8;E=2.74;R=2.92;rawP=0.0066;adjP=0.0278	*C3; C5; LRP5; CACNA1F; VEGFA; WFS1; PCDH15; GPR98*
Dermoid Cyst	C=8;O=2;E=0.13;R=15.69;rawP=0.0067;adjP=0.0281	*PTCH1; MUC16*
Brain Diseases	C=411;O=14;E=6.55;R=2.14;rawP=0.0068;adjP=0.0282	*MMP9; DMBT1; PC; ATP7B; PEX16; SERPINA3; LRRK2; AQP1; COL4A1; ECE1; DLST; AGER; SETX; SLC1A2*
Eclampsia	C=105;O=6;E=1.67;R=3.59;rawP=0.0068;adjP=0.0282	*MMP9; HSPA4; VEGFA; AGER; KLRC2; PAPPA2*
Renal diabetes	C=105;O=6;E=1.67;R=3.59;rawP=0.0068;adjP=0.0282	*TNNT2; VEGFA; AVP; AGER; WFS1; LBX2*
Neuroendocrine Tumors	C=210;O=9;E=3.35;R=2.69;rawP=0.0069;adjP=0.0283	*KRT20 ; MAGEA10; NF2; GNAQ; APAF1; CDH2; CEACAM1; PRPH; IGFBP7*
Nevi and Melanomas	C=210;O=9;E=3.35;R=2.69;rawP=0.0069;adjP=0.0283	*MAGEA10; MMP8; GNAQ; APAF1; ANTXR1; CDH2; CEACAM1; PRPH; IGFBP7*
CHARGE syndrome	C=48;O=4;E=0.76;R=5.23;rawP=0.0071;adjP=0.0286	*SMARCA4; GATAD2B; SALL3; SMARCC2*
Pleural Effusion	C=25;O=3;E=0.40;R=7.53;rawP=0.0071;adjP=0.0286	*MMP9; SAA1; VEGFA*
Food Hypersensitivity	C=25;O=3;E=0.40;R=7.53;rawP=0.0071;adjP=0.0286	*FOXP3; PTGDR2; CD14*
Airway Obstruction	C=48;O=4;E=0.76;R=5.23;rawP=0.0071;adjP=0.0286	*CD14; KCNRG; MUC5AC; DCBLD2*
Arrhythmias, Cardiac	C=107;O=6;E=1.70;R=3.52;rawP=0.0074;adjP=0.0297	*SLC35F1; AKAP9; TNNT2; CYP2D6; TRDN; NOS1AP*
Motor Neuron Disease	C=141;O=7;E=2.25;R=3.12;rawP=0.0076;adjP=0.0304	*GEMIN5; SPG7; VEGFA; SLC1A2; SETX; NOP56; PRPH*
Viremia	C=49;O=4;E=0.78;R=5.12;rawP=0.0077;adjP=0.0305	*FOXP3; CD34; TRIM5; SIGLEC1*
Signs and Symptoms	C=417;O=14;E=6.64;R=2.11;rawP=0.0077;adjP=0.0305	*SLC26A3; PCSK1; SLC9A6; WFS1; PCDH15; GPR98; TNNT2; PLIN4; KIF21A; VEGFA; AVP; FMR1; SETX; DFNA5*
Environmental allergy	C=77;O=5;E=1.23;R=4.08;rawP=0.0078;adjP=0.0307	*DISC1; CD14; FCER1A; CYP2D6; CCDC88C*
Chromosome Disorders	C=418;O=14;E=6.66;R=2.10;rawP=0.0078;adjP=0.0307	*C2CD2; SPG7; DOCK4; DISC1; PAK3; HLCS; SLC9A6; WFS1; ATP10A; FXR2; FMR1; ABCA13; MCPH1; DFNA5*
Endometrial Hyperplasia	C=26;O=3;E=0.41;R=7.24;rawP=0.0080;adjP=0.0308	*OVGP1; GLI1; MSH2*
Genital Neoplasms, Female	C=26;O=3;E=0.41;R=7.24;rawP=0.0080;adjP=0.0308	*RIC8A; PEG3; MSH2*
Respiratory Sounds	C=26;O=3;E=0.41;R=7.24;rawP=0.0080;adjP=0.0308	*PTGDR2; CD14; PHF11*

Adenoma, Pleomorphic	C=26;O=3;E=0.41;R=7.24;rawP=0.0080;adjP=0.0308	*ERBB2; MUC1; MCM2*
Diabetes Mellitus, Type 2	C=254;O=10;E=4.05;R=2.47;rawP=0.0080;adjP=0.0308	*VPS13C; INSR; LARS2; PCSK1; WFS1; GPR119; VEGFA; AGER; NOS1AP; BCL11A*
Carcinoma, Ductal, Breast	C=78;O=5;E=1.24;R=4.02;rawP=0.0082;adjP=0.0310	*ERBB2; DIRAS3; MUC1; VEGFA; LTBP4*
Hearing Disorders	C=216;O=9;E=3.44;R=2.62;rawP=0.0082;adjP=0.0310	*MYO1A; WFS1; CHML; PCDH15; ATP2B2; TRIOBP; GPR98; DIAPH1; DFNA5*
Breast Neoplasms	C=377;O=13;E=6.01;R=2.16;rawP=0.0081;adjP=0.0310	*MMP9; DIRAS3; LASP1; MUC1; CYP2D6; ABCC11; TACC2; CLCA2; ERBB2; VEGFA; SCGB1D; UHRF1; IGFBP7*
Liver Failure	C=78;O=5;E=1.24;R=4.02;rawP=0.0082;adjP=0.0310	*HGF; PTCD2; NUP210; PKD2L2; ATP7B*
Hemorrhage	C=109;O=6;E=1.74;R=3.46;rawP=0.0081;adjP=0.0310	*MMP9; COL4A1; FGA; VEGFA; MMP8; SERPINA3*
Ovarian hyperstimulation syndrome	C=9;O=2;E=0.14;R=13.95;rawP=0.0085;adjP=0.0316	*FSHR; VEGFA*
Mucoid otitis media NOS	C=9;O=2;E=0.14;R=13.95;rawP=0.0085;adjP=0.0316	*MUC2; MUC5AC*
Hematoma, Subdural	C=9;O=2;E=0.14;R=13.95;rawP=0.0085;adjP=0.0316	*VEGFA; AQP1*
Carcinoma, Large Cell	C=110;O=6;E=1.75;R=3.42;rawP=0.0085;adjP=0.0316	*PTCH1; ERBB2; KRT20; MUC1; MUC16; MCM2*
Diarrhea	C=79;O=5;E=1.26;R=3.97;rawP=0.0087;adjP=0.0322	*FOXP3; SLC26A3; ABCC2; PCSK1; SLC9A3*
Neoplasm Invasiveness	C=298;O=11;E=4.75;R=2.32;rawP=0.0088;adjP=0.0325	*MMP9; MDK; KRT20; MUC1; CDH17; ERBB2; HGF; VEGFA; CDH2; CEACAM1; GLI1*
Mood Disorders	C=258;O=10;E=4.11;R=2.43;rawP=0.0089;adjP=0.0327	*DISC1; CYP2D6; WFS1; PBRM1; HTR5A; AVP; FMR1; PCLO; NOS1AP; PDE11A*
Mental Retardation	C=426;O=14;E=6.79;R=2.06;rawP=0.0091;adjP=0.0333	*C2CD2; INPP5E; PAK3; HLCS; NF1; SLC9A6; ATP7B; AMMECR1; ATP10A; KIF21A; FXR2; FMR1; BRWD3; MCPH1*
Depressive Disorder	C=112;O=6;E=1.78;R=3.36;rawP=0.0092;adjP=0.0334	*DISC1; CYP2D6; AVP; PCLO; WFS1; PDE11A*
Narcolepsy	C=112;O=6;E=1.78;R=3.36;rawP=0.0092;adjP=0.0334	*OR10T2; CRADD; CWF19L2; PCLO; RIF1; OR8S1*
Loss of Heterozygosity	C=342;O=12;E=5.45;R=2.20;rawP=0.0094;adjP=0.0340	*KANK1; DIRAS3; LIMD1; TET2; DMBT1; NF2; NF1; MSH2; PTCH1; DOCK8; PDE11A; TARM1*
Hypertrophy	C=147;O=7;E=2.34;R=2.99;rawP=0.0095;adjP=0.0342	*TNNT2; TTN; CMYA5; ABRA; NFATC4; AGPAT2; TRDN*
Obsessive-Compulsive Disorder	C=81;O=5;E=1.29;R=3.88;rawP=0.0096;adjP=0.0343	*CNTNAP1; LEKR1; MZT1; HTR5A; SLC1A2*
Gestational hypertension	C=113;O=6;E=1.80;R=3.33;rawP=0.0096;adjP=0.0343	*MMP9; HSPA4; VEGFA; AGER; KLRC2; PAPPA2*
Lipodystrophy	C=28;O=3;E=0.45;R=6.73;rawP=0.0098;adjP=0.0345	*C3; INSR; AGPAT2*
Dementia, Vascular	C=28;O=3;E=0.45;R=6.73;rawP=0.0098;adjP=0.0345	*COL4A1; ECE1; SREBF2*
Avian Leukosis	C=28;O=3;E=0.45;R=6.73;rawP=0.0098;adjP=0.0345	*ELK1; TRIM5; ISG15*
Heroin Dependence	C=28;O=3;E=0.45;R=6.73;rawP=0.0098;adjP=0.0345	*CYP2D6; DUSP27; XIRP2*
Translocation, Genetic	C=431;O=14;E=6.87;R=2.04;rawP=0.0100;adjP=0.0351	*CREB3L2; KANK1; TET2; SLC26A3; RTP4; DISC1; SLC8A3; ATP7B; ATP10A; ABCC2; NFATC4; COG3; RNF139; SLC1A2*
Brain Stem Neoplasms	C=53;O=4;E=0.84;R=4.74;rawP=0.0101;adjP=0.0352	*PTCH1; DMBT1; GLI1; ZNF521*
Autonomic Nervous System Diseases	C=53;O=4;E=0.84;R=4.74;rawP=0.0101;adjP=0.0352	*ELP4; ECE1; IKBKAP; LRRK2*
Dermal cellular nevus	C=10;O=2;E=0.16;R=12.56;rawP=0.0105;adjP=0.0361	*GNAQ; PRPH*
Apraxias	C=10;O=2;E=0.16;R=12.56;rawP=0.0105;adjP=0.0361	*PIK3R5; SETX*

Inflammatory Bowel Diseases	C=225;O=9;E=3.58;R=2.51;rawP=0.0105;adjP=0.0361	*CCR6; FOXP3; DMBT1; SLC2A4RG; AMIGO3; CD14; IL22; IL18R1; HSPA2*
Ventilator-acquired pneumonia	C=10;O=2;E=0.16;R=12.56;rawP=0.0105;adjP=0.0361	*CSF2RB; C5*
Tremor	C=54;O=4;E=0.86;R=4.65;rawP=0.0108;adjP=0.0368	*LINGO4; FMR1; LRRK2; SETX*
Tachycardia	C=54;O=4;E=0.86;R=4.65;rawP=0.0108;adjP=0.0368	*TNNT2; MARCH10; TRDN; NOS1AP*
Gram-Negative Bacterial Infections	C=116;O=6;E=1.85;R=3.25;rawP=0.0109;adjP=0.0370	*MMP9; C5; MUC1; ATP4A; CD14; CEACAM1*
Kidney Failure	C=152;O=7;E=2.42;R=2.89;rawP=0.0113;adjP=0.0383	*ACTN4; C3; KIF3B; TNNT2; AVP; AGER; NAT8B*
Coronary Stenosis	C=55;O=4;E=0.88;R=4.57;rawP=0.0115;adjP=0.0387	*MMP9; CD14; PECAM1; AGER*
Newcastle Disease	C=55;O=4;E=0.88;R=4.57;rawP=0.0115;adjP=0.0387	*OR5H2; ZNF33B; MSH2; ZNF37A*
Celiac Disease	C=153;O=7;E=2.44;R=2.87;rawP=0.0116;adjP=0.0389	*ACSL5; FOXP3; TMEM120B; LONRF2; SH3TC1; IL18R1; CD3G*
Mitochondrial Diseases	C=269;O=10;E=4.29;R=2.33;rawP=0.0117;adjP=0.0391	*MRPS22; SPG7; ETFDH; TACO1; PC; LARS2; WFS1; SLC25A19; KIF21A; VDAC1*
Meningococcal infection NOS	C=30;O=3;E=0.48;R=6.28;rawP=0.0119;adjP=0.0394	*C5; CD14; CEACAM1*
Lipoma	C=30;O=3;E=0.48;R=6.28;rawP=0.0119;adjP=0.0394	*LASP1; CD34; MGEA5*
Glomerulonephritis, IGA	C=56;O=4;E=0.89;R=4.48;rawP=0.0122;adjP=0.0400	*C3; C5; FCAMR; PIGR*
Disorders of Excessive Somnolence	C=119;O=6;E=1.90;R=3.17;rawP=0.0122;adjP=0.0400	*OR10T2; CRADD; CWF19L2; PCLO; RIF1; OR8S1*
Periodontitis	C=56;O=4;E=0.89;R=4.48;rawP=0.0122;adjP=0.0400	*MMP9; CD14; CTSC; MMP8*
Sensation Disorders	C=231;O=9;E=3.68;R=2.45;rawP=0.0123;adjP=0.0402	*MYO1A; CACNA1F; WFS1; CHML; PCDH15; ATP2B2; GPR98; DIAPH1; DFNA5*
Varicocele	C=11;O=2;E=0.18;R=11.41;rawP=0.0127;adjP=0.0410	*HSPA2; USP26*
Tachycardia, Supraventricular	C=11;O=2;E=0.18;R=11.41;rawP=0.0127;adjP=0.0410	*TNNT2; TRDN*
Thrombosis	C=120;O=6;E=1.91;R=3.14;rawP=0.0127;adjP=0.0410	*F13B; GNAQ; FGA; PECAM1; PDE3A; ARHGEF10*
Plexiform neurofibroma	C=11;O=2;E=0.18;R=11.41;rawP=0.0127;adjP=0.0410	*WISP1; NF1*
Critical Illness	C=31;O=3;E=0.49;R=6.08;rawP=0.0130;adjP=0.0415	*DISC1; CD14; AGER*
Lumbar Disc Herniation	C=31;O=3;E=0.49;R=6.08;rawP=0.0130;adjP=0.0415	*MMP9; IL18R1; ACAN*
Insulin Resistance	C=194;O=8;E=3.09;R=2.59;rawP=0.0130;adjP=0.0415	*C3; INSR; AGPAT2; PCSK1; WFS1; OGT; PRKAG3; IGFBP7*
Uterine Cervical Neoplasms	C=121;O=6;E=1.93;R=3.11;rawP=0.0132;adjP=0.0420	*MMP9; ERBB2; VEGFA; WAPAL; LHX6; TKTL1*
Deafness,nonsyndromic	C=88;O=5;E=1.40;R=3.57;rawP=0.0134;adjP=0.0425	*DIAPH1; WFS1; PCDH15; DFNA5; TRIOBP*
Neoplasms, Squamous Cell	C=236;O=9;E=3.76;R=2.39;rawP=0.0140;adjP=0.0442	*MMP9; MDK; PTPRD; FOXN1; ERBB2; MCM2; VEGFA; CDH2; TKTL1*
Hypoxia-Ischemia, Brain	C=32;O=3;E=0.51;R=5.89;rawP=0.0142;adjP=0.0443	*YTHDC1; VEGFA; AGER*
Respiratory Distress Syndrome, Adult	C=32;O=3;E=0.51;R=5.89;rawP=0.0142;adjP=0.0443	*CCDC40; HGF; VEGFA*
Hypotension	C=32;O=3;E=0.51;R=5.89;rawP=0.0142;adjP=0.0443	*CD34; AVP; SERPINA6*
Hemoglobinuria	C=32;O=3;E=0.51;R=5.89;rawP=0.0142;adjP=0.0443	*CD14; PIGA; KLRC2*
Preterm rupture of membranes	C=160;O=7;E=2.55;R=2.75;rawP=0.0146;adjP=0.0451	*MMP9; F13B; COL4A1; CD14; PECAM1; COL4A2; MMP8*
Ocular Motility Disorders	C=59;O=4;E=0.94;R=4.26;rawP=0.0146;adjP=0.0451	*CACNA1F; KIF21A; LRRK2; SETX*

Diabetic Angiopathies	C=59;O=4;E=0.94;R=4.26;rawP=0.0146;adjP=0.0451	*MMP9; VEGFA; AGER; NOS1AP*
Wounds and Injuries	C=90;O=5;E=1.43;R=3.49;rawP=0.0147;adjP=0.0452	*MMP9; LRP5; CD14; CD34; VEGFA*
Hypoalbuminaemia	C=12;O=2;E=0.19;R=10.46;rawP=0.0150;adjP=0.0457	*PIK3R5; SETX*
Tinnitus	C=12;O=2;E=0.19;R=10.46;rawP=0.0150;adjP=0.0457	*NF2; WFS1*
Acanthosis Nigricans	C=12;O=2;E=0.19;R=10.46;rawP=0.0150;adjP=0.0457	*INSR; AGPAT2*
Cholestasis	C=60;O=4;E=0.96;R=4.19;rawP=0.0154;adjP=0.0459	*ABCC2; NUP210; SP100; UGT1A3*
Rhinitis, Allergic, Perennial	C=33;O=3;E=0.53;R=5.71;rawP=0.0154;adjP=0.0459	*FOXP3; CD14; MUC5AC*
Heart Arrest	C=60;O=4;E=0.96;R=4.19;rawP=0.0154;adjP=0.0459	*SLC35F1; TNNT2; MARCH10; NOS1AP*
Drug Hypersensitivity	C=60;O=4;E=0.96;R=4.19;rawP=0.0154;adjP=0.0459	*PTGDR2; FCER1A; FANCM; DCBLD2*
Albuminuria	C=60;O=4;E=0.96;R=4.19;rawP=0.0154;adjP=0.0459	*VEGFA; AGER; PCSK1; LRP2*
Hemolysis	C=60;O=4;E=0.96;R=4.19;rawP=0.0154;adjP=0.0459	*C3; C5; PIGA; C8B*
Heart Diseases	C=366;O=12;E=5.83;R=2.06;rawP=0.0153;adjP=0.0459	*MMP9; TTN; CMYA5; MIA3; TNNT2; NFATC4; FGA; CD14; PECAM1; AGER; NOS1AP; PDE3A*
Congenital Abnormalities	C=643;O=18;E=10.24;R=1.76;rawP=0.0161;adjP=0.0478	*CDK5RAP2; PCDH15; ATP10A; PTCH1; FXR2; KRT10; IKBKAP; NOS1AP; TRPS1; SPG7; WFS1; SLC9A6; PEX16; GPR98; FGD4; CCDC40; FMR1; MCPH1*
Lymphopenia	C=61;O=4;E=0.97;R=4.12;rawP=0.0163;adjP=0.0480	*WIPF3; IL2RG; DOCK8; CD3G*
Periodontal Diseases	C=61;O=4;E=0.97;R=4.12;rawP=0.0163;adjP=0.0480	*MMP9; CD14; CTSC; MMP8*
Carcinoma, Squamous Cell	C=242;O=9;E=3.85;R=2.33;rawP=0.0162;adjP=0.0480	*MMP9; MDK; FOXN1; ERBB2; MCM2; VEGFA; KRT10; CDH2; TKTL1*
Organ Transplantation	C=128;O=6;E=2.04;R=2.94;rawP=0.0169;adjP=0.0496	*C3; FOXP3; HGF; TNNT2; ABCC2; VEGFA*

Pathway Commons	**Parameters**	**Genes**
Olfactory Signaling Pathway	C=370;O=40;E=5.89;R=6.79;rawP=2.19e-21;adjP=7.42e-19	*OR10A3; OR52A5; OR10C1; OR52N4; OR6S1; OR2G3; OR1J4; OR7C1; OR51D1; OR8B3; OR11A1; OR4P4; OR1F12; OR7G2; OR7D4; OR9A4; OR10T2; OR1D2; OR6C75; OR2T2; OR4C3; OR1L1; OR51B2; OR7A5; OR51I2; OR8S1; OR4A47; OR1L3; OR2T33; OR52J3; OR14A16; OR6K3; OR1S1; OR6C4; OR6M1; OR5H2; OR5M10; OR2B6; OR1G1; OR1L8*
Signaling by GPCR	C=823;O=57;E=13.11;R=4.35;rawP=2.30e-20;adjP=2.60e-18	*EMR2; OR10A3; OR52A5; OR10C1; CHRM1; C5; OR52N4; OR6S1; FSHR; OR2G3; PDE3A; OR1J4; CCR6; RAC2; OR7C1; OR51D1; OR8B3; OR11A1; OR4P4; PIK3R5; OR1F12; OR7G2; OR7D4; OR9A4; OR10T2; PTGDR2; OR1D2; AVP; OR6C75; C3; OR2T2; OR4C3; OR1L1; PTCH1; OR51B2; CHRM4; HTR5A; OR7A5; OR8S1; OR51I2; PDE11A; OR4A47; DAGLA; OR1L3; OR2T33; OR14A16; OR52J3; OR6K3; OR6M1; OR1S1; OR6C4; OR5H2; GNAQ; OR5M10; OR2B6; OR1G1; OR1L8*
GPCR downstream signaling	C=553;O=47;E=8.81;R=5.34;rawP=1.60e-20;adjP=2.60e-18	*OR10A3; OR52A5; OR10C1; CHRM1; OR52N4; OR6S1; OR2G3; PDE3A; OR1J4; OR7C1; RAC2; OR51D1; OR8B3; OR11A1; OR4P4; PIK3R5; OR1F12; OR7G2; OR7D4; OR9A4; OR10T2; OR1D2; OR6C75; OR2T2; OR4C3; OR1L1; OR51B2; OR7A5; OR8S1; OR51I2; PDE11A; OR4A47; DAGLA; OR1L3; OR2T33; OR14A16;*

		OR52J3; OR6K3; OR6M1; OR1S1; OR6C4; OR5H2; GNAQ; OR5M10; OR2B6; OR1G1; OR1L8
Signal Transduction	C=1231;O=66;E=19.61;R=3.37;rawP=1.25e-17;adjP=1.06e-15	*OR10A3; OR52A5; PDE3A; OR1J4; CCR6; OR7C1; OR1F12; OR7G2; OR7D4; AVP; C3; PTCH1; OR51B2; CHRM4; MAG; HTR5A; OR7A5; OR8S1; PRKAG3; OR51I2; OR4A47; OR14A16; OR52J3; PSMD3; OR6K3; OR6M1; OR1S1; OR6C4; OR5H2; OR5M10; OR1L8; EMR2; MMP9; OR10C1; CHRM1; C5; OR6S1; OR52N4; FSHR; PECAM1; OR2G3; RAC2; OR8B3; OR51D1; PIK3R5; OR4P4; OR11A1; OR9A4; PTGDR2; OR10T2; OR1D2; OR6C75; OR4C3; OR2T2; OR1L1; ELK1; PDE11A; DAGLA; OR1L3; OR2T33; INSR; SH3GL2; GNAQ; FGA; OR2B6; OR1G1*
Ion transport by P-type ATPases	C=29;O=6;E=0.46;R=12.99;rawP=5.56e-06;adjP=0.0004	*ATP10A; ATP4A; ATP2A3; ATP7B; ATP2C1; ATP2B2*
Transmembrane transport of small molecules	C=379;O=18;E=6.04;R=2.98;rawP=4.62e-05;adjP=0.0022	*SLC8A3; ABCC11; SLC13A3; ATP2A3; ATP2C1; ATP2B2; AQP1; ATP10A; ABCC2; ATP4A; SLC26A3; NUP210; ATP7B; SLC9A6; SLC9A3; AVP; SLC47A2; SLC1A2*
Hemostasis	C=376;O=18;E=5.99;R=3.01;rawP=4.17e-05;adjP=0.0022	*F13B; KIF3B; KIF3C; TTN; CD48; SLC8A3; ATP2A3; ATP2B2; PECAM1; DAGLA; RAC2; ESAM; PIK3R5; PPIL2; GNAQ; FGA; GATA5; DOCK8*
Ion channel transport	C=46;O=6;E=0.73;R=8.19;rawP=8.71e-05;adjP=0.0037	*ATP10A; ATP4A; ATP2A3; ATP7B; ATP2C1; ATP2B2*
Reduction of cytosolic Ca++ levels	C=10;O=3;E=0.16;R=18.83;rawP=0.0004;adjP=0.0151	*SLC8A3; ATP2A3; ATP2B2*
Meiotic Synapsis	C=47;O=5;E=0.75;R=6.68;rawP=0.0009;adjP=0.0254	*SYCP2; SYCE1; SYCP1; TERF1; HSPA2*
Integrin family cell surface interactions	C=1378;O=38;E=21.95;R=1.73;rawP=0.0009;adjP=0.0254	*MMP9; SSH1; NF1; CD3G; ERBB2; SMARCA4; ZFP36L1; PECAM1; PIGR; BIRC2; DDIT4; CSF2RB; ZYX; MSH2; COL4A1; IPCEF1; C3; FOXP3; DOCK4; RHOQ; IL2RG; KALRN; CD14; SMARCC2; VEGFA; ELK1; DIAPH1; MDK; INSR; HSPA4; PAK3; SH3GL2; ACTN4; GNAQ; HGF; APAF1; FGANA; CDH2*
Activation of C3 and C5	C=3;O=2;E=0.05;R=41.85;rawP=0.0008;adjP=0.0254	*C5; C3*
Meiosis	C=72;O=6;E=1.15;R=5.23;rawP=0.0010;adjP=0.0261	*SYCP2; SYCE1; MSH4; SYCP1; TERF1; HSPA2*
Regulation of RAC1 activity	C=205;O=10;E=3.27;R=3.06;rawP=0.0018;adjP=0.0358	*FARP1; SH3GL2; NF1; SLC9A3; ECT2; ARHGEF10; KALRN; ARHGEF2; DIAPH1; DEF6*
Regulation of RhoA activity	C=205;O=10;E=3.27;R=3.06;rawP=0.0018;adjP=0.0358	*FARP1; SH3GL2; NF1; SLC9A3; ECT2; ARHGEF10; KALRN; ARHGEF2; DIAPH1; DEF6*
Platelet calcium homeostasis	C=16;O=3;E=0.25;R=11.77;rawP=0.0019;adjP=0.0358	*SLC8A3; ATP2A3; ATP2B2*
RhoA signaling pathway	C=205;O=10;E=3.27;R=3.06;rawP=0.0018;adjP=0.0358	*FARP1; SH3GL2; NF1; SLC9A3; ECT2; ARHGEF10; KALRN; ARHGEF2; DIAPH1; DEF6*
RAC1 signaling pathway	C=205;O=10;E=3.27;R=3.06;rawP=0.0018;adjP=0.0358	*FARP1; SH3GL2; NF1; SLC9A3; ECT2; ARHGEF10; KALRN; ARHGEF2; DIAPH1; DEF6*
Beta1 integrin cell surface interactions	C=1351;O=36;E=21.52;R=1.67;rawP=0.0021;adjP=0.0375	*MMP9; SSH1; NF1; CD3G; ERBB2; SMARCA4; ZFP36L1; PIGR; BIRC2; DDIT4; CSF2RB; ZYX; MSH2; COL4A1; IPCEF1; FOXP3; DOCK4; RHOQ; IL2RG; KALRN; CD14; SMARCC2; VEGFA; ELK1; DIAPH1; MDK; INSR; HSPA4; PAK3; SH3GL2; ACTN4; GNAQ; HGF; APAF1; FGA; CDH2*
Immune System	C=532;O=18;E=8.47;R=2.12;rawP=0.0025;adjP=0.0404	*C3; IL2RG; CD3G; ECSIT; C5; CD14; PECAM1; IFIT3; ELK1; ISG15; C8B; CSF2RB; SP100; PSMD3; ERAP1; SPTBN5; IFIT1; AGER*

