From candidate genes to causal variants:

Strategies to identify (or not) genes and sequence variants in rodent populations

Webinar 4

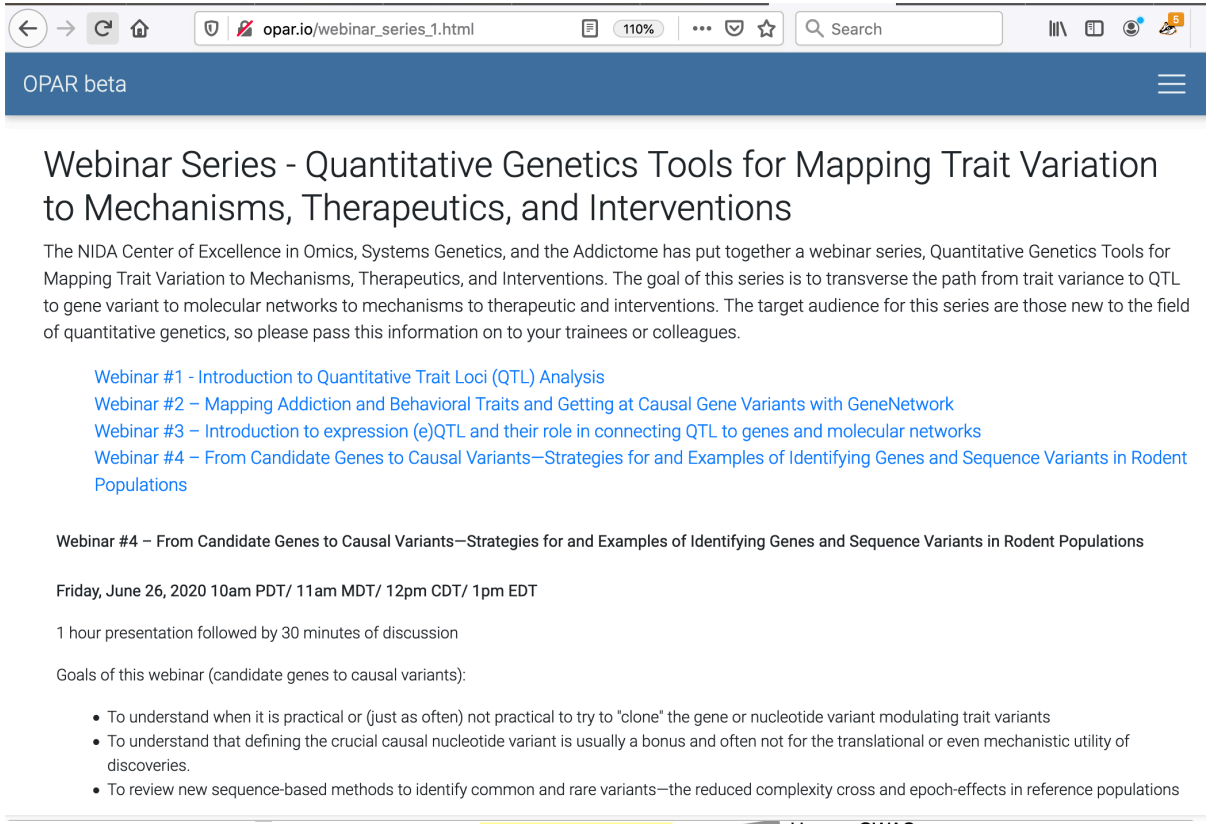
Williams RW, Ashbrook DG, Mulligan MK, Lu L, Williams EG, Chen H, Prins J, Saba LM, Sen S, Sloan Z, Centeno A, and the P30 and GN teams

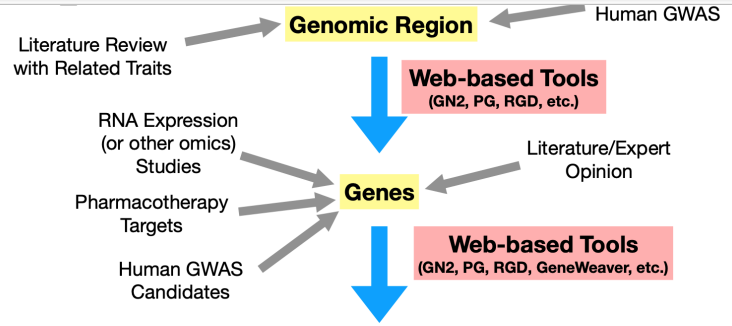
Sponsored by the NIDA Center of Excellence in Omics, Systems Genetics, and the Addictome (NIDA P30 DA044223), with support from NIGMS Systems Genetics and Precision Medicine Project (R01 GM123489)

Presented by Rob Williams
rwilliams@uthsc.edu

Where to find the presentations in this series

OPAR.io





1 Mapping resources and comparisons

2 When is it practical or not to define genes and causal variants associated with loci? Key factors: Map precision, genetic complexity of locus (numbers of variants and genes in interval), robustness and GXE sensitivity

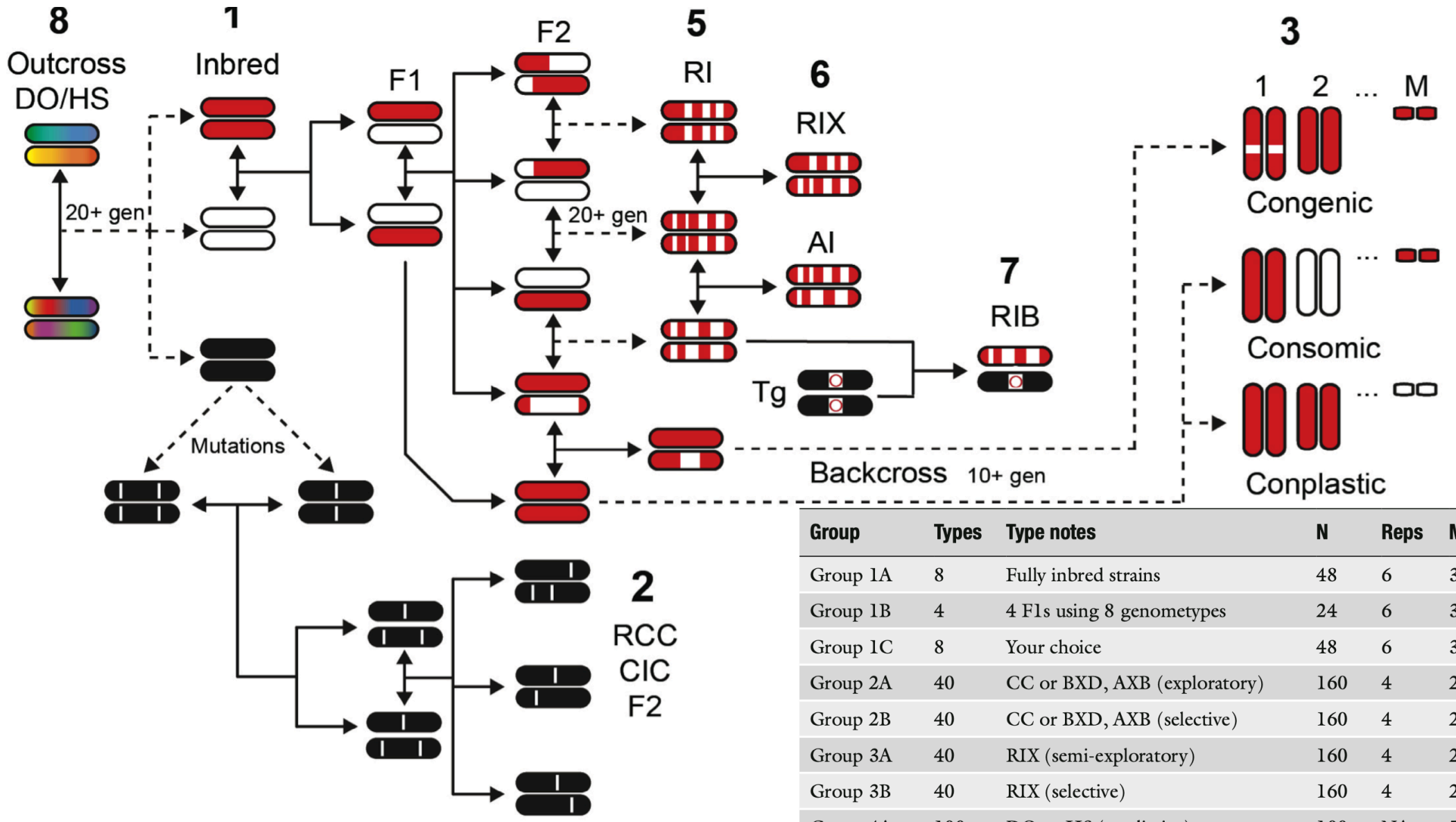
3 Classic forward genetic methods that have worked, but almost always with a caveat or two. QT nucleotide are icing-on-cake, not essential for mechanistic or tranlational studies.

4 New reverse genetic methods and phenome-wide association studies (PheWas). Epoch effects as a hybrid methods (forward and reverse)

5 Demo of some ways to cherry pick. (1) Finding strong loci worth converting to gene variants. (2) Finding impactful variants worth converting to phenotypes

Main types of mapping resources

Williams RW, Williams EG (2017) Resources for Systems Getnetics. *Systems Genetics: Methods and Protocols*, Methods in Molecular Biology, vol. 1488: 3–29



Group	Types	Type notes	N	Reps	M	F	Months
Group 1A	8	Fully inbred strains	48	6	3	3	2.4
Group 1B	4	4 F1s using 8 genometypes	24	6	3	3	1.2
Group 1C	8	Your choice	48	6	3	3	2.4
Group 2A	40	CC or BXD, AXB (exploratory)	160	4	2	2	8
Group 2B	40	CC or BXD, AXB (selective)	160	4	2	2	8
Group 3A	40	RIX (semi-exploratory)	160	4	2	2	8
Group 3B	40	RIX (selective)	160	4	2	2	8
Group 4A	100	DO or HS (predictive)	100	NA	50	50	5
Group 4B	100	DO or your choice	100	NA	50	50	5
Sums	380		960				48

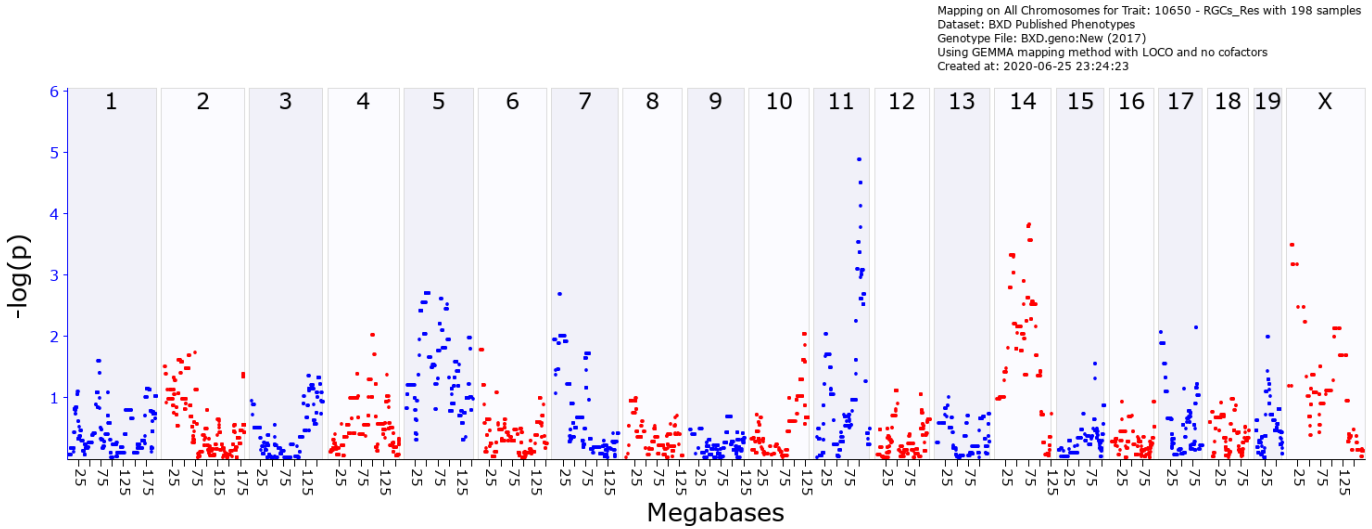
Some comparisons of resources

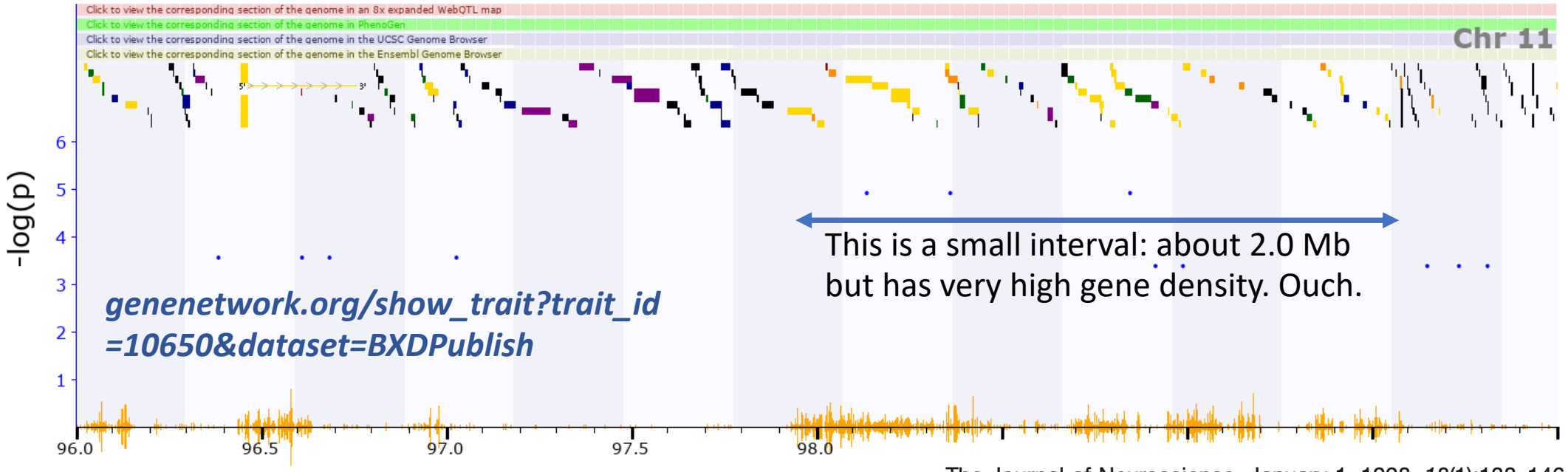
Williams RW, Williams EG (2017) Resources for Systems Getnetics. *Systems Genetics: Methods and Protocols*, Methods in Molecular Biology, vol. 1488: 3–29

Type of cross	Recs/case	LOD Threshold	$/Geno typing	$/Case[a]	Isogenic	Inbred	Phen-ome	GXE	Breeding	References
Consomic and congenic sets	1	1–2	0	140	Yes	Yes	Yes	Easy	Variable	[14, 55]
Reduced complexity cross	25	1–2	25	20	Almost	Almost	Hard	Hard	Easy	[44, 45]
F2 intercross, 2-way or 4-way	25	2.5–3	25	15	No	No	Hard	Hard	Easy	[8, 16]
Advanced intercross	100	4–5	100	100	No	No	Hard	Hard	Hard	[9, 10]
RI strains and advanced RI Strains	50 to 80	3–4	0	140	Yes	Yes	Yes	Easy	Variable	[4, 8, 22]
Advanced intercross RI strains	80	4–5	0	140	Yes	Yes	Yes	Easy	Variable	[4, 8]
RI Intercross F1s (RIX, RIB)	100 to 200	4–6	0	50	Yes	No	Hard	Easy	Easy	[36, 38, 40]
Hybrid diversity panel (HDP)	1000	6+	0	20–150	Yes	Yes	Yes	Yes	Easy	[18, 19]
Collaborative cross (8-way RI)	135	4–6	0	195	Yes	Yes	Yes	Easy	Variable	[13, 17]
Diversity outcross (DO HS)	400+	5–7	100	55	No	No	Hard	Hard	Easy	[84, 85]
Outbred stock (e.g., CD-1, CF-1)	1000	6+	100	7	No	No	Hard	Hard	Easy	[68, 79]

Strategies to improve chances of defining causal genes

1 Maximize the ***effective*** precision by reducing the product of *[QTL length]* X *[n of polymorphics genes and DNA variants]*. 10 Mb can be great effective precision if there are few variants or genes (see ***Comt*** example to follow). Conversely, 2 Mb can be problematic if a regions is very gene-rich and polymorphic.





