



ESALQ

Variantes genéticas e sua importância na ciência moderna

LGN5809 – Genética Molecular

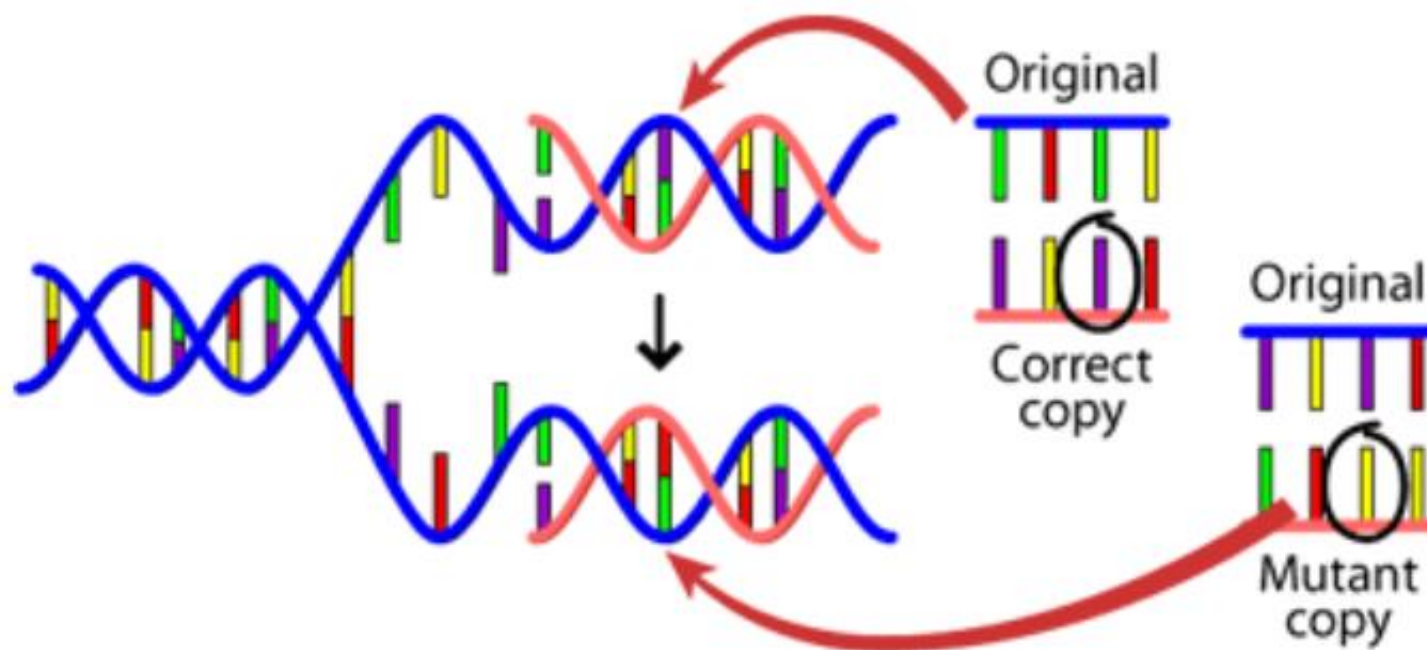
LUAN CLEMENTE

LUIZ FIRMINO

MARCELO SANTOS

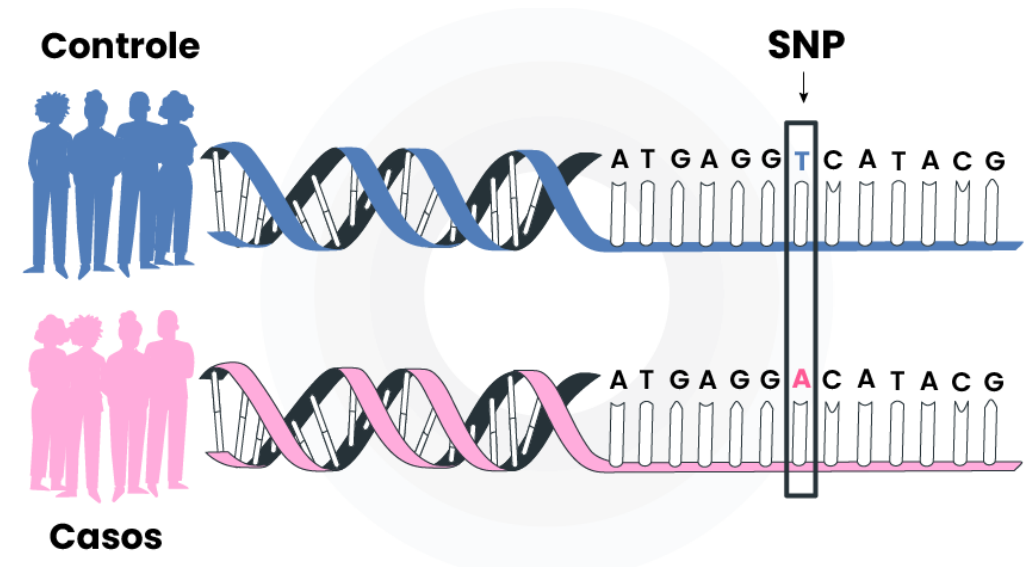
2023

Variantes genéticas e sua importância na ciência moderna



GWAS - Genome-wide Association Studies

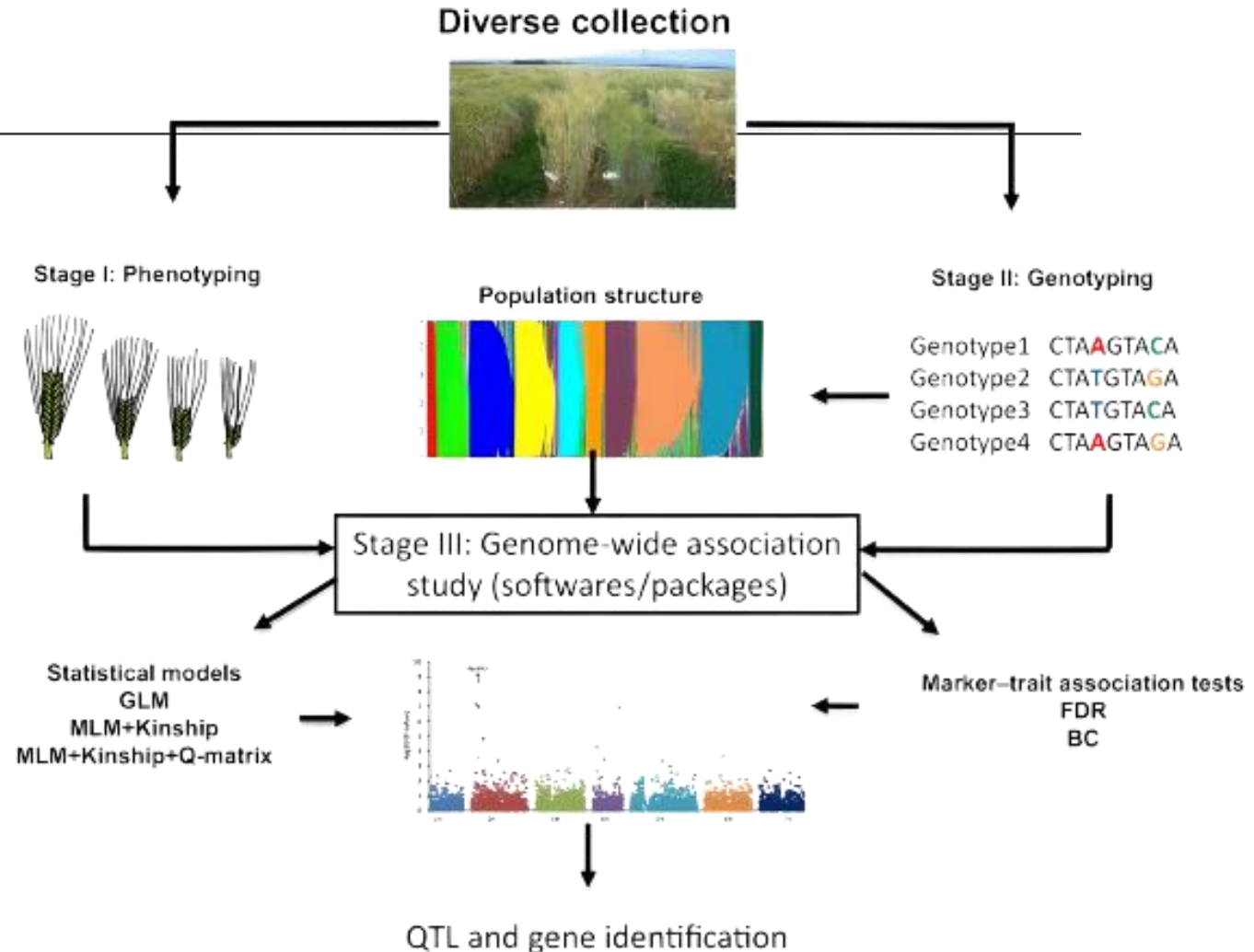
- Metodologia usada para detectar associações entre variações genéticas e uma determinada características de interesse
- SNPs - Amplamente distribuídos pelo genoma. Representam a maior colaboração para a variabilidade genética



Fonte: Varsomics, 2021

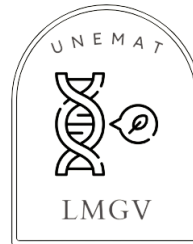
GWAS - Genome-wide Association Studies

- População diversa geneticamente
- Sequenciamento do genoma
- Análises estatísticas