Muscarinic acetylcholine receptors	C=5;O=2;E=0.08;R=25.11;rawP=0.0025;adjP=0.0404	*CHRM4; CHRM1*
Cell-Cell communication	C=117;O=7;E=1.86;R=3.76;rawP=0.0028;adjP=0.0431	*ACTN4; FLNC; SPTBN5; CDH2; TESK1; KIRREL3; INADL*
Glypican pathway	C=1338;O=35;E=21.31;R=1.64;rawP=0.0032;adjP=0.0472	*MMP9; SSH1; NF1; CD3G; ERBB2; SMARCA4; ZFP36L1; PIGR; BIRC2; DDIT4; ZYX; MSH2; IPCEF1; FOXP3; DOCK4; RHOQ; IL2RG; KALRN; PTCH1; LRP5; VEGFA; ELK1; SMARCC2; DIAPH1; MDK; INSR; HSPA4; PAK3; SH3GL2; ACTN4; GNAQ; HGF; APAF1; FGA; CDH2*

KEGG	Parameters	Gene
Olfactory transduction	C=388;O=46;E=6.18;R=7.44;rawP=3.95e-26;adjP=4.46e-24	*OR10A3; OR8G1; OR52A5; OR10C1; OR52N4; OR6S1; OR2G3; OR1J4; OR7C1; OR51D1; OR11A1; OR4P4; OR7G2; OR11H6; OR7D4; OR9A4; OR10T2; OR1D2; OR6C75; OR13C3; OR2T2; OR4C3; OR11L1; OR1L1; OR51B2; OR7A5; OR11H4; OR51I2; OR8S1; OR4A47; OR1L3; OR2T33; OR14A16; OR52J3; OR12D2; CLCA2; OR6K3; OR6M1; OR1S1; OR6C4; OR5H2; OR5M10; OR2B6; OR1G1; OR13A1; OR1L8*
Focal adhesion	C=200;O=15;E=3.19;R=4.71;rawP=9.01e-07;adjP=5.09e-05	*FLNC; RAC2; ZYX; PAK3; COL4A2; PIK3R5; ERBB2; ACTN4; HGF; COL4A1; VEGFA; ELK1; DIAPH1; BIRC2; LAMB4*
Amoebiasis	C=106;O=9;E=1.69;R=5.33;rawP=5.21e-05;adjP=0.0020	*COL4A2; PIK3R5; ACTN4; MUC2; GNAQ; COL4A1; CD14; C8B; LAMB4*
Complement and coagulation cascades	C=69;O=6;E=1.10;R=5.46;rawP=0.0008;adjP=0.0151	*F13B; C5; CR2; C3; FGA; C8B*
Pathways in cancer	C=326;O=14;E=5.19;R=2.70;rawP=0.0008;adjP=0.0151	*MMP9; RAC2; COL4A2; PIK3R5; MSH2; PTCH1; ERBB2; HGF; COL4A1; VEGFA; APC2; BIRC2; LAMB4; GLI1*
ABC transporters	C=44;O=5;E=0.70;R=7.13;rawP=0.0007;adjP=0.0151	*ABCC2; ABCC12; ABCC11; ABCA13; ABCA10*
Pancreatic secretion	C=101;O=7;E=1.61;R=4.35;rawP=0.0012;adjP=0.0194	*SLC26A3; GNAQ; CPA3; PLA2G1B; CLCA2; ATP2A3; ATP2B2*
Cell adhesion molecules (CAMs)	C=133;O=8;E=2.12;R=3.78;rawP=0.0014;adjP=0.0198	*CNTNAP1; ESAM; MAG; PECAM1; CD34; CDH2; SIGLEC1; PTPRM*
Regulation of actin cytoskeleton	C=213;O=10;E=3.39;R=2.95;rawP=0.0024;adjP=0.0226	*RAC2; PAK3; SSH1; PIK3R5; CHRM1; ACTN4; CHRM4; CD14; APC2; DIAPH1*
Protein digestion and absorption	C=81;O=6;E=1.29;R=4.65;rawP=0.0019;adjP=0.0226	*COL4A1; SLC8A3; COL4A2; CPA3; COL15A1; SLC9A3*
Calcium signaling pathway	C=177;O=9;E=2.82;R=3.19;rawP=0.0023;adjP=0.0226	*CACNA1F; SLC8A3; ATP2A3; ATP2B2; CHRM1; ERBB2; GNAQ; VDAC1; HTR5A*
Small cell lung cancer	C=85;O=6;E=1.35;R=4.43;rawP=0.0024;adjP=0.0226	*COL4A1; APAF1; COL4A2; PIK3R5; BIRC2; LAMB4*
MAPK signaling pathway	C=268;O=11;E=4.27;R=2.58;rawP=0.0041;adjP=0.0356	*FLNC; CACNB1; CACNA1F; RAC2; NF1; PLA2G1B; ECSIT; NFATC4; CD14; ELK1; HSPA2*

Molecular Function	Parameters	Gene
olfactory receptor activity	C=419;O=48;E=15.49;R=3.10;rawP=2.86e-12;adjP=1.24e-09	*OR10A3; OR8G1; OR52A5; OR10C1; OR52N4; OR6S1; OR2G3; OR1J4; OR7C1; OR51D1; OR8B3; OR1F12; OR11A1; OR4P4; OR5AC1; OR7G2; OR7D4; OR11H6; OR9A4; OR10T2; OR1D2; OR6C75; OR13C3; OR2T2; OR4C3; OR11L1; OR1L1; OR51B2; OR11H4; OR7A5; OR51I2; OR8S1; OR4A47; OR1L3; OR2T33; OR52J3; OR14A16; OR12D2; OR6K3; OR6C4; OR6M1; OR1S1; OR5H2; OR5M10; OR2B6; OR13A1; OR1G1; OR1L8*

transmembrane signaling receptor activity	C=1231;O=86;E=45.51;R=1.89;rawP=5.37e-09;adjP=1.17e-06	*OR10A3; OR8G1; OR52A5; ERBB2; VN1R1; CD5; CCR6; OR1J4; OR7C1; OR1F12; TAS2R8; OR7G2; OR7D4; GPR98; GPR119; KLRC2; OR13C3; PTPRM; GPR110; OR51B2; PTCH1; CR2CHRM4; LRP5; CD14; HTR5A; OR7A5; OR8S1; OR51I2; OR4A47; OR52J3; OR14A16; MUSK; OR6K3; OR6C4; OR1S1; OR6M1; OR5H2; OR5M10; OR1L8; EMR2; PTPRD; OR10C1; CD3G; CHRM1; OR52N4; OR6S1; FSHR; IL18R1; OR2G3; CSF2RB; OR8B3; OR51D1; OR4P4; OR11A1; OR5AC1; OR11H6; OR9A4; OR10T2; PTGDR2; FCER1A; OR1D2; AGER; OR6C75; OR4C3; OR2T2; SIGMAR1; OR11L1; IL2RG; OR1L1; OR11H4; ANTXR1; OR1L3; PTCHD2; INSR; OR2T33; LTBP4; FRZB; GPR124; OR12D2; GPR133; OSMR; FCAMR; OR2B6; OR13A1; OR1G1*
signaling receptor activity	C=1325;O=88;E=48.98;R=1.80;rawP=3.78e-08;adjP=5.47e-06	*OR10A3; OR8G1; OR52A5; ERBB2; VN1R1; CD5; CCR6; OR1J4; OR7C1; DMBT1; OR1F12; TAS2R8; OR7G2; OR7D4; GPR98; GPR119; KLRC2; OR13C3; PTPRM; GPR110; OR51B2; PTCH1; CR2; CHRM4; LRP5; CD14; HTR5A; OR7A5; OR8S1; OR51I2; OR4A47; OR52J3; OR14A16; MUSK; OR6K3; OR6C4; OR1S1; OR6M1; OR5H2; OR5M10; OR1L8; EMR2; PTPRD; OR10C1; CD3G; CHRM1; OR52N4; OR6S1; FSHR; IL18R1; OR2G3; CSF2RB; OR8B3; OR51D1; OR4P4; OR11A1; OR5AC1; KCNH4; OR11H6; OR9A4; OR10T2; PTGDR2; FCER1A; OR1D2; AGER; OR6C75; OR4C3; OR2T2; SIGMAR1; OR11L1; IL2RG; OR1L1; OR11H4; ANTXR1; OR1L3; PTCHD2; INSR; OR2T33; LTBP4; FRZB; GPR124; OR12D2; GPR133; OSMR; FCAMR; OR2B6; OR13A1; OR1G1*
receptor activity	C=1531;O=97;E=56.60;R=1.71;rawP=6.84e-08;adjP=7.42e-06	*OR10A3; OR8G1; OR52A5; ERBB2; VN1R1; CD5; CCR6; OR1J4; OR7C1; DMBT1; OR1F12; TAS2R8; OR7G2; OR7D4; GPR98; GPR119; KLRC2; SEMA6C; OR13C3; PTPRM; GPR110; OR51B2; PTCH1; CR2; CHRM4; LRP5; CD14; HTR5A; OR7A5; OR8S1; OR51I2; OR4A47; OR52J3; OR14A16; MUSK; OR6K3; OR6C4; OR1S1; OR6M1; OR5H2; OR5M10; OR1L8; EMR2; PTPRD; OR10C1; CD3G; CHRM1; OR52N4; OR6S1; FSHR; IL18R1; OR2G3; CSF2RB; OR51D1; OR8B3; OR4P4; OR11A1; OR5AC1; KCNH4; FCGRT; OR11H6; OR9A4; OR10T2; PTGDR2; FCER1A; OR1D2; AMOT; AGER; OR6C75; OR2T2; OR4C3; CD48; LILRB4; SIGMAR1; OR11L1; IL2RG; OR1L1; PKHD1; OR11H4; ANTXR1; CNTNAP1; OR1L3; PTCHD2; INSR; OR2T33; LTBP4; FRZB; GPR124; OR12D2; GPR133; OSMR; TG; FCAMR; RNF139; OR2B6; OR13A1; OR1G1*
G-protein coupled receptor activity	C=862;O=63;E=31.87;R=1.98;rawP=1.57e-07;adjP=1.36e-05	*EMR2; OR10A3; OR8G1; OR52A5; OR10C1; CHRM1; VN1R1; OR52N4; OR6S1; FSHR; OR2G3; CCR6; OR1J4; OR7C1; OR51D1; OR8B3; OR1F12; OR11A1; OR4P4; TAS2R8; OR5AC1; OR7G2; OR7D4; GPR98; OR11H6; OR9A4; OR10T2; PTGDR2; OR1D2; GPR119; OR6C75; OR13C3; OR2T2; OR4C3; SIGMAR1; OR11L1; GPR110; OR1L1; OR51B2; CHRM4; OR11H4; OR7A5; HTR5A; OR51I2; OR8S1; OR4A47; OR1L3; OR2T33; OR52J3; OR14A16; GPR124; OR12D2; OR6K3; OR6C4; GPR133; OR1S1; OR6M1; OR5H2; OR5M10; OR2B6; OR1G1; OR13A1; OR1L8*

molecular transducer activity	C=1628;O=96;E=60.18;R=1.60;rawP=2.15e-06;adjP=0.0001	*OR10A3; OR8G1; OR52A5; ERBB2; VN1R1; CD5; CCR6; OR1J4; OR7C1; DMBT1; OR1F12; TAS2R8; OR7G2; OR7D4; GPR98; AVP; GPR119; KLRC2; OR13C3; PTPRM; GPR110; ECT2; OR51B2; PTCH1; CR2; CHRM4; LRP5; CD14; HTR5A; OR7A5; OR8S1; OR51I2; OR4A47; OR52J3; OR14A16; MUSK; OR6K3; OR6C4; OR1S1; BST2; OR6M1; OR5H2; OR5M10; OR1L8; EMR2; PTPRD; OR10C1; CD3G; CHRM1; OR52N4; OR6S1; FSHR; IKBKAP; IL18R1; OR2G3; CSF2RB; ASZ1; OR51D1; OR8B3; OR4P4; OR11A1; OR5AC1; KCNH4; OR11H6; OR9A4; OR10T2; PTGDR2; FCER1A; OR1D2; AGER; OR6C75; OR2T2; OR4C3; SIGMAR1; OR11L1; ATP2C1; IL2RG; OR1L1; OR11H4; ANTXR1; OR1L3; PTCHD2; INSR; OR2T33; LTBP4; FRZB; GPR124; OR12D2; GPR133; VAPA; OSMR; GNAQ; FCAMR; OR2B6; OR13A1; OR1G1*
signal transducer activity	C=1628;O=96;E=60.18;R=1.60;rawP=2.15e-06;adjP=0.0001	*OR10A3; OR8G1; OR52A5; ERBB2; VN1R1; CD5; CCR6; OR1J4; OR7C1; DMBT1; OR1F12; TAS2R8; OR7G2; OR7D4; GPR98; AVP; GPR119; KLRC2; OR13C3; PTPRM; GPR110; ECT2; OR51B2; PTCH1; CR2; CHRM4; LRP5; CD14; HTR5A; OR7A5; OR8S1; OR51I2; OR4A47; OR52J3; OR14A16; MUSK; OR6K3; OR6C4; OR1S1; BST2; OR6M1; OR5H2; OR5M10; OR1L8; EMR2; PTPRD; OR10C1; CD3G; CHRM1; OR52N4; OR6S1; FSHR; IKBKAP; IL18R1; OR2G3; CSF2RB; ASZ1; OR51D1; OR8B3; OR4P4; OR11A1; OR5AC1; KCNH4; OR11H6; OR9A4; OR10T2; PTGDR2; FCER1A; OR1D2; AGER; OR6C75; OR2T2; OR4C3; SIGMAR1; OR11L1; ATP2C1; IL2RG; OR1L1; OR11H4; ANTXR1; OR1L3; PTCHD2; INSR; OR2T33; LTBP4; FRZB; GPR124; OR12D2; GPR133; VAPA; OSMR; GNAQ; FCAMR; OR2B6; OR13A1; OR1G1*
Rho GTPase binding	C=49;O=8;E=1.81;R=4.42;rawP=0.0004;adjP=0.0217	*ARHGEF2; DOCK4; KIF3B; PAK3; DIAPH1; LRRK2; TRIOBP; ECT2*
cytoskeletal protein binding	C=638;O=40;E=23.58;R=1.70;rawP=0.0008;adjP=0.0386	*MYH4; LASP1; TTN; FARP1; SSH1; NF2; HOOK2; APC2; DST; MARK4; STARD9; GPR98; ARHGEF2; ABRA; CCDC88C; MICAL3; GLI1; FLNC; PPP1R18; LRRK2; CDK5RAP2; ARHGEF10; CROCC; TNS1; MYO7B; ANTXR1; STK38L; DIAPH1; XIRP2; MYO1A; FKBP15; WIPF3; TRIOBP; FGD4; ACTN4; SPTBN5; TNNT2; MTUS2; ERMN; TERF1*

Cellular Component	**Parameters**	**Genes**
cell periphery	C=4377;O=207;E=160.16;R=1.29;rawP=1.22e-05;adjP=0.0041	*SSH1; NF1; OR52A5; AQP1; VN1R1; PIGR; VNN3; HMCN1; BIRC2; CD5; RIC8A; CCR6; DST; OR7C1; INPP5E; ZYX; OR1F12; TRPV5; OR7G2; ABRA; COG3; USP14; KLRC2; CEACAM1; FLNC; C3; EQTN; BRPF1; PTPRM; GPR110; LRRK2; OR51B2; MGAM; CR2; AKAP12; LRP5; ABCC2; OR7A5; STK38L; DLST; OR51I2; OR4A47; MUC1; OR52J3; MUSK; ATP7B; OR6K3; OR6C4; OR1S1; OGT; BST2; ACTN4; OR5M10; SIGLEC1; MUC13; EMR2; LASP1; PTPRD; PCDH15; CD3G; C5; SLC6A17; ATP4A; OR52N4; OR6S1; FSHR; OR2G3; IL18R1; C8B; RAC2; OR8B3; IYD; OR4P4; ICAM5; OR5AC1; KCNH4; FCGRT; OR9A4; OR1D2; KIRREL3; OR6C75; DISC1; OR4C3; RHOQ; PODXL; SLC13A3; SIGMAR1; OR11L1; IL2RG; RASAL1; OR11H4; ANTXR1; DAGLA; OR1L3; INSR; SH3GL2; CDH17; OSMR; TG; GNAQ; FCAMR; FGA; CD34; CDH2; OR2B6; OR1G1; OR13A1; HAUS7; CACNA1F; OR8G1; OR10A3; NF2; TRDN; CCDC110; NAALADL1; CHD1L; ESYT3; ERBB2; OOEP; APC2; SYTL2; OR1J4; ESAM; MUC16; GPR98; OR7D4; PLIN4; ECE1; GPR119; HSPA2; SEMA6C; OR13C3; SLC8A3; ATP10A; PTCH1; CHRM4;*

		CROCC; CD14; MEGF11; MAG; ART3; VDAC1; HTR5A; RHBDL2; INADL; DCBLD2; OR8S1; DUOX1; BFSP1; OR14A16; SLC9A6; CLCA2; OR6M1; OR5H2; DSG1; SPTBN5; DYTN; SLC47A2; OR1L8; CAMKV; NEU3; ESRP1; OR10C1; LRP2; CHRM1; KHDC3L; PECAM1; DIRAS3; CSF2RB; HMGCS1; OR51D1; OR11A1; OR11H6; TCHH; ARHGEF2; OR10T2; PTGDR2; IPCEF1; FCER1A; AMOT; AGER; SLC1A2; KANK1; OR2T2; CD48; LILRB4; ERMAP; ABCC11; ATP2B2; OR1L1; PKHD1; MYO7B; DIAPH1; CNTNAP1; MYO1A; SLC26A3; CACNB1; OR2T33; FRZB; GPR124; IGSF11; OR12D2; SLC9A3; GPR133; VAPA; MTMR1; DEF6
plasma membrane	C=4289;O=201;E=156.94;R=1.28;rawP=3.22e-05;adjP=0.0055	*SSH1; NF1; OR52A5; AQP1; VN1R1; PIGR; VNN3; BIRC2; CD5; RIC8A; CCR6; DST; OR7C1; INPP5E; ZYX; OR1F12; TRPV5; OR7G2; ABRA; COG3; USP14; KLRC2; CEACAM1; FLNC; C3; EQTN; BRPF1; PTPRM; GPR110; LRRK2; OR51B2; MGAM; CR2; AKAP12; LRP5; ABCC2; OR7A5; STK38L; DLST; OR51I2; OR4A47; MUC1; OR52J3; MUSK; ATP7B; OR6K3; OR6C4; OR1S1; OGT; BST2; OR5M10; SIGLEC1; MUC13; EMR2; PTPRD; PCDH15; CD3G; C5; SLC6A17; ATP4A; OR52N4; OR6S1; FSHR; OR2G3; IL18R1; C8B; RAC2; OR8B3; IYD; OR4P4; ICAM5; OR5AC1; KCNH4; FCGRT; OR9A4; OR1D2; KIRREL3; OR6C75; DISC1; OR4C3; RHOQ; PODXL; SLC13A3; SIGMAR1; OR11L1; IL2RG; RASAL1; OR11H4; ANTXR1; DAGLA; OR1L3; INSR; SH3GL2; CDH17; OSMR; TG; GNAQ; FCAMR; FGA; CD34; CDH2; OR2B6; OR1G1; OR13A1; HAUS7; CACNA1F; OR8G1; OR10A3; NF2; TRDN; CCDC110; NAALADL1; CHD1L; ESYT3; ERBB2; APC2; SYTL2; OR1J4; ESAM; MUC16; GPR98; OR7D4; PLIN4; ECE1; GPR119; HSPA2; SEMA6C; OR13C3; SLC8A3; ATP10A; PTCH1; CHRM4; CROCC; CD14; MEGF11; MAG; ART3; VDAC1; HTR5A; RHBDL2; INADL; DCBLD2; OR8S1; DUOX1; BFSP1; OR14A16; SLC9A6; CLCA2; OR6M1; OR5H2; DSG1; DYTN; SLC47A2; OR1L8; CAMKV; NEU3; ESRP1; OR10C1; LRP2; CHRM1; PECAM1; DIRAS3; CSF2RB; HMGCS1; OR51D1; OR11A1; OR11H6; TCHH; ARHGEF2; OR10T2; PTGDR2; IPCEF1; FCER1A; AMOT; AGER; SLC1A2; KANK1; OR2T2; CD48; LILRB4; ERMAP; ABCC11; ATP2B2; OR1L1; PKHD1; MYO7B; DIAPH1; CNTNAP1; MYO1A; SLC26A3; CACNB1; OR2T33; FRZB; GPR124; IGSF11; OR12D2; SLC9A3; GPR133; VAPA; MTMR1; DEF6*
apical plasma membrane	C=224;O=21;E=8.20;R=2.56;rawP=7.86e-05;adjP=0.0070	*PODXL; NAALADL1; ATP2B2; LRP2; AQP1; MGAM; ERBB2; PKHD1; ABCC2; ATP4A; MYO7B; INADL; MYO1A; SLC26A3; MUC1; DUOX1; TRPV5; SLC9A3; BST2; CD34; MUC13*
basal plasma membrane	C=30;O=7;E=1.10;R=6.38;rawP=8.29e-05;adjP=0.0070	*DST; MYO1A; ERBB2; CD34; AGER; CLCA2; AQP1*
basal part of cell	C=37;O=7;E=1.35;R=5.17;rawP=0.0003;adjP=0.0127	*DST; MYO1A; ERBB2; CD34; AGER; CLCA2; AQP1*

non-membrane-bounded organelle	C=3800;O=175;E=139.05;R=1.26;rawP=0.0003;adjP=0.0127	*MYH4; RP1L1; SSH1; TRIM24; TRIM5; PTPN23; SUDS3; ATF5; FAM188A; DST; INPP5E; ZYX; GOLGA3; ABRA; COG3; SRRM2; KY; DNMT1; FLNC; CMYA5; CENPA; SYCP1; GPR110; KALRN; AKAP12; TNS1; PBRM1; SMC2; KRT10; MCM2; STK38L; KIF16B; C2CD3; MUC1; HLCS; WIPF3; HTATSF1; OGT; ACTN4; RNF168; KRTAP9-4; MTUS2; FAM60A; SLC25A26; LIMD1; LRRFIP1; LASP1; TTN; EHHADH; TFB2M; PCDH15; FAM71B; RPTN; LYAR; ASZ1; SYCP2; INTS7; RPL39; PPIL2; MSH2; INO80C; MCPH1; UHRF1; REPIN1; PRPH; GATAD2B; KRT20; DISC1; RHOQ; ECSIT; WRN; MBIP; CCDC86; EHMT2; SLU7; WAPAL; TNKS2; FGD4; TNNT2; HAUS7; AKAP9; KIF3C; NF2; URB2; CCDC110; CHD1L; SMARCA4; CWC22; HOOK2; MDC1; APC2; URB1; PCLO; TRIM66; NOP56; MRPL50; DECR1; SP100; TACC2; AURKC; RRP9; MSH4; CCDC15; KDM3B; DOCK4; RPL29; PPP1R18; RPL35; SLC8A3; ARHGEF10; ECT2; CHRM4; CROCC; VDAC1; MRPL14; PEG3; NIN; PES1; HELLS; HSPA4; BFSP1; DHX16; TRIOBP; SPTBN5; FAU; CBX7; CBX4; SETX; KIF3B; FARP1; MZT1; SPAG16; ESRP1; RIF1; CHRM1; HMGN3; IKBKAP; SAPCD2; MRPS22; MARK4; STARD9; HMGCS1; PPIG; TCHH; ARHGEF2; GCN1L1; DIS3; AMOT; FMR1; MICAL3; CDK5RAP2; PARP4; PKHD1; FXR2; SMARCC2; MYO7B; DIAPH1; SMTNL1; UTP20; MYO1A; MRPL4; FKBP15; NCAPD3; ZNF77; NFATC4; KIF21A; SYCE1; APAF1; KRTAP16-; ERMN; TERF1; ESPL1; DEF6; IFI16; DDX52*
intracellular non-membrane-bounded organelle	C=3800;O=175;E=139.05;R=1.26;rawP=0.0003;adjP=0.0127	*MYH4; RP1L1; SSH1; TRIM24; TRIM5; PTPN23; SUDS3; ATF5; FAM188A; DST; INPP5E; ZYX; GOLGA3; ABRA; COG3; SRRM2; KY; DNMT1; FLNC; CMYA5; CENPA; SYCP1; GPR110; KALRN; AKAP12; TNS1; PBRM1; SMC2; KRT10; MCM2; STK38L; KIF16B; C2CD3; MUC1; HLCS; WIPF3; HTATSF1; OGT; ACTN4; RNF168; KRTAP9-4; MTUS2; FAM60A; SLC25A26; LIMD1; LRRFIP1; LASP1; TTN; EHHADH; TFB2M; PCDH15; FAM71B; RPTN; LYAR; ASZ1; SYCP2; INTS7; RPL39; PPIL2; MSH2; INO80C; MCPH1; UHRF1; REPIN1; PRPH; GATAD2B; KRT20; DISC1; RHOQ; ECSIT; WRN; MBIP; CCDC86; EHMT2; SLU7; WAPAL; TNKS2; FGD4; TNNT2; HAUS7; AKAP9; KIF3C; NF2; URB2; CCDC110; CHD1L; SMARCA4; CWC22; HOOK2; MDC1; APC2; URB1; PCLO; TRIM66; NOP56; MRPL50; DECR1; SP100; TACC2; AURKC; RRP9; MSH4; CCDC15; KDM3B; DOCK4; RPL29; PPP1R18; RPL35; SLC8A3; ARHGEF10; ECT2; CHRM4; CROCC; VDAC1; MRPL14; PEG3; NIN; PES1; HELLS; HSPA4; BFSP1; DHX16; TRIOBP; SPTBN5; FAU; CBX7; CBX4; SETX; KIF3B; FARP1; MZT1; SPAG16; ESRP1; RIF1; CHRM1; HMGN3; IKBKAP; SAPCD2; MRPS22; MARK4; STARD9; HMGCS1; PPIG; TCHH; ARHGEF2; GCN1L1; DIS3; AMOT; FMR1; MICAL3; CDK5RAP2; PARP4; PKHD1; FXR2; SMARCC2; MYO7B; DIAPH1; SMTNL1; UTP20; MYO1A; MRPL4; FKBP15; NCAPD3; ZNF77; NFATC4; KIF21A; SYCE1; APAF1; KRTAP16-; ERMN; TERF1; ESPL1; DEF6; IFI16; DDX52*
nuclear chromosome	C=285;O=23;E=10.43;R=2.21;rawP=0.0003;adjP=0.0127	*TTN; CENPA; TRIM24; SYCP1; SUDS3; SMARCA4; PBRM1; SMC2; MCM2; SYCP2; MUC1; NCAPD3; WAPAL; TNKS2; MSH2; INO80C; MSH4; SYCE1; FAM60A; TERF1; CBX7; UHRF1; REPIN1*
apical part of cell	C=294;O=23;E=10.76;R=2.14;rawP=0.0005;adjP=0.0189	*PODXL; NAALADL1; ATP2B2; LRP2; AQP1; MGAM; ERBB2; PKHD1; KHDC3L; OOEP; ABCC2; ATP4A; MYO7B; INADL; MYO1A; SLC26A3; MUC1; DUOX1; TRPV5; SLC9A3; BST2; CD34; MUC13*
intrinsic to membrane	C=5437;O=236;E=198.95;R=1.19;rawP=0.0008;adjP=0.0255	*TCTN3; CCDC136; LRRTM4; NF1; OR52A5; WDR83OS; AQP1; VN1R1; PIGR; VNN3; BIRC2; TARM1; CD5; CCR6; DST; TLCD2; OR7C1; ZYX; NUP210; CEACAM20; SREBF2; OR1F12; TRPV5; PEX16; OR7G2; FUT11; KLRC2; CEACAM1; EQTN; GALNT5; PTPRM; GPR110; OR51B2;*