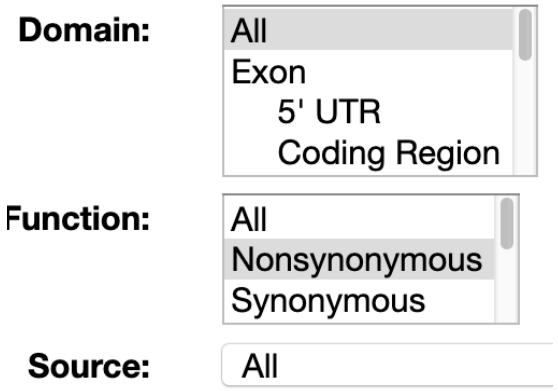
Natural Variation in Neuron Number in Mice Is Linked to a Major Quantitative Trait Locus on Chr 11

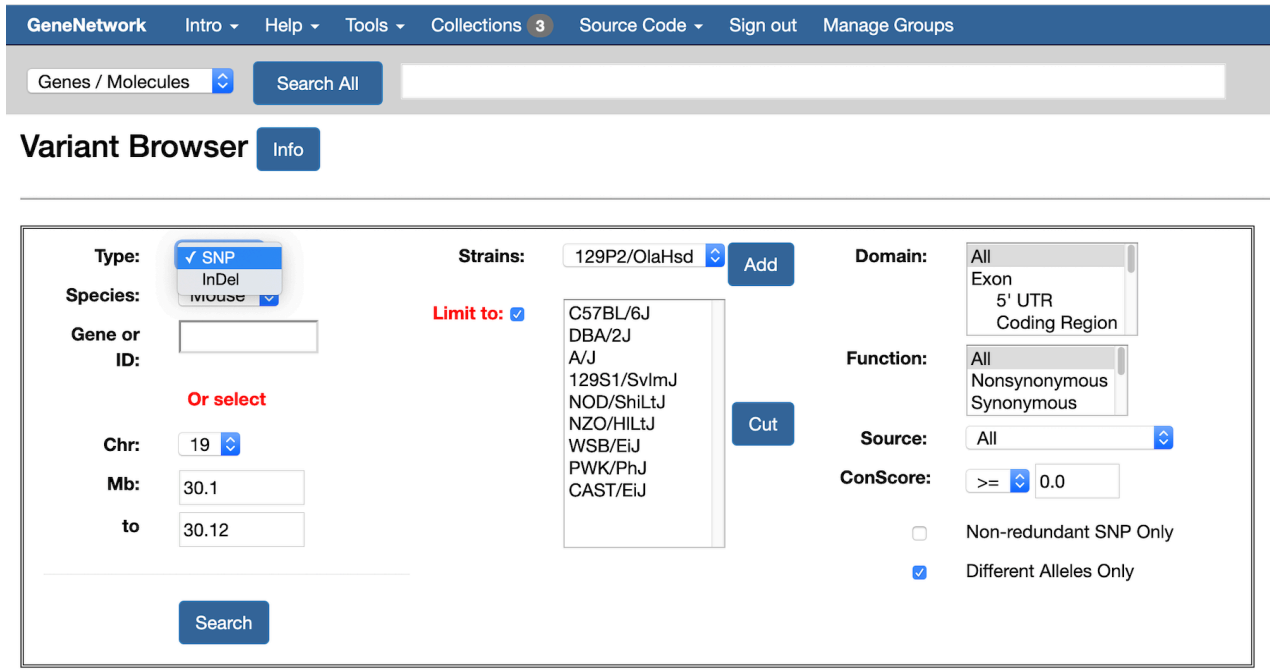
The Journal of Neuroscience, January 1, 1998, *18*(1):138–146

Part 2: Slide 5

Strategy 2

2 Maximize your use of sequence data and eQTL data. There are only two major ways a DNA variant can operate—protein-coding variants and isoform expression level variatiants. This seems painfully obvious, but if it is so obvious then why are rat and mouse sequence data still so sadly incomplete? Why are there not validated compendia of sequence variants segregating in major crosses? Where are the all of the great data on splice isoforms in brain or other tissues and cells? Why are maps still expressed in centiMorgans rather than basepairs?





Med1 NCBI	ENSMUST00000018304	17	Exon 17	Coding	Nonsynonymous	Biotype: Protein Coding, T -> S, aCt -> aGt, 945	C	G
Cdk12 NCBI	ENSMUST00000107538	1	Exon 1	Coding	Nonsynonymous	Biotype: Protein Coding, L -> I, Ctc -> Atc, 346	A	C
Cdk12 NCBI	ENSMUST00000092733	2	Exon 2	Coding	Nonsynonymous	Biotype: Protein Coding, L -> I, Ctc -> Atc, 74	A	C
Cdk12 NCBI	ENSMUST00000003203	1	Exon 1	Coding	Nonsynonymous	Biotype: Protein Coding, L -> I, Ctc -> Atc, 346	A	C
Cdk12 NCBI	ENSMUST00000107539	1	Exon 1	Coding	Nonsynonymous	Biotype: Protein Coding, L -> I, Ctc -> Atc, 346	A	C
Erbb2 NCBI	ENSMUST00000058295	3	Exon 3	Coding	Nonsynonymous	Biotype: Protein Coding, A -> T, Gcc -> Acc, 130	A	G
Ikzf3 NCBI	ENSMUST00000103141	4	Exon 4	Coding	Nonsynonymous	Biotype: Protein Coding, G -> S, Ggt -> Agt, 114	T	C
Nr1d1 NCBI	ENSMUST00000064941	2	Exon 2	Coding	Nonsynonymous	Biotype: Protein Coding, A -> T, Gca -> Aca, 84	T	C
Casc3 NCBI	ENSMUST00000017384	2	Exon 2	Coding	Nonsynonymous	Biotype: Protein Coding, E -> G, gAg -> gGg, 64	G	A
Casc3 NCBI	ENSMUST00000169695	1	Exon 1	Coding	Nonsynonymous	Biotype: Protein Coding, E -> G, gAg -> gGg, 64	G	A

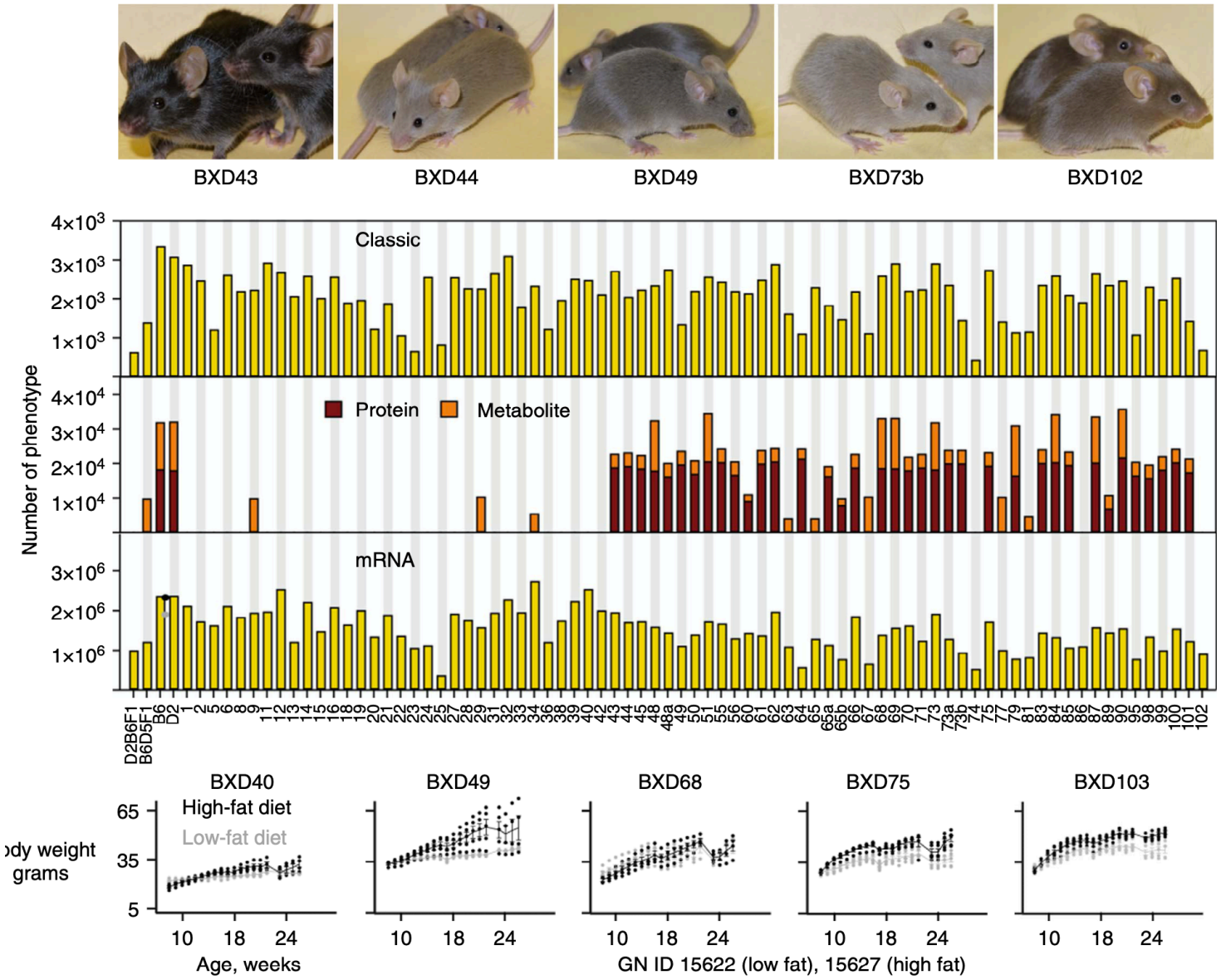
Strategy 3

3 Study populations for which you can generate or inherit a deep phenome. Use either reference populations or work as part of a larger collaboration. HRDP, HMDP, HS, CC, BXDs—all of these work well from this perspective.

When possible get complementary omics data, and if at all possible, biobank tissues.

The power of replicability can be huge when heritability is low. We have strong lifespan loci with only 70 BXD strains, but 10 replicates of each. Replicability also enables GXE analysis.

While all of this is obvious, it goes against the grain of ”doing your own thing” and being innovative.



Part 2: Slide 5

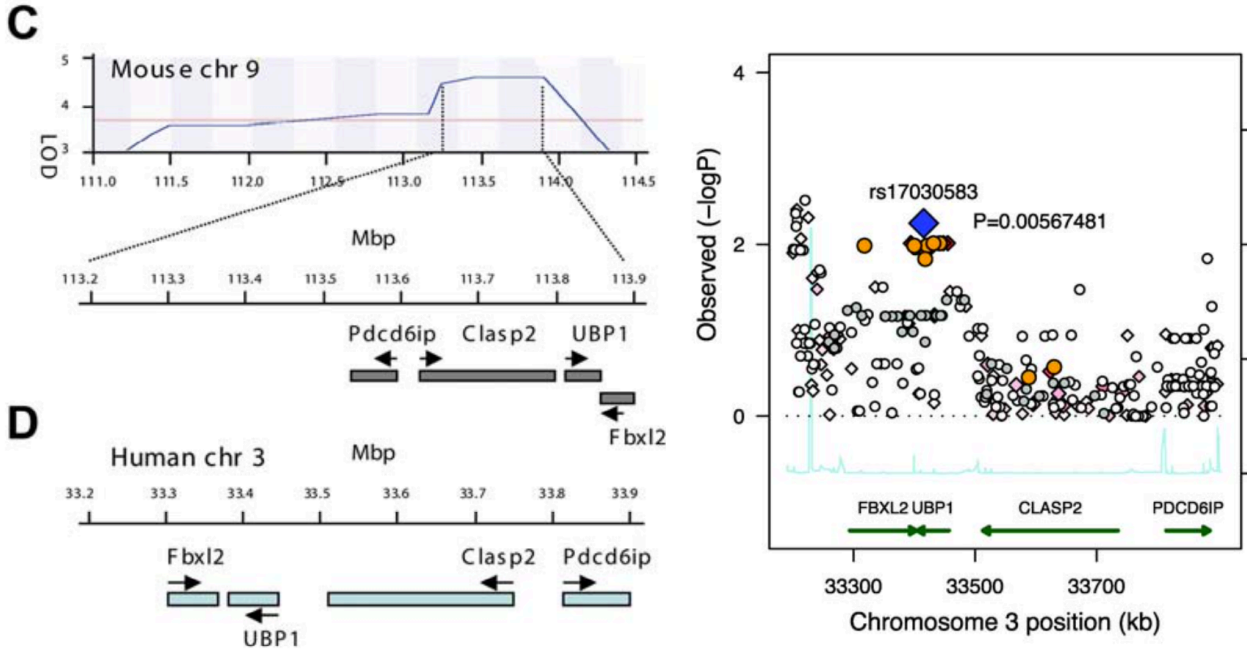
NATURE COMMUNICATIONS | 7:10464 | DOI: 10.1038/ncomms1046

Strategy 4

4 Use multiple crosses and mulitple species to “fine-map” loci that modulate trait values or variance.

OPEN ACCESS Freely available online PLoS GENETICS

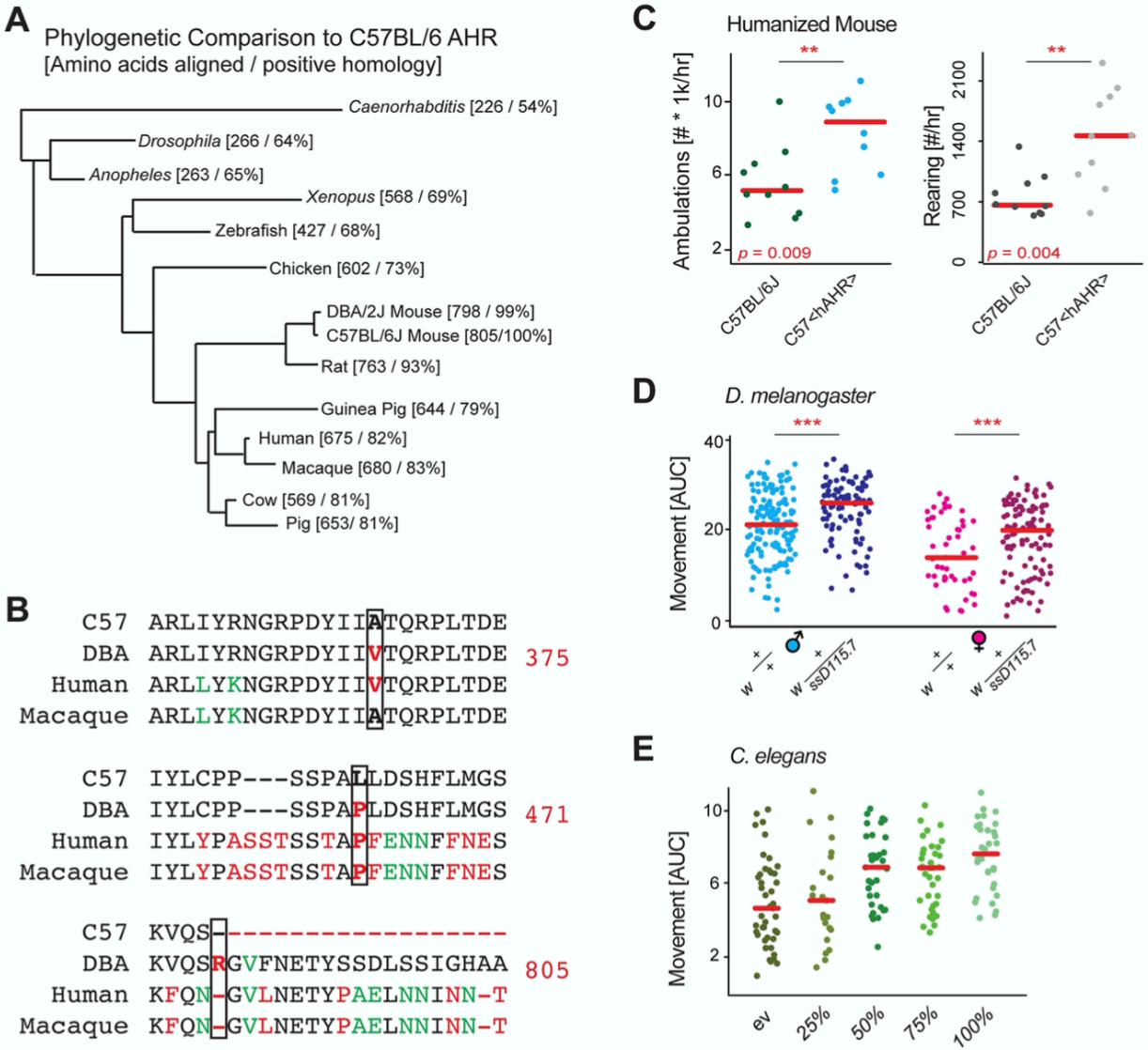
Identification of the *UBP1* Locus as a Critical Blood Pressure Determinant Using a Combination of Mouse and Human Genetics



OPEN ACCESS Freely available online PLOS GENETICS

An Evolutionarily Conserved Role for the Aryl Hydrocarbon Receptor in the Regulation of Movement

Evan G. Williams[1], Laurent Mouchiroud[1], Michael Frochaux[2], Ashutosh Pandey[3], Pénélope A. Andreux[1], Bart Deplancke[2], Johan Auwerx[1]*



Part 2: Slide 7

Strategy 5

5 Study molecular endophenotypes. They tend to be eaiser to measure, may have less complex genetic architecture, often have known partners and pathways, and can be easier clone and put into a mechanistic context.