GWAS - Genome-wide Association Studies

ESTUDO DO MAPEAMENTO ASSOCIATIVO DA RESISTÊNCIA FISIOLÓGICA AO MOFO BRANCO NO FEIJÃO COMUM

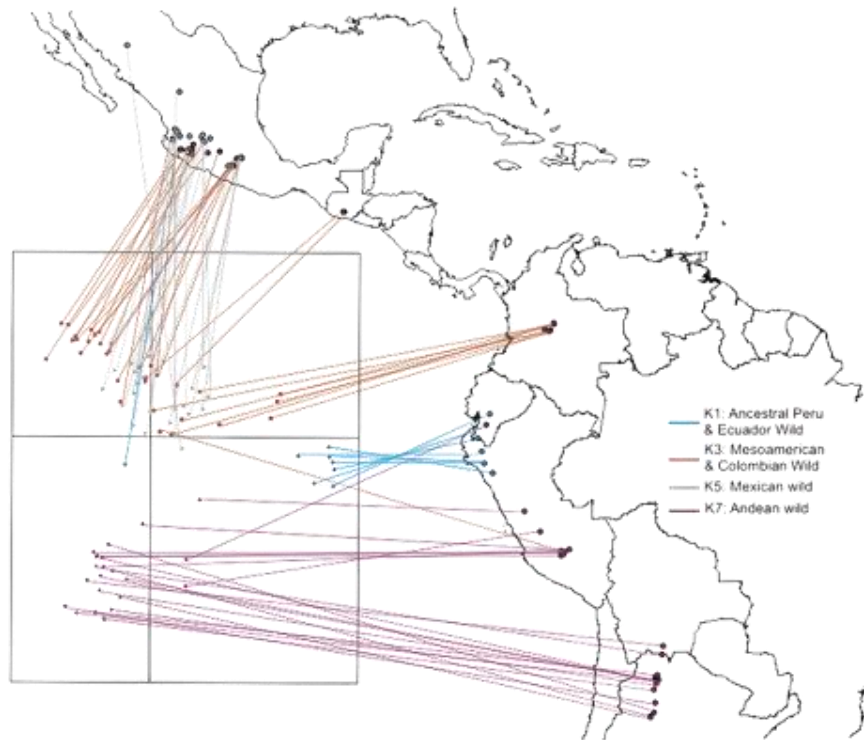


Avaliados 112 acessos de feijão
comum



- UFV-S.s463
- UFV-S.s 486
- UFV-S.s597

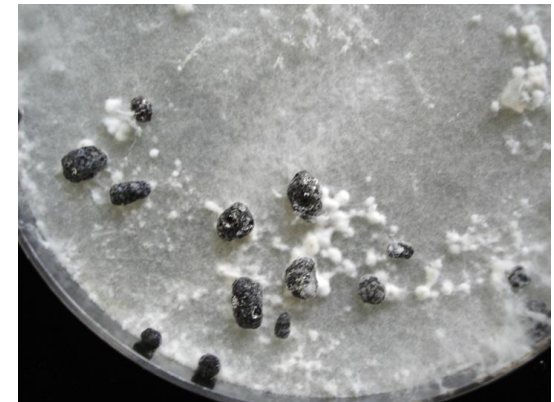
GWAS - Genome-wide Association Studies



Fonte: Kwak e Gepts, 2009



Fonte: Embrapa, 2016



Fonte: UFV, 2018

GWAS - Genome-wide Association Studies

FENOTIPAGEM

- Avaliação do comportamento vegetal em resposta a característica de interesse



GENOTIPAGEM



Genoma de referência do *P. vulgaris*
Cultivar crioula G19833

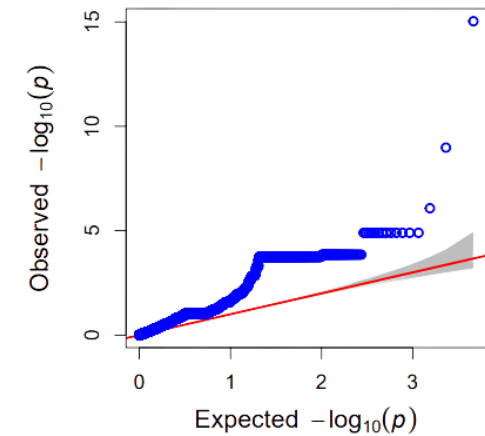
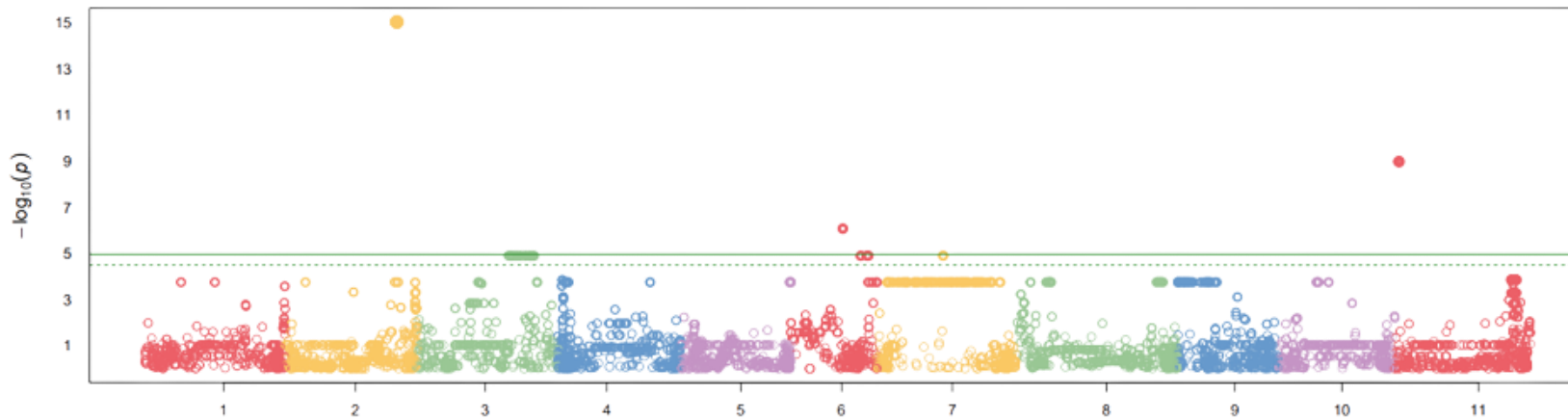
GWAS - Genome-wide Association Studies

- 4.632 SNPs foram utilizados para a característica do nível de resistência
- FarmCPU (modelo fixo e aleatório de unificação de probabilidade circulante) (Liu et al., 2016).
- BLINK (modelo com informações Bayesianas e desequilíbrio de ligação) (Huang et al, 2019).