		MGAM; CR2; LRP5; ABCC2; OR7A5; SLC10A4; SEC14L2; FAM205A; OR51I2; OR4A47; MUC1; OR52J3; AGPAT2; MUSK; ATP7B; OR6K3; COL15A1; OR6C4; OR1S1; BST2; CHPT1; OR5M10; SIGLEC1; SLC25A26; MUC13; XXYLT1; PCSK5; EMR2; TMEM63C; PTPRD; ATP2A3; IMPG2; PCDH15; CD3G; C5; SLC6A17; ATP4A; OR52N4; OR6S1; FSHR; OR2G3; IL18R1; UGT1A3; C8B; OR8B3; IYD; DPEP3; OR4P4; ICAM5; OR5AC1; KCNH4; FCGRT; OR9A4; OR1D2; PIGA; KIRREL3; OR6C75; OR4C3; PODXL; PIGT; TM9SF1; SLC13A3; SIGMAR1; OR11L1; IL2RG; GYLTL1B; RASAL1; OR11H4; ANTXR1; HERPUD1; DAGLA; OR1L3; INSR; CDH17; OSMR; TG; FCAMR; CD34; CDH2; OR2B6; RNF139; OR1G1; OR13A1; CREB3L2; RTP4; CACNA1F; OR8G1; OR10A3; TRDN; CLEC17A; GALNT12; NAALADL1; ESYT3; ERBB2; TMEM104; PDE3A; OR1J4; DPCR1; ESAM; MUC16; TMEM129; SLC25A34; TAS2R8; KIAA0195; AMIGO3; GPR98; OR7D4; ECE1; GPR119; TMEM120B; HSPA2; KIAA1549; SEMA6C; TMBIM1; OR13C3; ACSL5; OMA1; SLC8A3; ATP10A; TMEM144; PTCH1; CHRM4; TOR1AIP2; CD14; MEGF11; MAG; ART3; VDAC1; HTR5A; NAT8B; SPPL3; RHBDL2; DCBLD2; OR8S1; TMEM184C; DUOX1; PKD2L2; OR14A16; WFS1; SLC9A6; CLCA2; OR6M1; OR5H2; DSG1; SLC47A2; OR1L8; ABCC12; LRFN4; NEU3; OR10C1; LRP2; CHRM1; PECAM1; ABCA13; MFSD10; CSF2RB; SLC35F1; OR51D1; OR11A1; SLITRK4; OR11H6; OR10T2; TEDDM1; PTGDR2; FCER1A; AMOT; AGER; TPCN1; FAM134C; SLC1A2; OR2T2; LINGO4; CD48; LILRB4; FOCAD; ERMAP; ABCC11; ATP2C1; MIA3; MRS2; ATP2B2; OR1L1; ABCA10; PKHD1; VSTM5; SPG7; CNTNAP1; PTCHD2; ETFDH; SLC26A3; CACNB1; NAALAD2; OR2T33; GPR124; IGSF11; OR12D2; SYT7; SLC25A40; SLC9A3; GPR133; VAPA; ERAP1; SLC25A19
paranode region of axon	C=6;O=3;E=0.22;R=13.66;rawP=0.0009; adjP=0.0255	*CNTNAP1; MAG; ERMIN*
integral to membrane	C=5321;O=231;E=194.70;R=1.19;rawP= 0.0009;adjP=0.0255	*TCTN3; CCDC136; LRRTM4; OR52A5; WDR83OS; AQP1; VN1R1; PIGR; BIRC2; TARM1; CD5; CCR6; DST; TLCD2; OR7C1; ZYX; NUP210; CEACAM20; SREBF2; OR1F12; TRPV5; PEX16; OR7G2; FUT11; KLRC2; CEACAM1; EQTN; GALNT5; PTPRM; GPR110; OR51B2; MGAM; CR2; LRP5; ABCC2; OR7A5; SLC10A4; SEC14L2; FAM205A; OR51I2; OR4A47; MUC1; OR52J3; AGPAT2; MUSK; ATP7B; OR6K3; COL15A1; OR6C4; OR1S1; BST2; CHPT1; OR5M10; SIGLEC1; SLC25A26; MUC13; XXYLT1; PCSK5; EMR2; TMEM63C; PTPRD; ATP2A3; IMPG2; PCDH15; CD3G; C5; SLC6A17; ATP4A; OR52N4; OR6S1; FSHR; OR2G3; IL18R1; UGT1A3; C8B; OR8B3; IYD; OR4P4; ICAM5; OR5AC1; KCNH4; FCGRT; OR9A4; OR1D2; PIGA; KIRREL3; OR6C75; OR4C3; PODXL; PIGT; TM9SF1; SLC13A3; SIGMAR1; OR11L1; IL2RG; GYLTL1B; OR11H4; ANTXR1; HERPUD1; DAGLA; OR1L3; INSR; CDH17; OSMR; TG; FCAMR; CD34; CDH2; OR2B6; RNF139; OR1G1; OR13A1; CREB3L2; RTP4; CACNA1F; OR8G1; OR10A3; TRDN; CLEC17A; GALNT12; NAALADL1; ESYT3; ERBB2; TMEM104; PDE3A; OR1J4; DPCR1; ESAM; MUC16; TMEM129; SLC25A34; TAS2R8; KIAA0195; AMIGO3; GPR98; OR7D4; ECE1; GPR119; TMEM120B; HSPA2; KIAA1549; SEMA6C; TMBIM1; OR13C3; ACSL5; OMA1; SLC8A3; ATP10A; TMEM144; PTCH1; CHRM4; TOR1AIP2; MEGF11; MAG; ART3; VDAC1; HTR5A; NAT8B; SPPL3; RHBDL2; DCBLD2; OR8S1; TMEM184C; DUOX1; PKD2L2; OR14A16; WFS1; SLC9A6; CLCA2; OR6M1; OR5H2; DSG1; SLC47A2; OR1L8; ABCC12; LRFN4; NEU3; OR10C1; LRP2; CHRM1; PECAM1; ABCA13; MFSD10; CSF2RB; SLC35F1; OR51D1; OR11A1; SLITRK4; OR11H6; OR10T2; TEDDM1; PTGDR2;*

		FCER1A; AMOT; AGER; TPCN1; FAM134C; SLC1A2; OR2T2; LINGO4; CD48; LILRB4; FOCAD; ERMAP; ABCC11; ATP2C1; MIA3; MRS2; ATP2B2; OR1L1; ABCA10; PKHD1; VSTM5; SPG7; CNTNAP1; PTCHD2; ETFDH; SLC26A3; CACNB1; NAALAD2; OR2T33; GPR124; IGSF11; OR12D2; SYT7; SLC25A40; SLC9A3; GPR133; VAPA; ERAP1; SLC25A19
synaptonemal complex	C=24;O=5;E=0.88;R=5.69;rawP=0.0015; adjP=0.0364	*SYCP2; SYCE1; MSH4; SYCP1; WAPAL*
nuclear chromosome	C=242;O=19;E=8.86;R=2.15;rawP=0.0015;adjP=0.0364	*CENPA; TRIM24; SYCP1; SUDS3; SMARCA4; MCM2; SYCP2; MUC1; NCAPD3; WAPAL; TNKS2; SYCE1; MSH4; INO80C; FAM60A; TERF1; UHRF1; CBX7; REPIN1*
condensed nuclear chromosome	C=62;O=8;E=2.27;R=3.53;rawP=0.0018; adjP=0.0408	*TTN; SYCP2; CENPA; NCAPD3; SYCP1; WAPAL; MSH4; SYCE1*

References

1. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159 (1987).
2. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).
3. Parra, G., Bradnam, K. & Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23,** 1061–1067 (2007).
4. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12,** 491 (2011).
5. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5,** 59 (2004).
6. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. WindowMasker: Window-based masker for sequenced genomes. *Bioinformatics* **22,** 134–141 (2006).
7. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-3.0. *RepeatMasker Open-3.0* www.repeatmasker.org (1996).
8. Jurka, J. Repbase Update: A database and an electronic journal of repetitive elements. *Trends in Genetics* **16,** 418–420 (2000).
9. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13,** 1028–1040 (2006).
10. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580 (1999).
11. Tamazian, G. *et al.* Annotated features of domestic cat - Felis catus genome. *Gigascience* **3,** 13 (2014).
12. Gardner, P. P. *et al.* Rfam: Updates to the RNA families database. *Nucleic Acids Res.* **37,** (2009).
13. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **25,** 1335–1337 (2009).
14. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964 (1997).
15. Walters-Conte, K., Johnson, D., Allard, M. & Pecon-Slattery, J. Carnivore-Specific SINEs (Can-SINEs): Distribution, Evolution, and Genomic Impact. *J. Hered.* **102,** S2–S10 (2011).
16. Vassetzky, N. S. & Kramerov, D. A. CAN - A pan-carnivore SINE family. *Mamm. Genome* **13,** 50–57 (2002).
17. Chan, P. P. & Lowe, T. M. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37,** (2009).
18. Frith, M. C., Hamada, M. & Horton, P. Parameters for accurate genome alignment. *BMC Bioinformatics* **11,** 80 (2010).
19. Li, H. *et al.* The Sequence Alignment / Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
20. Cho, Y. S. *et al.* The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat. Commun.* **4,** 2433 (2013).

21. Montague, M. J. *et al.* Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc. Natl. Acad. Sci.* **111,** 17230–17235 (2014).

22. Cho, Y. S. *et al.* The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat. Commun.* **4,** 2433 (2013).

23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows – Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

24. Li, G., Davis, B. W., Eizirik, E. & Murphy, W. J. Pervasive signals of ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* 1–11 (2016). doi:10.1101/gr.186668.114.4

25. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

26. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).

27. Johnson, W. E. *et al.* The late Miocene radiation of modern Felidae: a genetic assessment. *Science (80-.).* **311,** 73–77 (2006).

28. Hayden, S. *et al.* Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res.* **20,** 1–9 (2010).

29. Hughes, G. M., Gang, L., Murphy, W. J., Higgins, D. G. & Teeling, E. C. Using Illumina next generation sequencing technologies to sequence multigene families in de novo species. *Mol. Ecol. Resour.* **13,** 510–521 (2013).

30. Altschul, S. F. *et al.* Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

31. Hayden, S. *et al.* A cluster of olfactory receptor genes linked to frugivory in bats. *Mol. Biol. Evol.* **31,** 917–927 (2014).

32. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** (2011).

33. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest-HPC: Fast selection of best-fit models of protein evolution. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **6586 LNCS,** 177–184 (2011).

34. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475,** 493–496 (2011).

35. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1,** 18 (2012).

36. Green, R. *et al.* A draft sequence of the Neandertal genome. *Science (80-.).* **328,** 710 (2010).

37. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41,** W77–83 (2013).

38. Blake, J. A. *et al.* Gene ontology consortium: Going forward. *Nucleic Acids Res.* **43,** D1049–D1056 (2015).

39. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

40. Cerami, E. G. *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39,** 685–690 (2011).

41. Kanehisa, M. & Goto, S. Yeast Biochemical Pathways. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28,** 27–30 (2000).

42. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44,** D457–D462 (2016).

43. Köhler, S. *et al.* The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* **42,** (2014).

44. Casarini, A. P. M. *et al.* Acromegaly: correlation between expression of somatostatin receptor subtypes and response to octreotide-lar treatment. *Pituitary* **12,** 297–303 (2009).

45. Greco, D. S. Feline Acromegaly. *Top. Companion Anim. Med.* **27,** 31–35 (2012).

46. Bebee, T. W. *et al.* The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife* **4,** 1–27 (2015).

47. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell* **139,** 871–890 (2009).

48. Wilkie, A. O. Craniosynostosis: genes and mechanisms. *Hum Mol Genet* **6,** 1647–1656 (1997).

49. Mai, S. *et al.* The Missense Mutation W290R in Fgfr2 Causes Developmental Defects From Aberrant IIIb and IIIc Signaling. 1888–1900 (2010). doi:10.1002/dvdy.22314

50. Hosokawa, R., Deng, X., Takamori, K. & Xu, X. U. N. Epithelial-Specific Requirement of FGFR2 Signaling During Tooth and Palate Development. **350,** 343–350 (2009).

51. Yang, S. K. & Chen, C. Involvement of somatostatin receptor subtypes in membrane ion channel modification by somatostatin in pituitary somatotropes. *Clin. Exp. Pharmacol. Physiol.* **34,** 1221–1227 (2007).

52. Gastambide, F., Viollet, C., Lepousez, G., Epelbaum, J. & Guillou, J. L. Hippocampal SSTR4 somatostatin receptors control the selection of memory strategies. *Psychopharmacology (Berl).* **202,** 153–163 (2009).

53. Sunquist, F. & Sunquist, M. The wild cat book: everything you ever wanted to know about cats. *Univ. Chicago Press* 208 (2014).

54. Gonyea, W. J. Adaptative differences in the body proportions of large felids. *Acta Anat. (Basel).* **96,** 81–96 (1976).

55. Seymour, K. L. Panthera onca. *Mamm. Species* **11,** 1–9 (1989).

56. Emmons, L. Jaguar predation on chelonians. *J. Herpetol.* **23,** 311–314 (1989).

57. Weckel, M., Giuliano, W. & Silver, S. Jaguar (Panthera onca) feeding ecology : distribution of predator and prey through time and space. *J. Zool.* **270,** 25–30 (2006).

58. Segura, V., Prevosti, F. & Cassini, G. Cranial ontogeny in the puma lineage, puma concolor, herpailurus yagouaroundi, and acinonyx jubatus (carnivora: Felidae): A three-dimensional geometric morphometric approach. *Zool. J. Linn. Soc.* **169,** 235–250 (2013).

59. Wroe, S. & Milne, N. Convergence and remarkably consistent constraint in the evolution of carnivore skull shape. *Evolution (N. Y).* **61,** 1251–1260 (2007).

60. Christiansen, P. Evolution of skull and mandible shape in cats (Carnivora: Felidae). *PLoS One* **3,** (2008).

61. Li, T. B. W. Æ. M. J., Lei, Æ. Y. J. J. Æ. C. Z. & Wang, J. Q. A coding SNP of LHX4 gene is associated with body weight and body length in bovine. 417–422 (2010). doi:10.1007/s11033-009-9486-6

62. Mayr, J. a. *et al.* Lipoic acid synthetase deficiency causes neonatal-onset epilepsy, defective mitochondrial energy metabolism, and glycine elevation. *Am. J. Hum. Genet.* **89,** 792–797 (2011).

63. Spong, G., Stone, J., Creel, S. & Björklund, M. Genetic structure of lions (*Panthera leo* L.) in the Selous game reserve: Implications for the evolution of sociality. *J. Evol. Biol.* **15,** 945–953 (2002).

64. Mosser, A. & Packer, C. Group territoriality and the benefits of sociality in the African lion, Panthera leo. *Anim. Behav.* **78,** 359–370 (2009).

65. Schaller, G. B. The Serengeti Lion: A study of predator-prey relations. *Wildl. Behav. Ecol. Ser.* 504 (1972).

66. Cowley, M. *et al.* Developmental Programming Mediated by Complementary Roles of Imprinted Grb10 in Mother and Pup. *PLoS Biol.* **12,** (2014).

67. Garfield, A. S. *et al.* Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature* **469,** 534–538 (2011).

68. Lavdas, a a, Grigoriou, M., Pachnis, V. & Parnavelas, J. G. The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex. *J. Neurosci.* **19,** 7881–7888 (1999).

69. Alifragis, P. Lhx6 Regulates the Migration of Cortical Interneurons from the Ventral Telencephalon But Does Not Specify their GABA Phenotype. *J. Neurosci.* **24,** 5643–5648 (2004).

70. Denaxa, M., Sharpe, P. T. & Pachnis, V. The LIM homeodomain transcription factors Lhx6 and Lhx7 are key regulators of mammalian dentition. *Dev. Biol.* **333,** 324–336 (2009).

71. Zhang, Y. *et al.* Comparison of the expression patterns of two LIM-homeodomain genes, Lhx6 and L3/Lhx8, in the developing palate. *Orthod. Craniofac. Res.* **5,** 65–70 (2002).

72. Choi, G. B. *et al.* Lhx6 delineates a pathway mediating innate reproductive behaviors from the amygdala to the hypothalamus. *Neuron* **46,** 647–660 (2005).

73. Weiss, L. A. *et al.* Identification of EFHC2 as a quantitative trait locus for fear recognition in Turner syndrome. **16,** (2007).

74. Kim, H.-I. *et al.* Association of CLOCK, ARNTL , and NPAS2 gene polymorphisms and seasonal variations in mood and behavior. *Chronobiol. Int.* **0528,** 1–7 (2015).

75. Han, G., Sun, J., Wang, J., Song, F. & Lei, H. Genomics in Neurological Disorders. *Genomics. Proteomics Bioinformatics* **12,** 156–163 (2014).

76. Ollila, H. M. *et al.* Genome-wide association study of sleep duration in the Finnish population. *J. Sleep Res.* **6,** 609–618 (2014).

77. Maxwell, M. *et al.* Pex13 inactivation in the mouse disrupts peroxisome biogenesis and leads to a Zellweger syndrome phenotype. *Mol. Cell. Biol.* **23,** 5947–57 (2003).

78. Sakamoto, M. & Ruta, M. Convergence and divergence in the evolution of cat skulls: Temporal and spatial patterns of morphological diversity. *PLoS One* **7,** e39752 (2012).

79. Stein, A. B. & Hayssen, V. Panthera pardus (Carnivora: Felidae). *Mamm. Species* **45,** 30–48 (2013).

80. Zhao, Y. *et al.* LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. *Proc Natl Acad Sci U S A* **104,** 13182–13186 (2007).

81. Agulnick, A. D. *et al.* Interactions of the LIM-domain-binding factor Ldbl with LIM homeodomain proteins. *Nature* **384,** 270–272 (1996).

82. Mukhopadhyay, M. *et al.* Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* **130,** 495–505 (2003).

83. Almaidhan, A. *et al.* Neural crest-specific deletion of Ldb1 leads to cleft secondary palate with impaired palatal shelf elevation. *BMC Dev. Biol.* **14,** 3 (2014).

84. Park, J.-A. & Kim, K. C. Expression patterns of PRDM10 during mouse embryonic development. *BMB Rep.* **43,** 29–33 (2010).

85. Tzchori, I. *et al.* LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development* **136,** 1375–1385 (2009).

86. Vitureira, N. *et al.* Podocalyxin is a novel polysialylated neural adhesion protein with multiple roles in neural development and synapse formation. *PLoS One* **5,** (2010).

87. Knoll, J. H. M. *et al.* Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am. J. Med. Genet.* **32,** 285–290 (1989).

88. Shishkov, R., Chervenkov, T., Yamashima, T. & Tonchev, A. B. Expression of Cyclon/CCDC86, a novel nuclear protein, in the hippocampus of adult non-human primates. *J. Neuroimmunol.* **258,** 96–99 (2013).

89. Yamada, K. *et al.* Genome-wide association study of schizophrenia in Japanese population. *PLoS One* **6,** e20468 (2011).

90. Sokolowski, M., Wasserman, J. & Wasserman, D. Genome-wide association studies of suicidal behaviors: A review. *Eur. Neuropsychopharmacol.* **24,** 1567–1577 (2014).

91. Kovanen, L., Saarikoski, S. T., Aromaa, A., Lönnqvist, J. & Partonen, T. ARNTL (BMAL1) and NPAS2 gene variants contribute to fertility and seasonality. *PLoS One* **5,** e10007 (2010).

92. Hemmer, H. Uncia uncia. *Mamm. Species* 1 (1972). doi:10.2307/3503882

93. Fox, J. L. A review of the status and ecology of the snow leopard. *International Snow Leopard Trust* 1–40 (1989).

94. Doherty, D., Millen, K. J. & Barkovich, A. J. Midbrain and hindbrain malformations: Advances in clinical diagnosis, imaging, and genetics. *Lancet Neurol.* **12,** 381–393 (2013).

95. İncecik, F., Hergüner, M. Ö., Altunbaşak, Ş. & Gleeson, J. G. Joubert syndrome: report of 11 cases. *Turk J Pediatr* 605–611 (2015).

96. Gallant, N. M., Baldwin, E., Salamon, N., Dipple, K. M. & Quintero-Rivera, F. Pontocerebellar hypoplasia in association with de novo 19p13.11p13.12 microdeletion. *Am. J. Med. Genet. Part A* **155,** 2871–2878 (2011).

97. Ban, D. M., Zhang, B., Wang, Z. X., Zhang, H. & Wu, C. X. Differential gene expression of epigenetic modifying enzymes between Tibet pig and Yorkshire in high and low altitudes. *Genet. Mol. Res.* **14,** 3274–3280 (2015).

98. Twigg, S. R. F. *et al.* Frontorhiny, a distinctive presentation of frontonasal dysplasia caused by recessive mutations in the ALX3 homeobox gene. *Am. J. Hum. Genet.* **84,** 698–705 (2009).

99. Sunquist, M. & Sunquist, F. *Wild cats of the world*. *Wild Cats of the World* **85,** (University of Chicago Press, 2002).

100. Montgomery, R. K., Mulberg, A. E. & Grand, R. J. Development of the human gastrointestinal tract: Twenty years of progress. *Gastroenterology* **116,** 702–731 (1999).

101. De Caro, J. *et al.* Occurrence of pancreatic lipase-related protein-2 in various species and its relationship with herbivore diet. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **150,** 1–9 (2008).

102. Mazak. Panthera tigris. *Mammalian Species* **152,** 1–8 (1981).

103. Eizirik, E. *et al.* Defining and Mapping Mammalian Coat Pattern Genes: Multiple Genomic Regions Implicated in Domestic Cat Stripes and Spots. *Genetics* **184,** 267–U422 (2010).

104. Niemann, C. & Schneider, M. R. Hair type specific function of canonical Wnt activity in adult mouse skin. *Exp. Dermatol.* 881–883 (2014). doi:10.1111/exd.12509

105. Dunn, K. J., Williams, B. O., Li, Y. & Pavan, W. J. Neural crest-directed gene transfer demonstrates Wnt1 role in melanocyte expansion and differentiation during mouse development. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10050–10055 (2000).

106. Schmidt-Ullrich, R. & Paus, R. Molecular principles of hair follicle induction and morphogenesis. *BioEssays* **27,** 247–261 (2005).

107. Ye, J. *et al.* Wnt10b promotes differentiation of mouse hair follicle melanocytes. *Int. J. Med. Sci.* **10,** 691–698 (2013).

108. Gao, C. & Chen, Y. G. Dishevelled: The hub of Wnt signaling. *Cell. Signal.* **22,** 717–727 (2010).

109. Saito, H. *et al.* Microphthalmia-associated transcription factor in the wnt signaling pathway. *Pigment Cell Res* **16,** 261–265 (2003).

110. Sick, S., Reinker, S., Timmer, J. & Schlake, T. Reaction-Diffusion Mechanism. **314,** 1447–1450 (2006).

111. Kondo, S., Miura, T. & Turing, T. Reaction-Diffusion Model as a. **329,** 1616–1620 (2010).

112. Dijksterhuis, J. P. *et al.* Systematic mapping of WNT-FZD protein interactions reveals functional selectivity by distinct WNT-FZD pairs. *J. Biol. Chem.* **290,** 6789–6798 (2015).

113. Ouji, Y. *et al.* Wnt-10b, uniquely among Wnts, promotes epithelial differentiation and shaft growth. *Biochem. Biophys. Res. Commun.* **367,** 299–304 (2008).

114. Ouji, Y., Yoshikawa, M., Shiroi, A. & Ishizaka, S. Promotion of hair follicle development and trichogenesis by Wnt-10b in cultured embryonic skin and in reconstituted skin. *Biochem. Biophys. Res. Commun.* **345,** 581–587 (2006).

115. Huh, S., Na, K. & Mikkola, M. L. Fgf20 governs formation of primary and secondary dermal condensations in developing hair follicles. 450–458 (2013). doi:10.1101/gad.198945.112.on

116. Filmus, J., Capurro, M. & Rast, J. Glypicans. *Genome Biol.* **9,** 224 (2008).

117. Filmus, J. Glypicans in growth control and cancer. *Glycobiology* **11,** 19R–23R (2001).

CAPÍTULO III

EXOME SEQUENCING REVEALS SIGNATURES OF LOCAL ADAPTATION IN NATURAL JAGUAR POPULATIONS

Molecular Ecology

Manuscript, February 29, 2016.

Exome sequencing reveals signatures of local adaptation in natural jaguar populations

Henrique Vieira Figueiró[1], Cristine S. Trinca[1], Maíra R. Rodrigues[1], Tyler Linderoth[2], Lucas Gonçalves da Silva[1], Fernanda J. Trindade[1], Ke Bi[2], Leandro Silveira[3], Ronaldo G. Morato[7], Dênis A. Sana[9], Laury Cullen[4], Peter Crawshaw[7], Emiliano E. Ramalho[5], Fernando Tortato[6], Daniel Kantek[7], William Murphy[8], Rasmus Nielsen[2], Eduardo Eizirik[1]

[1] Faculdade de Biociências, PUCRS. Porto Alegre, RS 90619-900, Brazil

[2] Institute of Biology, University of California Berkeley. Berkeley, California, United States

[3] Jaguar Conservation Fund/Instituto Onça-Pintada. Mineiros, GO 75830-00, Brazil

[4] Instituto de Pesquisas Ecológicas – IPÊ. Nazaré Paulista, SP 12960-00, Brazil

[5] Instituto de Desenvolvimento Sustentável Mamirauá. Tefé, AM 69553-225, Brazil

[6] Panthera Foundation. New York, NY, United States

[7] Instituto Chico Mendes de Conservação da Biodiversidade. Brasília, DF 70670-350, Brazil

[8] Texas A&M University. College Station, TX, United States

[9] Secretaria Estadual do Meio Ambiente do Rio Grande do Sul – SEMA. Porto Alegre, RS 90030-020, Brazil

Abstract

Local adaptation is one of the most important drivers of genetic divergence among populations and speciation. From the genetic standpoint, it can be defined as a change of regional allele frequencies mediated by natural selection, which increases the fitness of individuals in a given environment. The jaguar is the largest Neotropical felid, occurring in several biomes throughout South and Central America and southern North America, and exhibiting marked body size variation among regions. Using exome capture and high-throughput sequencing, the present study aimed to assess genetic diversity and signatures of local adaption throughout most of the Brazilian portion of its distribution. We sampled five widespread geographic populations,

representing the Atlantic Forest, Pantanal, Caatinga, Cerrado and Amazonia biomes, and demonstrated that they exhibit low but detectable levels of genome-wide differentiation. In addition, we identified signature of selection in several loci related to body size and color pattern. Most of the genes related to body size variation were associated with energetic metabolism, which corroborates the hypothesis that body size variation in the jaguar is driven by dietary differences among biomes. Moreover, we identified candidate loci that could help explain some color variation, especially on the Pantanal. This is the first assessment of genetic diversity, population structure and adaptive evolution in any felid using exome capture and population genomic approaches.