Type: Phenotypes [Info]

Dataset: BXD Published Phenotypes

Get Any: dopamine

BXD_10265	Central nervous system, pharmacology, protein expression: Dopamine receptor 2 and 3 (DRD2/DRD3) binding maximum (Bmax) in membrane fragments in the dorsal striatum (caudate putamen) of females (125I-epidepride ligand) [fmol/mg wet weight]	236.137	Jones BC, Tarantino LM, Rodriguez LA, Reed CL, McClearn GE, Plomin R, et al.	1999	25.3	Chr15: 87.476581	90.157
BXD_10725	Central nervous system, metabolism, nutrition: Zinc level in medial prefrontal cortex of females [nmol/g]	224.933	Jones LC, McCarthy KA, Beard JL, Keen CL, Jones BC	2006	24.9	Chr1: 153.969506	28.575
BXD_17033	Central nervous system, pharmacology, toxicology: Effect of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) on homovanilic acid (HVA) concentration in caudate-putamen in females 48h after injection (saline-MPTP group) [ug/mg wet weight]	0.219	Jones BC, Miller DB, O'Callaghan JP, Unger EL, Lu L, Alam G, et al.	2014	24.3	Chr11: 65.756786	-0.153
BXD_10234	Central nervous system, pharmacology, protein expression: Dopamine transporter (DAT, SLC6A3) protein density in the dorsal striatum (caudate putamen) [Bmax, pmol/mg]	3.118	Janowsky A, Mah C, Johnson RA, Cunningham CL, Phillips TJ, Crabbe JC, et al.	2001	23.8	Chr19: 15.292517	0.939

› Cell. 2012 Sep 14;150(6):1287-99. doi: 10.1016/j.cell.2012.08.012. Epub 2012 Aug 30.

Systems Genetics of Metabolism: The Use of the BXD Murine Reference Panel for Multiscalar Integration of Traits

Pénélope A Andreux [1], Evan G Williams, Hana Koutnikova, Riekelt H Houtkooper, Marie-France Champy, Hugues Henry, Kristina Schoonjans, Robert W Williams, Johan Auwerx

Index	Record	Description	Mean	Authors	Year	Max LRS?	Max LRS Location	Additive Effect?
122	BXD_12942	Blood chemistry, cardiovascular system: Mean cell volume (red blood cells) of 14-week old males (mean corpuscular volume, MCV) [fl]	45.680	Andreux P, Williams EG, Koutnikova H, Houtkooper RH, Champy MF, Henry H, et al.	2012	32.0	Chr7: 104.149021	1.685
74	BXD_12894	Blood chemistry: Alkaline phosphatase of 14-week old males (ALPL gene product) [U/l]	137.285 137.285	Andreux P, Williams EG, Koutnikova H, Houtkooper RH, Champy MF, Henry H, et al.	2012	31.0	Chr4: 131.999242	31.848
69	BXD_12889	Central nervous system, metabolism, behavior: Water intake of 13-week old females [ml/mouse/unit time]	2.082	Andreux P, Williams EG, Koutnikova H, Houtkooper RH, Champy MF, Henry H, et al.	2012	30.8	Chr9: 41.851653	-0.692
111	BXD_12931	Blood chemistry, cardiovascular system: Hematocrit of 14-week old males [%]	44.690	Andreux P, Williams EG, Koutnikova H, Houtkooper RH, Champy MF, Henry H, et al.	2012	30.6	Chr7: 109.979743	3.437
4	BXD_12824	Blood chemistry: Alkaline phosphatase of 14-week old females, (ALPL gene product) [U/l]	178.500	Andreux P, Williams EG, Koutnikova H, Houtkooper RH, Champy MF, Henry H, et al.	2012	30.4	Chr4: 131.999242	39.076

Part 2: Slide 9

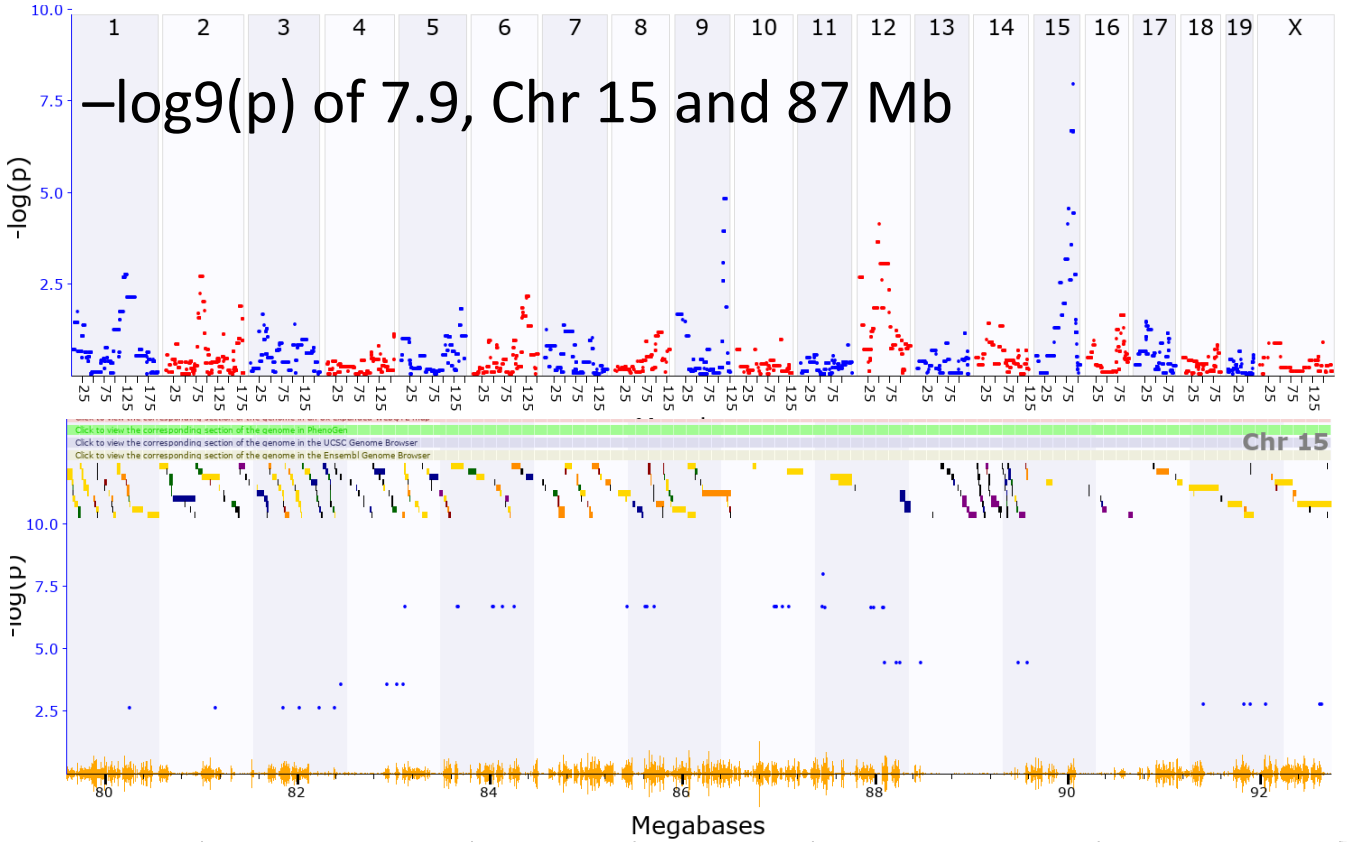
Does strategy 5 work?

BXD_10265	Central nervous system, pharmacology, protein expression: Dopamine receptor 2 and 3 (DRD2/DRD3) binding maximum (Bmax) in membrane fragments in the dorsal striatum (caudate putamen) of females (125I-epidepride ligand) [fmol/mg wet weight]	236.137	Jones BC, Tarantino LM, Rodriguez LA, Reed CL, McClearn GE, Plomin R, et al.	1999	25.3	Chr15: 87.476581	90.157

Ataxin 10 has very high expression in MSN in striatum

Ataxin 10 is a strong cis eQTL in striatum

Ataxin 1 reported to modulate *Drd2* expression in cerebellum



Down-regulation of the dopamine receptor D2 in mice lacking ataxin 1

FREE

Robert Goold, Michael Hubank, Abigail Hunt, Janice Holton, Rajesh P. Menon, Tamas Revesz, Massimo Pandolfo, Antoni Matilla-Dueñas

Human Molecular Genetics, Volume 16, Issue 17, 1 September 2007, Pages 2122–2134, https://doi.org/10.1093/hmg/ddm162

☑	6	1422576_at	Atxn10	spinocerebellar ataxia 10 homolog; three exons	Chr15: 85.393336	14.419	24.2	Chr15: 87.454918	-0.203

Pharmacogenetics. 1999 Oct;9(5):607-17.

Quantitative-trait Loci Analysis of Cocaine-Related Behaviours and Neurochemistry

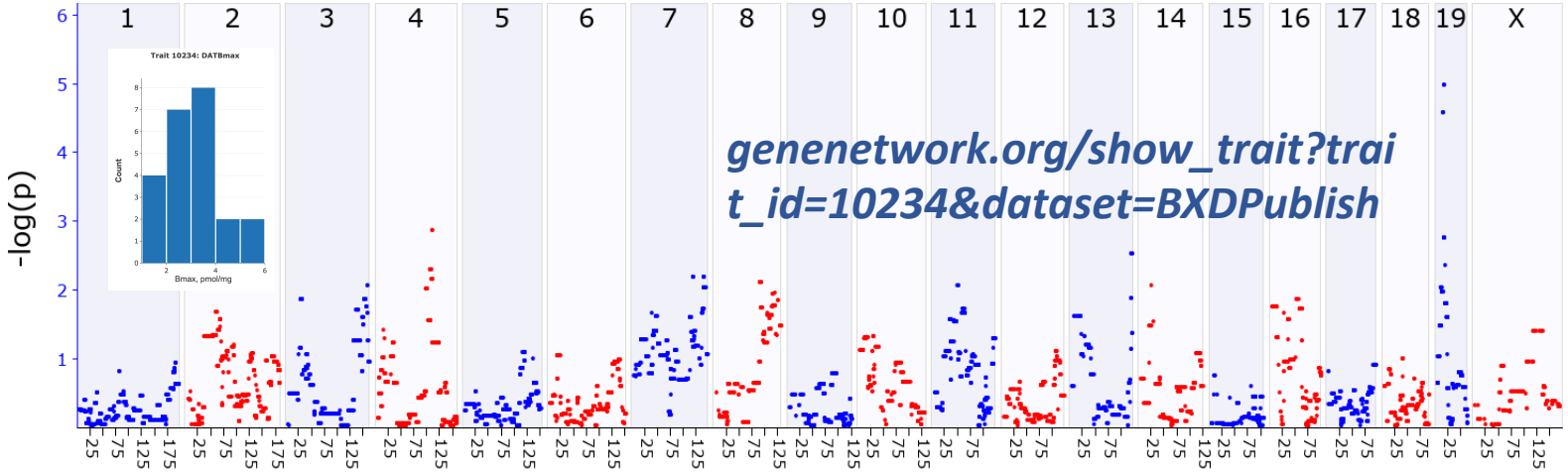
B C Jones [1], L M Tarantino, L A Rodriguez, C L Reed, G E McClearn, R Plomin, V G Erwin

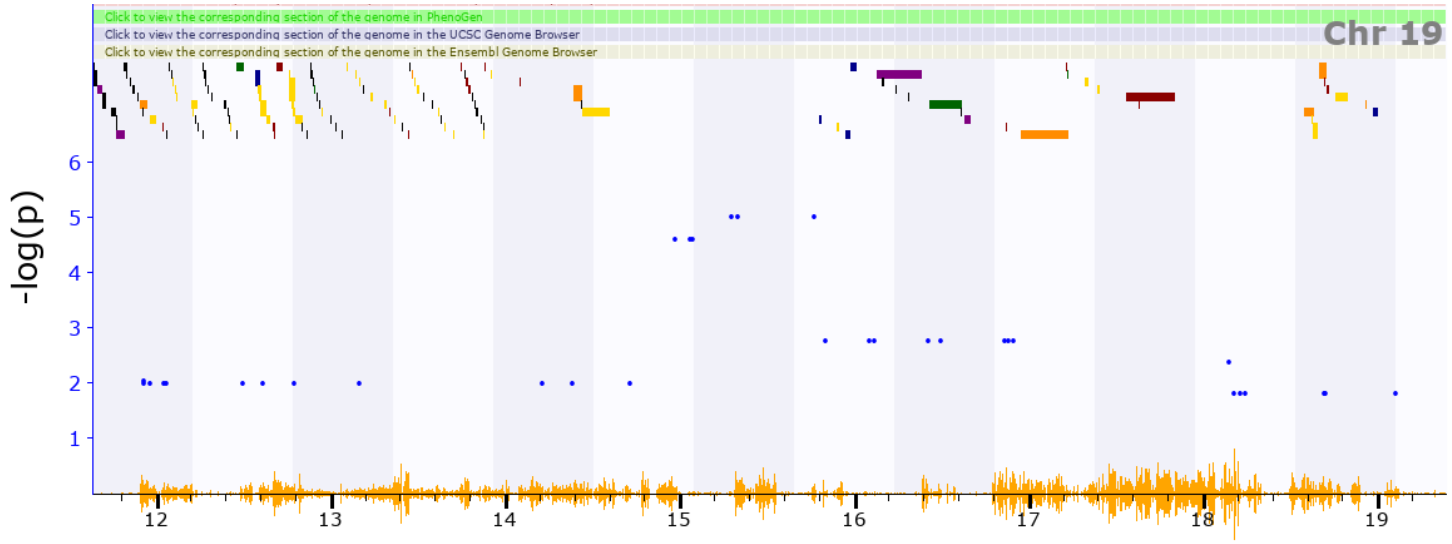
Try strategy 5 with 2nd trait

BXD_10234	Central nervous system, pharmacology, protein expression: Dopamine transporter (DAT, SLC6A3) protein density in the dorsal striatum (caudate putamen) [Bmax, pmol/mg]	3.118	Janowsky A, Mah C, Johnson RA, Cunningham CL, Phillips TJ, Crabbe JC, et al.	2001	23.8	Chr19: 15.292517	0.939

The ***Slc6a3*** gene (DAT) is located on Chr 13 at 73.6 Mb, but DAT protein activity maps to Chr 19 at 15–16 Mb with a –log(*p*) of 5.0.

Psat1 is almost the only candidate in this gene sparse region.





Case Reports > Neurosci Res. 2011 Feb;69(2):154-60. doi: 10.1016/j.neures.2010.10.003. Epub 2010 Oct 16.

A Novel Balanced Chromosomal Translocation Found in Subjects With Schizophrenia and Schizotypal Personality Disorder: Altered L-Serine Level Associated With Disruption of PSAT1 Gene Expression

Yuji Ozeki [1], Benjamin S Pickard, Shin-ichi Kano, Mary P Malloy, Mariela Zeledon, Daniel Q Sun, Kumiko Fujii, Keiko Wakui, Yukihiko Shirayama, Yoshimitsu Fukushima, Hiroshi Kunugi, Kenji Hashimoto, Walter J Muir, Douglas H Blackwood, Akira Sawa

> J Pharmacol Exp Ther. 2001 Aug;298(2):634-43.

Mapping Genes That Regulate Density of Dopamine Transporters and Correlated Behaviors in Recombinant Inbred Mice

A Janowsky [1], C Mah, R A Johnson, C L Cunningham, T J Phillips, J C Crabbe, A J Eshleman, J K Belknap

Part 3: Slide 11 Classic forward methods

Phenotypes | Search All | nicotine cocaine morphine alcohol THC amphetamine impulsivity

GN searched for the term(s) "nicotine" in 51 datasets and 13763 traits across 10 species and found 105 results that match your query.
You can filter these results by adding key words in the fields below and you can also sort results on most columns.