GWAS - Genome-wide Association Studies

Associação dos SNPs com a característica

UFV-S.s597 FarmCPU – Nota ao 7º dia

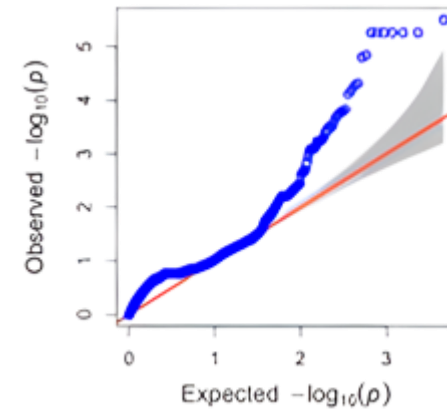
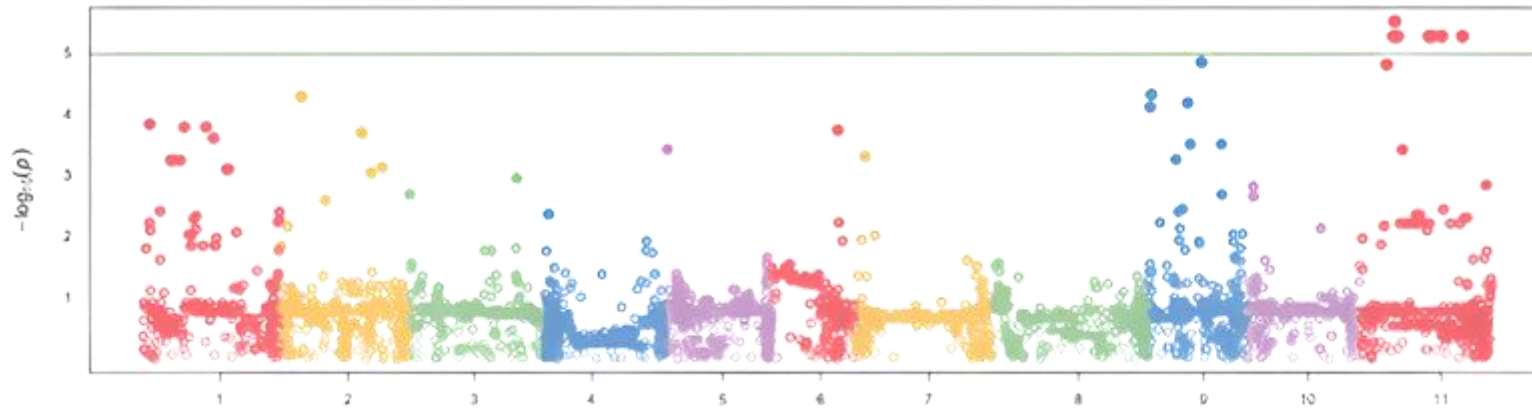


Modelo	SNP	Pv	Posição SNP (Pb)	Varição SNPs	p -valor
FarmCPU	715645010	11	1.531.456	C/A	1.04E-09
	715647642	6	19.318.348	A/G	8.41E-07
	715639729	2	41.385.245	C/A	9.31E-16

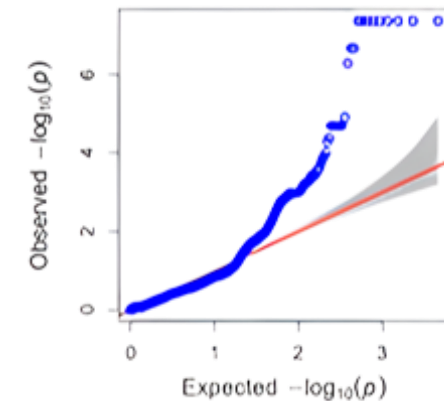
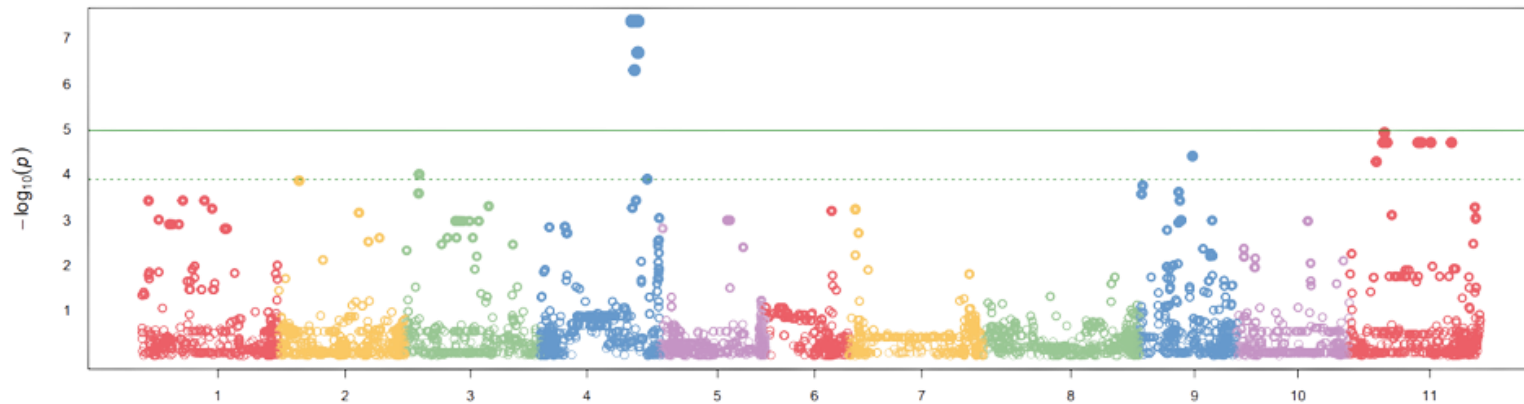
GWAS - Genome-wide Association Studies

Associação dos SNPs com a característica

UFV-S.s486 BLINK – AACPD



UFV-S.s486 FarmCPU – AACPD



GWAS - Genome-wide Association Studies

Associação dos SNPs com a característica

Modelo	SNP	Pv	Posição SNP (Pb)	Variação SNPs	p-valor
BLINK	715641520	11	13.000.676	C/T	5,51E-06
	715642155	11	13.533.435	T/C	3,17E-08
	715650419	11	14.481.893	C/A	5,51E-06
	715643922	11	26.453.670	G/T	5,51E-06
	715640930	11	27.708.187	G/A	5,51E-06
	715643799	11	31.275.831	A/G	5,51E-06
	715641250	11	39.235.446	C/T	5,51E-06
FarmCPU	715641721	4	34.996.996	A/G	4,23E-08
	715641963	4	35.152.997	C/T	4.23E-08
	715650775	4	35.160.033	C/T	4.23E-08
	715644663	4	35.262.492	C/A	4,23E-08
	715641343	4	35.507.879	G/A	4.23E-08
	715641342	4	35.509.821	T/C	4.23E-08
	715639648	4	36.034.366	A/G	5.09E-07
	715639634	4	36.633.380	C/A	4.23E-08
	715639632	4	36.811.261	T/C	4.23E-08
	715642830	4	37.226.669	T/G	2.09E-07
	715639507	4	37.382.175	T/C	4.23E-08
	715647336	4	37.491.486	G/T	2.09E-07