Introduction

Local adaptation can be defined as a change of allele frequencies in a given population that increases its fitness in its particular environment (Kawecki & Ebert 2004). The study of local adaptation on a genomic scale is still in its infancy, with no more than a handful of biological systems possessing enough resources to allow the assessment of this process in natural populations (for a review, see Stapley et al., 2010). Until recently, this type of study has mostly depended on the possibility of manipulating organisms and doing common garden experiments (Stapley *et al.* 2010; Savolainen *et al.* 2013). Although these approaches are indeed very useful, they are not feasible for most species. In that sense, with new methods arising from the flood of new genomic resources, novel methods to study local adaptation began to appear. The main advantage of these methods is that their requirements are much easier to achieve than experimental approaches. They rely on the genetic variation present across the genome of individuals sampled in different populations to identify genomic regions under selection (Savolainen *et al.* 2013).

Jaguars (*Panthera onca*) present remarkable morphological and ecological characteristics, deriving from a unique evolutionary history among the great cats. It is the only big cat still inhabiting the American continent. Is a top predator with one of the broadest ranges of prey items among the big cats (Hayward *et al.* 2016). It possesses two very distinct coat color phenotypes, the wild type with a yellow to orange background and black rosettes, and a melanistic type (dark brown or black background coloration). Several early authors described the jaguar as comprising several subspecies (Seymour 1989). Larson (1997) reviewed its taxonomy and concluded that none of the subspecies was valid on the basis of morphological characters. This conclusion was supported by a phylogeographic study based on molecular data (Eizirik *et al.* 2001), which

reported high historical connectivity across the species' distribution. In spite of this low level of genetic structure, some geographic patterns of phenotypic differentiation have been observed. For example, jaguars present a wide variation in their body size, with adult male individuals ranging from ~70kg to ~150kg. Some authors argue that the difference in size is caused by differences in diet on each region, more precisely biomass of the prey, especially when you compare floodplain and forests areas (Kiltie 1984; Hoogesteijn & Mondolfi 1996; Hayward *et al.* 2016). However, so far no studies have addressed this question systematically. Recent results indicate that the stocky build of the species, one of its most striking characteristics and one of the few differences between the jaguar and the leopard, is possibly caused by positive selection acting on a few important developmental genes (Figueiró *et al.*, *in prep*).

Using microsatellites, some studies had previously focused on the genetic structure of jaguars. Eizirik *et al* (2001) has showed an absence of historical structure in populations located south of the Amazon River. That is probably caused by the long-distance dispersal ability of the species. Other studies focused on specific populations throughout the jaguar distribution. They reported different results, with some highly isolated populations, such as those of the Atlantic Forest, and the Caatinga (Haag *et al.* 2010; Roques *et al.* 2014), and others with low differentiation, such as in the southern Pantanal (Valdez *et al.* 2015). However, some populations still lack any genetic information to assess their current situation, such as those located in the Cerrado.

Thus, the goals of this study were to: (i) characterize the jaguar exome, allowing genome-wide assessments of its genetic diversity; and (ii) use large-scale exome sequence data to investigate the jaguar's demographic history, population structure and signatures of local adaptation.

Methods

Sampling

We selected 110 individuals sampled in five different biomes for use in this study (Figure 1). All samples came from blood or tissue material collected during ecology fieldwork and deposited at the LBGM and CENAP collections (see Supplementary Table 1). Blood samples were kept in an EDTA solution and tissue samples were fixed and stored in ethanol 96%. Whole genomic DNA was extracted with the DNeasy Tissue and Blood Kit (Qiagen).

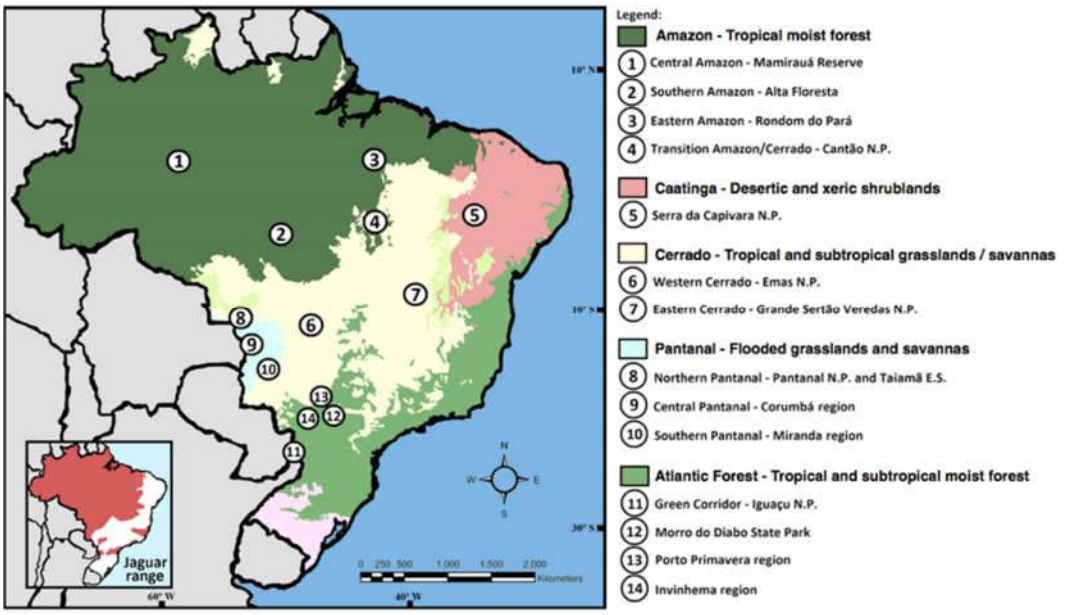


Figure 1: Sampling geographic distribution – Jaguar populations sampled for the present study. Number of individuals: 16 from the Amazon; 19 from the Atlantic Forest; two from the Caatinga; four from the Cerrado; 67 from the Pantanal.

Exome Probe design

For probe design, we parsed the data available for the five *Panthera* genomes (see Supplementary Table 2), aiming to produce an effective array for exome capture across the whole genus. The *Panthera* lineage has a recent common origin, making it likely that the same probe set would work for the whole group (Johnson *et al.* 2006; Bi *et al.* 2012). For the annotation and selection of regions, we followed a previously published analytical pipeline (Bi *et al.* 2012). Based on the cat genome, we annotated the coding sequences for each species and, based on coverage and completeness, we chose which sequence to use for the probes. When sequences of equal quality were available or more than one species, we prioritized CDS sequences in the following order: Jaguar > Tiger > Leopard = Lion = Snow Leopard > Domestic cat. This priority was chosen based on the goals of the study and completion of the available data. We also included in the probes 500bp of UTRs on both flanks the coding region of each gene.

Library construction and sequencing

Genomic libraries for 110 jaguars and eight libraries from five other closely related species (*Panthera tigris*, *P. leo*, *P. pardus*, *P. uncia* and *Neofelis nebulosa*) were constructed according to the protocol outlined by Meyer & Kirchner (2010) with some modifications. Exome capture was performed using a custom-designed Nimblegen Capture Kit (Roche) and sequencing was

performed in two lanes of the Illumina HiSeq 2500 platform, with 100bp reads and *ca.* 300bp insert size.

Read filtering and mapping

Sequence quality filtering was performed using a previously published analytical pipeline (Bi *et al.* 2012). The parameters involved sequence trimming and removal based on the quality score (phred score > 30) with Trimmomatic (Bolger *et al.* 2014), removal of sequencing adapters with CutAdapt (Martin 2011), and preliminary assembly of overlapping paired reads using FLASH (Magoc & Salzberg 2011). This preliminary assembly facilitated further steps, where we mapped the whole exome against our reference.

Read mapping was performed using NovoAlign with the exome probes as reference. This enabled us to have a complete mitochondrial genome for all samples, and to evaluate potential mapping problems as well. Coverage was checked using ANGSD *(-doDepth* parameter) (Korneliussen *et al.* 2014) and in-house python scripts.

SNP calling and estimation of sample allele frequencies

After assessing coverage, we trimmed the data removing individuals and sites with discrepant coverage using SNPcleaner (https://github.com/tplinderoth/SNPcleaner). Based on depth average per individual and per site, we decided to keep up to 70% of individuals with site depth between 2x and 30x (see Supplementary Figure 1). The filtered information was used for all downstream analyses. In order to generate summary statistics and to have a preliminary assessment of data trends, we generated the site frequency spectrum (Figure 2). The site frequency spectrum is an efficient metric that enables a general overview of signatures of demographic events, selection and paralogy issues (Nielsen *et al.* 2012).

Population genetic inferences

For all analyses, we removed sites exhibiting linkage disequilibrium using PLINK (Purcell *et al.* 2007), as well as related individuals using the relatedness algorithm implemented in VCFtools (Yang *et al.* 2010; Danecek *et al.* 2011). We then calculated the following summary statistics: nucleotide diversity (π), Tajima's D (Tajima 1989), F_{ST} (Weir & Cockerham 1984) and heterozygosity, for all populations, using VCFtools. We used these statistics both to asses population structure and to detect outlier loci that could be under selection (Akey *et al.* 2002).

To investigate patterns of population structure, we performed a Principal Component Analysis (PCA) using Eigensoft SmartPCA (Price *et al.* 2006). PCA plots (Patterson *et al.* 2006; Zheng *et al.* 2012) were generated for principal components with p-values below 0.05. In order to look for ancestry relationships, we performed a maximum likelihood estimation of individual ancestries from SNP genotypes, as implemented in the software ADMIXTURE (Alexander *et al.* 2009). The latter software uses the same statistical model as STRUCTURE (Pritchard *et al.* 2000) but has a better performance with large datasets. The algorithm was executed in the unsupervised mode and using 5-fold cross-validation to assess model error. Admixture plots were generated using R statistical software.

Selection analysis

Only sites with extreme outlier values of F_{ST} were considered. For that purpose, we also calculated the average site depth per locus in order to look for paralogous sites. Sites with discrepant coverage (2 standard deviations) were removed from further analysis. Since F_{ST} is affected by sample sizes, we selected only three out of five populations to perform this analysis: Pantanal, split between North Pantanal (35 individuals) and South Pantanal (32 individuals), Amazon (15 individuals) and Atlantic Forest (19 individuals). We have tested all populations against each other and metapopulations considering the type of habitat (forests against open areas). Another method we applied was to look for sites with higher explanatory power in the PCA analysis, especially in cases where the variation was not explained by population structure, *i.e.* clusters that did not differentiate populations by distance, for example. We also used Bayescan, which uses the posterior distribution to look for outlier sites (Foll & Gaggiotti 2008). We used the following parameters for this analysis: Sample size = 5000/50000, Thinning interval = 10, Pilot runs = 20, Pilot run length = 5000, Additional burn in = 50000/500000.

For all genes under selection for a given population, we performed enrichment tests with the following databases: Gene Ontology (GO), KEGG Pathway, Pathway Commons and Disease, through the web application WebGestalt (Wang *et al.* 2013) using the human annotation as reference. We also developed in-house scripts to perform enrichment analyses, with similar databases as above (GO and GWAS studies), but using our own dataset (probes) as the background, instead of the complete list of human genes as in the WebGestalt application. For that, we applied the Fisher exact test with a correction for multiple testing. For each analysis, only enriched terms with a false discovery rate (FDR) ≤ 0.05 were kept.

Results

Probe design and SNP calling

We were able to annotate 19,000 genes spanning a total 36 Mb in size for our probe experiment. This outcome enabled a highly comprehensive assessment of the genetic diversity for the jaguar and the other species belonging to the *Panthera* lineage.

We observed an average coverage of 8x for all individuals. The coverage was evenly split among all populations and no individuals needed to be discarded. Since the observed coverage for our samples was not adequate for reliable genotype calling based on depth alone, we used a different approach. This method relies on SNP calling and allele frequency estimation based on genotype likelihoods. We used information for all individuals from all sites simultaneously (Pool *et al.* 2010; Nielsen *et al.* 2011). With this method, we have increased the statistical power to detect SNPs and we were able to detect 160,000 SNPs with high confidence. For the downstream analyses, we filtered our dataset in order to remove sites under linkage disequilibrium. Such filtering enabled us to attain a final dataset of 67,397 SNPs. To evaluate the allelic distribution at this sites, we generated the site frequency spectrum for the data (Figure 2) in order to look for deviations from the expected distribution (Nielsen *et al.* 2011).

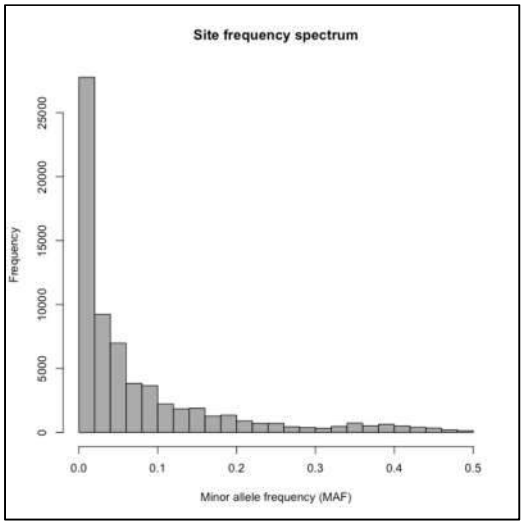


Figure 2: Site Frequency Spectrum – Minor allele frequency for all samples in the present study.

Population structure

The estimated genetic diversity varied across the analyzed populations (Table 1): it was generally moderate to high, and considerably lower in the Caatinga population. A related trend was observed for the differentiation between populations (Table 2), with pairwise divergences (F_{ST}) involving the Caatinga showing the highest values across all comparisons.

Table 1: Genetic diversity – Genetic diversity and observed and expected heterozygosity on each analyzed population. All sites are under linkage equilibrium.

Population	Nucleotide Diversity (π)	Observed Het (HO)	Expected Het (HE)
Amazon (16)	0.090	0.803	0.807
Atlantic Forest (19)	0.084	0.783	0.798
Caatinga (2)	0.074	0.360	0.448
Cerrado (4)	0.085	0.639	0.642
South Pantanal (32)	0.086	0.812	0.821
North Pantanal (35)	0.083	0.821	0.828

Table 2: Pairwise F_{ST} – Weighted pairwise F_{ST} comparing each population for all sites under linkage equilibrium.

Population	Atlantic Forest	South Pantanal	Amazon	North Pantanal	Cerrado	Caatinga
Atlantic Forest	-	0.052	0.043	0.043	0.042	0.104
South Pantanal	-	-	0.048	0.025	0.052	0.116
Amazon	-	-	-	0.034	0.025	0.077
North Pantanal	-	-	-	-	0.040	0.098
Cerrado	-	-	-	-	-	0.087
Caatinga	-	-	-	-	-	-

Ancestry analysis using ADMIXTURE indicated K=2 as the best explanatory model for the studied populations (Figure 3). This indicated that the populations from the Cerrado, Caatinga and Atlantic Forest were highly connected with the Amazon, while the Pantanal was considered a genetically distinct, unsubdivided population. However, when results assuming higher K values were inspected (see Supplementary Figure 2-5), we observed a northern/southern structure in the Pantanal, which fits the results from other analyses (see below).

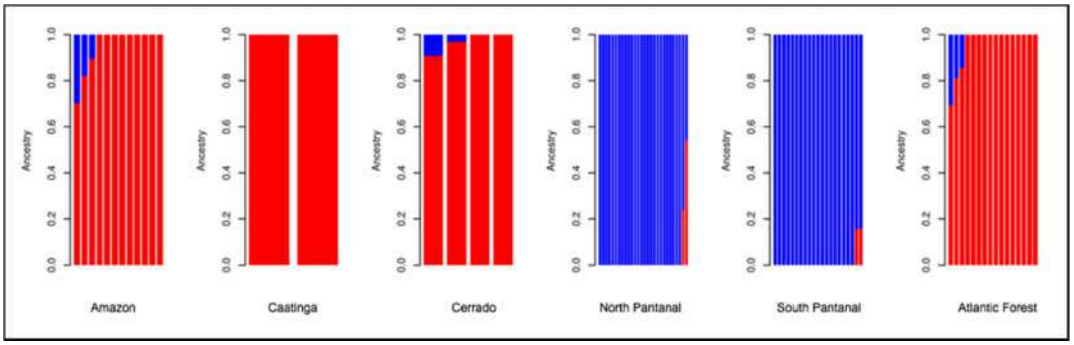


Figure 3: Admixture plot – Ancestry proportions of the 82 individuals from six extant South American jaguar populations with K = 2.

The PCA plots were able to differentiate each population, and to highlight some striking patterns of diversity among the different biomes (Figure 4 and Supplementary Figure 6). PC1 highlighted the difference between the Pantanal and the other populations, the same pattern observed in the Admixture plot. In addition, PC2 and PC3 show a split between Northern and Southern populations and the split between South and North Pantanal respectively.

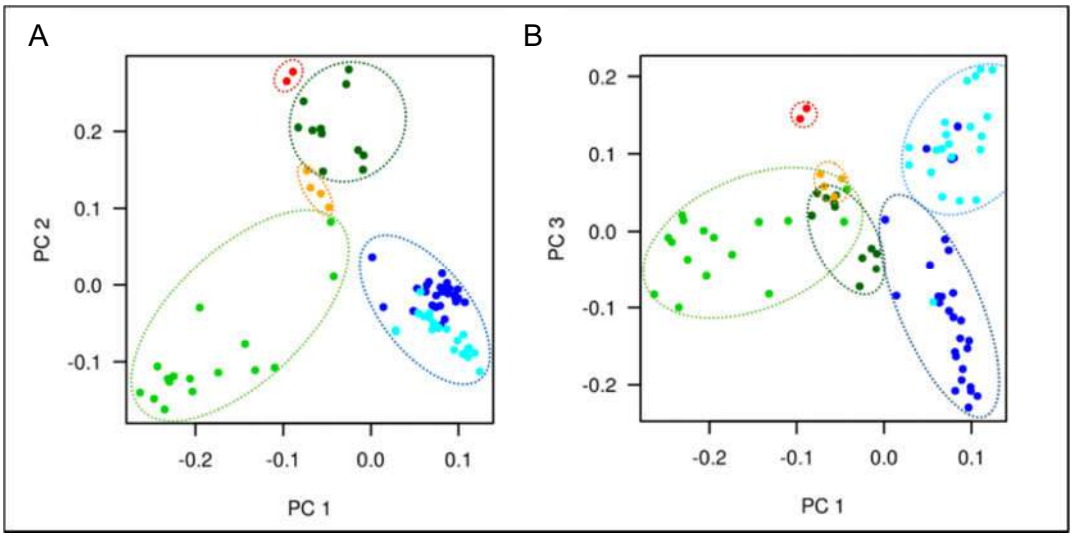


Figure 4: PCA Plots – Principal Component Analysis with 60,183 SNPS for 82 individuals. Populations are identified by different colors. North Pantanal – dark blue; South Pantanal – light blue; Cerrado – orange; Caatinga – red; Atlantic forest – light green; Amazon – dark green. **(A)** PC1 against PC2. **(B)** PC1 against PC3.

Selection

We found 330 candidate genes potentially under positive selection based on our F_{ST} outlier approach (Figure 5). To investigate region-specific cases of local adaptation, we kept only genes found to be under selection in one population and not the others used in our comparison. Using the same principle for the PCA outliers, we found 516 outliers for PC1, 533 outliers for PC2 and 527 outliers for PC3.

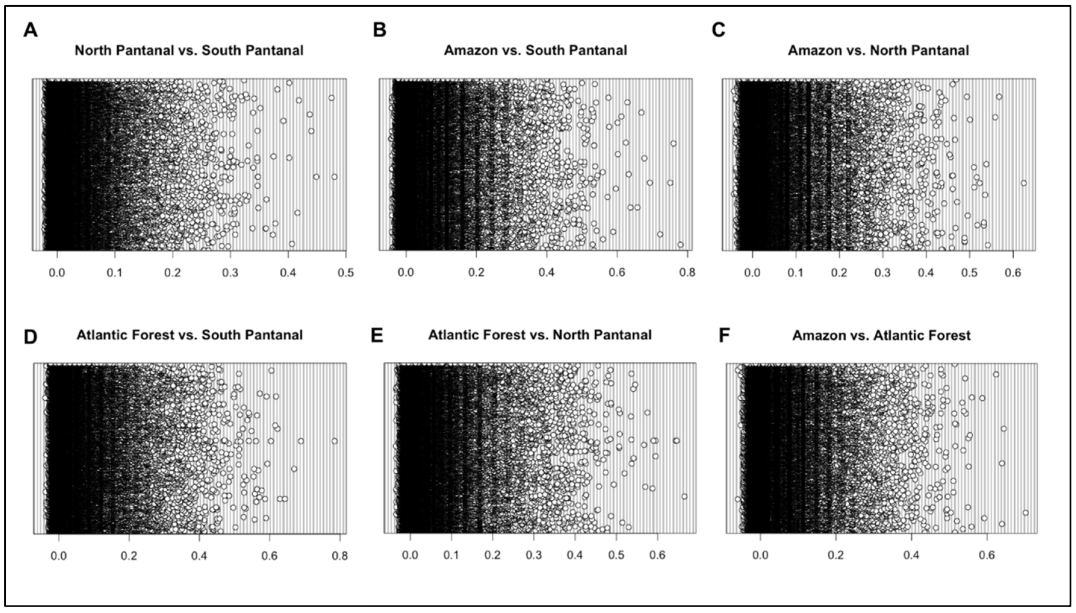


Figure 5: Outlier F_{ST} analysis – Empirical distribution for 67,387 loci sampled among four jaguar populations. Only the 1% highest F_{ST} values were considered as outliers in each comparison. **(A)** Comparison between individuals located in the North and South Pantanal populations. **(B)** Comparison between the Amazon and South Pantanal populations. **(C)** Comparison between Amazon and North Pantanal populations. **(D)** Comparison between Atlantic Forest and South Pantanal populations. **(E)** Comparisons between Atlantic Forest and North Pantanal populations. **(F)** Comparison between Amazon and Atlantic Forest populations.

Using a gene enrichment approach (Supplementary Table 4-6), we were able to find some interesting relationships among the genes we found for each population without overlap with the others. Gene enrichment analysis retrieved concordant patterns among different analysis, especially between the F_{ST} outlier analysis and the PCA outlier analysis, strengthening our confidence on their results. For the Amazon, we obtained 63 enriched terms, which were related with several different phenotypes. Among them is worth mentioning glycogen and fat metabolism, visual perception and reproduction. The Atlantic Forest population had 104 enriched terms, most of them associated to immunity and head development. Since we obtained a better sampling throughout the Pantanal populations, we analyzed the population as whole and also divided it into its southern and northern parts. The Pantanal population presented enrichment for brain function-related terms and several developmental routes, such as the skeletal system, skin, and central nervous system.

The Bayesian outlier analysis performed with Bayescan (Figure 6) retrieved four genes for the comparison between Pantanal and Forest biomes, and 47 genes for the comparison among all populations. The comparison A retrieved genes related to reproduction, lipid metabolism and immunity (Figure 6A). Comparison B retrieved genes related body growth and energetic metabolism (Figure 6B).

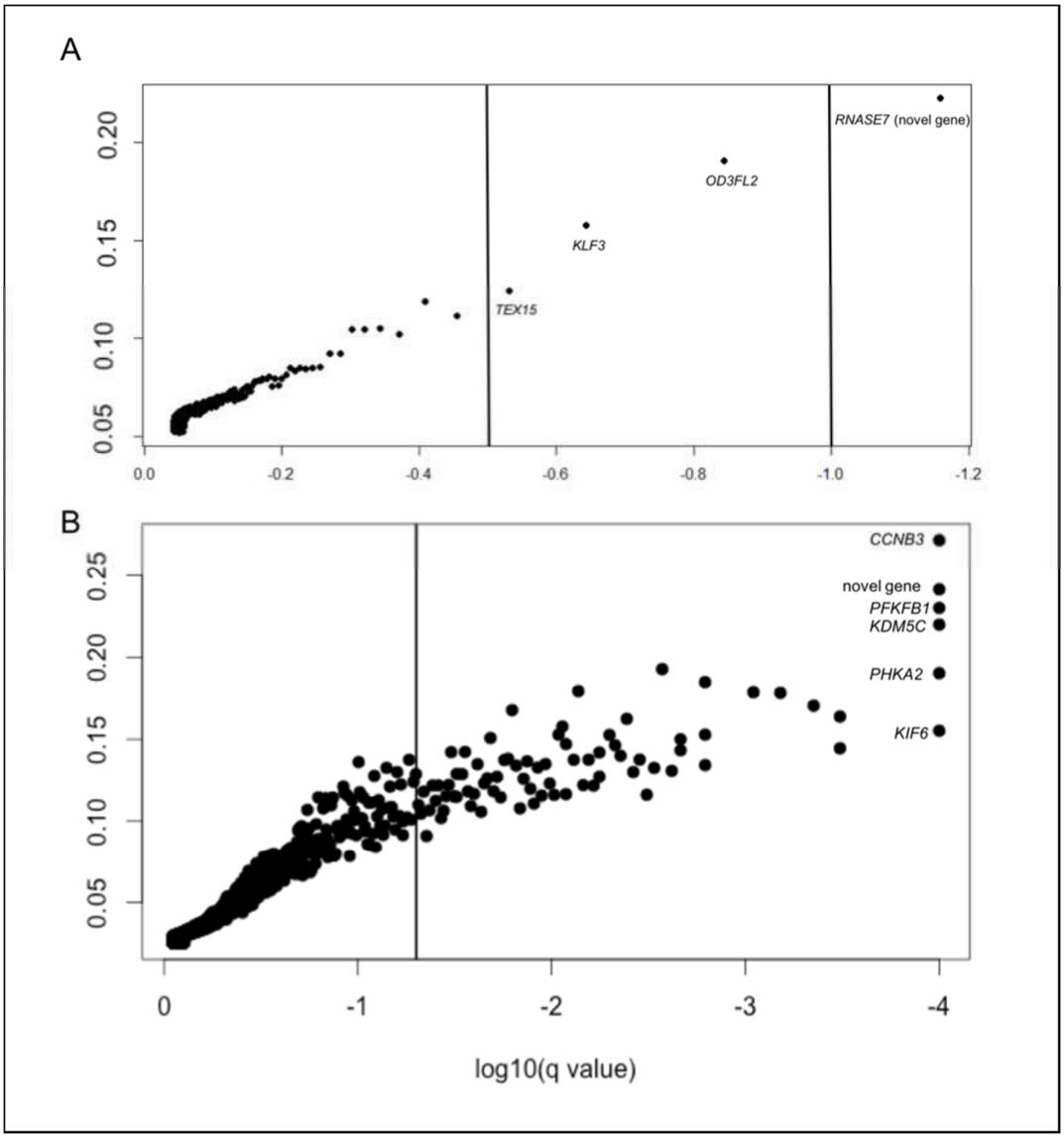


Figure 6: Bayescan FST analysis – Bayescan plot for 67,387 loci sampled in *Panthera onca* populations, testing: **(A)** outliers observed in a comparison between populations located in open areas (Pantanal) *vs.* populations located in forests (Amazon and Atlantic Forest); and **(B)** outliers observed in a global comparison including all the sampled populations. Lines indicate regions with extreme outliers: values > 0.5 indicate substantial evidence of selection, values > 1.0 indicate strong evidence of selection, and values > 1.5 indicate gene with enough evidence to confidently infer selection (Foll & Gaggiotti 2008).