When is it practical to identify an almost "sure thing" candidate gene

Select and search

Species: Mouse (mm10)

Group: BXD Family

Type: Phenotypes | Info

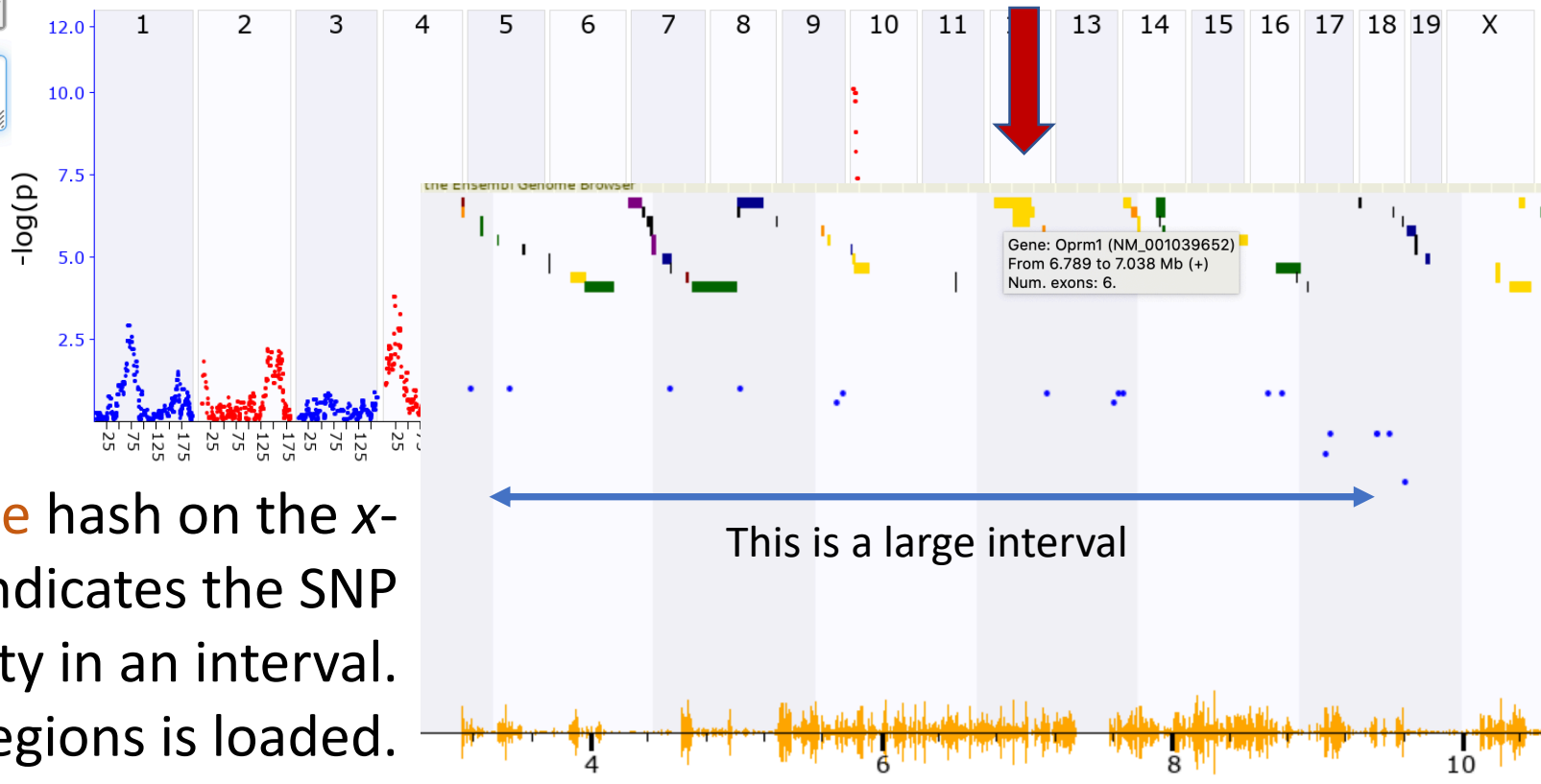
Dataset: BXD Published Phenotypes

Get Any: morphine

Showing 1 to 778 of 778 entries

	Index	Record	Description	Mean	Authors	Year	Max LRS?	Max LRS Location	Additive Effect?
☑	552	BXD_11845	Central nervous system, pharmacology, behavior: Morphine response (50 mg/kg ip), locomotion (open field) from 45-60 min after injection in an activity chamber for males and females [cm]	6427.696	Philip VM, Ansah TA, Blaha CD, Cook MN, Hamre KM, Lariviere WR, et al.	2010	39.3	Chr10: 3.180582	-3597.846

Mapping on All Chromosomes for Trait: 11845 - AMDIST60 with 198 samples
Dataset: BXD Published Phenotypes
Genotype File: BXD.geno:New (2017)
Using GEMMA mapping method with LOCO and no cofactors
Created at: 2020-06-25 21:14:44



Serendipity: A strong correspondence between well understood biology and known functions of a candidate gene. Risk is *narrative potenial* again

The orange hash on the *x*-axis indicates the SNP density in an interval. *Oprm1* regions is loaded.

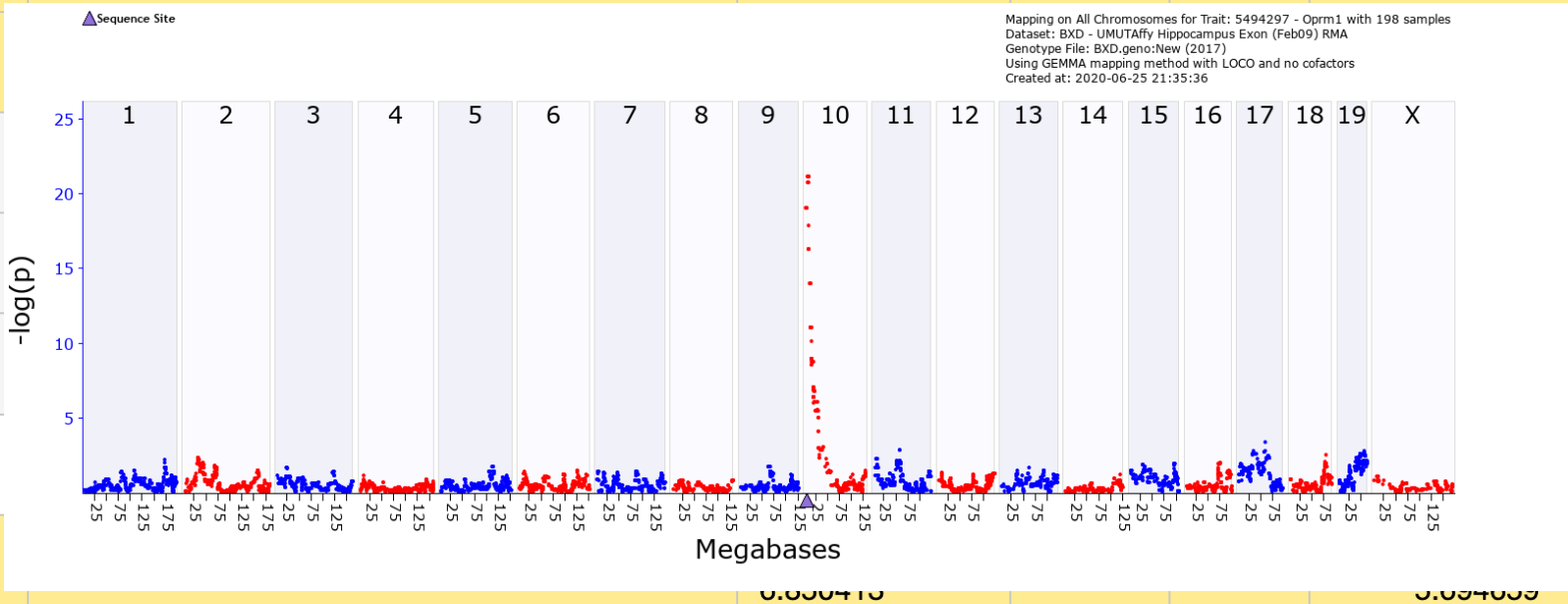
Part 3: Slide 12

GeneNetwork Intro Help Tools Collections 3 Source Code Sign out Manage Groups

Genes / Molecules | Search All | OPRM1

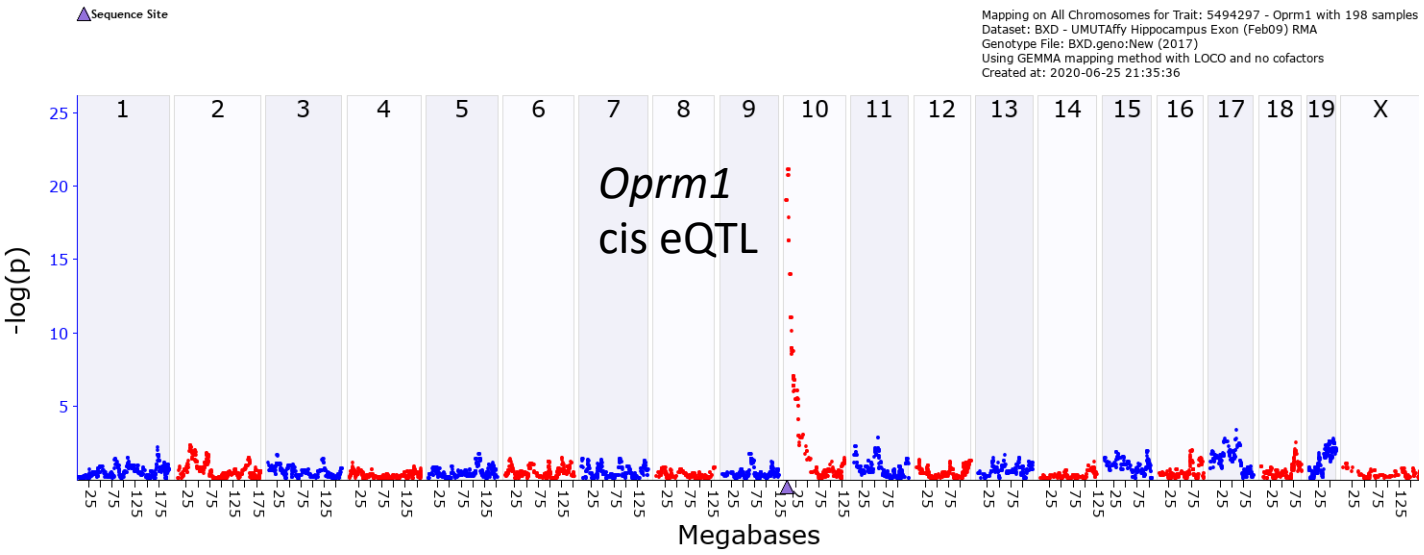
Oprm1 has multiple strong cis eQTLs in relevant brain regions

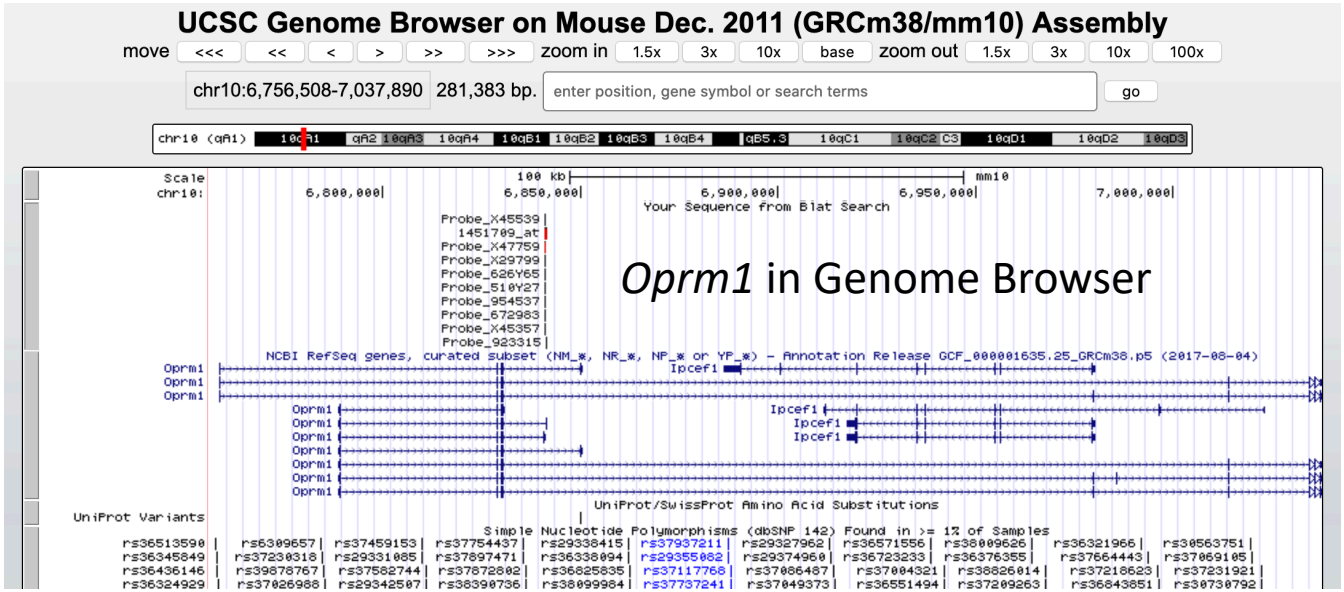
Dataset	Symbol	Description	Location	Mean	Max LRS?	Max LRS Location	Additive Effect?
UMUTAffy Hippocampus Exon (Feb09) R...	Oprm1	opioid receptor, mu 1	Chr10: 6.853588	8.670	96.3	Chr10: 5.737023	-0.712
HQF Striatum Affy Mouse Exon 1.0ST ...	Oprm1						-1.204
VCU BXD VTA EtOH M430 2.0 (Jun09) R...	Oprm1						0.531
Hippocampus Consortium M430v2 (Jun0...	Oprm1						0.209
Hippocampus Consortium M430v2 (Jun0...	Oprm1						0.643
Hippocampus Consortium M430v2 (Jun0...	Oprm1						0.182
INIA Adrenal Affy MoGene 1.0ST (Jun...	Oprm1						-0.798

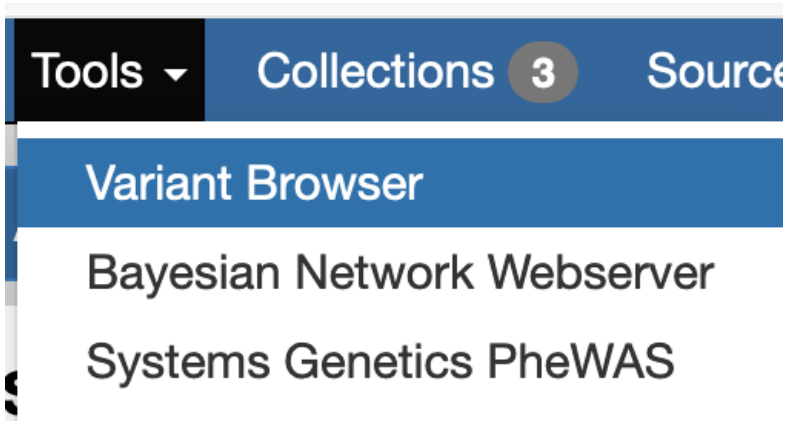


The bad news is finding **THE** causal nucleotide variants associated with *Oprm1* cis eQTLs is impractical because there are almost certainly multiple variants between *B* and *D* haplotypes at work.

Oprm1 has no known missense variants







Oprm1 has no known mis-sense variants or small indels between *B* and *D* haplotypes

Chr	Mb	Alleles	Source	ConScore	Gene	Transcript	Domain 1	Domain 2	C57BL/6J	DBA/2J
10	6.793780	G/C	SangerUCLA	1.0	*lyd* NCBI	ENSMUST00000019896	Intron	Nonsplice Site	G	C
10	6.813389	C/G	SangerUCLA	0.245	*Oprm1* NCBI		Oprm1		G	C
10	6.813923	A/T	SangerUCLA	1.0	*Oprm1* NCBI		Oprm1		A	T
10	6.813924	T/C	SangerUCLA	1.0	*Oprm1* NCBI		Oprm1		T	C
10	6.813932	C/T	SangerUCLA	1.0	*Oprm1* NCBI		Oprm1		T	C
10	6.813938	A/C	SangerUCLA	1.0	*Oprm1* NCBI		Oprm1		C	A
10	6.813939	T/A	SangerUCLA	1.0	*Oprm1* NCBI		Oprm1		A	T
10	6.813940	A/C	SangerUCLA	1.0	*Oprm1* NCBI		Oprm1		C	A

Phenotypes | Search All | nicotine cocaine morphine alcohol THC amphetamine impulsivity

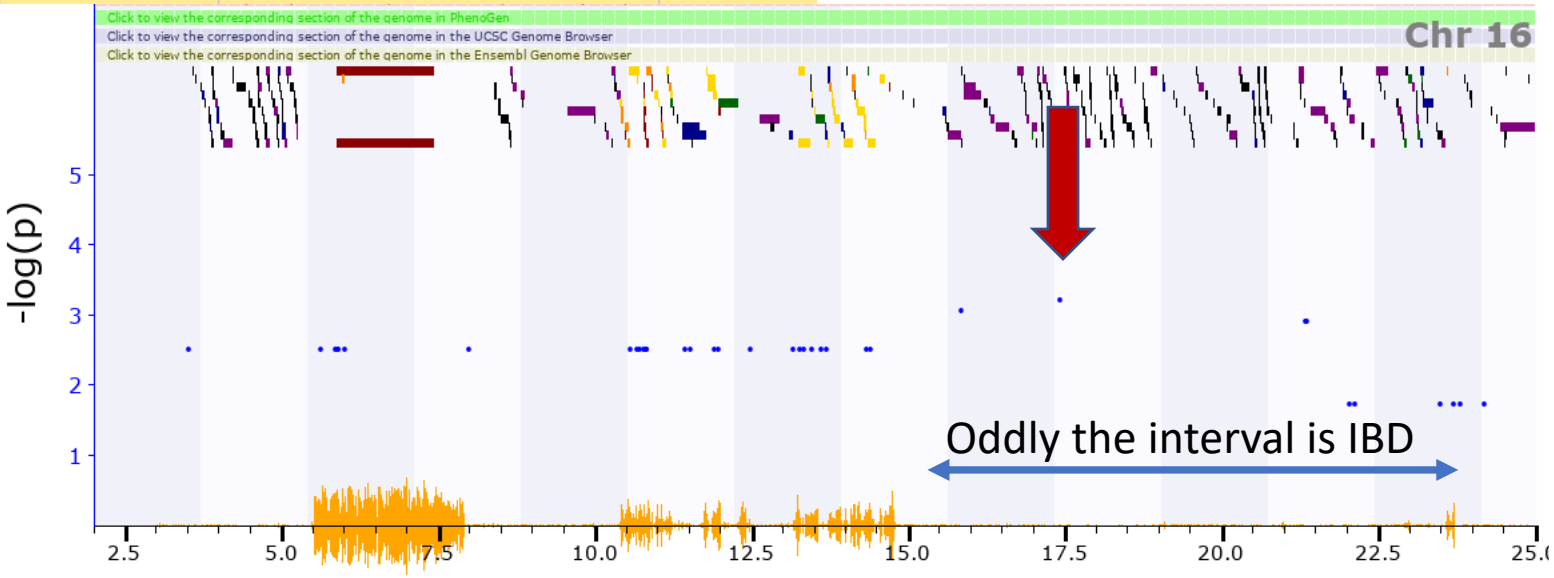
GN searched for the term(s) "nicotine" in 51 datasets and 13763 traits across 10 species and found 105 results that match your query.
You can filter these results by adding key words in the fields below and you can also sort results on most columns.