Rahmanzadeh et al., (2022)
Análise de meta-QTL

Pv 11 ($n = 6$) Pv 7 ($n = 24$)

GWAS - Genome-wide Association Studies

Tabela 6. SNPs significativos, posição física da região candidata e anotação funcional dos genes candidatos para a reação de resistência fisiológica ao isolado UFV-S.s597 em feijão comum identificados pelo modelo FarmCPU

Caract.	SNP	Gene candidato	PV	Posição do Gene (Pb)	Anotação funcional
	715639729	PHAVU_002G247100g	2	41.385.242 - 41.385.247	Protein kinase domain-containing protein

TCP	715647642	PHAVU_006G073600g	6	19.318.347 - 19.318.349	Na_H_Exchanger domain-containing protein

	715645010	PHAVU_011G019200g	11	1.531.452 -1.531.457	Transcription factor MYB41-like

Transmissão de sinais extracelulares para o núcleo através de uma cascata de sinalização.

Regula volume celular e absorção de sódio. Interação entre NHE e cálcio (Qiu et al., 2004).

Síntese de suberina alifática e deposição de lamelas semelhantes a suberina associadas à parede celular (Franke et al., 2009).

Transcriptome-wide association studies - TWAS

- GWAS associaram de forma robusta milhares de loci genômicos com características complexas
- Apesar desse sucesso, os loci GWAS são muitas vezes difíceis de interpretar
- Desequilíbrio de ligação (LD): obscurece as variantes causais que conduzem a associação
- Genes causais que mediam os efeitos variantes na característica raramente são determinados apenas pelos dados GWAS
- Este desafio de interpretação motivou o desenvolvimento de métodos para priorizar genes causais em loci GWAS

TWAS - Introdução

- Estudos de associação do transcriptoma em larga escala (TWASs), como uma abordagem prática e prevalente para detectar as associações entre genes geneticamente regulados e características, estão agora levando a uma melhor compreensão dos complexos mecanismos de variantes genéticas na regulação de várias doenças e características
- Uma dessas famílias de métodos é o TWAS, que utiliza painéis de referência de expressão (cohorts eQTL com dados de expressão e genótipo) para descobrir associações gene-traits de conjuntos de dados GWAS

TWAS - Introdução

PERSPECTIVE

<https://doi.org/10.1038/s41588-019-0385-z>

nature
genetics

Opportunities and challenges for transcriptome-wide association studies

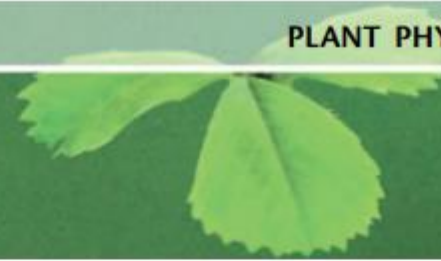
Michael Wainberg¹, Nasa Sinnott-Armstrong ², Nicholas Mancuso ³, Alvaro N. Barbeira ⁴,
David A. Knowles ^{5,6}, David Golan², Raili Ermel⁷, Arno Ruusalepp^{7,8}, Thomas Quertermous ⁹,
Ke Hao ¹⁰, Johan L. M. Björkegren ^{8,10,11,12*}, Hae Kyung Im ^{4*}, Bogdan Pasaniuc ^{3,13,14*},
Manuel A. Rivas ^{15*} and Anshul Kundaje ^{1,2*}

Transcriptome-wide association studies - TWAS

doi:10.1093/plphys/kiab161

PLANT PHYSIOLOGY 2021: 186: 1800–1811

Plant Physiology®



TWAS results are complementary to and less affected by linkage disequilibrium than GWAS

Delin Li ,^{1,2,3} Qiang Liu⁴ and Patrick S. Schnable  ^{1,4,*†}

TWAS - Introdução

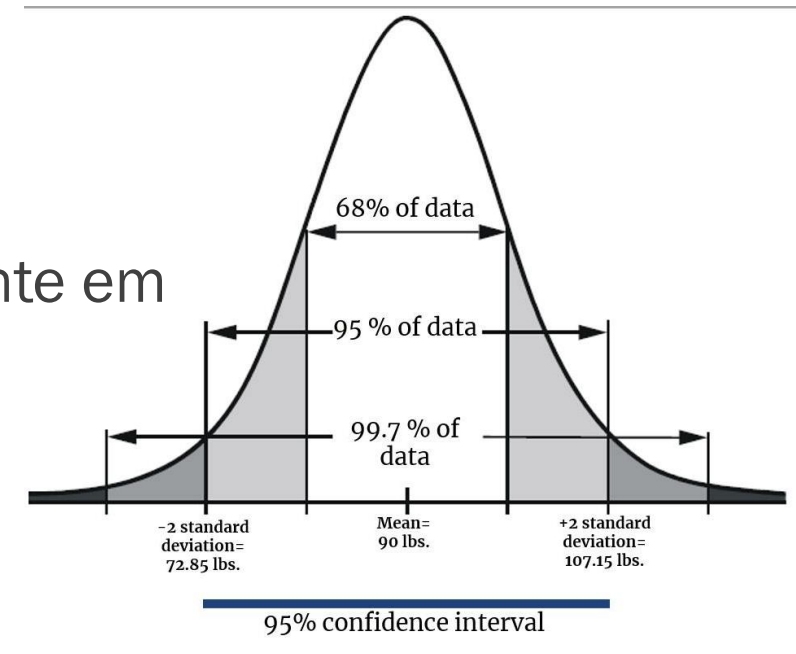
- Devido ao desequilíbrio de ligação (LD), muitas vezes não é possível determinar de forma inequívoca qual dos múltiplos genes ligados a um marcador genético associado a uma característica via GWAS é de fato o gene causal
- O estudo de associação transcriptoma ampla (TWAS) detecta associações entre variação nos níveis de expressão gênica e variação fenotípica
- Comparou-se os resultados do GWAS e do TWAS na característica qualitativa cor da pubescência na soja (*Glycine max*), que possui um LD médio de aproximadamente 100 kb