Discussion

This is the first study to assess the genetic diversity and evolution of a Neotropical carnivore using high throughput sequencing and population genomic analyses. Our population structure results, derived from the most complete sampling of Brazil jaguar achieved so far, revealed some interesting patterns. The observed pairwise F_{ST} among populations showed little

differentiation among them. The most isolated population was the Caatinga, with differentiation values similar to those observed by Roques et al (2014) with microsatellite data. A similar pattern was observed for the PCA analysis (PC1 vs. PC2) (Figure 4A), but the small sample size makes it difficult to interpret the reason of this differentiation. Based on the results from the other analyzed populations, selection cannot be discarded as an explanation. However, demographic factors such as accelerated drift due to human-induced fragmentation and population decline (see Haag et al. 2010 for an example) is a more plausible hypothesis. According to recent surveys, there are fewer than 250 sexually mature individuals remaining in fragmented areas within this biome, with most of the suitable habitat being restricted to protected areas (Paula *et al.* 2013). Additional sampling of Caatinga jaguars is thus critically needed to further investigate this observed pattern of genetic differentiation.

The Cerrado biome seems to operate as a transitional habitat, with much of its current diversity overlapping with that observed in other populations (Figure 4A: PC1). As in the Caatinga, jaguars from this biome have still been poorly sampled, hampering the development of in-depth analysis focusing on this biome. In both the PCA and Admixture analysis, the Amazon emerged as a highly diverse biome, which it expected since there are likely over 10,000 sexually mature individuals in this region (Paula *et al.* 2013). Moreover, is relevant to note that the genetic diversity observed in the Caatinga, Cerrado and Atlantic Forest overlapped to some extent with that of the Amazon, suggesting historical connections and/or similar selective regimes in these biomes. With the best K value estimated with the Admixture analysis (k=2), the main genetic division was between these populations and the Pantanal, a pattern that was also observed in the PCA (PC1). Our ancestry comparison showed intense gene flow within the Pantanal, although each sub-population (northern and southern Pantanal) could be recognized as distinct units by subsequent analyses. This differentiation between northern and southern Pantanal samples, especially visible in the PCA (PC3) (Figure 4B), suggests that the biome cannot be considered to be homogenous for jaguar populations, and that demographic and/or selective factors drive some degree of population structure within it.

We observed relevant differences regarding selection patterns among the analyzed populations. Using a Bayesian approach, we searched for candidate genes for adaptation to different biomes by comparing an open-vegetation biome (Pantanal) against forested regions. Bayescan results revealed four candidate genes: *KLF3*, associated with body fat metabolism; *RNASE7*, associated with skin immunity; and *OD3FL2* and *TEX15*, associated with spermatogenesis. Body fat is an important component of energy storage and protection system against trauma. *KLF3* is a transcription factor described as a negative regulator of white adipocyte growth and maintenance. White adipocytes are responsible for storing triacylglycerol, which can be converted

342 into energy in periods of scarcity. They play a major role in body insulation and secrete adipokines
343 such as leptin and resistin, thus being an important endocrine organ (Friedman & Halaas 1998;
344 Sue *et al.* 2008). Recent studies using knockout mice observed a 20% body size reduction when
345 this gene was non-functional, mostly due to the lack of white adipocytes and associated changes
346 (Sue *et al.* 2008). One of the candidate loci in the Amazon was the leptin receptor (*LEPR*). Leptin
347 is a hormone secreted by the adipose tissue and is responsible by the satiety feeling, which makes
348 it an important factor for energy homeostasis. The leptin receptor is expressed in the
349 hypothalamus, and obesity is the main health issue associated with it. Leptin is also associated
350 with bone density, and could be involved in mechanisms evolved to cope with body weight
351 decrease (Friedman & Halaas 1998). Other gene associated with energy metabolism that
352 presented significant changes in frequency was *PFKFB1*. This gene is associated with response to
353 starvation (GO: 0042594) and gluconeogenesis (GO: 0006094).
354 Our F_{ST} outlier analysis retrieved additional genes that could be associated with body size
355 in jaguars. One of the clearest results is the vertebrae development associated gene (*VRTN*). In
356 domestic breeds of pigs, this gene is associated with thoracic vertebrae number (Yang *et al.* 2016),
357 so it is plausible to hypothesize that vertebrae number or size could be a factor influencing the
358 larger body size observed in the Pantanal population. Additionally, we observed a striking change
359 in allele frequencies for the *IGSF1* gene in the southern Pantanal population. This gene has been
360 associated with the thyroid, and knockout mice have serious developmental problems, including
361 an increased body mass (Sun *et al.* 2012). Other genes that could be associated with body size
362 were: *CTR9*, an imprinting regulator; the parathyroid hormone *PTHLH*; *CENPF* and *KDM5C*, related
363 to short stature in an association analysis (Hanson *et al.* 2014); and in addition, *HEXB*, associated
364 with skeletal development (GO: 0001501).
365 The genes we have found to be under selection seem to corroborate the hypotheses that
366 body size variation in jaguar is regulated by dietary constraints (Hoogesteijn & Mondolfi 1996).
367 Smaller jaguars might have a higher energetic efficiency than larger ones, and should require less
368 biomass for their energetic needs. This can be associated with the Kleiber's law, which postulates
369 that body mass is correlated with the basal metabolic rate (Kleiber 1947). A lower metabolic rate
370 in forests could help to cope with the smaller prey mass available, when compared to the Pantanal.
371 We have also found evidence of selection in genes linked to sexual maturation. In humans, fast
372 growth and early sexual maturation is associated with the pygmy's short stature (Migliano *et al.*
373 2007; Perry & Dominy 2009). This could be also one of the factors related to body size variation
374 in jaguars. Due to the high diversity of pathogens and other threats, fast growth and early sexual
375 maturation could be an alternative to leave more descendants faster in forests than in open
376 habitats. This hypothesis can be investigating by collecting and statistically comparing life history

data for jaguars in these different environments. Although this is a challenging endeavor for such an elusive species, some long-term field studies have been recently established, and can be directed to target this type of question to probe into the physiological and ecological processes underlying the genomic signatures observed here.

All populations presented coat color-related genes with signatures of natural selection. The reason for this might be a result of the pleiotropic effects that these genes have, participating in several different traits, such as development and immunity (Reissmann & Ludwig 2013). Based on previous data, one exception may be hypothesized (Silva 2014). The Pantanal is described as being the only place in the jaguar distribution where there are no melanistic individuals. Silva (2014) raised the possibility that this pattern might occur due to selection. In the present study, we have found a gene, *TPCN2*, that has been found to be associated with blond hair in humans (Sulem *et al.* 2007a). *TPCN2* is a calcium pump, with a similar function to other coloration genes such as *SLC24A4* and *SLC24A5* (Lamason *et al.* 2005; Sulem *et al.* 2007b), and may be involved in the control of eumelanin and pheomelanin production in the cell. However, this gene is also expressed in the kidneys and liver, and its signature of selection could thus be associated with other traits. Interestingly, when we compared the allele frequencies among all populations, we observed that in the Pantanal the ancestral allele is fixed (see Supplementary Table 7), while in the other populations it occurs at lower frequencies. The same pattern was observed in the southern Pantanal population with the *ATRN* gene, also related to coat color and development. These results open up interesting avenues to further investigate the biological role of these genes in jaguars, and indicate that careful assessments of pelage variation across regions (including quantitative and systematic measurements of color, using standardized methods) will be important to test whether this phenotype is indeed the driver of our observed signatures of selection.

In the Amazon population, we observed a significant signature of selection for the *LYST* gene in the FST outlier test. This gene is responsible for controlling the size of lysosomes, and it is associated with the beige coat color phenotype (Barbosa *et al.* 1996). Mutations in this gene can cause the Chediak-Higashi syndrome, and its upregulation can prevent infection by *Leishmania amazonensis* (Wilson *et al.* 2008). Leishmaniosis is a tropical disease, with high prevalence in the Amazon. Our data demonstrated a high frequency of the derived allele in this population, while in the other two, Atlantic Forest and Pantanal, it was almost null (see Supplementary Table 7). This results leads to the hypothesis that Amazonian jaguars have evolved improved resistance to leishmaniosis via positive selection on this gene, indicating that further scrutiny of this locus in the jaguar may not only be revealing in terms of ecological and evolutionary processes, but also relevant from a biomedical standpoint.

412 Outlier methods can be robust to retrieve information about positive selection when there
413 is a considerable amount of loci in the analysis. However, this method is highly sensitive to type
414 II errors, *i.e.* an increase in the number of false positives (Kelley *et al.* 2006), and demographic
415 effects such as population contractions and expansions that can be mistaken for selection (Bierne
416 *et al.* 2013). For this reason, we supplemented the outlier F_{ST} test with other types of selection
417 scans that are more robust to this type of error and allow a more reliable statistical framework.
418 However, the outlier detection methods can still be an effective way to assess genome-wide
419 patterns of variation among different populations, and to retrieve candidate genes with functional
420 importance (Savolainen *et al.* 2013).
421 The lack of strong population structure observed here and also in previous studies (e.g.
422 Eizirik et al. 2001; Valdez *et al.* 2015), associated with the presence of divergent selection in the
423 different jaguar populations, should be taken into account when establishing conservation
424 priorities for the species. Our results indicate that populations that inhabit forests (Atlantic Forest
425 and Amazon) share genomic similarities that likely reflect common adaptive processes, involving
426 genes associated with energetic metabolism, reproduction and immunity at similar frequencies
427 (Figure 6A). The Atlantic Forest is considered a biodiversity hotspot and the most endangered
428 biome in Brazil, with only 10 to 14% of its original area remaining (Ribeiro *et al.* 2009). Jaguar
429 populations in this biome are extremely isolated, with some populations having an effective size
430 of only five individuals (Haag *et al.* 2010). On the other hand, the Amazon is considered to harbor
431 a stable population that is essential for jaguar conservation in the long term (Paula *et al.* 2013).
432 Even though further studies using both genetic and phenotypic data are required, our results
433 indicate that that translocation projects aimed at repopulating viable habitat areas in the Atlantic
434 Forest might use Amazon populations as the source of individuals, given the evidence for similar
435 adaptive signatures at the genomic level. The strategy of using environmental and phenotypic
436 data, in addition to neutral molecular markers, to guide conservation and management actions,
437 has been suggested previously (Eizirik *et al.* 2006), and can now be complemented and extended
438 dramatically with the incorporation of genome-wide information revealing signals of local
439 adaptation.
440 This is the first assessment of genetic diversity, population structure and signatures of
441 selection in the jaguar using high-throughput sequencing techniques and population genomic
442 approaches. The results obtained here provide unprecedented genomic resources for future
443 studies targeting the jaguar and the other *Panthera* species, potentially being applicable to all
444 wild cats globally. This methodology can be applied with a wide variety of samples, such as ancient
445 DNA from museum samples and non-invasive approaches such as those based on field-collected
446 feces or hair (Bi *et al.* 2012). Future use of these materials for genome-wide analyses, especially if

447 combined with improved collection of phenotypic data (*e.g.* morphology from museum
448 specimens and ecological data from field samples) hold immense potential for unravelling
449 detailed aspects the biology and evolutionary dynamics of natural populations.
450
451 **Acknowledgments**
452
453 Authors would like to thank all the institutions and people listed in supplementary table 2, as well
454 Lydia Smith (IB – UC Berkeley). The study was supported by Conselho Nacional de
455 Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and Fundação de Apoio à Pesquisa do
456 Rio Grande do Sul (FAPERGS), Brazil.
457

458 ## References

459

460 Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for
461 signatures of natural selection. *Genome research*, **12**, 1805–14.
462
463 Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated
464 individuals. *Genome Research*, **19**, 1655–1664.
465
466 Barbosa MD, Nguyen QA, Tchernev VT *et al.* (1996) Identification of the homologous beige and
467 Chediak-Higashi syndrome genes. *Nature*, **382**, 262–5.
468
469 Bi K, Vanderpool D, Singhal S *et al.* (2012) Transcriptome-based exon capture enables highly cost-
470 effective comparative genomic data collection at moderate evolutionary scales. *BMC*
471 *Genomics*, **13**, 403.
472
473 Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it.? why are FST outliers sometimes so
474 frequent? *Molecular Ecology*, **22**, 2061–2064.
475
476 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data.
477 *Bioinformatics*, **30**, 2114–2120.
478
479 Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*,
480 **27**, 2156–2158.
481
482 Eizirik E, Johnson WE, O'Brien SJ *et al.* (2006) Definindo unidades evolutivamente significativas e
483 unidades de manejo para a conservação de carnívoros neotropicais. In: *Manejo e*
484 *Conservação de Carnívoros Neotropicais*, pp. 47–66.
485
486 Eizirik E, Kim JH, Menotti-Raymond M *et al.* (2001) Phylogeography, population history and
487 conservation genetics of jaguars (Panthera onca, Mammalia, Felidae). *Molecular ecology*, **10**,
488 65–79.
489
490 Figueiró HV, Li G, Trindade F *et al.* (2016) Jaguar genome sheds lights in the complex evolution of
491 the big cats. *in prep*.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.

Friedman JM, Halaas JL (1998) Leptin and the regulation of body weight in mammals. *Nature*, **395**, 763–770.

Haag T, Santos a S, Sana D a *et al.* (2010) The effect of habitat fragmentation on the genetic structure of a top predator: loss of diversity and high differentiation among remnant populations of Atlantic Forest jaguars (Panthera onca). *Molecular ecology*, **19**, 4906–21.

Hanson D, Stevens A, Murray PG, Black GCM, Clayton PE (2014) Identifying biological pathways that underlie primordial short stature using network analysis. *Journal of Molecular Endocrinology*, **52**, 333–344.

Hayward MW, Kamler JF, Montgomery RA *et al.* (2016) Prey Preferences of the Jaguar Panthera onca Reflect the Post-Pleistocene Demise of Large Prey. *Frontiers in Ecology and Evolution*, **3**.

Hoogesteijn R, Mondolfi E (1996) Body mass and skull measurements in four jaguar populations and observations on their prey base. *Bulletin of the Florida Museum of Natural History*, **39**, 195–219.

Johnson WE, Eizirik E, Pecon-Slattery J *et al.* (2006) The late Miocene radiation of modern Felidae: a genetic assessment. *Science*, **311**, 73–77.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM (2006) Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome research*, **16**, 980–9.

Kiltie RA (1984) Size ratios among sympatric neotropical cats. *Oecologia*, **61**, 411–416.

Kleiber M (1947) Body size and metabolic rate. *Physiol Rev*, **27**, 511–541.

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, **15**, 356.

Lamason RL, Mohideen M-APK, Mest JR *et al.* (2005) SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. *Science*, **310**, 1782–1786.

Larson SE (1997) Taxonomic Re-Evaluation of the Jaguar. *ZooBiology*, **120**, 107–120.

Magoc T, Salzberg SL (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.

Meyer M, Kircher M (2010) Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, **2010**, pdb.prot5448–pdb.prot5448.

Migliano AB, Vinicius L, Lahr MM (2007) Life history trade-offs explain the evolution of human pygmies. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 20216–9.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE*, **7**.

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature reviews. Genetics*, **12**, 443–51.

Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. *PLoS Genetics*, **2**, e190.

Paula RC, Desbiez A, Cavalcanti SMC (2013) *Plano de açao nacional para a conservaçao da onça pintada*.

Perry GH, Dominy NJ (2009) Evolution of the human pygmy phenotype. *Trends in Ecology and Evolution*, **24**, 218–225.

Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome research*, **20**, 291–300.

Price A, Patterson NJ, Plenge RM *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, **38**, 904–9.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, **81**, 559–575.

Reissmann M, Ludwig A (2013) Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Seminars in Cell and Developmental Biology*, **24**, 576–586.

Ribeiro MC, Metzger JP, Martensen AC, Ponzoni FJ, Hirota MM (2009) The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, **142**, 1141–1153.

Roques S, Furtado M, Jácomo a. T a. *et al.* (2014) Monitoring jaguar populations Panthera onca with non-invasive genetics: a pilot study in Brazilian ecosystems. *Oryx*, 1–9.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.

Seymour KL (1989) Panthera onca. *Mammalian Species*, **11**, 1–9.

Silva LG da (2014) Análise da distribuição espacial do melanismo na família felidae em função de condicionantes ambientais. Porto Alegre.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: The next generation. *Trends in Ecology and Evolution*, **25**, 705–712.

596 Sue N, Jack BH a, Eaton S a *et al.* (2008) Targeted disruption of the basic Krüppel-like factor gene
597 (Klf3) reveals a role in adipogenesis. *Molecular and cellular biology*, **28**, 3967–3978.
598
599 Sulem P, Gudbjartsson DF, Stacey SN *et al.* (2007a) Genetic determinants of hair, eye and skin
600 pigmentation in Europeans. *Nature genetics*, **39**, 1443–52.
601
602 Sun Y, Bak B, Schoenmakers N *et al.* (2012) Loss-of-function mutations in IGSF1 cause an X-linked
603 syndrome of central hypothyroidism and testicular enlargement. *Nature genetics*, **44**, 1375–
604 81.
605
606 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA
607 polymorphism. *Genetics*, **123**, 585–595.
608
609 Valdez FP, Haag T, Azevedo FCC *et al.* (2015a) Population genetics of Jaguars (*Panthera onca*) in
610 the Brazilian Pantanal: Molecular evidence for demographic connectivity on a regional scale.
611 *Journal of Heredity*, **106**, 503–511.
612
613 Valdez FP, Haag T, Azevedo FCC *et al.* (2015b) Population Genetics of Jaguars (Panthera onca) in
614 the Brazilian Pantanal: Molecular Evidence for Demographic Connectivity on a Regional
615 Scale. *Journal of Heredity*, **106**, 503–511.
616
617 Wang J, Duncan D, Shi Z, Zhang B (2013) WEB-based GEne SeT AnaLysis Toolkit (WebGestalt):
618 update 2013. *Nucleic acids research*, **41**, W77–83.
619 Weir B, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure
620 Author (s): B . S . Weir and C . Clark Cockerham. *Evolution*, **38**, 1358–1370.
621
622 Wilson J, Huynh C, Kennedy KA *et al.* (2008) Control of parasitophorous vacuole expansion by
623 LYST/Beige restricts the intracellular growth of Leishmania amazonensis. *PLoS Pathogens*, **4**.
624
625 Yang J, Benyamin B, McEvoy BP *et al.* (2010) Common SNPs explain a large proportion of the
626 heritability for human height. *Nature Genetics*, **42**, 565–569.
627
628 Yang J, Huang L, Yang M *et al.* (2016) Possible introgression of the VRTN mutation increasing
629 vertebral number, carcass length and teat number from Chinese pigs into European pigs.
630 *Scientific Reports*, **6**, 19240.
631
632 Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and
633 principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.
634
635
636

Supplementary Material

Supplementary Table 1: Detailed information for the samples used in this study – Sample identification, geographic location, assigned population, collectors and mean depth.

ID	City	State/Province	Population	Collector	Mean Depth
bPon-002	Foz do Iguaçu	Paraná	Atlantic Forest	Peter Crawshaw	5.58
bPon-012	Puerto Iguazú	Misiones	Atlantic Forest	Peter Crawshaw	3.61
bPon-016	Bataguassu	Mato Grosso do Sul	Atlantic Forest	Dênis Sana	7.11
bPon-017	Anaurilândia	Mato Grosso do Sul	Atlantic Forest	Ronaldo Morato	6.41
bPon-020	Anaurilândia	Mato Grosso do Sul	Atlantic Forest	Ronaldo Morato	3.59
bPon-024	Alto Paraná	Paraná	Atlantic Forest	Laury Cullen	15.38
bPon-027	Anaurilândia	Mato Grosso do Sul	Atlantic Forest	Dênis Sana	5.11
bPon-030	Anaurilândia	Mato Grosso do Sul	Atlantic Forest	Dênis Sana	5.59
bPon-035	Ivinhema	Mato Grosso do Sul	Atlantic Forest	Dênis Sana	17.36
bPon-040	Ivinhema	Mato Grosso do Sul	Atlantic Forest	Dênis Sana	15.17
bPon-045	Anaurilândia	Mato Grosso do Sul	Atlantic Forest		9.27
bPon-047	Ivinhema	Mato Grosso do Sul	Atlantic Forest	Dênis Sana	5.95
bPon-048	Teodoro Sampaio	São Paulo	Atlantic Forest	Laury Cullen	8.28
bPon-050	Teodoro Sampaio	São Paulo	Atlantic Forest	Laury Cullen	6.42
bPon-052	Teodoro Sampaio	São Paulo	Atlantic Forest	Laury Cullen	5.29
bPon-053	Teodoro Sampaio	São Paulo	Atlantic Forest	Laury Cullen	11.84
bPon-068	Miranda	Mato Grosso do Sul	South Pantanal	Fernando Azevedo	5.91
bPon-069	Miranda	Mato Grosso do Sul	South Pantanal	Fernando Azevedo	4.61
bPon-071	Miranda	Mato Grosso do Sul	South Pantanal	Fernando Azevedo	12.73
bPon-074	Miranda	Mato Grosso do Sul	South Pantanal	Fernando Azevedo	10.64
bPon-076	Miranda	Mato Grosso do Sul	South Pantanal	Fernando Azevedo	4.60
bPon-078	Ivinhema	Mato Grosso do Sul	Atlantic Forest	Dênis Sana	15.33

Supplementary Table 1: Continued.

ID	City	State/Province	Population	Collector	Mean Depth
bPon-084		Pará	Amazon	Marcos Renato Mattos	9.93
bPon-085		Pará	Amazon	Marcos Renato Mattos	12.14
bPon-130			Atlantic Forest		5.39
bPon-144	Corumbá	Mato Grosso do Sul	North Pantanal	Fernando Azevedo	8.71
bPon-145	Corumbá	Mato Grosso do Sul	North Pantanal	Fernando Azevedo	9.71
bPon-148	Corumbá	Mato Grosso do Sul	North Pantanal	Fernando Azevedo	8.61
bPon-155	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	3.68
bPon-157	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	5.05
bPon-158	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	4.61
bPon-159	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	4.11
bPon-161	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	7.83
bPon-162	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	7.01
bPon-163	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	9.25
bPon-167	Tefé	Amazonas	Amazon	Emiliano Ramalho	5.57
bPon-170	Corumbá	Mato Grosso do Sul	North Pantanal	Fernando Azevedo	10.53
bPon-171	Corumbá	Mato Grosso do Sul	North Pantanal	Fernando Azevedo	12.67
bPon-309*			South Pantanal	Peter Crawshaw	5.34
bPon-318	Manacapuru	Amazonas	Amazon		8.23
bPon-319	Manacapuru	Amazonas	Amazon		6.61
bPon-320	Grande Sertão Veredas	Minas Gerais	Cerrado		8.82
bPon-333	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	9.74
bPon-334	Lambari D'Oeste	Mato Grosso	North Pantanal	Rodrigo Jorge	2.88
bPon-338	Chapadão do Céu	Goiás	Cerrado	Leandro Silveira	9.03
bPon-339	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	9.77
bPon-340	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	8.32

Supplementary Table 1: Continued.

ID	City	State/Province	Population	Collector	Mean Depth
bPon-342	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	13.19
bPon-343	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	9.89
bPon-344	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	11.02
bPon-349	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	10.11
bPon-350	Miranda	Mato Grosso do Sul	South Pantanal	Leandro Silveira	10.93
bPon-353	Caseara	Tocantins	Amazon	Leandro Silveira	14.71
bPon-354	Caseara	Tocantins	Amazon	Leandro Silveira	7.01
bPon-361	Aquidauana	Mato Grosso do Sul	South Pantanal	Sandra Cavalcanti	8.03
bPon-376	Tucuruí	Pará	Amazon	Joares May	11.49
bPon-377	Rondon do Pará	Pará	Amazon	Joares May	10.76
bPon-378	Rondon do Pará	Pará	Amazon	Joares May	8.13
bPon-379	Aquidauana	Mato Grosso do Sul	South Pantanal	Sandra Cavalcanti	4.78
bPon-381	Corumbá	Mato Grosso do Sul	North Pantanal	Sandra Cavalcanti	3.54
bPon-384	Aquidauana	Mato Grosso do Sul	South Pantanal	Sandra Cavalcanti	5.89
bPon-390	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	5.83
bPon-394	Chapada Gaúcha	Minas Gerais	Cerrado	Joares May	5.60
bPon-395	Foz do Iguaçu	Paraná	Atlantic Forest	Raphael Xavier	10.5
bPon-397	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	18.15
bPon-399	Poconé	Mato Grosso	North Pantanal	Peter Crawshaw	5.78
bPon-400	Poconé	Mato Grosso	North Pantanal	Peter Crawshaw	6.03
bPon-401	Uarini	Amazonas	Amazon	Emiliano Ramalho	10.53
bPon-402	Uarini	Amazonas	Amazon	Emiliano Ramalho	10.32
bPon-403	Uarini	Amazonas	Amazon	Emiliano Ramalho	9.56
bPon-404	Uarini	Amazonas	Amazon	Emiliano Ramalho	9.46
bPon-405	Uarini	Amazonas	Amazon	Emiliano Ramalho	6.30

Supplementary Table 1: Continued.

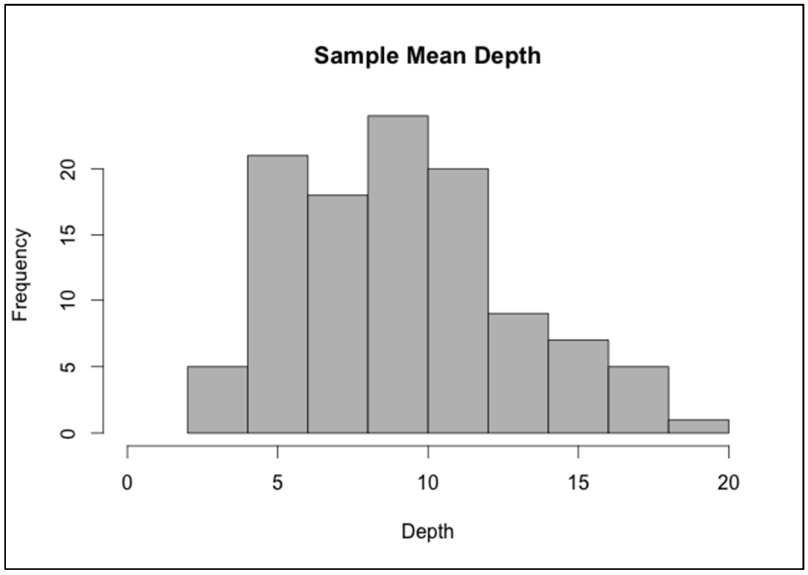
ID	City	State/Province	Population	Collector	Mean Depth
bPon-406	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	15.22
bPon-407	Alta Floresta	Mato Grosso	Amazon	Maria Emília	7.56
bPon-408	Corumbá	Mato Grosso	North Pantanal	Joares May	5.20
bPon-409	Corumbá	Mato Grosso do Sul	South Pantanal	Joares May	6.93
bPon-410	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	11.64
bPon-411	Cocos	Bahia	Cerrado	Joares May	10.67
bPon-412	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	15.84
bPon-413	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	16.21
bPon-414	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	7.02
bPon-415	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	7.46
bPon-416	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	6.19
bPon-417	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	7.89
bPon-418	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	10.77
bPon-419	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	10.63
bPon-421	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	16.15
bPon-423	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	17.90
bPon-424	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	13.64
bPon-426	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	10.44
bPon-427	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	9.85
bPon-428	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	13.86
bPon-429	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	11.59
bPon-430	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	12.69
bPon-432	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	10.03
bPon-433	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	4.2
bPon-435	Corumbá	Mato Grosso do Sul	North Pantanal	Joares May	15.63

Supplementary Table 1: Continued.