When is it practical to identify a single DNA variant?

Rarely—even in rodent models and in GWASs with high precision. Most successes are NOT pure forward genetic studies, but rely on known and rare sequence variants and reverse genetic methods. Examples include a B2 SINE in ***Comt*** and a deletion in ***Gabra2***—both in the "wildtype" B6 mouse strains (Mulligan and colleagues, 2012, 2019)

Record	Description	Mean	Authors	Year
BXD_10252	Central nervous system, pharmacology, protein expression: Dopamine receptor 1 (DRD1) binding maximum (Bmax) in membrane fragments in the dorsal striatum (caudate putamen) of males (Schering compound 3H-23390 ligand) [fmol/mg wet weight]	1530.133	Jones BC, Tarantino LM, Rodriguez LA, Reed CL, McClearn GE, Plomin R, et al.	1999

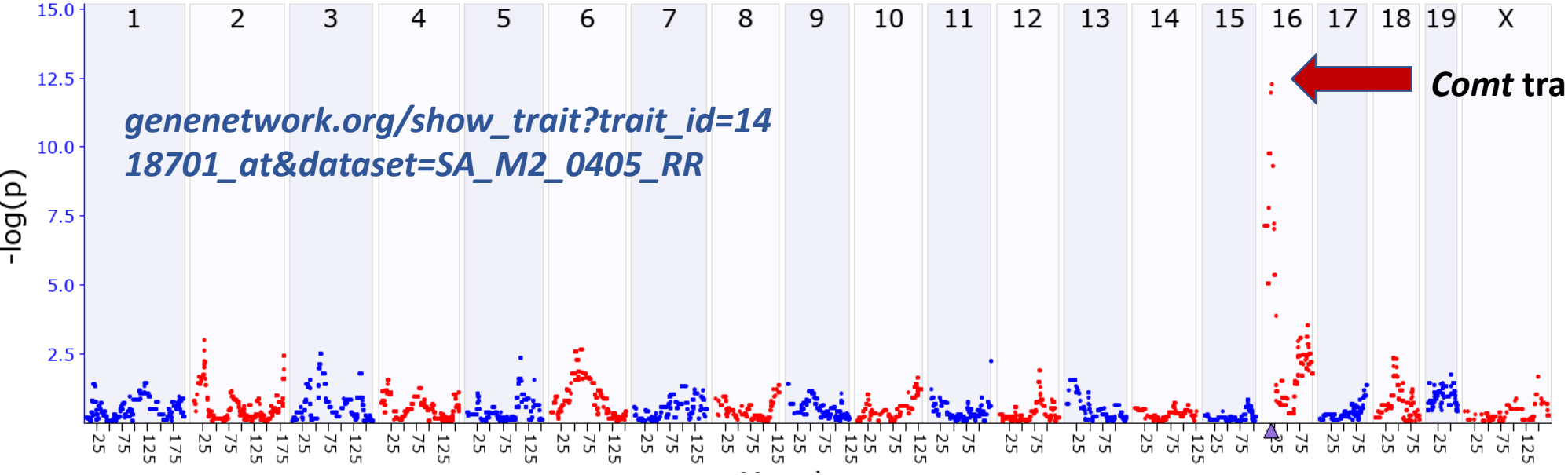
genenetwork.org/show_trait?trait_id=10252&dataset=BXDPublish



Part 4: Slide 15 Reverse genetics

Comt is a strong cis eQTL despite being in the middle of an IBD region

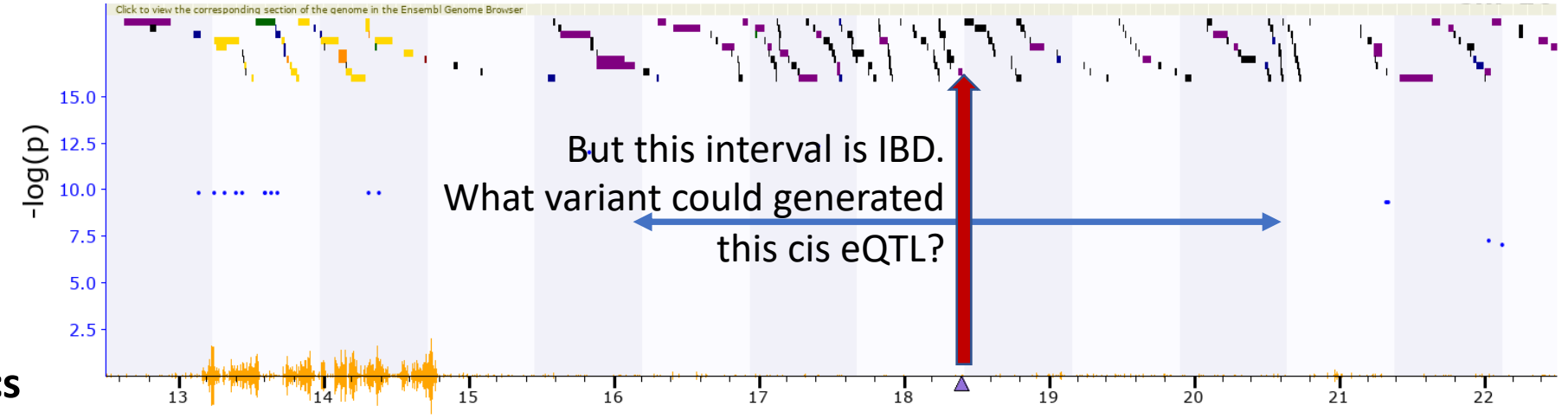
Serendipity again: The Chr 16 QTL D1R binding contains the catechol-O-methyltransferase gene (***Comt***)



Comt **transcript cis eQTL**

genenetwork.org/show_trait?trait_id=10252&dataset=BXDPublish

Record	Description	Mean	Authors	Year
BXD_10252	Central nervous system, pharmacology, protein expression: Dopamine receptor 1 (DRD1) binding maximum (Bmax) in membrane fragments in the dorsal striatum (caudate putamen) of males (Schering compound 3H-23390 ligand) [fmol/mg wet weight]	1530.133	Jones BC, Tarantino LM, Rodriguez LA, Reed CL, McClearn GE, Plomin R, et al.	1999

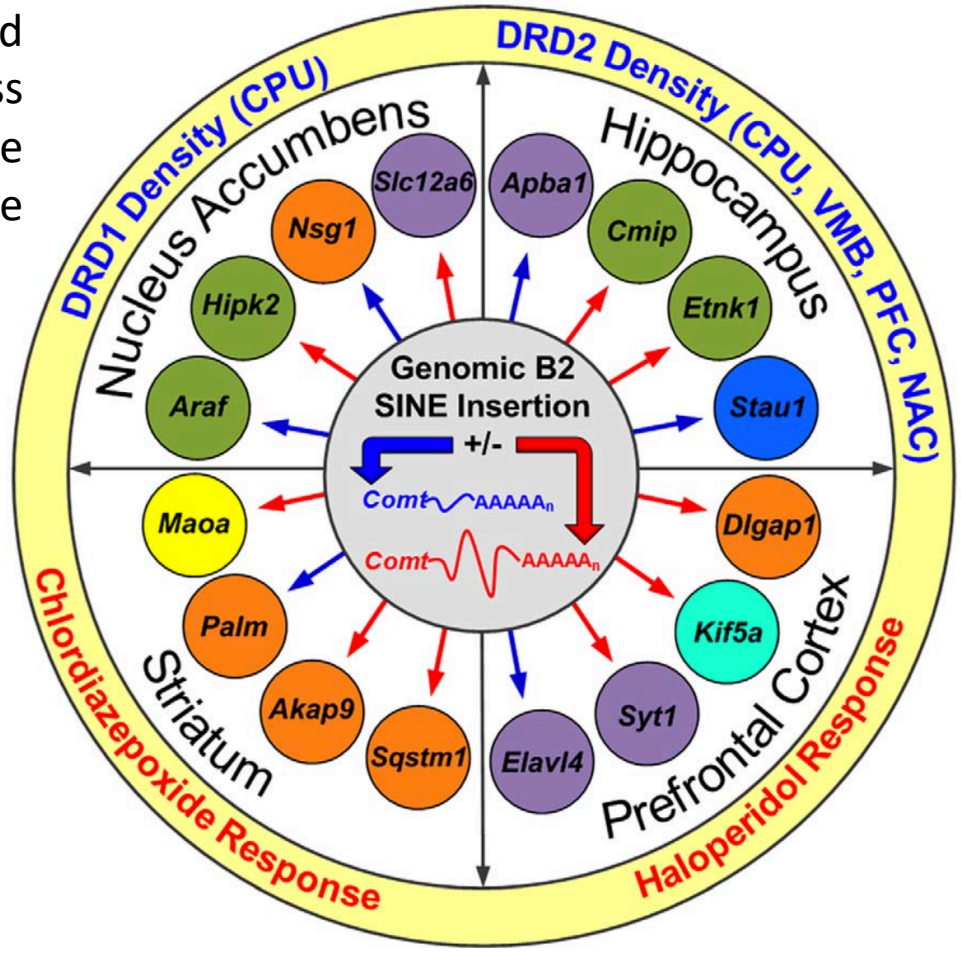


Genes / Molecules | Search All | comt

Part 3: Reverse Genetics and PheWAS

FAIR data: Start of a multiscalar phenome

Find
Access
Interoperate
R*euse*/Replicate



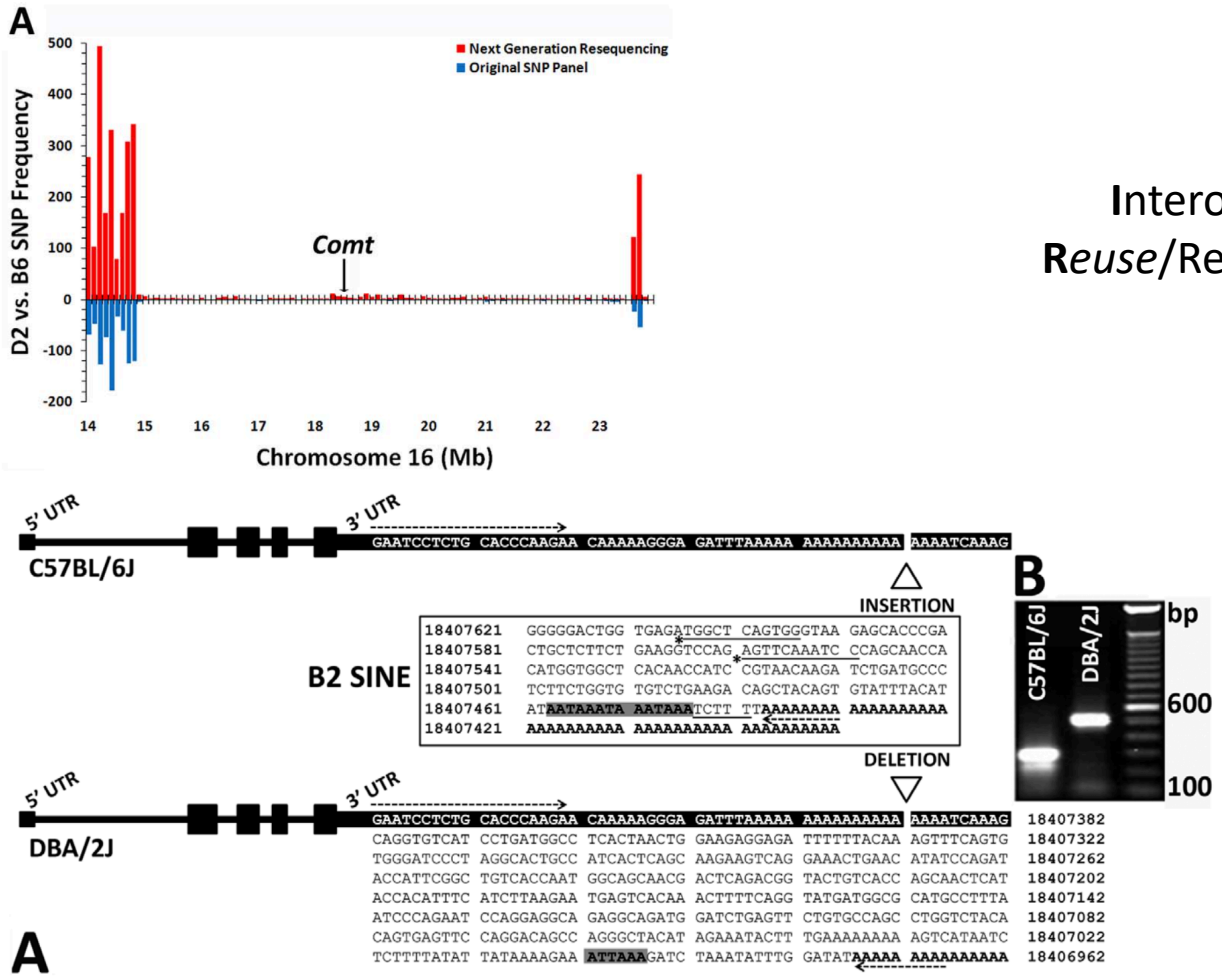


Figure 5. Insertion of a B2 SINE containing an alternative polyadenylation site into the *Comt* gene leads to the production of mRNA containing a shorter 3′ UTR in B6. (A) Gel electrophoresis of 3′ RACE products shows that D2 produces mRNA containing a 3′ UTR that is

› PLoS One. 2010 Aug 17;5(8):e12181. doi: 10.1371/journal.pone.0012181.

2010

A Transposon in Comt Generates mRNA Variants and Causes Widespread Expression and Behavioral Differences Among Mice

Zhengsheng Li [1], Megan K Mulligan, Xusheng Wang, Michael F Miles, Lu Lu, Robert W Williams