TWAS – Materiais e Métodos

- Fontes dos dados: Os SNPs do The 1001 Genome Project
- RNA-seq
- SNPs bialélicos foram identificados dentro de cada amostra
- Quantidade mínima de leituras [> 2 para cada alelo] e $< 20\%$ de dados perdidos
- Os SNPs foram filtrados (MAF de $> 5\%$)
- 75.289 SNPs retidos para 102 linhagens de soja
- Imputação dos SNPs usando Beagle e genótipos imputados foram usados para GWAS

TWAS – Materiais e Métodos

- Os valores de expressão normalizados foram transformados (GAPIT):
 - Intervalo numérico de 0 a 2 para cada gene (lidar com valores extremos de expressão ou outliers)
 - Valores de expressão menores que o quantil 5: 0
 - Maiores que o quantil 95: 2
 - Valores de expressão restantes: transformados linearmente em valores entre 0 e 2.



TWAS – Materiais e Métodos

- O GWAS e o TWAS: Modelo Linear Misto Comprimido implementado no pacote R GAPIT
- Corte MAF adicional de $\geq 5\%$ em amostras com um fenótipo foi aplicado para GWAS
- 3 primeiros componentes da análise de componentes principais derivados de SNPs de entrada (GWAS) e expressão gênica (TWAS) foram usados para controlar separadamente a estrutura da população
- Valores -P resultantes foram ajustados para controlar testes múltiplos
- SNPs ou genes que exibiram um FDR de $< 0,05$ foram definidos como traço associado

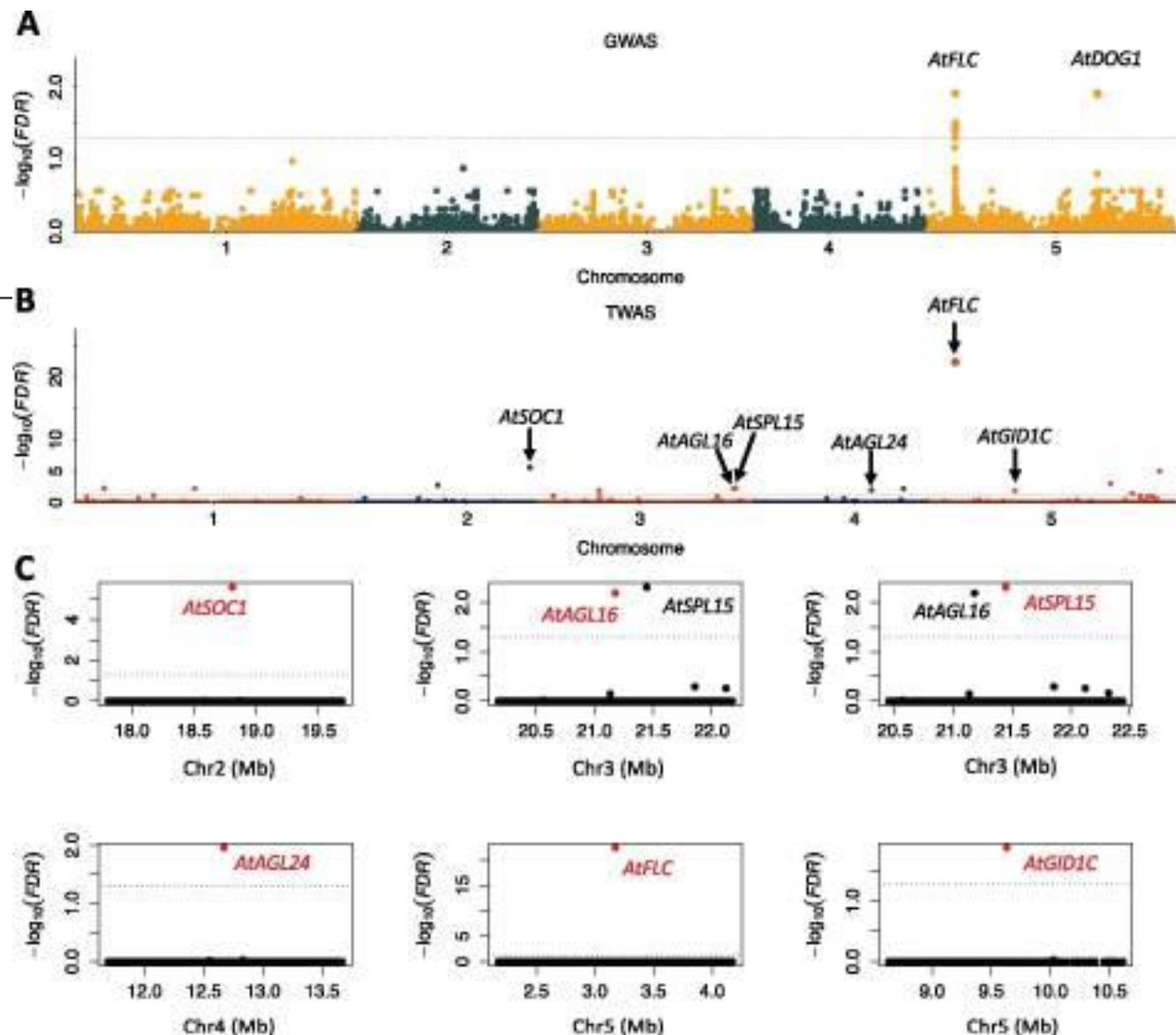
TWAS – Materiais e Métodos

- A diferença entre os níveis médios de expressão do gene T nos dois grupos de cor da pubescência da soja (amarelado e cinza) foi analisada usando o teste t de duas amostras de Welch (Welch, 1947)
- Um teste exato de Fisher unilateral (Fisher, 1922) foi usado para verificar se há mais sobreposição do que o esperado por acaso entre dois grupos de genes
- $P < 0,01$ desses testes foi considerado significativo