ID	City	State/Province	Population	Collector	Mean Depth
bPon-437	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	9.86
bPon-438	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	7.67
bPon-440	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	9.98
bPon-441	Corumbá	Mato Grosso do Sul	North Pantanal	Gediendson Araújo	12.21
bPon-445	Miranda	Mato Grosso do Sul	South Pantanal	Joares May	9.74
bPon-446	Corumbá	Mato Grosso do Sul	North Pantanal	Gediendson Araújo	8.26
bPon-449	Corumbá	Mato Grosso do Sul	North Pantanal	Gediendson Araújo	9.31
bPon-460			Caatinga		10.55
bPon-461			Caatinga		7.30
bPon-462	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	6.07
bPon-463	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	16.09
bPon-464	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	10.16
bPon-465	Cáceres	Mato Grosso	North Pantanal	Daniel Kantek	13.48

Supplementary Table 2: Genomic resources used in this study – Genome resources from the five *Panthera* species used for probe design.

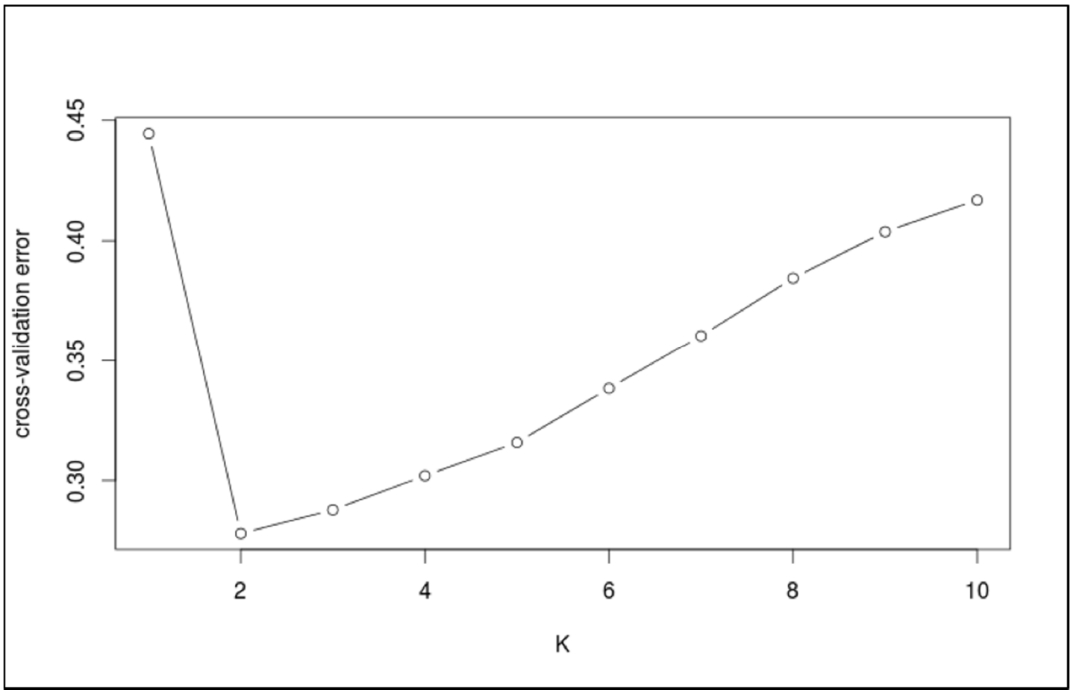
Common name	Scientific Name	Type of Data	Accession Number	Reference
Jaguar	*Panthera onca*	High-coverage genome	-	Figueiró et al, *in prep*
Leopard	*Panthera pardus*	Low-coverage genome	-	Figueiró et al, *in prep*
Tiger	*Panthera tigris*	High-coverage genome	ATCQ01	Cho et al. 2013
Lion	*Panthera leo*	Low-coverage genome (raw reads)	SRX273034	Cho et al. 2013
Snow Leopard	*Panthera uncia*	Low-coverage genome (raw reads)	SRX273036	Cho et al. 2013



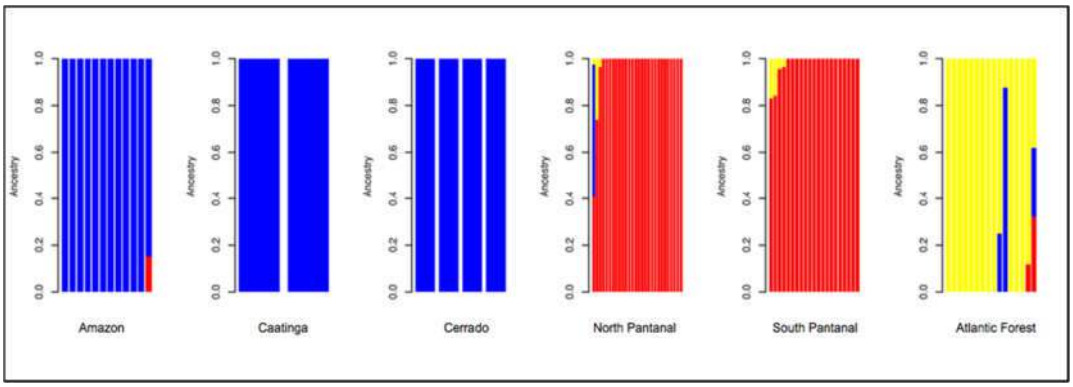
Supplementary Figure 1: Mean individual depth for all samples in the present study.

Supplementary Table 3: Relatedness among individuals in each population – Relatedness comparison between individuals in each population. Values > 0.15 indicate close relatives.

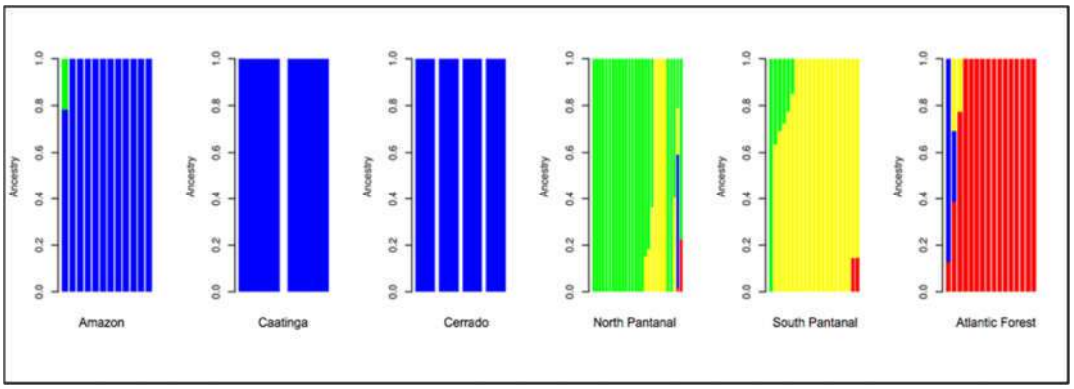
Population	Individual 1	Individual 2	Relatedness
Atlantic Forest	bPon017	bPon024	0.140077
Atlantic Forest	bPon050	bPon052	0.151488
Amazon	bPon401	bPon404	0.176210
Amazon	bPon401	bPon405	0.126517
Amazon	bPon404	bPon405	0.363307
South Pantanal	bPon068	bPon074	0.236139
South Pantanal	bPon069	bPon071	0.157694
South Pantanal	bPon157	bPon162	0.189538
South Pantanal	bPon159	bPon339	0.160219
South Pantanal	bPon163	bPon424	0.189701
South Pantanal	bPon163	bPon440	0.201479
South Pantanal	bPon340	bPon361	0.161819
South Pantanal	bPon340	bPon384	0.170515
South Pantanal	bPon342	bPon343	0.201404
South Pantanal	bPon342	bPon349	0.199664
South Pantanal	bPon343	bPon349	0.274443
South Pantanal	bPon361	bPon379	0.160054
South Pantanal	bPon416	bPon432	0.161878
South Pantanal	bPon424	bPon445	0.190362
North Pantanal	bPon145	bPon171	0.284739
North Pantanal	bPon381	bPon441	0.166160
North Pantanal	bPon390	bPon413	0.205808
North Pantanal	bPon397	bPon427	0.152657
North Pantanal	bPon399	bPon408	0.153084
North Pantanal	bPon408	bPon410	0.198432
North Pantanal	bPon414	bPon446	0.204343
North Pantanal	bPon418	bPon430	0.278328
North Pantanal	bPon419	bPon438	0.220122
North Pantanal	bPon428	bPon437	0.124346
North Pantanal	bPon428	bPon438	0.112581
North Pantanal	bPon430	bPon465	0.114092
North Pantanal	bPon437	bPon438	0.132825
North Pantanal	bPon441	bPon446	0.197968



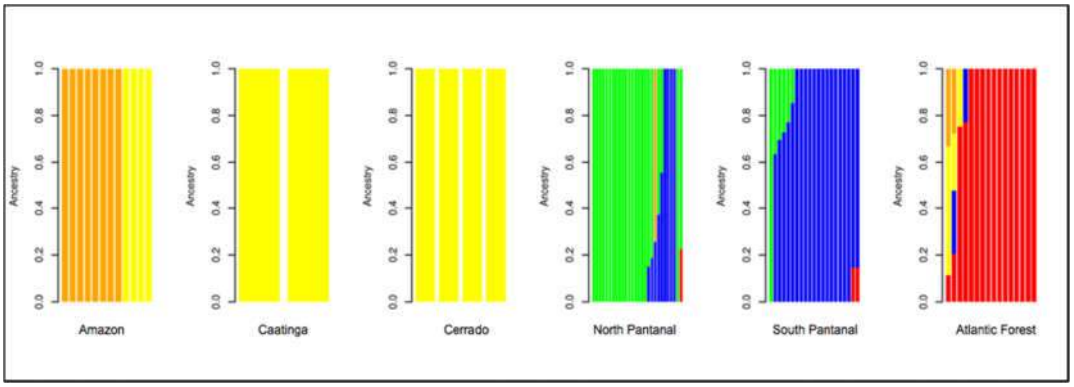
Supplementary Figure 2: Admixture cross validation – Cross-validation error showing the best estimated K value for the Admixture plot. K = 3-5 are shown in Supplementary Figure 3-5.



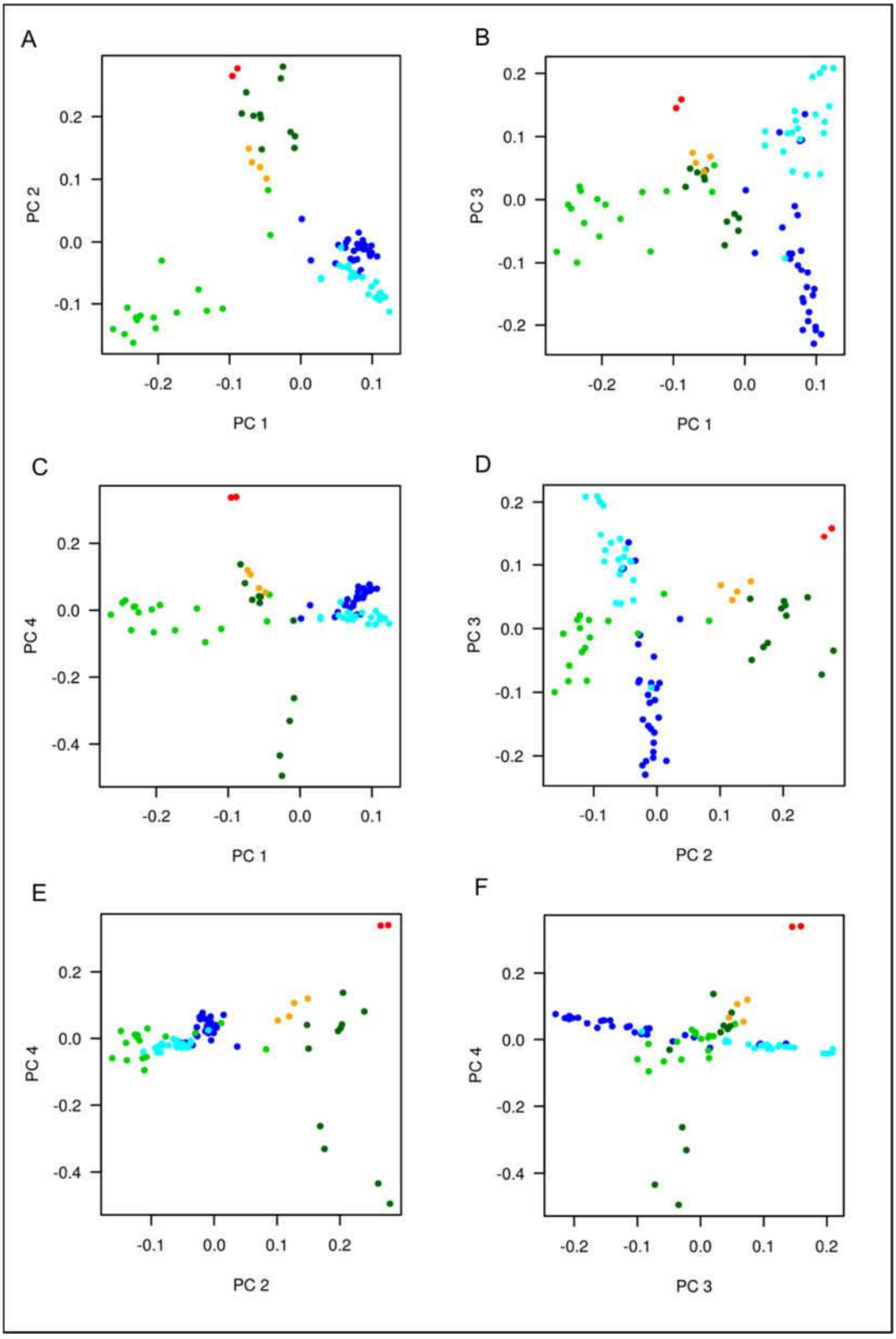
Supplementary Figure 3: Admixture plot – Ancestry proportions of the 82 individuals from six extant South America populations with K = 3.

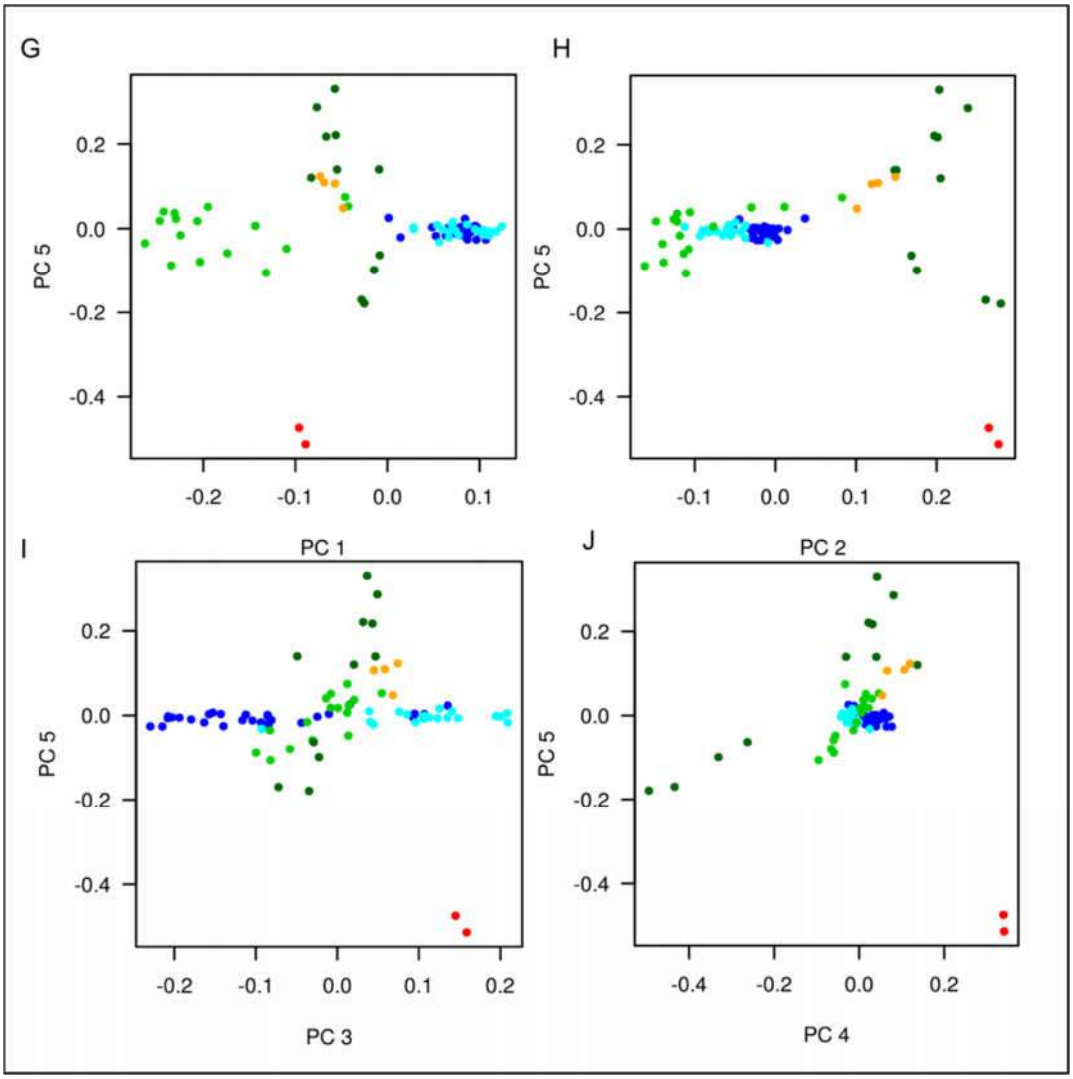


Supplementary Figure 4: Admixture plot – Ancestry proportions of the 82 individuals from six extant South American populations with K = 4.



Supplementary Figure 5: Admixture plot – Ancestry proportions of the 82 individuals from six extant South America populations with K = 5.


A
PC 2
PC 1
B
PC 3
PC 1
C
PC 4
PC 1
D
PC 3
PC 2
E
PC 4
PC 2
F
PC 4
PC 3



Supplementary Figure 6: PCA Plots – Principal Component Analysis with 60,183 SNPS for 82 individuals. Populations are identified by different colors. North Pantanal – dark blue; South Pantanal – light blue; Cerrado – orange; Caatinga – red; Atlantic forest – light green; Amazon – dark green. **(A-J)** PCA plots among principal components 1-5.

Supplementary Table 4: List of enriched terms for the candidate genes for the Atlantic Forest population – Enriched terms for the F_{ST} outlier analysis with two or more genes. Annotation Term with Gene Ontoloy and Disease terms. O: Observed, with number of genes under selection; and E: Expected, with number of genes related to the term. False Discovery rate value (FDR) > 0.05. Gene list with gene symbol.

Annotation Term	O	E	q-value	Gene List
GO:0005515(protein binding)	33	8670	1,65E-04	*RFC1; CASP3; MC5R; TEC; PRDM4; PPP4R2; NLRC4; PLEKHF1; DDO; GNE; SYTL5; SEMA4G; TDRD7; OCIAD1; INTS2; S1PR3; ANKLE2; CORO2B; CBX7; NOD2; HPSE; TRAF6; KLB; PRDM1; ADAMTS12; ARF1; ZNF350; DTL; ABCG8; NOTCH2; LIG4; SPRY1; FBXO7*
GO:0046872(metal ion binding)	15	1870	1,65E-04	*THAP9; POLR3B; TEC; ZNF829; PARP12; FREM1; PLEKHF1; GNE; ZNF608; ZNF527; STEAP1; PRDM1; ZNF350; EXTL3; LIG4*
GO:0005524(ATP binding)	12	1467	8,81E-04	*RFC1; XYLB; TEC; NLRC4; MAP3K19; NARS2; GNE; BMS1; NOD2; MCM9; ABCG8; LIG4*
GO:0003677(DNA binding)	11	1453	2,06E-03	*RFC1; POLR3B; ZNF829; PRDM4; TLR8; JRKL; ZNF527; TBPL2; ZNF350; MCM9; LIG4*
GO:0045087(innate immune response)	**8**	**881**	**5,58E-03**	***POLR3B; TEC; TLR8; NLRC4; NOD2; TRAF6; KLB; TLR10***
GO:0006351(transcription, DNA-templated)	8	1809	3,04E-02	*RFC1; ZNF829; ZNF527; CBX7; CCNT2; PRDM1; ZNF350; RLIM*
GO:0006355(regulation of transcription, DNA-templated)	7	1290	2,49E-02	*RFC1; ZNF829; PRDM4; TSC22D2; ZNF527; ZNF350; NOTCH2*
GO:0004842(ubiquitin-protein transferase activity)	6	284	1,03E-03	*KLHL13; NLRC4; TRAF6; DTL; RLIM; FBXO7*
GO:0000122(negative regulation of transcription from RNA polymerase II promoter)	6	719	1,57E-02	*CBX7; TRAF6; PRDM1; ZNF350; RLIM; NOTCH2*
GO:0008270(zinc ion binding)	6	1176	3,54E-02	*PRDM4; RNF26; ZDHHC14; TRAF6; ADAMTS12; RLIM*
GO:0007155(cell adhesion)	5	417	1,57E-02	*GNE; CDH11; COL15A1; KIAA1462; WISP1*
GO:0005509(calcium ion binding)	5	683	2,49E-02	*CSN2; SYTL5; PCDH18; CDH11; NOTCH2*
GO:0007186(G-protein coupled receptor signaling pathway)	5	883	3,71E-02	*HTR5A; OR52B6; S1PR3; OR2AE1; OR51E1*
Crohn's disease(23128233)	**4**	**74**	**1,03E-03**	***TEC; TEC; NOD2; NOTCH2***
GO:0002755(MyD88-dependent toll-like receptor signaling pathway)	**4**	**85**	**1,47E-03**	***TLR8; NOD2; TRAF6; TLR10***
GO:0002224(toll-like receptor signaling pathway)	**4**	**114**	**3,50E-03**	***TLR8; NOD2; TRAF6; TLR10***

Supplementary Table 4: Continued.

Annotation Term	O	E	q-value	Gene List
GO:0005975(carbohydrate metabolic process)	4	346	2,03E-02	*XYLB; MANBA; HPSE; KLB*
GO:0016567(protein ubiquitination)	4	363	2,07E-02	*KLHL13; NLRC4; RLIM; FBXO7*
GO:0006954(inflammatory response)	**4**	**358**	**2,07E-02**	***TLR8; NLRC4; S1PR3; TLR10***
GO:0006281(DNA repair)	4	379	2,07E-02	*RFC1; PPP4R2; DTL; LIG4*
GO:0048011(neurotrophin TRK receptor signaling pathway)	**4**	**395**	**2,24E-02**	***CASP3; PRDM4; TRAF6; KLB***
GO:0045892(negative regulation of transcription, DNA-templated)	4	483	3,04E-02	*TBPL2; TRAF6; ZNF350; RLIM*
GO:0004930(G-protein coupled receptor activity)	4	640	4,58E-02	*OR52B6; S1PR3; OR2AE1; OR51E1*
GO:0035872(nucleotide-binding domain, leucine rich repeat containing receptor signaling pathway)	**3**	**43**	**3,78E-03**	***NLRC4; NOD2; TRAF6***
Attention deficit hyperactivity disorder and conduct disorder(18951430)	3	52	5,58E-03	*CDH11; LIG4; ABHD13*
GO:0034166(toll-like receptor 10 signaling pathway)	**3**	**69**	**9,73E-03**	***NOD2; TRAF6; TLR10***
GO:0034146(toll-like receptor 5 signaling pathway)	**3**	**69**	**9,73E-03**	***NOD2; TRAF6; TLR10***
GO:0034162(toll-like receptor 9 signaling pathway)	**3**	**75**	**1,15E-02**	***TLR8; NOD2; TRAF6***
GO:0051092(positive regulation of NF-kappaB transcription factor activity)	**3**	**130**	**1,57E-02**	***NLRC4; NOD2; TRAF6***
GO:0042742(defense response to bacterium)	**3**	**134**	**1,57E-02**	***NLRC4; NOD2; NOTCH2***
GO:0007160(cell-matrix adhesion)	3	93	1,57E-02	*FREM1; HPSE; ADAMTS12*
Rheumatoid arthritis(24390342)	3	171	2,07E-02	*TEC; TEC; TRAF6*
GO:0004872(receptor activity)	3	197	2,36E-02	*TLR8; EXTL3; NOTCH2*
GO:0006974(cellular response to DNA damage stimulus)	**3**	**197**	**2,36E-02**	***CASP3; DTL; MCM9***

Supplementary Table 4: Continued.

Annotation Term	O	E	q-value	*Gene List*
GO:0016874(ligase activity)	3	254	3,04E-02	*TRAF6; RLIM; LIG4*
GO:0043065(positive regulation of apoptotic process)	3	299	3,71E-02	*NLRC4; TRAF6; NOTCH2*
GO:0038095(Fc-epsilon receptor signaling pathway)	3	315	3,95E-02	*TEC; TRAF6; KLB*
GO:0051301(cell division)	3	320	4,01E-02	*ANKLE2; CCNT2; LIG4*
GO:0042493(response to drug)	3	327	4,19E-02	*CASP3; CBX7; ABCG8*
GO:0008283(cell proliferation)	3	363	4,72E-02	*PRDM4; LMNTD1; LIG4*
GO:0019901(protein kinase binding)	3	381	4,95E-02	*NOD2; TRAF6; FBXO7*
GO:0004553(hydrolase activity, hydrolyzing O-glycosyl compounds)	2	6	3,50E-03	*GNE; KLB*
GO:0050707(regulation of cytokine secretion)	**2**	**10**	**7,23E-03**	***TLR8; TLR10***
Formal thought disorder in schizophrenia(22648509)	2	12	9,73E-03	*HTR5A; JRKL*
GO:0016045(detection of bacterium)	**2**	**15**	**1,15E-02**	***NLRC4; NOD2***
GO:0005123(death receptor binding)	2	15	1,15E-02	*CASP3; PRDM4*
GO:0030889(negative regulation of B cell proliferation)	**2**	**16**	**1,25E-02**	***CASP3; PRDM1***
GO:0045089(positive regulation of innate immune response)	2	18	1,51E-02	*POLR3B; TLR8*
GO:0006297(nucleotide-excision repair, DNA gap filling)	2	24	1,57E-02	*RFC1; LIG4*
GO:0004857(enzyme inhibitor activity)	2	30	1,57E-02	*CSN2; FRY*
GO:0003727(single-stranded RNA binding)	2	36	1,57E-02	*TLR8; CBX7*
GO:0010165(response to X-ray)	2	23	1,57E-02	*CASP3; LIG4*
GO:0043011(myeloid dendritic cell differentiation)	**2**	**21**	**1,57E-02**	***TRAF6; NOTCH2***

Supplementary Table 4: Continued.

Annotation Term	O	E	q-value	*Gene List*
GO:0007281(germ cell development)	2	34	1,57E-02	*TDRD7; PRDM1*
GO:0045736(negative regulation of cyclin-dependent protein serine/threonine kinase activity)	2	22	1,57E-02	*CASP3; FBXO7*
GO:0070423(nucleotide-binding oligomerization domain containing signaling pathway)	2	25	1,57E-02	*NOD2; TRAF6*
GO:0042246(tissue regeneration)	2	22	1,57E-02	*TEC; NOTCH2*
GO:0042769(DNA damage response, detection of DNA damage)	2	36	1,57E-02	*RFC1; DTL*
GO:0046835(carbohydrate phosphorylation)	2	21	1,57E-02	*XYLB; GNE*
GO:0051402(neuron apoptotic process)	2	38	1,60E-02	*CASP3; LIG4*
Erectile dysfunction and prostate cancer treatment(20932654)	2	40	1,76E-02	*SYTL5; ZNF608*
Very long-chain saturated fatty acid levels (fatty acid 20:0)(25378659)	2	41	1,82E-02	*ZNF608; CDH11*
GO:0009411(response to UV)	**2**	**43**	**1,98E-02**	***CASP3; DTL***
GO:0007187(G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger)	2	46	2,07E-02	*HTR5A; MC5R*
Crohn's disease(18587394)	2	50	2,07E-02	*NOD2; PRDM1*
Ulcerative colitis(23128233)	2	53	2,07E-02	*JRKL; MANBA*
GO:0019985(translesion synthesis)	2	48	2,07E-02	*RFC1; DTL*
GO:0007254(JNK cascade)	2	50	2,07E-02	*NOD2; TRAF6*
GO:0050727(regulation of inflammatory response)	2	54	2,13E-02	*NOD2; ADAMTS12*
GO:0007249(I-kappaB kinase/NF-kappaB signaling)	2	58	2,37E-02	*TLR8; TRAF6*
GO:0021766(hippocampus development)	**2**	**58**	**2,37E-02**	***CASP3; HTR5A***

Supplementary Table 4: Continued.