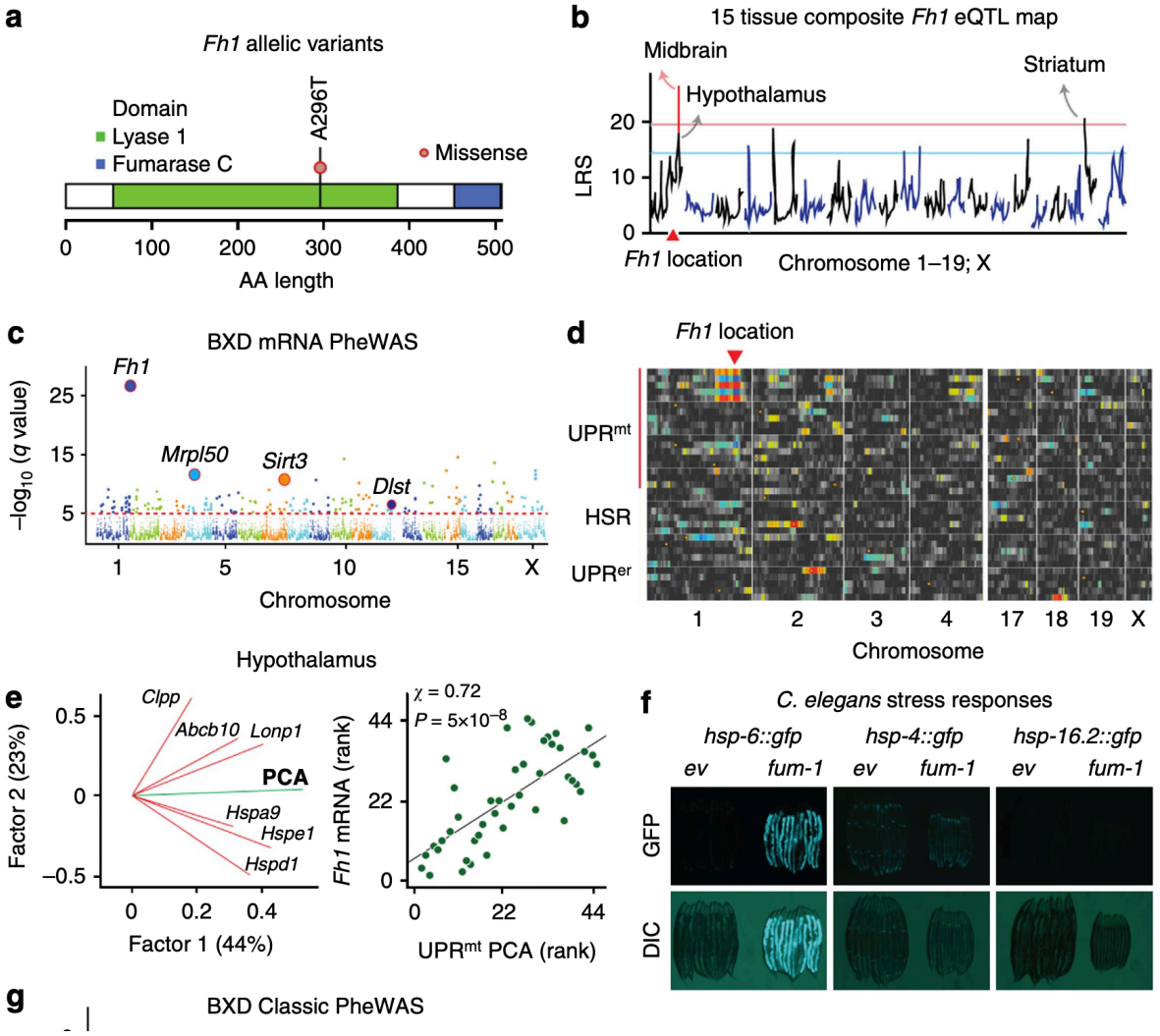
Part 4: Slide 17 or Experimental PheWAS

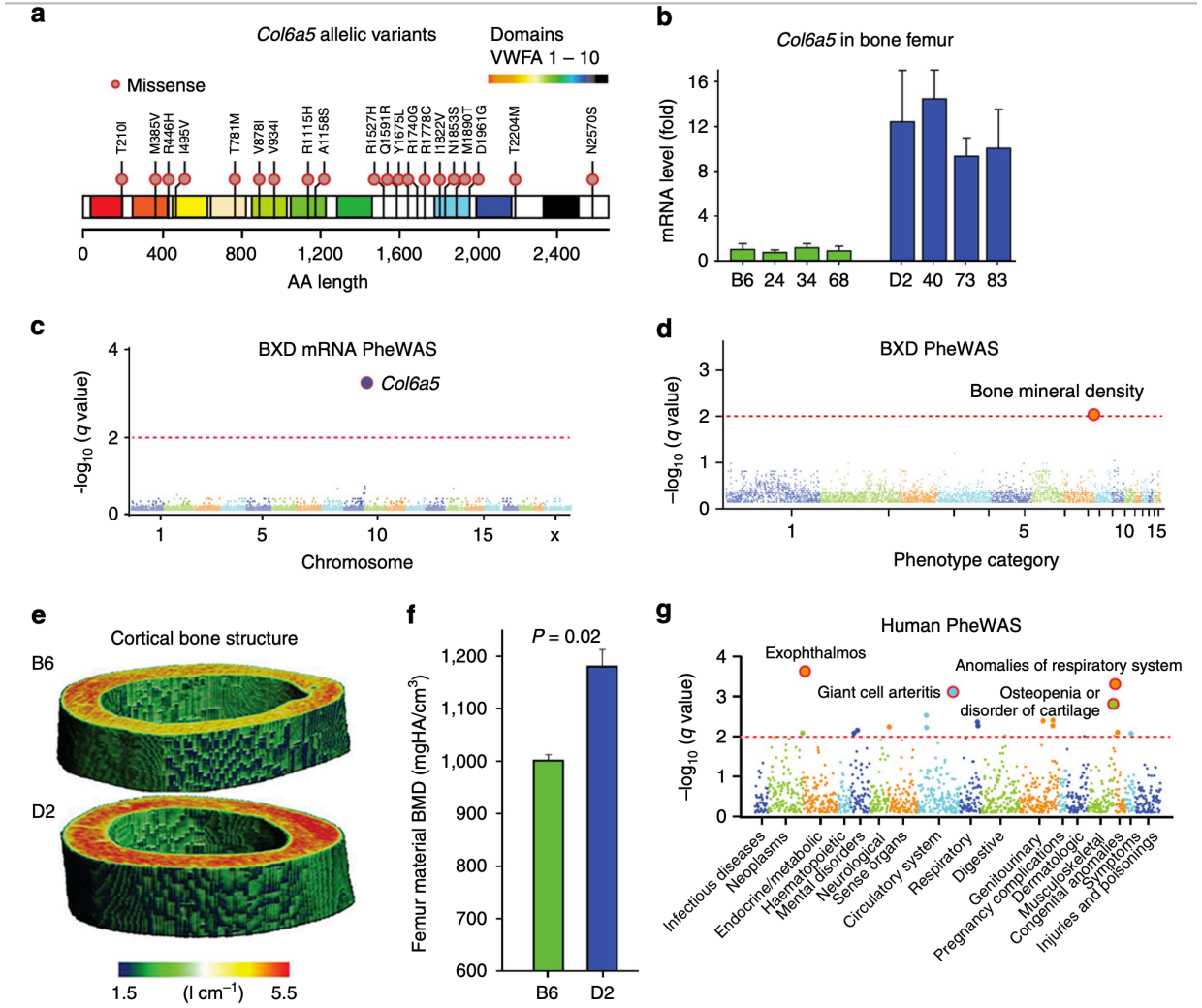
Two problems: Generating a phenome and seeing through the LD

Received 11 Jun 2015 | Accepted 11 Dec 2015 | Published 2 Feb 2016 DOI: 10.1038/ncomms10464 OPEN

Joint mouse–human phenome-wide association to test gene function and disease risk

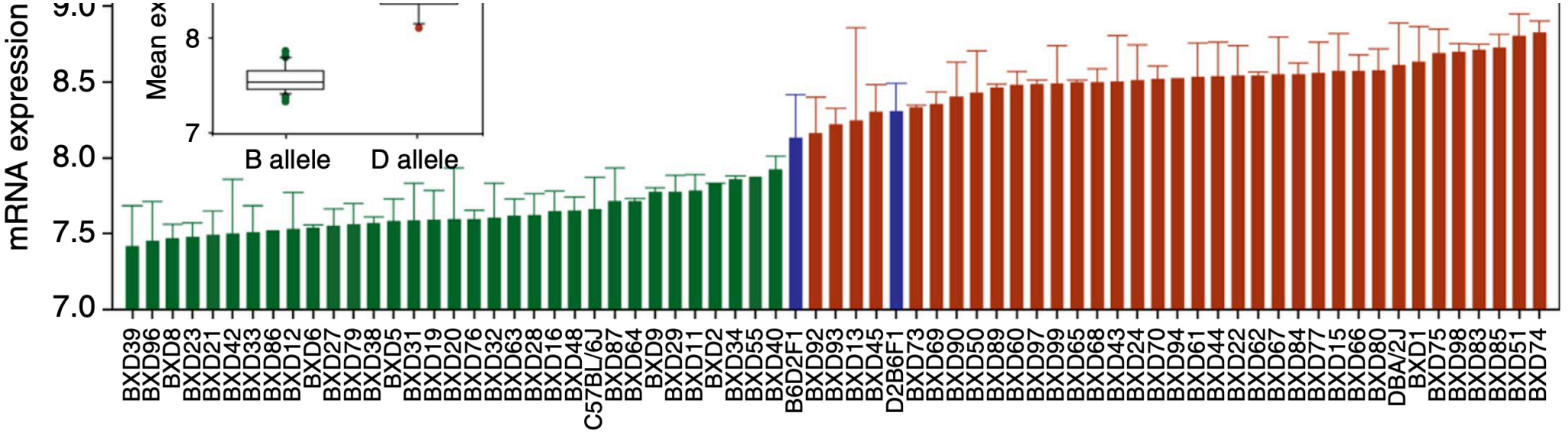
Xusheng Wang[1,2,*], Ashutosh K. Pandey[1,*], Megan K. Mulligan[1], Evan G. Williams[3], Khyobeni Mozhui[1],

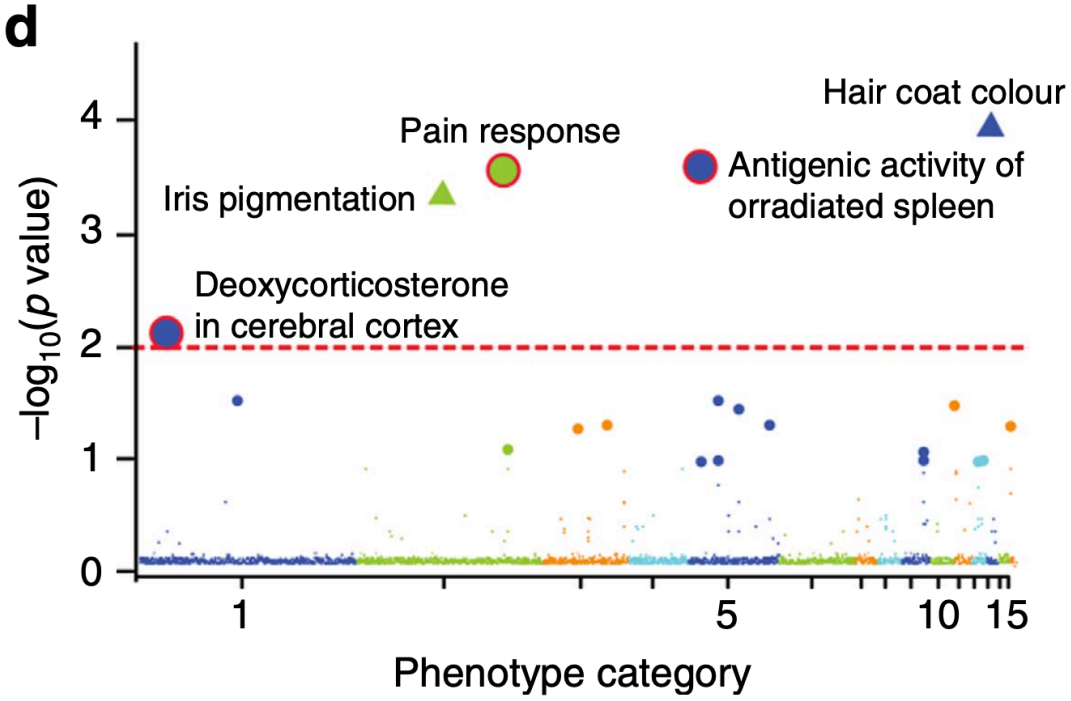


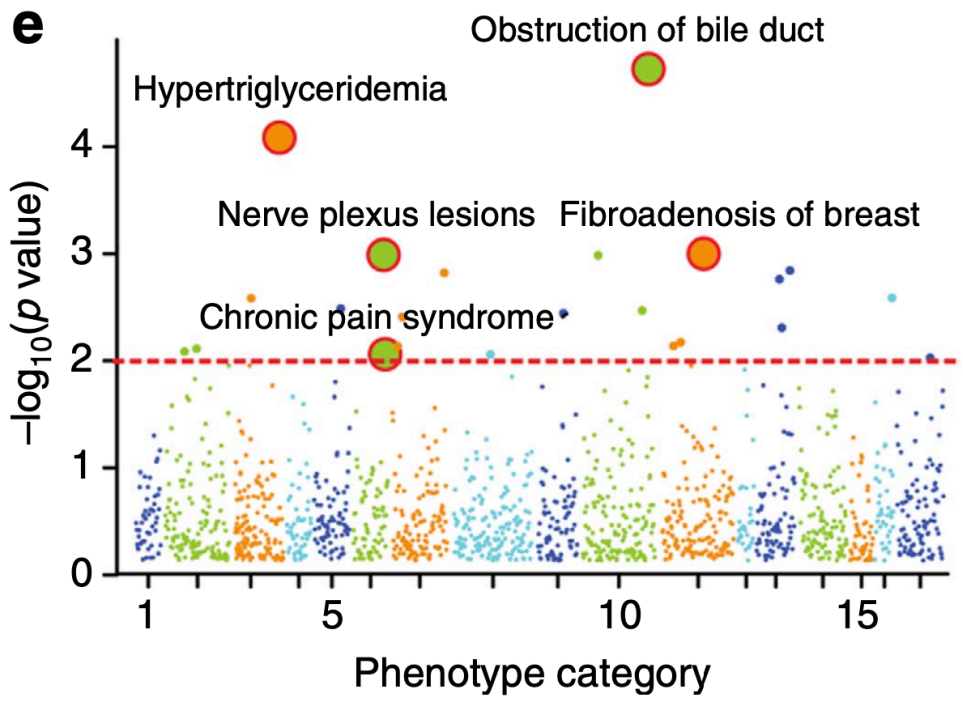


Part 4: Slide 18: PheWAS in rodents requires a phenome

AHR PheWAS in mouse and human (BioVU, Josh Denny and team)







Part 4: Slide 19

Catalog of 12,000 missense variants segregating in the BXD family

This is now practical for HXB/BXH rat family and will soon be possible using HRDP.
Here are about ~12,000 missense mutations segregating in the BXDs in 2015.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Authors: Wang X, Pandey AK, Mulligan MK, Williams EG, Mozhui K, Li Z, Jovaisaite V, Quarles LD, Xiao Z, Huang J, Capra JA, Chen Z, Taylor WL, Bastarache L, N												
2		Title: Joint mouse-human phenome-wide association to test gene function and disease risk												
3		Journal: Nature Communications 7:10464												
4		Date last updated or submitted: 5/9/2015												
5		Contact emails: labwilliams@gmail.com, xushengwang78@gmail.com, ashutoshmits@gmail.com												
6		PMID: 26833085												
7		These missense SNPs can be found at URL: http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=UTHSCSEQ												
8		Grantham score and Effect were calculated using in-house python scripts. (Amino acid difference formula to help explain protein evolution, Science, 1974, PMID: 48												
9														
10														
11	**Supplementary Table 4.** Nonsynonymous SNPs segregating in the BXD family													
12	**Index**	**Chr**	**Position bp (mm10)**	***B* allele**	***D* allele**	**Codon change**	**Position of AA change**	**Gene symbol**	**Transcript ID**	**Grantham score**	**Grantham Effect**	**SIFT**	**Polyphen**	**SIFT and Polyphen**
13	1	1	4,344,820	T	C	aAt/aGt	N2023S	*Rp1*	ENSMUST00000027032	46	Conservativ	-	Yes	-
14	2	1	4,344,992	C	T	Gat/Aat	D1966N	*Rp1*	ENSMUST00000027032	23	Conservativ	-	Yes	-
15	3	1	4,349,357	C	T	Ggt/Agt	G511S	*Rp1*	ENSMUST00000027032	56	Moderately	-	-	-
16	4	1	4,352,525	T	C	Aag/Gag	K101E	*Rp1*	ENSMUST00000027032	56	Moderately	-	-	-
17	5	1	5,070,062	A	C	aTt/aGt	I39S	*Rgs20*	ENSMUST00000118000	142	Moderately	-	-	-
18	6	1	6,214,740	G	C	Ccg/Gcg	P185A	*4732440D04Rik*	ENSMUST00000097832	27	Conservativ	-	-	-
19	7	1	6,214,823	G	C	cCa/cGa	P157R	*4732440D04Rik*	ENSMUST00000097832	103	Moderately	-	-	-
20	8	1	6,215,064	G	C	Ccg/Gcg	P77A	*4732440D04Rik*	ENSMUST00000097832	27	Conservativ	-	-	-
21	9	1	6,240,127	A	T	Acg/Tcg	T250S	*Rb1cc1*	ENSMUST00000027040	58	Moderately	-	Yes	-
11985	11973	X	153,558,848	G	A	cGt/cAt	R79H	*Cypt3*	ENSMUST00000112573	29	Conservativ	Yes	-	-
11986	11974	X	164,948,084	G	A	Cca/Tca	P288S	*Mospd2*	ENSMUST00000004715	74	Moderately	-	-	-
11987	11975	X	164,991,671	A	G	Atg/Gtg	M435V	*Fancb*	ENSMUST00000101082	21	Conservativ	-	-	-
11988	11976	X	164,991,672	T	A	aTg/aAg	M435K	*Fancb*	ENSMUST00000101082	95	Moderately	-	-	-
11989	11977	X	166,180,639	G	A	cGt/cAt	R97H	*Gemin8*	ENSMUST00000130880	29	Conservativ	-	-	-
11990	11978	X	170,673,769	G	C	Ggg/Cgg	G78R	*AB512673.1*	ENSMUST00000178693	125	Moderately	-	-	-
11991	11979	X	170,676,481	T	C	Tgc/Cgc	C242R	*AB512673.1*	ENSMUST00000178693	180	Radical	-	-	-

ReadMe Summary | Supplementary Table 1 | Supplementary Table 2 | Supplementary Table 3 | **Supplementary Table 4** | Supplementary Table 5 | Supplementary Table 6 | Supplementa

Part 4: Slide 20 *(show **Supplement table 4** live from Wang et al. 2016)*

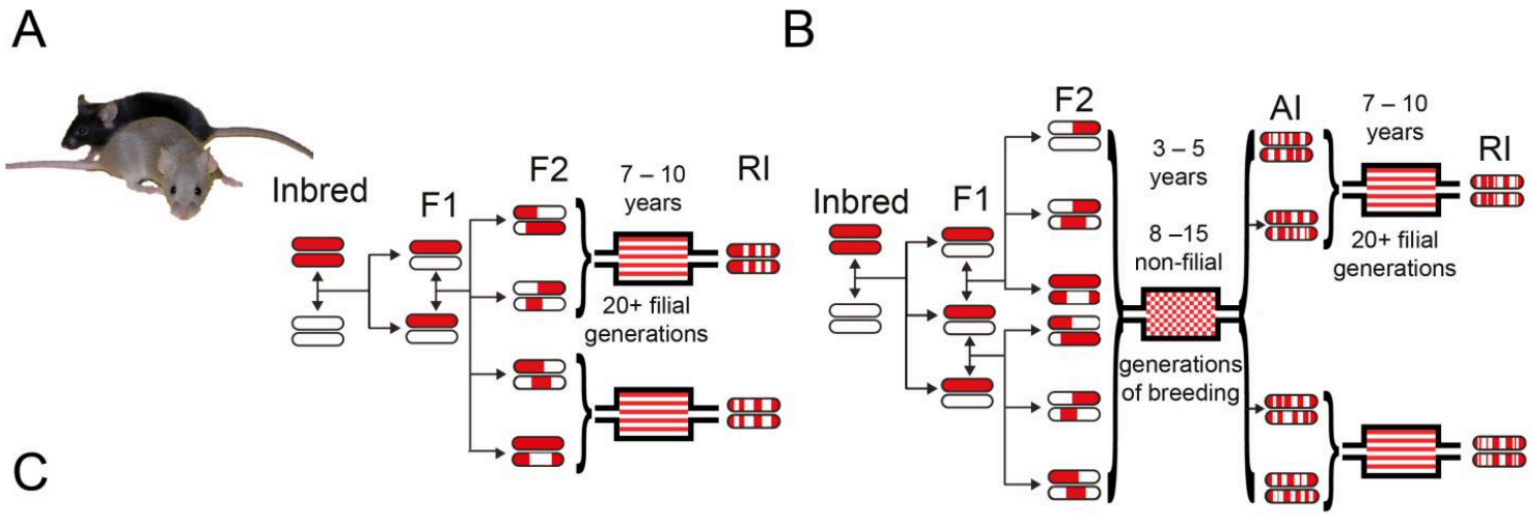
Interlude: a set of ~20 identified genes