TWAS

Resultados

- GWAS baseado SNP para tempo de floração a 16°C (FT16) em um painel de 970 acessos de Arabidopsis: dois loci (FLOWERING LOCUS C (AtFLC) e ATRASO DE GERMINAÇÃO1 (AtDOG1))
- TWAS foi realizado no subconjunto de acessos deste painel para dados de RNA-seq de tecido foliar e fenótipos FT16 (N = 690 amostras)
 - Identificou 14 genes associados a características incluindo apenas um dos dois genes identificados via GWAS: AtFLC



TWAS - Resultados

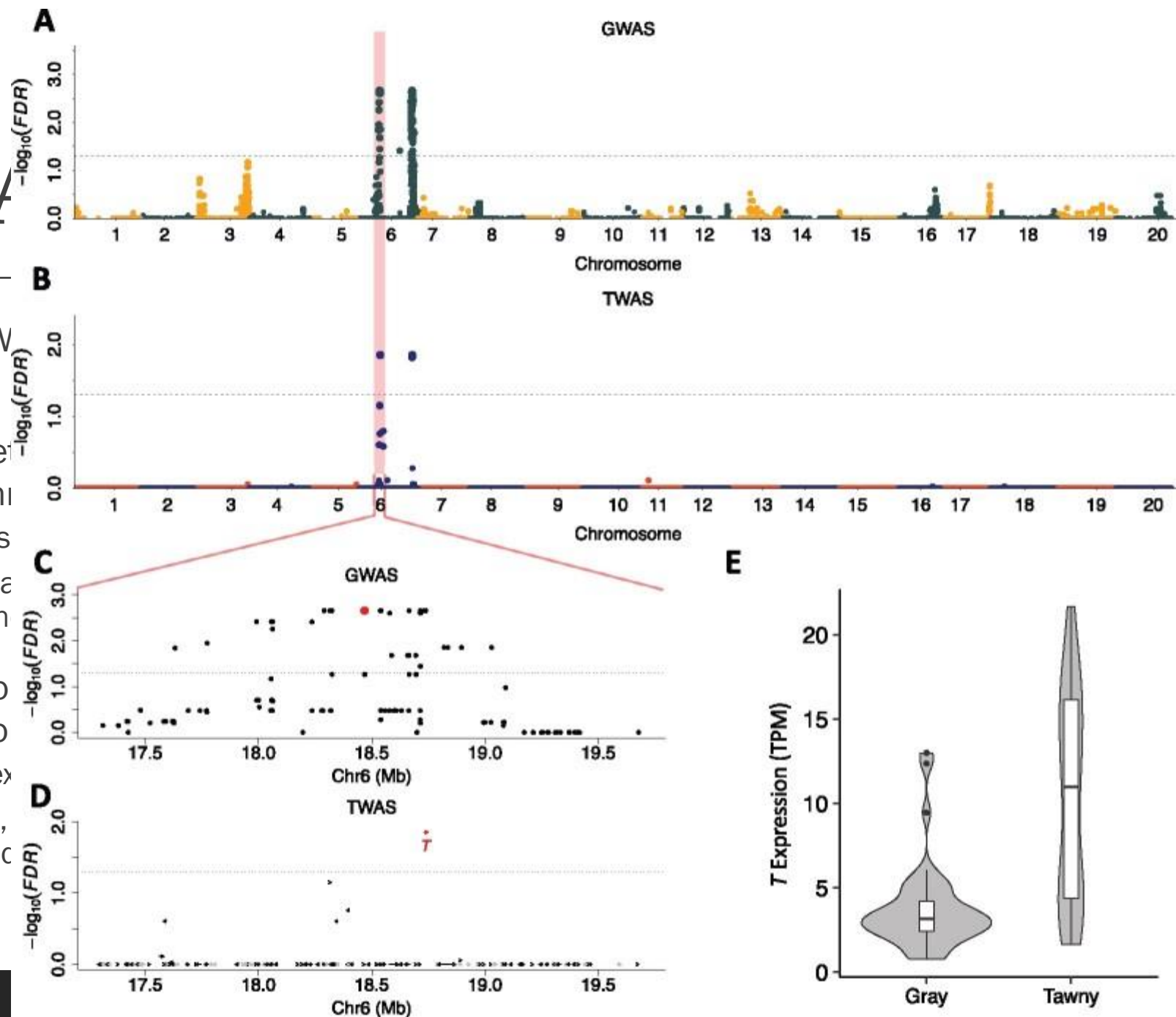
- Embora o outro gene associado ao FT16 via GWAS, AtDOG1 , seja expresso no tecido foliar, ele não foi associado ao tempo de floração via TWAS
- Um total de 6 dos 14 genes associados a características de TWAS estão incluídos no banco de dados FLOR-ID, que inclui 306 genes relacionados à floração de *Arabidopsis* selecionados manualmente
- Isso é substancialmente mais sobreposição do que o esperado por acaso (teste exato de Fisher unilateral P -valor 5E-8)