Annotation Term	O	E	q-value	*Gene List*
GO:0008168(methyltransferase activity)	**2**	**61**	**2,49E-02**	***PRDM4; PRDM1***
GO:0007368(determination of left/right symmetry)	**2**	**60**	**2,49E-02**	***ARMC4; NOTCH2***
GO:0051403(stress-activated MAPK cascade)	2	61	2,49E-02	*NOD2; TRAF6*
GO:0032147(activation of protein kinase activity)	2	63	2,49E-02	*MAP3K19; TRAF6*
GO:0050790(regulation of catalytic activity)	2	60	2,49E-02	*PPP4R2; ANKLE2*
GO:0043086(negative regulation of catalytic activity)	2	73	2,87E-02	*CSN2; FRY*
GO:0043085(positive regulation of catalytic activity)	2	71	2,87E-02	*RFC1; ARF1*
GO:0038123(toll-like receptor TLR1:TLR2 signaling pathway)	**2**	**76**	**2,89E-02**	***NOD2; TRAF6***
GO:0038124(toll-like receptor TLR6:TLR2 signaling pathway)	**2**	**76**	**2,89E-02**	***NOD2; TRAF6***
GO:0001503(ossification)	**2**	**77**	**2,95E-02**	***CDH11; TRAF6***
GO:0034134(toll-like receptor 2 signaling pathway)	**2**	**78**	**3,00E-02**	***NOD2; TRAF6***
GO:0035666(TRIF-dependent toll-like receptor signaling pathway)	**2**	**81**	**3,04E-02**	***NOD2; TRAF6***
GO:0002756(MyD88-independent toll-like receptor signaling pathway)	2	83	3,04E-02	*NOD2; TRAF6*
GO:0050731(positive regulation of peptidyl-tyrosine phosphorylation)	2	86	3,13E-02	*TEC; NOD2*
GO:0034138(toll-like receptor 3 signaling pathway)	**2**	**87**	**3,19E-02**	***NOD2; TRAF6***
GO:0019886(antigen processing and presentation of exogenous peptide antigen via MHC class II)	**2**	**93**	**3,33E-02**	***TRAF6; ARF1***
GO:0006464(cellular protein modification process)	2	92	3,33E-02	*PPP4R2; MANBA*
GO:0047485(protein N-terminus binding)	2	99	3,54E-02	*TDRD7; TRAF6*
GO:0000187(activation of MAPK activity)	2	103	3,69E-02	*NOD2; TRAF6*

Supplementary Table 4: Continued.

Annotation Term	**O**	**E**	**q-value**	***Gene List***
GO:0042826(histone deacetylase binding)	2	104	3,71E-02	*TRAF6; PRDM1*
GO:0034142(toll-like receptor 4 signaling pathway)	**2**	**104**	**3,71E-02**	***NOD2; TRAF6***
GO:0005178(integrin binding)	2	106	3,71E-02	*S1PR3; WISP1*
GO:0007417(central nervous system development)	**2**	**118**	**4,19E-02**	***JRKL; LIG4***
GO:0032355(response to estradiol)	**2**	**118**	**4,19E-02**	***CASP3; HTR5A***
GO:0022617(extracellular matrix disassembly)	2	120	4,26E-02	*CASP3; COL15A1*
GO:0097190(apoptotic signaling pathway)	2	120	4,26E-02	*CASP3; TRAF6*
GO:0009653(anatomical structure morphogenesis)	**2**	**128**	**4,46E-02**	***KRT35; S1PR3***
GO:0006260(DNA replication)	2	131	4,58E-02	*DTL; MCM9*
GO:0005215(transporter activity)	2	140	4,76E-02	*CSN2; STEAP1*
GO:0000724(double-strand break repair via homologous recombination)	2	141	4,81E-02	*PPP4R2; MCM9*

Supplementary Table 5: List of enriched terms for the candidate genes for the Amazon population: Enriched terms for the F_{ST} outlier analysis with two or more genes. Annotation Term with Gene Ontoloy and Disease terms. O: Observed, with number of genes under selection; and E: Expected, with number of genes related to the term. False Discovery rate value (FDR) > 0.05. Gene list with gene symbol.

Annotation Term	O	E	q-value	Gene List
GO:0005515(protein binding)	27	8670	1,23E-04	*AMPH; ZNF169; DEDD; NDP; GPANK1; PRRC2A; CENPF; TEKT1; SSH1; LEPR; PIP5K1B; CCNB3; PCID2; ZSCAN16; LYST; MTERF4; LMO1; RASAL2; RPAP3; HEXIM2; ATG14; TNNT2; RBCK1; SERPINA5; POMGNT2; COG5; SIT1*
GO:0044281(small molecule metabolic process)	7	1706	2,08E-02	*EPHX2; DLST; PIP5K1B; PHKA2; ACSBG2; UST; AOC1*
GO:0007165(signal transduction)	6	1033	1,09E-02	*NDP; SCARF1; TG; RASAL2; RPS6KL1; SIT1*
GO:0042803(protein homodimerization activity)	6	719	1,09E-02	*EPHX2; NDP; TG; CENPF; CR2; AOC1*
Obesity-related traits(23251661)	**6**	**1228**	**1,96E-02**	***RP1L1; GREB1; RASAL2; RPAP3; FSCB; POMGNT2***
GO:0003677(DNA binding)	6	1453	2,68E-02	*ZNF169; DEDD; KDM5C; SSH1; ZNF710; CR2*
GO:0006351(transcription, DNA-templated)	6	1809	3,96E-02	*ZNF169; DEDD; KDM5C; ZNF710; ZSCAN16; HEXIM2*
GO:0007283(spermatogenesis)	**5**	**380**	**9,76E-03**	***DEDD; ACSBG2; DAZAP1; CABS1; SERPINA5***
GO:0043547(positive regulation of GTPase activity)	4	480	1,78E-02	*AMPH; ELMOD2; RASAL2; ARHGAP36*
GO:0007596(blood coagulation)	**4**	**492**	**1,80E-02**	***NBEAL2; LYST; LRRC16A; SERPINA5***
GO:0004888(transmembrane signaling receptor activity)	3	200	1,42E-02	*SCARF1; LEPR; CR2*
GO:0032403(protein complex binding)	3	227	1,78E-02	*TG; LRRC16A; AOC1*
GO:0005096(GTPase activator activity)	3	262	2,08E-02	*ELMOD2; RASAL2; ARHGAP36*
GO:0051301(cell division)	3	320	2,66E-02	*CENPF; CCNB3; TRIOBP*
GO:0042493(response to drug)	3	327	2,77E-02	*CENPF; LYST; AOC1*
GO:0008283(cell proliferation)	3	363	3,05E-02	*NDP; CENPF; DAZAP1*

Supplementary Table 5: Continued.

Annotation Term	O	E	q-value	*Gene List*
GO:0030154(cell differentiation)	3	441	3,66E-02	*CENPF; ACSBG2; DAZAP1*
GO:0045892(negative regulation of transcription, DNA-templated)	3	483	4,14E-02	*KDM5C; CENPF; HEXIM2*
GO:0051346(negative regulation of hydrolase activity)	2	6	5,14E-03	*LEPR; SERPINA5*
GO:0010389(regulation of G2/M transition of mitotic cell cycle)	2	11	9,76E-03	*CENPF; CCNB3*
Migraine with aura(23793025)	2	31	1,09E-02	*RPAP3; KIAA1462*
Vertical cup-disc ratio(25241763)	**2**	**37**	**1,09E-02**	***RPAP3; TRIOBP***
GO:1900026(positive regulation of substrate adhesion-dependent cell spreading)	2	28	1,09E-02	*TRIOBP; LRRC16A*
Idiopathic membranous nephropathy(21323541)	2	31	1,09E-02	*GPANK1; PRRC2A*
Schizophrenia(23894747)	2	21	1,09E-02	*PRRC2A; GRIK3*
GO:0005044(scavenger receptor activity)	2	45	1,14E-02	*SCARF1; ENDOU*
GO:0006091(generation of precursor metabolites and energy)	2	52	1,46E-02	*DLST; PHKA2*
GO:0003690(double-stranded DNA binding)	2	74	2,08E-02	*MTERF4; RBCK1*
GO:0005543(phospholipid binding)	2	78	2,15E-02	*AMPH; NBEAL2*
GO:0051087(chaperone binding)	2	81	2,28E-02	*TG; DLST*
Blood metabolite ratios(24816252)	2	82	2,32E-02	*PRRC2A; AOC1*
GO:0006897(endocytosis)	2	130	3,30E-02	*AMPH; CLEC4F*
GO:0016887(ATPase activity)	2	155	3,66E-02	*DNAH11; TNNT2*
GO:0006898(receptor-mediated endocytosis)	2	152	3,66E-02	*SCARF1; ENDOU*
GO:0008201(heparin binding)	2	154	3,66E-02	*SERPINA5; AOC1*

Supplementary Table 5: Continued.

Annotation Term	O	E	q-value	*Gene List*
GO:0051607(defense response to virus)	**2**	**157**	**3,67E-02**	***ELMOD2; LYST***
GO:0008083(growth factor activity)	2	160	3,79E-02	*NDP; ENDOU*
GO:0006644(phospholipid metabolic process)	2	176	4,09E-02	*PIP5K1B; LYST*
GO:0006805(xenobiotic metabolic process)	2	178	4,14E-02	*EPHX2; AOC1*
GO:0007507(heart development)	**2**	**184**	**4,23E-02**	***DNAH11; MTERF4***
GO:0007601(visual perception)	**2**	**197**	**4,54E-02**	***RP1L1; NDP***
Amyotrophic lateral sclerosis (sporadic)(24529757)	2	219	4,97E-02	*RP1L1; TP53I11*
Menarche (age at onset)(25231870)	2	219	4,97E-02	*LEPR; RPAP3*

Supplementary Table 6: List of enriched terms for the candidate genes for the Pantanal population: Enriched terms for the F_{ST} outlier analysis with two or more genes. Annotation Term with Gene Ontoloy and Disease terms. O: Observed, with number of genes under selection; and E: Expected, with number of genes related to the term. False Discovery rate value (FDR) > 0.05. Gene list with gene symbol.

Annotation	**O**	**E**	**q-value**	***Gene list***
GO:0005515(protein binding)	38	8670	6,63E-09	*METTL21A; UNC13B; RB1CC1; MAL; RAB32; AFG3L2; ALMS1; FEN1; CUBN; SEMA4G; ARHGAP21; KLK5; NCBP1; ANKLE2; CORO2B; SIRT3; FOCAD; TPCN2; TMEM19; SYNPO; NAA38; ACSS1; ITGA6; MARCH10; CTR9; TIAM1; ATN1; ATP2B4; ADAMTS12; TSHZ2; SPAG8; PKD1; DPM2; RNMTL1; C4orf19; AP2B1; FAM208B; LIG3*
GO:0046872(metal ion binding)	8	1870	2,91E-02	*UNC13B; ZFP1; ENPP4; MSS51; ITGA6; ATP2B4; TSHZ2; ZNF174*
GO:0008270(zinc ion binding)	6	1176	3,17E-02	*TRIM62; AFG3L2; SIRT3; MARCH10; ADAMTS12; LIG3*
GO:0005524(ATP binding)	6	1467	4,24E-02	*ANKK1; AFG3L2; EPHA1; ACSS1; ATP2B4; LIG3*
GO:0007160(cell-matrix adhesion)	5	93	8,44E-05	*NID1; ITGA6; TIAM1; ADAMTS12; PKD1*
GO:0007268(synaptic transmission)	**5**	**410**	**1,77E-02**	***UNC13B; SLC5A7; GRIK3; KCNH4; AP2B1***
GO:0005509(calcium ion binding)	5	683	2,40E-02	*CALML6; CUBN; CDH11; NID1; CLGN*
GO:0045944(positive regulation of transcription from RNA polymerase II promoter)	5	966	3,89E-02	*HEXB; ITGA6; CTR9; SPAG8; PKD1*
GO:0006996(organelle organization)	4	295	1,77E-02	*ALMS1; MRPS25; SIRT3; PKD1*
GO:0016567(protein ubiquitination)	4	363	2,10E-02	*KLHL13; TRIM62; MARCH10; WDSUB1*
GO:0019901(protein kinase binding)	4	381	2,40E-02	*RB1CC1; TPCN2; EPHA1; PKD1*
GO:0007283(spermatogenesis)	**4**	**380**	**2,40E-02**	***PIWIL4; ODF3; TEX15; ATP2B4***
GO:0055085(transmembrane transport)	4	643	4,02E-02	*MAL; SLC5A7; TPCN2; ATP2B4*
GO:0042803(protein homodimerization activity)	4	719	4,43E-02	*AADAT; HEXB; CUBN; ZNF174*
GO:0042552(myelination)	**3**	**42**	**7,46E-03**	***HEXB; MAL; AFG3L2***

Supplementary Table 6: Continued.

Annotation	O	E	q-value	*Gene list*
GO:0032092(positive regulation of protein binding)	3	57	1,41E-02	*TIAM1; SPAG8; PKD1*
GO:0048013(ephrin receptor signaling pathway)	3	94	1,77E-02	*EPHA1; TIAM1; AP2B1*
GO:0001501(skeletal system development)	**3**	**136**	**1,77E-02**	***PTHLH; HEXB; CDH11***
GO:0006874(cellular calcium ion homeostasis)	3	96	1,77E-02	*HEXB; TPCN2; ATP2B4*
GO:0004872(receptor activity)	3	197	2,40E-02	*UNC13B; TLR8; CUBN*
GO:0034220(ion transmembrane transport)	3	238	2,94E-02	*GRIK3; TPCN2; ATP2B4*
GO:0004842(ubiquitin-protein transferase activity)	3	284	3,55E-02	*KLHL13; TRIM62; WDSUB1*
GO:0003779(actin binding)	3	286	3,60E-02	*SYNPO2L; CORO2B; SYNPO*
GO:0035556(intracellular signal transduction)	3	372	4,24E-02	*STYXL1; UNC13B; AKAP11*
GO:0000278(mitotic cell cycle)	3	428	4,74E-02	*ALMS1; FEN1; ANKLE2*
GO:0007275(multicellular organismal development)	3	437	4,95E-02	*PIWIL4; ODF3; TSHZ2*
Uterine fibroids(21460842)	2	11	1,63E-02	*ODF3; SIRT3*
GO:0051492(regulation of stress fiber assembly)	**2**	**12**	**1,63E-02**	***ALMS1; SYNPO***
GO:0010506(regulation of autophagy)	2	39	1,77E-02	*RB1CC1; TPCN2*
GO:0032331(negative regulation of chondrocyte differentiation)	**2**	**17**	**1,77E-02**	***PTHLH; ADAMTS12***
Platelet counts(24026423)	2	16	1,77E-02	*ODF3; SIRT3*
GO:0007259(JAK-STAT cascade)	2	32	1,77E-02	*CTR9; PKD1*
GO:2001237(negative regulation of extrinsic apoptotic signaling pathway)	2	36	1,77E-02	*RB1CC1; ITGA6*
GO:0007528(neuromuscular junction development)	**2**	**33**	**1,77E-02**	***UNC13B; AFG3L2***

Supplementary Table 6: Continued.

Annotation	O	E	q-value	*Gene list*
GO:0048565(digestive tract development)	2	39	1,77E-02	*ITGA6; PKD1*
GO:0043588(skin development)	**2**	**38**	**1,77E-02**	***ITGA6; PKD1***
GO:0090630(activation of GTPase activity)	2	37	1,77E-02	*EPHA1; TIAM1*
GO:0007040(lysosome organization)	2	34	1,77E-02	*HEXB; TPCN2*
GO:0008138(protein tyrosine/serine/threonine phosphatase activity)	2	32	1,77E-02	*STYXL1; PTPDC1*
GO:0006284(base-excision repair)	2	48	2,40E-02	*FEN1; LIG3*
GO:0050790(regulation of catalytic activity)	2	60	2,74E-02	*ANKLE2; DPM2*
GO:0007269(neurotransmitter secretion)	2	66	2,76E-02	*UNC13B; SLC5A7*
GO:0034329(cell junction assembly)	2	78	2,99E-02	*CDH11; ITGA6*
GO:0042391(regulation of membrane potential)	2	78	2,99E-02	*GRIK3; KCNH4*
GO:0007005(mitochondrion organization)	2	74	2,99E-02	*SIRT3; LIG3*
Platelet counts(22139419)	**2**	**81**	**3,08E-02**	***FEN1; AP2B1***
GO:0008544(epidermis development)	**2**	**83**	**3,18E-02**	***PTHLH; KLK5***
Breast size(22747683)	2	91	3,50E-02	*PTHLH; ATP2B4*
GO:0001889(liver development)	2	96	3,60E-02	*RB1CC1; PKD1*
GO:0051082(unfolded protein binding)	2	101	3,74E-02	*AFG3L2; CLGN*
GO:0004222(metalloendopeptidase activity)	2	107	3,77E-02	*AFG3L2; ADAMTS12*
GO:0034765(regulation of ion transmembrane transport)	2	112	3,86E-02	*KCNH4; TPCN2*
Bone mineral density(22504420)	**2**	**114**	**3,86E-02**	***PTHLH; AKAP11***

Supplementary Table 6: Continued.

Annotation	O	E	q-value	*Gene list*
Breast cancer(23535729)	2	114	3,86E-02	*PTHLH; ITGA6*
GO:0006461(protein complex assembly)	2	115	3,90E-02	*AFG3L2; CLGN*
GO:0031047(gene silencing by RNA)	2	116	3,94E-02	*PIWIL4; NCBP1*
GO:0007417(central nervous system development)	**2**	**118**	**4,02E-02**	***MAL; ATN1***
GO:0070588(calcium ion transmembrane transport)	2	129	4,20E-02	*ATP2B4; PKD1*
GO:0000724(double-strand break repair via homologous recombination)	2	141	4,24E-02	*FEN1; LIG3*
GO:0006302(double-strand break repair)	2	151	4,52E-02	*FEN1; LIG3*
GO:0007156(homophilic cell adhesion via plasma membrane adhesion molecules)	2	154	4,57E-02	*CDH11; PKD1*
GO:0006886(intracellular protein transport)	2	168	4,94E-02	*EXPH5; AP2B1*

Supplementary Table 7: Allele Frequencies of candidate loci mentioned in the discussion – Genes associated with body size, reproduction, pigmentation and righting reflex. Each gene contains the position with the outlier F_{ST} and its frequency on each analyzed population. Ancestral alleles are indicated in bold.

Trait	Gene	Position	Alleles	Populations		
				Pantanal	Amazon	Atlantic Forest
Body size	*KLF3*	509	**G**	0.38	0.72	0.66
			T	0.62	0.28	0.34
	LEPR	2800	**A**	0.98	0.68	0.97
			G	0.02	0.32	0.03
	CTR9	3354	**G**	0.96	0.68	0.92
			A	0.04	0.32	0.08
		4122	**G**	0.94	0.75	0.92
			A	0.06	0.25	0.08
	PTHLH	16	**C**	0.97	0.66	0.66
			T	0.03	0.34	0.34
	VRTN	1329	**G**	0.95	0.6	0.55
			A	0.05	0.4	0.45
		1445	**C**	0.93	0.56	0.55
			T	0.07	0.44	0.45
	HEXB	1548	**C**	0.98	0.78	0.66
			G	0.02	0.22	0.34
	CENPF	8917	**A**	0.96	0.56	0.95
			G	0.04	0.44	0.05
Reproduction	*TEX15*	1509	T	0.98	0.625	0.6
			C	0.02	0.375	0.4
	ODF3L2	1178	**T**	0.37	0.72	0.68
			C	0.63	0.28	0.32
Pigmentation	*RAB32*	864	**C**	0.99	0.72	0.95
			T	0.01	0.28	0.05
	TPCN2	2509	**G**	1	0.84	0.76
			A	0	0.16	0.24
	ATRN	424	**G**	0.88	0.69	0.74
			A	0.12	0.31	0.26
	LYST	2765	**T**	0.98	0.66	0.97
			C	0.02	0.44	0.03
Righting reflex	*AF3GL2*	1896	**C**	0	0.16	0.18
			T	1	0.84	0.82

CAPÍTULO IV | **DISCUSSÃO GERAL**

1. O genoma da onça-pintada e comparações com outros felinos

O presente projeto foi a primeira iniciativa em direção à caracterização do genoma da onça-pintada. O sequenciamento do genoma completo de uma espécie é um processo complexo e contínuo, que se beneficia da constante criação de novas tecnologias e métodos analíticos para a sua descrição. Essa primeira etapa auxiliou a identificar genes envolvidos em alguns processos esperados para os grandes felinos (tamanho corporal, por exemplo). Ao mesmo tempo, novas questões foram levantadas e serão discutidas neste capítulo.

A partir do sequenciamento de quatro bibliotecas genômicas, duas delas com tamanho de inserto de 180pb e demais de 3kb e 8kb, obtivemos um total de 2.606.456.270 *reads* (segmentos curtos de DNA, base para montagem de fragmentos maiores). Com tamanho estimado do genoma de 2.4 Gb, os dados gerados forneceram, após passar por filtros de qualidade, uma cobertura média de aproximadamente 84x. A montagem realizada utilizando o software Allpaths-LG (Butler *et al.* 2008) gerou por fim 7521 *scaffolds* com N50 de 1,52Mb. O sequenciamento do genoma da onça-pintada já se encontra atualmente em sua segunda fase seguinte, com o objetivo de aumentar a contiguidade do genoma e iniciar analises de arquitetura genômica e variação intraespecífica, com a adição de novos indivíduos.

Apesar do número pequeno de espécies no gênero, *Panthera* sempre foi alvo de um intenso debate sobre as relações filogenéticas de suas espécies. Trabalhos anteriores discutiram a questão à exaustão e apenas recentemente foi possível ter uma indicação dos processos que levaram a isso (Johnson *et al.* 2006; Davis *et al.* 2010; Li *et al.* 2016). O presente trabalho avança nessa direção e demonstra, através de análises de genômica comparativa, a complexa relação entre as espécies do gênero. A origem das linhagens atuais de *Panthera* aconteceu há pelo menos 4 milhões de anos e, devido à grande mobilidade das espécies do gênero e grandes áreas de simpatria em sua distribuição histórica, ocorreram de diversos contatos secundários entre as espécies do gênero, além da segregação incompleta de linhagens devido a rápidos episódios sequenciais de especiação. Estes eventos foram cruciais durante o processo de especiação no grupo e moldaram as espécies aos diferentes tipos de habitat que elas ocupam atualmente.

Nesse sentido, com as análises filogenéticas realizadas ao longo do genoma, foi possível observar diferentes topologias de árvore. Essa variação demonstra a ocorrência de eventos de hibridação, com grandes blocos genômicos mostrando uma filogenia diferente daquela reconhecida como 'species tree' para o gênero. Notadamente, o leão é a espécie com o maior número de trocas de posição, sendo que a segunda e a terceira topologias de árvore

:

mais frequentes envolvem mudanças nas relações filogenéticas com a espécie e a onça-pintada e sendo grupo externo do clado onça-pintada + leopardo. Outro ponto importante foi a observação de um forte sinal de mistura (*admixture*) em análises que buscam quantificar o fluxo gênico entre as espécies do grupo. São resultados como este que ajudam a explicar a grande dificuldade para determinação de uma filogenia precisa para o grupo, além de servir de reflexo da biogeografia das espécies do gênero, relativamente bem conhecida através de registros fósseis e arqueológicos.

No entanto, além do efeito dos processos demográficos, a seleção foi um fator importante para moldar as espécies do gênero. Com base na razão entre mutações sinônimas e não sinônimas, através da aplicação de análises estatísticas para testar sítios sob seleção, foi observado um conjunto de genes importante que contribuíram na evolução da forma, comportamento e fisiologia do grupo. Além do método tradicional de análise de seleção, que utiliza a árvore da espécie como padrão explicativo, também aplicamos uma abordagem inovadora, que utiliza a árvore de cada gene para o cálculo de verossimilhança. Como exposto acima, as relações evolutivas do gênero ainda são alvo de intenso debate e nosso resultado mostrou um alto grau de discordância filogenética ao longo do genoma. Por esse motivo, a utilização da árvore da espécie poderia gerar uma grande quantidade falso positivos. Com a utilização da árvore de gene, podemos observar a flutuação das relações filogenéticas ao longo do genoma, refletindo um cenário mais realista. Até o momento, esse é o primeiro estudo a realizar tal comparação.

Os genes encontrados auxiliam na compreensão de características que contribuíram para formar as espécies como as conhecemos hoje, como um maior tamanho corporal em relação aos outros felídeos e um conjunto diverso de receptores olfativos. Esta lista de genes pode servir como ponto de partida para estudos mais detalhados, na caracterização evolutiva de famílias gênicas bem como na descrição funcional e associação direta com o fenótipo dos mesmos. Além de uma análise de seleção global para o gênero, também foram realizadas análises de seleção espécie-especificas. Essas análises trouxeram resultados interessantes e possibilitaram a identificação de sinais de seleção em características marcantes de cada espécie.

A onça-pintada é um animal robusto, com membros curtos, musculatura marcante e maxilares massivos. Apesar da grande semelhança com o leopardo no que tange à pelagem, o porte físico é muito mais semelhante ao leão e tigre, sendo a terceira maior espécie do gênero. A detecção de assinaturas de seleção em genes relacionados com desenvolvimento craniofacial e muscular na espécie provavelmente se relaciona a essas características. Dois dos genes apresentando assinaturas de seleção positiva nessa espécie (*ESRP1* e *SSTR4*) estão associados à formação da face, mais especificamente da região da mandíbula. *ESRP1* é um fator de

:

transcrição que regula diversas rotas metabólicas que são chave para o desenvolvimento, e desempenha um papel fundamental na formação da face. *SSTR4* é um receptor com diferentes níveis de expressão ao longo do desenvolvimento e também está relacionado com crescimento corporal e desenvolvimento da face. A seleção desses genes pode estar associada ao tipo de presa e comportamento de caça da espécie. Além disso, ainda em relação ao porte da espécie, foram encontrados genes associados à musculatura dos membros, que provavelmente contribuem para o fenótipo robusto da onça-pintada.

O leopardo das neves é uma espécie altamente associada a ambientes frios e altitudes elevadas (Sunquist & Sunquist 2002). Dentre o conjunto de genes apresentando assinaturas de seleção nessa espécie, foram observados dois loci associados com apneia, que provavelmente estão associados a hipóxia. Além disso, foram encontrados outros genes, associados com lactação e metabolismo de lipídios e também diretamente relacionado com o desenvolvimento das narinas, todas potenciais adaptações a ambientes frios. A espécie é uma das mais ameaçadas do gênero, pois já ocorre normalmente em baixas densidades. Além da perda de habitat devido à ocupação humana, retaliação e caça, um dos fatores que mais poderá prejudicar a espécie no futuro são as alterações climáticas. Estimativas recentes preveem uma perda de 30% do habitat da espécie nos Himalaias, provocando o isolamento de algumas populações causada pelas mudanças climáticas (Forrest *et al.* 2012). Essa informação, associada ao alto nível de especialização da espécie, podem ser fatores determinantes para uma redução no tamanho populacional nos próximos anos.