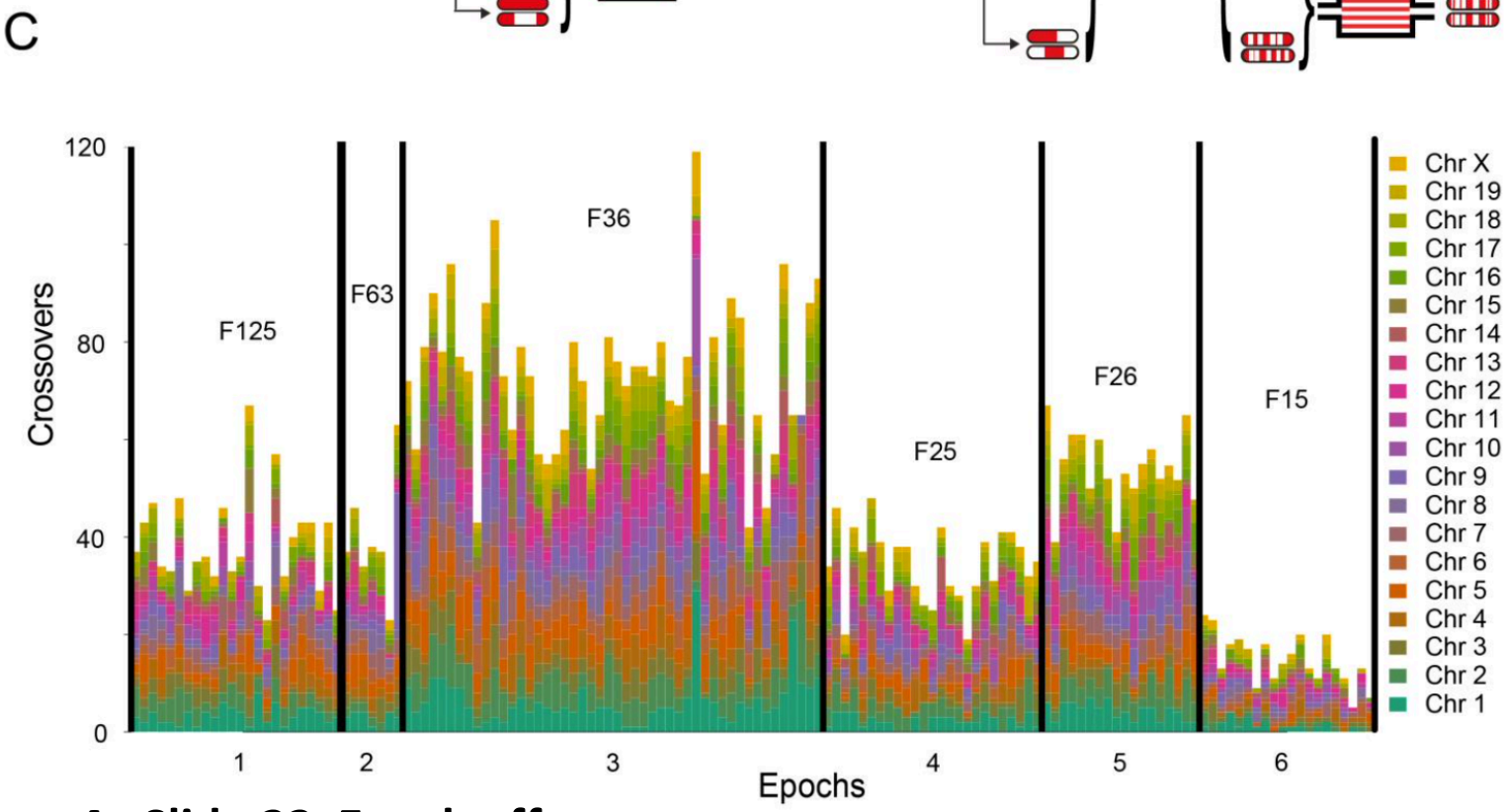
Pum2 for control of translation (Journal of Cellular Physiology, Scott RE et al 2005, PMID: 15617101)
Igpp2 and ***Irgb10*** for *Chlamydial* infection (Journal of Immunology, Miyari I et al 2007, PMID: 17641048, PMID: 22438999)
Fmn2 for control of tRNA synthases in brain (PLoS Genetics, Mozhui et al 2008, PMID: 19008955)
Ubp1 for blood pressure (PLoS Genetics, Koutnikova et al 2009, PMID: 19662162)
Comt for drug responses (PLoS One, Li, Mulligan et al 2010, PMID: 20808911)
Typr1 and ***Gpnmb*** for ocular phenotypes (Molecular Vision, Lu et al 2011)
Alpl for bone metabolism (Cell, Andreux et al 2012, PMID: 22939713)
Gabra2 for gene expression control and behavior (PLoS One and in submission, Mulligan et al 2012, PMID: 22506031 and Mulligan et al., 2019 in press)
Mrps5 for longevity and cognitive decline (Nature, Houtkooper et al 2013, PMID: 23698443)
Klrd1 for immune function (G3, Shin et al 2014, PMID: 25520036)
Ahr for locomotor active (PLoS Genetics, Williams EG et al 2014, PMID: 25255223)
Hp1bp3 and ***Mrp*** gene family for cognitive aging (Neurobiology of Aging, Neuner et al 2016, PMID: 27460150; Frontiers in Genetics, PMID: 28983317)
Dhtkd1 for metabolism (Cell, Wu et al 2014, PMID: 25215496)
D2hgdh for causal d-2-hydroxyglutaric acid metabolite effect (Science, Williams et al 2016, PMID: 27284200)
Mlycd for causal C3-dicarboxylcarnitine metabolite effect (unpublished, Houten et al)
Taar1 for methamphetamine addiction (PLoS One, Shi et al 2016, PMID: 27031617)
Echdc1 and ***Mmab*** for cholesterol metabolism (Science, Williams EG et al 2016, PMID: 27284200)
Bckdha, ***Bckdhb***, and ***Cox7a2l*** for amino acid degradation pathways and mitochondrial function (Science, Williams EG et al 2016, PMID: 27284200)
Cacna2d1 for intraocular pressure (Nature Communications, Chintalapudi et al 2017, PMID: 29176626)
Atf4 and ***Fh1*** for mitochondrial stress (Journal of Cell Biology, Quiros et al 2017, PMID: 28566324)

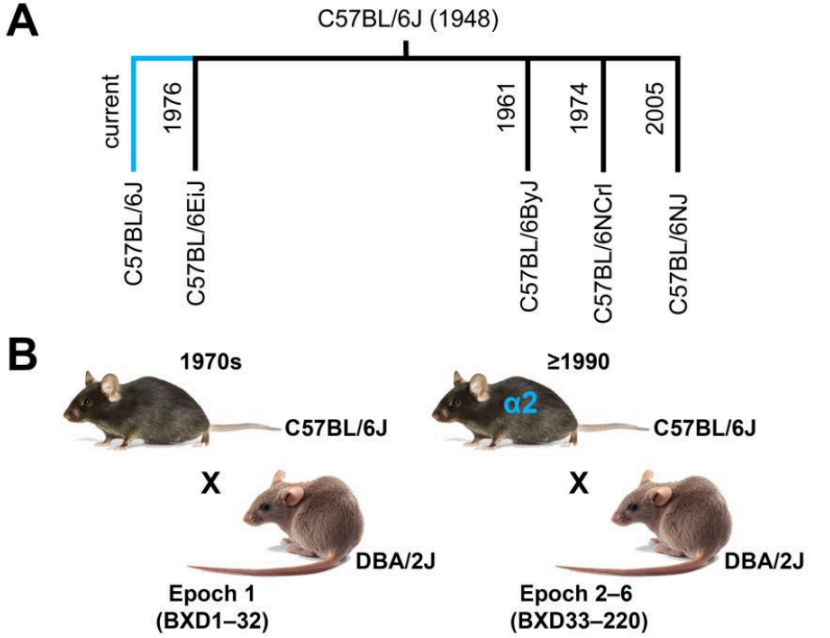
Part 4: Slide 21 Hits as of 2017

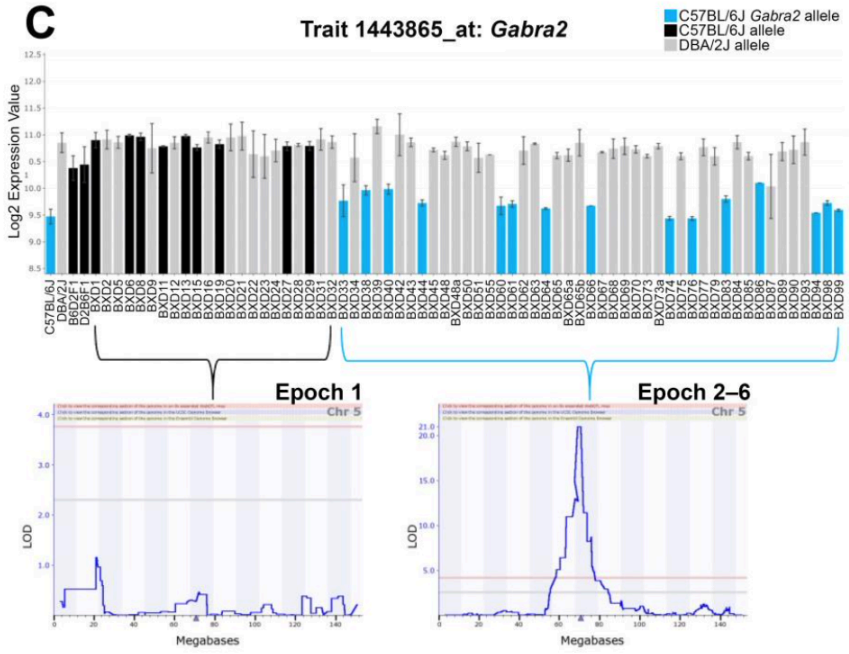
docs.google.com/document/d/1nZJydWEcFQSYdTcZwk_xwoxMjNVL_Ej9bLNEEbANkxs/edit

Epoch effects: A great resource rather than a nuisance (David Ashbrook)









Part 4: Slide 22 Epoch effect

Epoch effects and Gabra2 *(Mulligan and team)*

1443865_at	Gabra2	gamma-aminobutyric acid (GABA-A) receptor, subunit alpha 2 (alcoholism candidate gene); mid distal 3' UTR (long 3' UTR isoform, PM2L1 region)	Chr5: 70.959212	11.271	32.1	Chr5: 70.742059	0.497

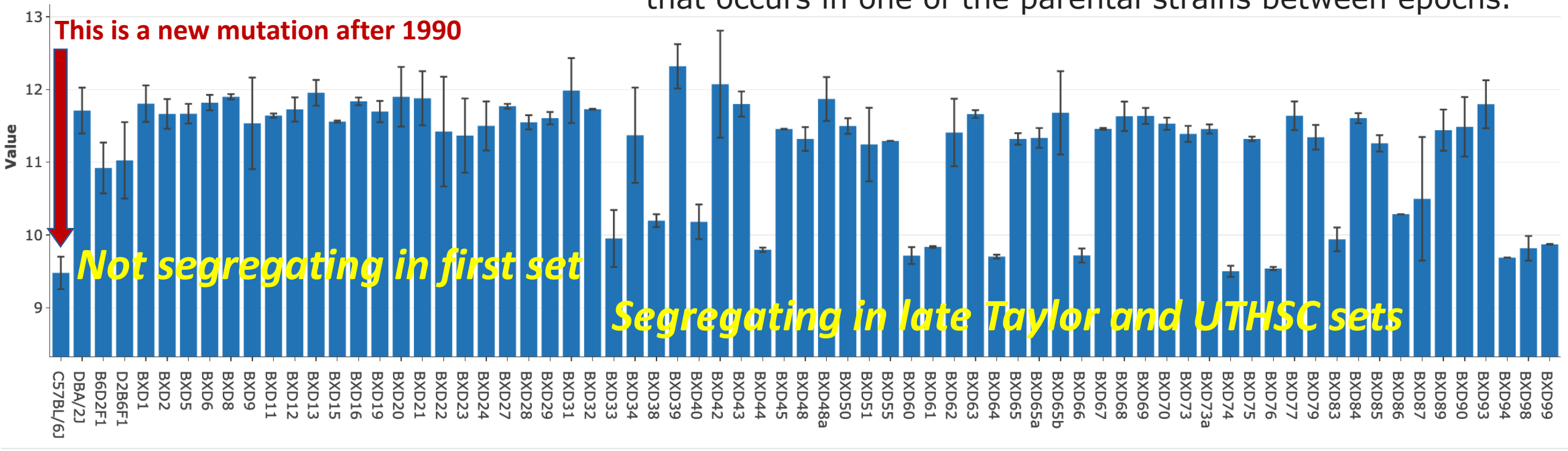
genenetwork.org/show_trait?trait_id=1443865_at&dataset=HC_M2_0606_P

Identification of a Functional Non-coding Variant in the GABA$_A$ Receptor α2 Subunit of the C57BL/6J Mouse Reference Genome: Major Implications for Neuroscience Research

Megan K. Mulligan[1], Timothy Abreo[1], Sarah M. Neuner[2,3], Cory Parks[1],*

frontiers in Genetics

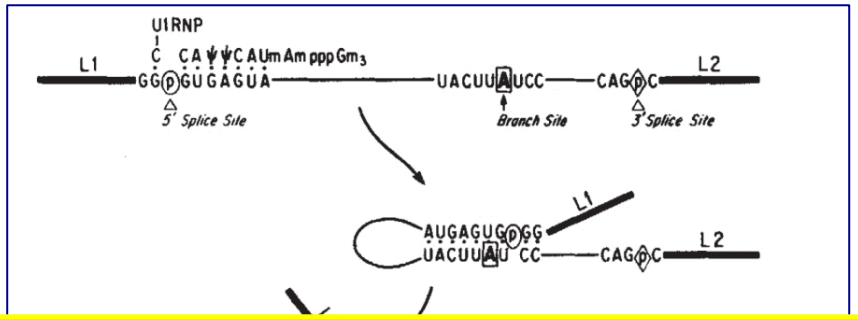
An "epoch effect" is due to fixed spontaneous mutations that occurs in one of the parental strains between epochs.



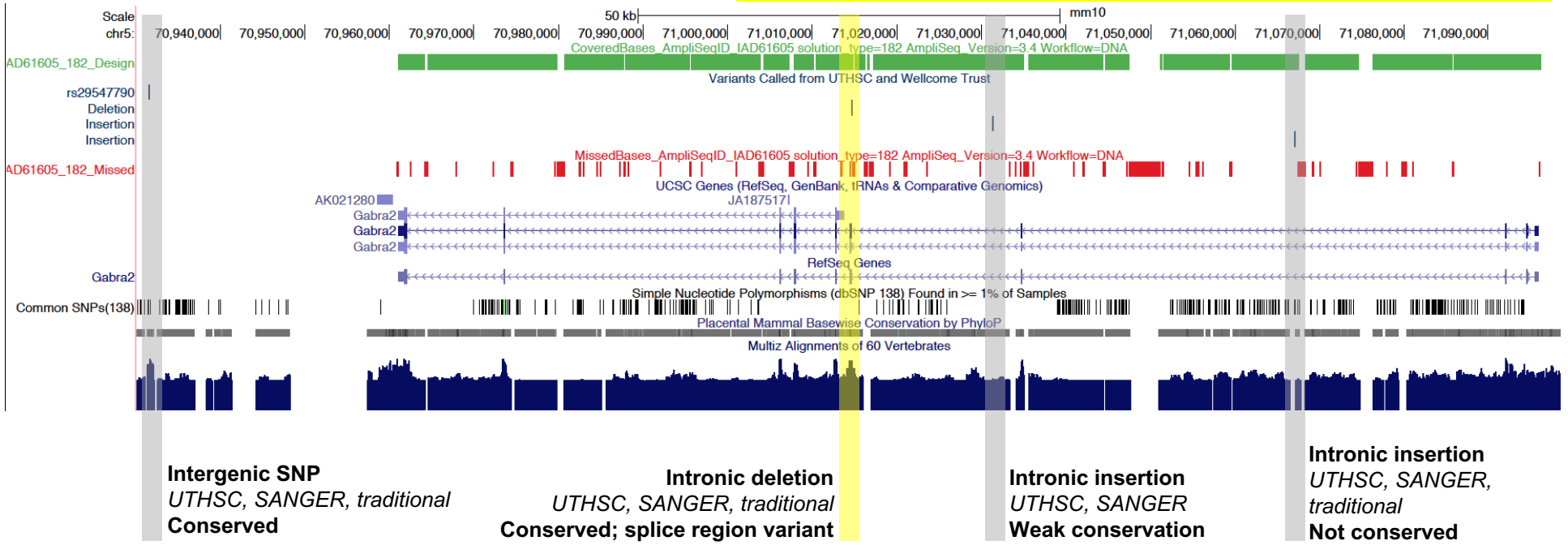
Part 4: Slide 23 A hybrid method to define DNA variants (forward) and to downstream phenotypes (reverse)

Sequencing of original B *haplotype and new mutant* B *haplotype*

Four putative variants but only one that was a strong contender—a 1 nt deletion adjacent to the exon 5 splice acceptor site.




```
Human: 5′-AACAUUUACUAUUUUCAAAAAAAUUCUA-UUUUUUAUUUCAGGAAUAUACA-3′ 50
Dog:   5′-AACAUUUACUAUUUUCCAAAGAAUUCUA-UUUUAUAUUUCAGGAAUAUACA-3′ 50
Rat:   5′-AACUUUUACUGUUUUCAAAAAGUUGUAAAAUUUAUAUUUUAGGAAUAUACA-3′ 51
B6J:   5′-AACAUGUACUGUUUUCAAAGGAAUUGUAAAUUUAUAUUU-AGGAGUAUACA-3′ 50
B6NJ:  5′-AACAUGUACUGUUUUCAAAGGAAUUGUAAAUUUAUAUUUUAGGAGUAUACA-3′ 51
          *** * **** ***** **    *   *  *** ***** **** ******
```



Intergenic SNP
UTHSC, SANGER, traditional
Conserved

Intronic deletion
UTHSC, SANGER, traditional
Conserved; splice region variant

Intronic insertion
UTHSC, SANGER
Weak conservation

Intronic insertion
UTHSC, SANGER, traditional
Not conserved

Why is this mutation valuable? (Mulligan and team)

Now possible to make a reduced complex cross between B6J (the mutant) and B6N (the wildtype) and evaluate linkage of gene expression (compensatory changes in expression of other GABA-A receptors), behavior, and drug phenotypes to this ***Gabra2***—the only functional polymorphism in this region.