TWAS - Resultados

- Os outros 8 genes associados a traços identificados via TWAS
 - A supressão antisense do FATOR DE INTERAÇÃO DE FITOCROMO3 (AtPIF3) leva ao florescimento precoce
 - 3 dos outros genes associados a características (ARABIDOPSIS CDK INHIBITOR1 [AtACK1], RNA HELICASE30 [AtRH30] e VITAMIN C DEFECTIVE5 [AtVTC5]) estão envolvidos em redes reguladoras de floração
 - Vários são fatores de transcrição etc
- GWAS x TWAS: estudos de associação foram realizados no subconjunto de amostras de Arabidopsis para os quais genótipos, dados de expressão e dados fenotípicos estavam todos disponíveis (N = 631)
 - Identificaram 1 locus e 10 genes associados a FT16 (1/1 de GWAS e 7/10 de TWAS): associados a característica tempo de floração

TWAS

- GWAS e TWAS em uma janela de 100 kb
- O GWAS de *T* identificou um locus em 1,4 Mb (Chr6) com 68 genes e 3 GWAS associados a distância
- O gene *T* foi definido por TWAS
 - *T* exibe expressão diferencial
 - Portanto, a associação de traços é causal



damente

imadamente

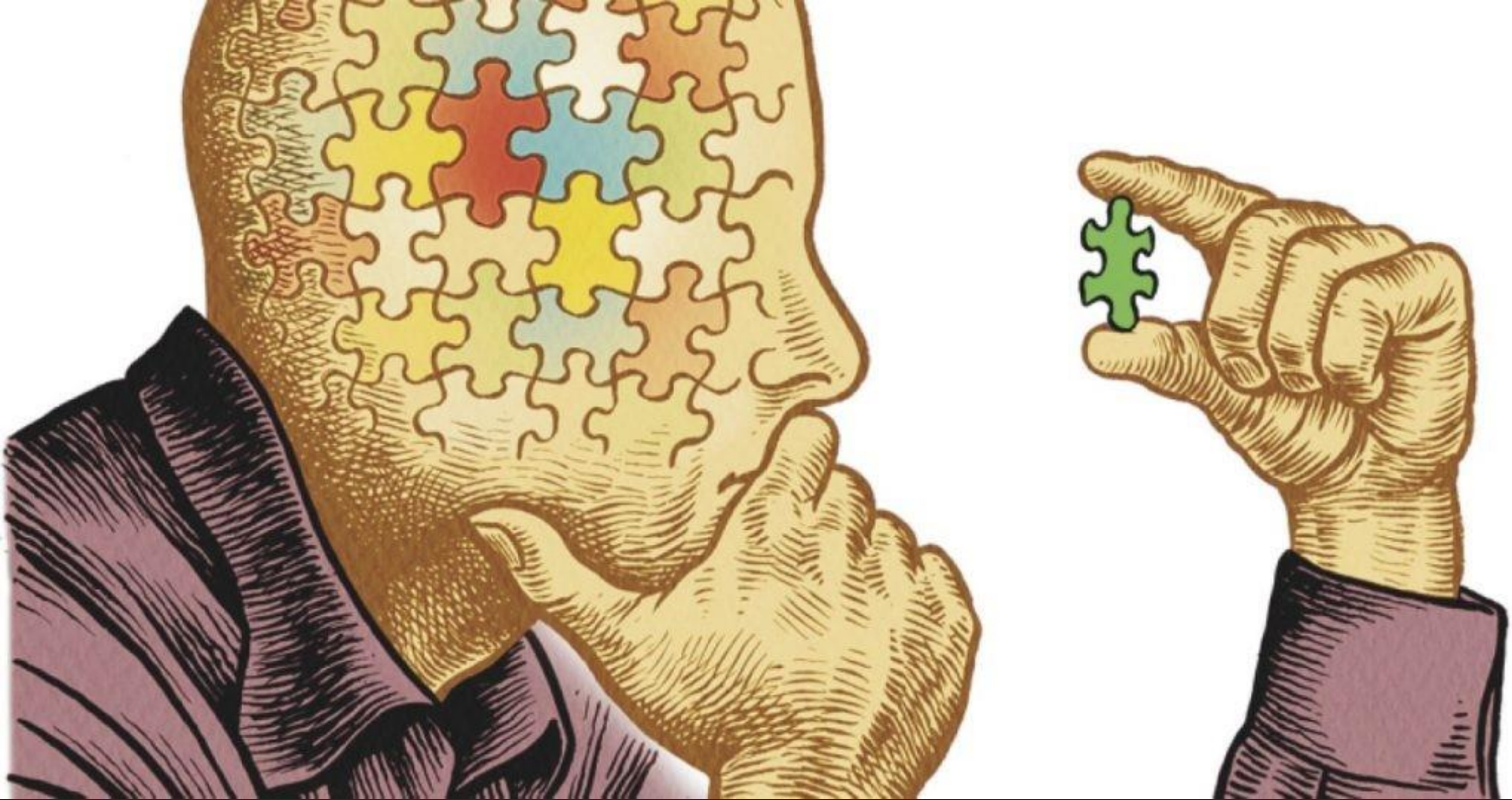
100 a 600 kb

,4 Mb

7)
iciados a
ia

TWAS - Conclusão

- Ao estudar características bem caracterizadas, demonstramos que o TWAS oferece altas taxas de resultados verdadeiros positivos e é menos afetado pelo LD do que o GWAS, o que o torna uma ferramenta promissora em espécies com alto LD.
- Embora o tecido relacionado ao traço seja preferido como fonte de dados de expressão, nossos resultados mostram que os dados de expressão de outros tecidos também podem ser usados para identificar genes associados ao traço.



*You can inherit good genes, not
greatness*

– RICHARD DAWKINS* (THE SELFISH GENE)