O leão é a única espécie da família Felidae que possui comportamento social. Um dos genes encontrados neste estudo como estando sob seleção positiva nesta espécie, *GRB10*, está associado com hierarquia social e comportamento dominante em roedores (Garfield et al, 2014). Esse é o primeiro registro de um gene candidato associado ao comportamento social em leões. Baseado nas características observadas em roedores, podemos postular este gene possa estar sendo expresso diferencialmente em machos dominantes do grupo. Outra característica única da espécie, a presença de juba, não apresentou relação clara com qualquer dos genes sob seleção na espécie. Isso não significa que outros genes não amostrados não possam estar relacionados com o fenótipo, uma vez que essa característica sofre seleção sexual.

Já o tigre é a única espécie da família Felidae que possui listras no corpo. Acredita-se que sua coloração peculiar esteja associada à camuflagem (Mazák 1981). No presente trabalho, foram encontrados diversos genes associado à melanogênese e possivelmente à formação das listras. As bases genéticas da delimitação das listras em felinos ainda são desconhecidas. Trabalhos com animais-modelo já exploraram essa questão e demonstraram que genes envolvidos na rota melanogênica também fazem parte de rotas importantes de

:

desenvolvimento corporal (Singh & Nüsslein-Volhard 2015). Com base nesse dado, é possível estabelecer uma relação entre a formação do padrão de listras com o desenvolvimento corporal. De qualquer forma, análises mais detalhadas serão necessárias para estabelecer a correta relação desses genes.

Esse é o primeiro estudo que se propôs a analisar o gênero *Panthera* em nível genômico. Os resultados aqui expostos poderão servir de base para diversos trabalhos funcionais com o objetivo de verificar o impacto dos genes candidatos no fenótipo das espécies. Visto isso, se sugere a realização de mais estudos comparativos com o grupo, especialmente em áreas como morfometria e fisiologia comparadas.

2. Genômica populacional

Este foi o primeiro estudo que teve como objetivo avaliar em escala genômica a estrutura genética de populações de onça-pintada em diferentes biomas. Com a utilização de técnicas de captura de exoma e sequenciamento de alto desempenho, foi possível sequenciar um volume de dados impossível há cerca de cinco anos para uma espécie não-modelo sem recursos genômicos previamente estabelecidos. Juntamente com o genoma completo da espécie, este é um recurso importantíssimo para a geração de conhecimentos aprofundados sobre a espécie. Além de informações atualizadas sobre os níveis atuais e históricos de conectividade entre populações de onça-pintada no Brasil, foi possível pela primeira vez identificar genes apresentando assinaturas de seleção natural indicando evidência de adaptação local. Outros aspectos ainda poderão ser explorados no conjunto de dados gerados neste estudo, pois além da onça-pintada foi sequenciado o exoma de outras cinco espécies (*P. leo*, *P. pardus*, *P. tigris*, *P. uncia* e *Neofelis nebulosa*), assim como o genoma mitocondrial dos 110 indivíduos de onça-pintada e 30kb adjacentes ao gene *MC1R*. Análises preliminares indicam que a eficácia de captura para essas espécies foi o mesmo obtido para os indivíduos de onça-pintada. Esse resultado indica a possiblidade de utilização do ensaio em outras espécies de Felidae com eficácia semelhante ao observado no presente estudo.

O nível de diversidade e diferenciação entre as populações analisadas está dentro do esperado. Populações do Pantanal e Amazônia apresentam alta diversidade genética. Mesmo populações da Mata Atlântica apresentaram valores de diversidade acima do esperado tendo em vista o grau de ameaça da espécie na região. Isso pode ter sido causado pela inclusão de amostras mais antigas em nosso estudo, obtidas em períodos onde havia maior conectividade entre os biomas, especialmente entre a Mata Atlântica e o Pantanal Sul. (Haag *et al.* 2010; Galetti

:

et al. 2013). No entanto, a análise de componentes principais (Capítulo 3 Figura 4A) mostra uma ampla área separando as populações da Mata Atlântica. Uma interpretação possível para esse resultado é a grande diferenciação entre as populações de onça-pintada no bioma, já observada em resultados anteriores (Haag *et al.* 2010). No futuro, análises demográficas mais robustas deverão ser adicionadas para avaliar esse cenário. Outro resultado importante foi a inclusão de indivíduos localizados no Cerrado e Catinga, biomas em que as populações de onça foram pouco estudadas até o momento (Paula *et al.* 2013), especialmente do ponto de vista genético. Os resultados obtidos aqui são semelhantes a resultados anteriores para a Caatinga (ref.). O alto nível de diferenciação em relação às outras populações (valores de F_{ST} entre 0,07 e 0,11) e baixo nível de heterozigosidade observada (0,34) são preocupantes. Um potencial problema é o número amostral díspar, que pode ter influenciado alguns resultados, sobretudo no que tange à estruturação populacional. Já o Cerrado parece ter sido um importante corredor entre populações do norte e sul, conectando a Amazônia e a Mata Atlântica historicamente. Nas análises de ancestralidade, o bioma mostra grande semelhança com a Amazônia e Caatinga. Essa é a primeira vez que se estudam populações do bioma, e assim como na Caatinga, um maior número amostral é necessário para que se atinja uma avaliação mais completa de suas populações.

Foi possível demonstrar a presença de genes sob seleção associados a alterações no tamanho corporal. Todas as populações apresentaram um conjunto de genes relacionados a rotas de desenvolvimento, altura e peso corporal, bem como metabolismo energético. Esse resultado está de acordo com hipóteses prévias, que relacionavam a variação do tamanho da espécie com a diversidade de presas e quantidade de biomassa disponível (Hoogesteijn & Mondolfi 1996). A Amazônia apresentou diversos genes associados a metabolismo energético, em especial metabolismo de carboidratos e uso e manutenção de reservas lipídicas. De acordo com relatos de pesquisadores de campo, onças na Amazônia pesam entre 50 e 70kg, enquanto animais do Pantanal podem chegar a pesar 120kg, com alguns relatos de animais pesando até 150kg. Análises morfométricas criteriosas nunca foram realizadas para quantificar essa variação. O único trabalho nesse sentido analisa um número pequeno de indivíduos e deve ser avaliado com parcimônia (Hoogesteijn & Mondolfi 1996). As outras populações também apresentaram genes sob seleção relacionados com desenvolvimento. Através de análises de enriquecimento, foi possível observar no Pantanal genes associados ao desenvolvimento do sistema ósseo, sistema circulatório e pele. Uma hipótese plausível para a ocorrência de seleção afetando essa característica é que essas rotas foram selecionadas para comportar o maior tamanho corporal. No entanto, com base nos dados encontrados, não é possível inferir a direcionalidade dos eventos de seleção, e novas análises com esse objetivo deverão ser adicionadas.

:

Dentre os resultados mais interessantes está a identificação de genes relacionados à pigmentação, os quais podem estar relacionados à ausência de indivíduos melânicos no Pantanal. O mecanismo que provoca o melanismo em onças-pintadas é bem conhecido e já foi aplicado em estudos populacionais (Haag *et al.* 2010):. No entanto, apenas agora começa a se investigar os mecanismos que levam certas regiões de ocorrência da espécie a não conterem indivíduos melânicos (Silva 2014). Pesquisadores de campo relatam uma pelagem mais clara em regiões abertas como o Pantanal, mas este dado nunca foi quantificado, de forma que estas observações devem no momento serem tratadas apenas como uma hipótese a ser testada. Ainda assim, a descoberta de genes sob seleção diferencial com provável envolvimento em rotas de pigmentação abre caminho para investigações detalhadas sobre os aspectos adaptativos deste fenótipo em diferentes populações de onça-pintada.

Além disso, foi possível observar uma grande quantidade de genes sob seleção que estão associados à imunidade, inclusive um gene diretamente ligado à Leishmaniose. Doenças tropicais encontram-se no grupo de doenças mais negligenciadas (Hotez *et al.* 2008), o que ocorre devido a diversos fatores, principalmente econômicos. Entender processos relacionados à doença em outras espécies poderá auxiliar na sua investigação do ponto de vista biomédico. A Mata Atlântica apresentou diversos genes sob seleção ligados à imunidade. Esses resultados devem ser tratados com cautela, devido ao alto grau de estruturação interna no bioma (Haag *et al.* 2010) e a influência que processos demográficos tem no método aplicado. Apesar disso, foram encontrados diversos genes associado ao sistema imune inato, responsável pela resposta imune inespecífica contra patógenos. Esse resultado é reforçado pela presença de um gene sob seleção encontrado em uma das análises (que busca corrigir efeitos demográficos) que confere proteção contra agentes externos na pele, e apresentou uma frequência diferencial em populações da Mata Atlântica.

A utilização de dados fenotípicos em genética de populações de grupos não-modelo não é algo usual. A principal razão é a dificuldade de coletar esse tipo de informação, mas também, a falta de comunicação entre pesquisadores de campo e geneticistas. Com o advento do sequenciamento de alto desempenho e a relativa facilidade de obtenção de dados, o grande desafio é tornar esse grande volume de informação também informativo em termos de interpretações relacionadas a fenótipos. A associação de dados fenotípicos e genéticos é usual em genética humana, uma tendência que deve ser adotada em outros grupos. Apenas dessa forma será possível realizar inferências mais robustas e detalhadas. Trabalhos com o objetivo de buscar assinaturas de seleção natural ao longo do genoma, geralmente tem êxito, encontrando um número considerável de genes afetados. O principal problema é encontrar sentido biológico para essas descobertas, e o maior uso de dados fenotípicos atende a essa

:

demanda. Por este motivo, sugere-se que ocorra a padronização e sistematização na obtenção de dados morfométricos e fisiológicos para estudos de associação em felinos.

3. Análise Global das Assinaturas de Seleção na Onça-pintada

Um dos principais temas abordados no projeto foi a detecção de assinaturas de seleção natural em diferentes níveis taxonômicos. Trata-se do primeiro trabalho com o objetivo de identificar genes sob seleção no gênero *Panthera* e em populações de onça-pintada em uma escala genômica. O presente estudo identificou genes de relevância para características que auxiliaram a moldar as espécies do gênero como as conhecemos atualmente. Do ponto de vista populacional, esta foi a primeira tentativa de responder questões sobre a variação fenotípica encontrada na onça-pintada de um ponto de vista seletivo, e não apenas demográfico. A adição desses resultados permitiu avanços no esclarecimento de questões sobre a evolução da forma e função do grupo, bem como abrir diferentes linhas de investigação.

A onça-pintada é um animal robusto, com membros curtos, musculatura marcante e maxilares massivos. Apesar da grande semelhança com o leopardo no que tange à pelagem, o porte físico é muito mais semelhante ao leão e tigre, sendo a terceira maior espécie do gênero. Os capítulos II e III da tese tinham com proposta avaliar e caracterizar assinaturas de seleção na espécie em diferentes níveis. O primeiro artigo realizou uma análise comparativa com os outros *Panthera*, onde buscou-se padrões mais gerais da espécie. O segundo artigo buscou assinaturas de seleção que explicassem a variação interpopulacional da onça-pintada. Os resultados encontrados indicam um forte regime de seleção, em ambos os casos relacionado a pressões exercidas pela dieta da espécie. Em uma escala interespecífica, genes sob seleção moldaram características do crânio e musculatura, componentes importantes que ajudam a definir o comportamento de caça do animal. Já em nível intraespecífico, o regime de seleção parece ser regulado pela disponibilidade de presas em cada ambiente.

Durante o final do Pleistoceno, com a extinção da megafauna no continente Americano, os carnívoros foram enormemente afetados. Um dos processos mais comuns foi a redução do tamanho corporal de predadores ou a extinção de espécies maiores, como por exemplo, o coiote (*Canis latrans*) (Meachen & Samuels 2012). Na onça-pintada, a redução de tamanho corporal foi de cerca de 15-20%, com alterações adicionais como redução do tamanho dos membros (Kurtén 1973). Os genes sob seleção encontrados nas análises interespecíficas parecem se relacionar a esse processo de redução de tamanho (musculatura dos membros) e adaptação novos tipos de presa, em particular répteis com tegumento resistente (desenvolvimento craniofacial). Esta adaptação da onça-pintada a uma dieta baseada em

:

répteis encouraçados havia sido postulada anteriormente(Emmons 1987), e encontra apoio em nossos resultados. Além disso, os genes candidatos indicam um segundo momento de adaptação fenotípica, refletindo os diferentes ambientes que a espécie ocupa atualmente. Essa variação também seria mediada pela dieta. Em ambientes de floresta, como a Amazônia, o tamanho médio das presas consumidas pela onça-pintada é de ± 15kg, enquanto em regiões do Pantanal o peso das presas pode passar dos 60kg (de Oliveira 2001). Ambos os resultados mostram a grande resiliência e capacidade de adaptação da espécie. As evidencias encontradas no presente trabalho indicam estruturação mediada pela adaptação local aos biomas que a espécie ocupa. Novos estudos, incorporando tanto abordagens genômicas quanto conjuntos de dados mais completos de aspectos fenotípicos, devem ser realizados no futuro para viabilizar uma compreensão mais detalhada dos processos históricos e atuais que influenciam a contínua adaptação da onça-pintada aos seus ambientes.

PERSPECTIVAS

O projeto Genoma da Onça-Pintada se iniciou em 2011 com o intuito de gerar recursos genômicos viabilizando análises em grande escala desta espécie. O arcabouço de informações obtidas durante o desenvolvimento deste projeto foi essencial para sua execução e irá auxiliar no delineamento de etapas futuras desta iniciativa. O Projeto Genoma está atualmente no início da segunda fase de caracterização em grande escala, em que estão adicionadas novas bibliotecas do indivíduo-referência, como cobertura adicional de sequenciamento e tecnologias de montagem que visam a aumentar a contiguidade do genoma. Há também a perspectiva de adição de novos indivíduos com genoma completo sequenciado, e análises especificas como a caracterização do cromossomo Y. Com relação à genômica de populações, serão adicionadas novas amostras e será iniciado o desenvolvimento de ferramentas para a aplicação em amostras não invasivas. Além disso, as técnicas desenvolvidas em ambos os artigos componentes desta tese poderão ser aplicadas em um futuro próximo em uma diversa gama de espécies cujos genomas estão atualmente sendo investigados.

:

REFERÊNCIAS

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome research*, **12**, 1805–14.

Amadio A (2009) The genome sequence of taurine cattle. a window to ruminant biology and evolution. *Science*, **324**.

Amemiya CT, Alföldi J, Lee AP *et al.* (2013) The African coelacanth genome provides insights into tetrapod evolution. *Nature*, **496**, 311–6.

Barnett R, Shapiro B, Barnes I *et al.* (2009) Phylogeography of lions (Panthera leo ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology*, **18**, 1668–1677.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Bhagavatula J, Singh L (2006) Genotyping faecal samples of Bengal tiger Panthera tigris tigris for population estimation: a pilot study. *BMC genetics*, **7**, 48.

Bi K, Linderoth T, Vanderpool D *et al.* (2013) Unlocking the vault: Next-generation museum population genomics. *Molecular Ecology*, **22**, 6018–6032.

Bi K, Vanderpool D, Singhal S *et al.* (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, **13**, 403.

Burger J, Rosendahl W, Loreille O *et al.* (2004) Molecular phylogeny of the extinct cave lion Panthera leo spelaea. *Molecular Phylogenetics and Evolution*, **30**, 841–849.

Busby GBJ, Gottelli D, Wacher T *et al.* (2009) Genetic analysis of scat reveals leopard Panthera pardus and cheetah Acinonyx jubatus in southern Algeria. *Oryx*, **43**, 412.

Butler J, MacCallum I, Kleber M *et al.* (2008) ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome research*, **18**, 810–20.

Cho YS, Hu L, Hou H *et al.* (2013) The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature communications*, **4**, 2433.

Christiansen P (2008) Cladistics Phylogeny of the great cats (Felidae : Pantherinae), and the influence of fossil taxa and missing characters. , **24**, 977–992.

Cosart T, Beja-Pereira A, Chen S *et al.* (2011) Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC genomics*, **12**, 347.

Davis BW, Li G, Murphy WJ (2010) Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, Panthera (Carnivora: Felidae). *Molecular Phylogenetics and Evolution*, **56**, 64–76.

Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. *Trends in Genetics*, **30**.

Driscoll CA, Yamaguchi N, Bar-Gal GK *et al.* (2009) Mitochondrial phylogeography illuminates the origin of the extinct caspian tiger and its relationship to the amur tiger. *PLoS ONE*, **4**, e4125.

Eisenberg JF, Redford KH (1999) *The Central Neotropics: Ecuador, Peru, Bolivia, Brazil*.

Eizirik E, Kim JH, Menotti-Raymond M *et al.* (2001) Phylogeography, population history and conservation genetics of jaguars (Panthera onca, Mammalia, Felidae). *Molecular ecology*, **10**, 65–79.

Eizirik E, Yuhki N, Johnson WEW *et al.* (2003) Molecular genetics and evolution of melanism in the cat family. *Current Biology*, **13**, 448–453.

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, **7**, 1026–1042.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.

Emmons LH (1987) feeding ecology of felids Comparative in a neotropical rainforest. *Behavioral Ecology and Sociobiology*, **20**, 271–283.

Emmons L (1989) Jaguar predation on chelonians. *Journal of Herpetology*, **23**, 311–314.

Forrest JL, Wikramanayake E, Shrestha R *et al.* (2012) Conservation and climate change: Assessing the vulnerability of snow leopard habitat to treeline shift in the Himalaya. *Biological Conservation*, **150**, 129–135.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Funston PJ, Mills MGL, Biggs HC, Richardson PRK (1998) Hunting by male lions: ecological influences and socioecological implications. *Animal Behaviour*, **56**, 1333–1345.

Galetti M, Eizirik E, Beisiegel B *et al.* (2013) Atlantic Rainforest's Jaguars in Decline. *Science*, **342**, 930.

Green R, Krause J, Briggs A *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, **328**, 710.

Haag T, Santos a S, Sana D a *et al.* (2010) The effect of habitat fragmentation on the genetic structure of a top predator: loss of diversity and high differentiation among remnant populations of Atlantic Forest jaguars (Panthera onca). *Molecular ecology*, **19**, 4906–21.

Haile J, Froese DG, MacPhee RDE *et al.* (2008) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 22352–22357.

Hast MH (1989) The larynx of roaring and non-roaring cats. *Journal of anatomy*, **163**, 117–121.

Heidtmann LM (2014) Caracterização do genoma mitocondrial de onça-pintada (Panthera onca) e elucidação da filogenia mitogenômica do gênero Panthera. Porto Alegre.

Hemmer H, Kahlke R-D, Vekua AK (2010) <I>Panthera onca georgica</I> ssp. nov. from the Early Pleistocene of Dmanisi (Republic of Georgia) and the phylogeography of jaguars (Mammalia, Carnivora, Felidae). *Neues Jahrbuch für Geologie und Paläontologie - Abhandlungen*, **257**, 115–127.

Henry P, Miquelle D, Sugimoto T *et al.* (2009) In situ population structure and ex situ representation of the endangered Amur tiger. *Molecular Ecology*, **18**, 3173–3184.

Hiby L, Lovell P, Patil N *et al.* (2009) A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins. *Biology letters*, **5**, 383–386.

Hoogesteijn R, Mondolfi E (1996) Body mass and skull measurements in four jaguar populations and observations on their prey base. *Bulletin of the Florida Museum of Natural History*, **39**, 195–219.

Hotez PJ, Bottazzi ME, Franco-Paredes C, Ault SK, Periago MR (2008) The neglected tropical diseases of Latin America and the Caribbean: A review of disease burden and distribution and a roadmap for control and elimination. *PLoS Neglected Tropical Diseases*, **2**.

IUCN (2015) IUCN Red List of Threatened Species. *Version 2015.3*, www.iucnredlist.org.

Janecka JE, Jackson R, Yuquang Z *et al.* (2008) Population monitoring of snow leopards using noninvasive collection of scat samples: A pilot study. *Animal Conservation*, **11**, 401–411.

Johnson WE, Eizirik E, Pecon-Slattery J *et al.* (2006) The late Miocene radiation of modern Felidae: a genetic assessment. *Science*, **311**, 73–77.

Jones MR, Good JM (2015) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, n/a–n/a.

Kawanishi K, Sunquist ME, Eizirik E *et al.* (2010) Near fixation of melanism in leopards of the Malay Peninsula. *Journal of Zoology*, **282**, 201–206.

Kim J, Antunes A, Luo S *et al.* (2006) Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. *Gene*, **366**, 292–302.

Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. *PLoS Genetics*, **4**.

Kurtén B (1973) *Pleistocene jaguars in North America*. Societas Scientiarum Fennica.

Larson SE (1997) Taxonomic Re-Evaluation of the Jaguar. *ZooBiology*, **120**, 107–120.

Li G, Davis BW, Eizirik E, Murphy WJ (2016) Pervasive signals of ancient hybridization in the genomes of living cats (Felidae). *Genome Research*, 1–11.

Li R, Fan W, Tian G *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**, 311–317.

Lindblad-Toh K, Wade CM, Mikkelsen TS *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–19.

Locke DP, Hillier LW, Warren WC *et al.* (2011) Comparative and demographic analysis of orang-utan genomes. (KJ Gutzwiller, Ed,). *Nature*, **469**, 529–533.

Lohmueller KE, Sparsø T, Li Q *et al.* (2013) Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *American Journal of Human Genetics*, **93**, 1072–1086.

Lopez J V., Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, **39**, 174–190.

Mazak (1981) Panthera tigris. *Mammalian Species*, **152**, 1–8.

Mcmahon BJ, Teeling EC, Höglund J (2014) How and why should we implement genomics into conservation? *Evolutionary Applications*, **7**, 999–1007.

Meachen J a., Samuels JX (2012) Evolution in coyotes (Canis latrans) in response to the megafaunal extinctions. *Proceedings of the National Academy of Sciences*, **109**, 11–16.

Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics*, **11**, 31–46.

Meyer M et al (2012) A high coverage genome sequence from an archaic Denisovan Individual. *Science*, **338**, 222–226.

Miththapala S, Seidensticker J, O'Brien SJ (1996) Phylogeographic subspecies recognition in leopards (Panthera pardus): Molecular genetic variation. *Conservation Biology*, **10**, 1115–1132.

Moraes E (2012) The status of the jaguar in the Cerrado. *CAT News*, 25–28.

Nelson AEW, Goldman EA, Nelson EW, Goldman A (1933) Revision of the Jaguars. *Journal of Mammalogy*, **14**, 221–240.

Nielsen R (2005) Molecular signatures of natural selection. *Annual review of genetics*, **39**, 197–218.

Nielsen R, Bustamante C, Clark AG *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS biology*, **3**, e170.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE*, **7**.

Nowell K, Jackson P (1996) *Wild cats. Status Survey and Conservation Action Plan.*

Oleksyk TK, Smith MW, Brien SJO, B PTRS (2010) Genome-wide scans for footprints of natural selection Genome-wide scans for footprints of natural selection. , 185–205.

de Oliveira TG (2001) Comparative feeding ecology of jaguar (Panthera onca) and puma (Puma concolor) in the neotropics. *Jaguars in the new millennium: a status assessment, priority detection, and recommendations for the conservation of jaguars in the Americas*.

Paula RC, Desbiez A, Cavalcanti SMC (2013) *Plano de açao nacional para a conservaçao da onça pintada*.

Quail M a, Smith M, Coupland P *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics*, **13**, 341.

Sakamoto M, Ruta M (2012) Convergence and divergence in the evolution of cat skulls: Temporal and spatial patterns of morphological diversity. *PLoS ONE*, **7**, e39752.

Der Sarkissian C, Ermini L, Schubert M *et al.* (2015) Evolutionary genomics and conservation of the endangered Przewalski's horse. *Current Biology*, **25**, 2577–2583.

Schneider A, David V a, Johnson WE *et al.* (2012) How the Leopard Hides Its Spots: ASIP Mutations and Melanism in Wild Cats. *PloS one*, **7**, e50386.

Schraiber JG, Akey JM (2015) Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics*.

Seymour KL (1989) Panthera onca. *Mammalian Species*, **11**, 1–9.

Sheldon JW, Reed G, Burnett AC, Li K, Crabtree RL (2009) Coyote, Canis latrans, predation on a Bison, Bison bison, calf in Yellowstone National Park. *Canadian Field-Naturalist*, **123**, 260–261.

Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135–1145.

Singh AP, Nüsslein-Volhard C (2015) Zebrafish stripes as a model for vertebrate colour pattern formation. *Current Biology*, **25**, R81–R92.

Silva LG da (2014) Análise da distribuição espacial do melanismo na família felidae em função de condicionantes ambientais. Porto Alegre.

Spong G, Stone J, Creel S, Björklund M (2002) Genetic structure of lions (Panthera leo L.) in the Selous game reserve: Implications for the evolution of sociality. *Journal of Evolutionary Biology*, **15**, 945–953.

Stander PE (1992) Cooperative hunting in lions: the role of the individual. *Behavioral Ecology and Sociobiology*, **29**, 445–454.

Stuart AJ, Lister AM (2011) Extinction chronology of the cave lion Panthera spelaea. *Quaternary Science Reviews*.

Sunquist M, Sunquist F (2002) *Wild cats of the world* (CAW Guggisberg, Ed,). University of Chicago Press.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tennessen JA, O'Connor TD, Bamshad MJ, Akey JM (2011) The promise and limitations of population exomics for human evolution studies. *Genome Biology*, **12**, 127.

Tseng ZJ, Wang X, Slater GJ *et al.* (2014) Himalayan fossils of the oldest known pantherine establish ancient origin of big cats. *Proc Biol Sci*, **281**, 20132686.

Uphyrkina O, Johnson WE, Quigley H *et al.* (2001) Phylogenetics, genome diversity and origin of modern leopard, Panthera pardus. *Molecular Ecology*, **10**, 2617–2633.

Uphyrkina O, Miquelle D, Quigley H, Driscoll C, O'Brien SJ (2002) Conservation genetics of the Far Eastern leopard (Panthera pardus orientalis). *Journal of Heredity*, **93**, 303–311.

Valdez FP, Haag T, Azevedo FCC *et al.* (2015) Population genetics of Jaguars (*Panthera onca*) in the Brazilian Pantanal: Molecular evidence for demographic connectivity on a regional scale. *Journal of Heredity*, **106**, 503–511.

Wade CM, Giulotto E, Sigurdsson S *et al.* (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science (New York, N.Y.)*, **326**, 865–7.

Weckel M, Giuliano W, Silver S (2006) Jaguar (Panthera onca) feeding ecology: distribution of predator and prey through time and space. *Journal of Zoology*, **270**, 060606025751038–???

Wei L, Wu X, Jiang Z (2009) The complete mitochondrial genome structure of snow leopard Panthera uncia. *Molecular Biology Reports*, **36**, 871–878.

Weissengruber GE, Forstenpointner G, Peters G, Kubber Heiss A, Fitch WT (2002) Hyoid apparatus and pharynx in the lion (Panthera leo), jaguar (Panthera onca), tiger (Panthera tigris), cheetah (Acinonyx jubatus) and domestic cat (Felis silvestris f. catus). *Journal of Anatomy*, **201**, 195–209.

West PM, Packer C (2002) Sexual selection, temperature, and the lion's mane. *Science*, **297**, 1339–1343.

Yang Z (2001) Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. , 908–917.

Yi X, Liang Y, Huerta-Sanchez E *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science (New York, N.Y.)*, **329**, 75–8.

Zhang G, Li C, Li Q *et al.* (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science (New York, N.Y.)*, **346**, 1311–20.