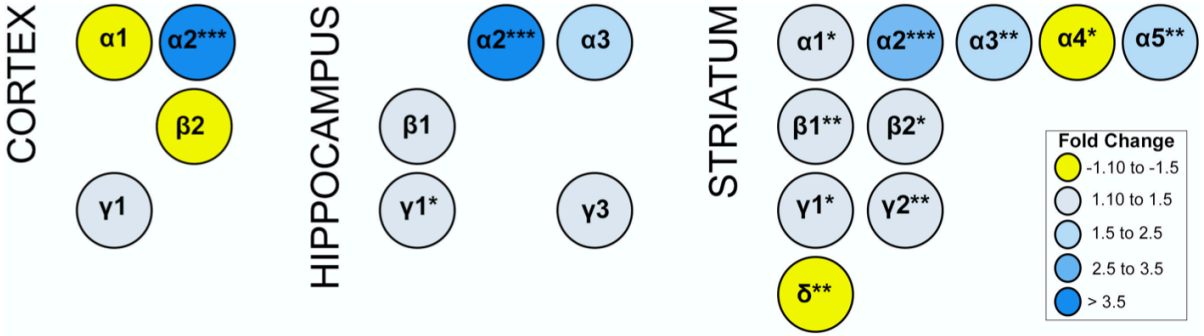


FIGURE 4 | Expression of GABA-A receptor subunit mRNA. Expression generated using the Affymetrix Clariom D Assay (microarray platform). Only subunits with significant or suggestive ($p < 0.1$) differential expression between B6J and KI *Gabra2* genotypes are shown. Significance defined as: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Fold change is indicated by color intensity with yellow representing increased expression in *Gabra2*$^{B6J/B6J}$ (B6J allele) mice relative to *Gabra2*$^{KI/KI}$ (KI allele) mice. In contrast, blue represents decreased expression in B6J allele mice relative to KI allele mice. Alterations in the mRNA levels of several alpha subunits,

frontiers in Genetics

Identification of a Functional Non-coding Variant in the GABA$_A$ Receptor α2 Subunit of the C57BL/6J Mouse Reference Genome: Major Implications for Neuroscience Research

Megan K. Mulligan[1], Timothy Abreo[1], Sarah M. Neuner[2,3], Cory Parks[1],*

What is a reduced complexity cross?

It is a new post-genomic cross type that requires full genome sequence for two very closely related substrains—for example C57BL/6J and C57BL/6NJ.

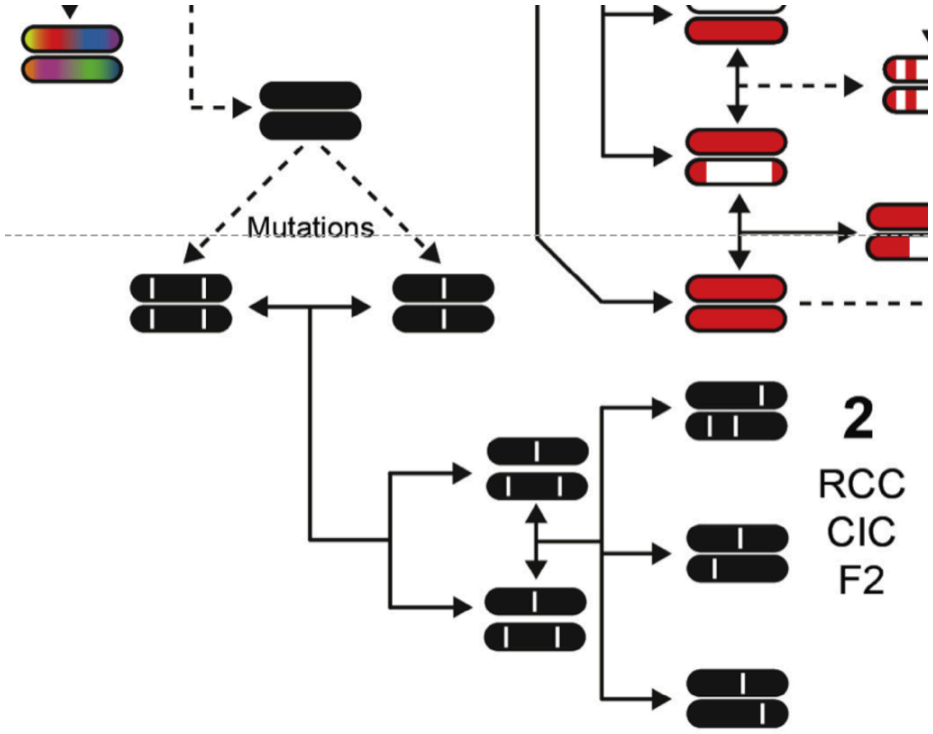
First used by Glenn Rosen, Megan Mulligan, and Vivek Kumar. An RCC was used by Kumar, Takahashi and colleagues to prove ***Cyfip2*** is involved in cocaine response.

Kumar V et al. (2013) C57BL/6N mutation in Cytoplasmic FMRP interacting protein 2 regulates cocaine response. Science 342: 1508

Part 4: Slide 25 The molecular cascade

The reduced complexity cross as a hybrid method

RCC is a post-genomic hybrid G2P method.
Why? In order to map an RCC you need to find 100+ very rare spontaneous mutations that distinguish pairs of substrains that you can use as markers for “classic” mapping of the RCC.



Facilitating Complex Trait Analysis via Reduced Complexity Crosses *Trends in Genetics* 2020

Camron D. Bryant,[1,*] Desmond J. Smith,[2] Kathleen M. Kantak,[3] Thaddeus S. Nowak, Jr[4] Robert W. Williams,[5] M. Imad Damaj,[6] Eva E. Redei,[7] Hao Chen,[8] and Megan K. Mulligan[5]

Mouse progenitor strains	Sequenced mouse substrains (on miniMUGA array)	Behavioral differences	Physiological, and/or disease model differences	Cellular differences	Molecular differences
A	A/J, A/JOlaHsd		Muscle dysfunction [T1]		
BALB/c	BALB/cJ, BALB/cByJ	Aggression [T2], alcohol preference [T3], anxiety-like behavior [T4], cognitive flexibility [T5], inhibitory control [T6], epilepsy and neuroanatomical abnormalities [28] [T7]	Allergic orchitis and encephalomyelitis [T8,T9], immune response to infection [T10], Grave’s hyperthyroidism [T11], experimental arthritis and spondylitis [T12], GABA transmission and anterior cingulate volume [T13,T14], cardiac calcinosis [T15], dexamethasone-induced osteonecrosis [T16], diet-induced fatty liver [T17], streptozotocin-induced diabetes [T18]	Sperm abnormalities [T19], antibody-mediated immunity [T20], hepatocyte invasion following infection [T21], virus-induced demyelination [T22]	Copy number variants [T23], amino acid and monoamine neurotransmitter content in caudate [T24]
C3H	C3H/HeJ, C3H/HeNCrl, C3H/HeNRj, C3H/HeH, C3H/HeNHsd, C3H/HeNTac	Nest building [T25], paw preference [T26]	Skeletal [T25], immune reactivity [T27], LPS responsiveness [T28], experimental leprosy [T29], spontaneous colitis [T30], experimental arthritis and spondylitis [T31], absence seizures [T32]	Cytotoxic activity of lymphocytes in cancer model [T33]	Toll-like receptor 4 [T34], Gpr179 [T35]
C57BL/6	C57BL/6NJ, C57BL/6NCrl, C57BL/6JBomTac, C57BL/6ByJ, C57BL/6JOlaHsd, C57BL/6N-Tyr<c>/BrdCrCrl, C57BL/6NJRj	Several; reviewed in [5], see also [6], [T36], and main text, corticosterone-induced depressive-like behaviors [T37]	Several; stroke [25], metabolic traits [T38], immune response [T39]; see also [6], kidney stones [T40], severity of Dravet syndrome model with Scn1a ± [T41],	Several [6]; cardiac fibrogenic response to angiotensin [T53], acetaminophen-induced hepatotoxicity [T54], hypoxic-ischemic brain	*Gabra2* [67], *Cyfip2* [10,11], Crb1 [T56], Nlrp12 [T57]

Some conclusions and suggestions

1. Most forward genetic studies of complex traits should aim to identify highy plausible gene variants using a variety of complementary resources. Obsessing about precision using one type of resource and one species is harder than using a variety of resources. Despirt this advice this is staill an uncommon approach.

2. Trying to identify the causal DNA variants is often not justifieds. Are you actually interested in mutation types and rate and their impact on phenotypes more generally? Usually not. While we know the precise cause of the *Comt* mutation that affects CNS monaminergic systems, this does not buy us much translational utility other than to be on the lookout in humans for mobile element polymorphisms in 3' UTRs that alter polyadenylation and therefore the metabolism and transport of mRNAs.

3. The most effective current methods to identify and QTL genes are often hybrids of forward and reverse genetic methods—now made possible by full genome sequencing. The ***Comt*** example ws actually catalyzed by a puzling cis eQTL that "should not have been". The ***Gabra2*** variant in C57BL/6J is an even better illustration that also highlights the value of eQTL data.

4. Take advantage of earlier studies and findings that interest you scientifically. These studies can often be reanalyzed/resuscitated using advanced genomics resources and high density marker maps.

Questions, suggestions, ideas

11:06:01 **Arthur Centeno** : The power point slides and pdf course material are now available at: ***http://opar.io/webinar_series_1.html***

11:35:47 **Joe Nadeau** : By using lots of different strains and resources, is there an implicit model that the QTL has a pretty constant and consistent effect across strains and backgrounds, I.e. not much epistasis, background, or context-dependent effects? Does strong independent effects limit the scope of discovery? QTLs whose action across backgrounds are of course pretty interesting. But the others are interesting too. It's probably important to understand the the strengths and limits of the underlying model that's being tested.
REPLY RWW: Yes, great point. This is even true within single complex crosses such as CC, DO, and HS were haplotype contrasts are an important part of mapping. If independent allele/haplotype effects have overlapping and opposite polarities, then they may fail to be detected using simple methods (e.g. only SNP genotypes rather than haplotypes). It is possible make the combination of resource types much easier by using complementary resources that segregated for the same haplotypes—e.g. an A/B F2 and A/B advanced intercross, A/B RI family and A/B RIX diallel, and the pair of A/B consomic and congenics.

11:41:25 From Camron Bryant : Don't tell people at your chalk talk that you don't need the nucleotide variant - you will not get the job.
11:42:20 From Saunak Sen : :)
11:59:47 From Camron Bryant : gotta go, sorry i couldn't be there for RCCs!
REPLY RWW: Presentation and Q&A recorded. Will be linked at opar.io soon

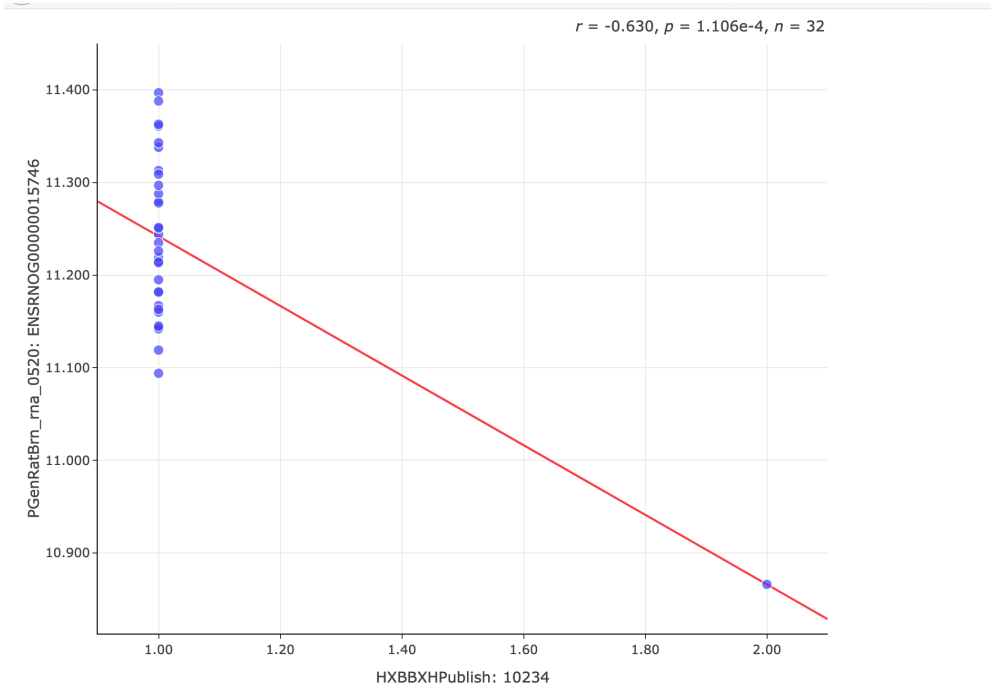
rwilliams@uthsc.edu

Time permitting: Cherry picking cases and rare mutations in HRDP and HMDP

1. We now have sequence data for the HXB, FXLE, and the eight parents of the HS rats. We have sequence data for BXD, CC, and about 25 other strains of mice. All traits will be remappable on the new assemblies and with comprehensive ”overlay” tracks of sequence variants and haplotypes in every locus.

2. Almost all rare variants willo be known and potentially usable for reverse genetics provided we have a deep phenome for the relevant strains. That is where the mouse and rat phenome projects become critical.

3. That is also why it is critical to use common rodent resources,



Statistic	Value
Number	32
Slope	-3.760E-01

X axis: rat HXBBXH HXBBXHPublish: 10234
PubMed: 2020 Cofactor, outlier detection: BN-Lx/Cub as outlier, where BN-Lx/Cub phenotype = 2, all

Y axis: rat HXBBXH Brain mRNA PGenRatBrn_rna_0520: ENSRNOG00000015746
[Kcng4 on Chr19: 52.270927 Mb] potassium voltage-gated channel modifier subfamily G member 4

Part 5: Slide 29 Live demo if possible. How to cherry pick strains or cases

Time permitting: Cherry picking traits and QTLs

Species: Mouse (mm10)

Group: BXD Family

Type: Phenotypes Info

Dataset: BXD Published Phenotypes

Get Any: behavior cocaine nicotine morphic

Enter terms, genes, ID numbers in the **Search** field.
Use * or ? wildcards (Cyp*a?, synap*).

BXD_10265	Central nervous system, pharmacology, protein expression: Dopamine receptor 2 and 3 (DRD2/DRD3) binding maximum (Bmax) in membrane fragments in the dorsal striatum (caudate putamen) of females (125I-epidepride ligand) [fmol/mg wet weight]	236.137	Jones BC, Tarantino LM, Rodriguez LA, Reed CL, McClearn GE, Plomin R, et al.	1999	25.3	Chr15: 87.476581	90.157
BXD_10725	Central nervous system, metabolism, nutrition: Zinc level in medial prefrontal cortex of females [nmol/g]	224.933	Jones LC, McCarthy KA, Beard JL, Keen CL, Jones BC	2006	24.9	Chr1: 153.969506	28.575
BXD_17033	Central nervous system, pharmacology, toxicology: Effect of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) on homovanilic acid (HVA) concentration in caudate-putamen in females 48h after injection (saline-MPTP group) [ug/mg wet weight]	0.219	Jones BC, Miller DB, O'Callaghan JP, Unger EL, Lu L, Alam G, et al.	2014	24.3	Chr11: 65.756786	-0.153
BXD_10234	Central nervous system, pharmacology, protein expression: Dopamine transporter (DAT, SLC6A3) protein density in the dorsal striatum (caudate putamen) [Bmax, pmol/mg]	3.118	Janowsky A, Mah C, Johnson RA, Cunningham CL, Phillips TJ, Crabbe JC, et al.	2001	23.8	Chr19: 15.292517	0.939

Part 5: Slide 30 Live demo on how to pick traits based on linkage or location