Exploration of Genomic Selection Strategies to Complement Wheat FHB Resistance Breeding

Emily Conley, Liangliang Gao, Jim Anderson

Dept. of Agronomy & Plant Genetics

University of Minnesota

St. Paul, MN

University of Minnesota Wheat Breeding Timeline - current

Year	Generation	No. (Loc.)	Lr, Sr	FHB	Quality	Markers
1	Crossing/F_1	300				✓
	F_2 (300 x 1000)	300,000	✓			
2	F_3 Winter Nursery	30,000				
	F_4 Headrows	24,000 (2)	✓			
3	F_5 Scab	2,400 (2)	✓	✓		
4	Winter Nursery	1,000			Prot/TWT/Mixo	✓
	Prelim. yield trial	550 (2–3)	✓	✓	+ flour color	
5–6	Adv. yield trial	170 (8–10)	✓	✓	+ bread bake	
7–9	MN Variety trial	10 (12)	✓	✓	Same as AYs	

University of Minnesota Wheat Breeding Timeline – with marker predictions

Year	Generation	No. (Loc.)	Lr, Sr	FHB	Quality	Markers
1	Crossing/F_1	300				☑
	F_2 (300 x 1000)	300,000	☑			
2	F_3 Winter Nursery	30,000				
	F_4 Headrows	24,000 (2)	☑			
~~3~~	~~F_5 Scab~~	~~2,400 (2)~~	☑	☑		
4	Winter Nursery	1,000			Prot/TWT/GPT	☑
	Prelim. yield trial	550 (2–3)	☑	☑	+ flour color	
5–6	Adv. yield trial	170 (8–10)	☑	☑	+ bread bake	
7–9	MN Variety trial	10 (12)	☑	☑	Same as AYs	

Use of Markers in the U of MN Wheat Breeding Program

1. Parental Characterization
2. Enrichment of BC_1F_1's and 3–way crosses

RR x rr	RR x rr
↓	↓
Rr x rr	Rr x RR
↓	↓
Rr or rr	RR or Rr

3. Screen all pre–yield trial lines (F_6's, ~1,000 lines) with 8 markers (done at USDA-ARS Fargo Genotyping Center)

Total: 10,000-15,000 datapoints per year

High Priority Genes

Fusarium Head Blight (*Fhb1* & FHB 5A)

High Molecular Weight Glutenins (2–3 genes)

Stem Rust (esp. Ug99–effective genes)

Leaf Rust (*Lr34*)

Grain Protein (*GPC–6B*)

Tan spot (*Tsn1*)

Reduced height (*Rht1* and *Rht2*)

Marker Enrichment Candidates

Pedigree	Fhb1	Ax2*	Ax1	Dx5	Lr34	tan spot 5B	FHB 5A
MN12193-6/Prosper//MN08165-8	2	1	1	0	1	1	2
MN12279-7/Linkert//MN10261-1	1	2	2	1	1	0	0
MN12279-7/MN07098-6-2//MN11325-7	2	1	2	0	0	0	0
MN12279-7/MN10281-1-98//MN11394-6	2	1	2	1	1	0	0
MN12345-3/MN11325-7//MN08165-8	2	1	2	1	0	1	2
MN12345-3/MN11394-6//MN10261-1	1	2	2	2	0	2	1
MN12345-3/Prosper//MN08165-8	2	1	2	1	1	2	2
MN13353-3/Linkert//MN13564-3	1	1	2	1	1	0	1
MN13353-3/MN07098-6-2//MN10261-1	2	0	2	0	0	0	1
MN13353-3/MN08165-8//MN11394-6	2	0	2	0.5	0	0	1
MN13424-3/MN08165-8//Linkert	2	0	2	1.5	0	1	1
MN13424-3/MN08165-8//MN10201-4-116	2	0	2	1.5	0	1	1
MN13424-3/MN10261-1//Linkert	2	0	2	2	0	1	1
MN13424-3/MN10261-1//MN10201-4-116	2	0	2	2	0	1	1
MN13618-7/MN07098-6-2//MN11394-6	1	0	1	0	0	0	0
MN13618-7/MN08165-8//MN10261-1	1	0	1	0.5	0	0	0
MN13618-7/MN10281-1-98//MN13564-3	1	0	1	1	1	0	0
SY-Rowyn/Linkert//MN10261-1	0	1	1	1	1	0	0
SY-Rowyn/MN07098-6-2//MN11394-6	1	0	1	0	0	0	0
SY-Rowyn/MN10281-1-98//MN13564-3	1	0	1	1	1	0	0
SY-Rowyn/Prosper//MN10281-1-95	1	0	1	0	1	1	1
KS14WGRC61/MN07098-6-2//MN07098-6-Lr34	1	0	1	0	0	0	0
KS14WGRC61/MN10261-1//MN10261-1	1	0	1	1	0	0	0
KS14WGRC61/MN11325-7//MN11325-7	1	0	1	0	0	0	1
KS14WGRC61/MN11394-6//MN11394-6	0	1	1	1	0	1	0
KS14WGRC61/Prosper//MN07098-6-Lr34	1	0	1	0	1	1	1
MN13573-1/MN10261-1//Linkert	2	0	2	1	0	0	0
MN13573-1/MN10261-1//MN10201-4-116	2	0	2	1	0	0	0

Objectives

1. Determine the best genome-wide prediction models for FHB traits
2. Determine if including known QTL can improve prediction accuracy

Short term: Use GS to discard MS and S lines & bypass F_5 generation

Long term: Prediction models to accelerate the breeding cycle.

Genomic Selection Panel

- 384 F_7-derived lines in advanced yield testing
- Representative of program wide genetic diversity
- 93 parents, 177 unique crosses represented in the pedigrees
- Parents include MN, ND, SD, AgriPro, WestBred, CIMMYT

FHB Evaluation

- **2 locations:** St. Paul and Crookston
- All preliminary and advanced yield lines are screened for FHB each year
  - **PYs:** 1 loc, 1 rep.
  - **AYs:** 2 loc, 2 reps.
  - **MN Variety Trials:** 2 loc, 3 reps

- **2013:** All 384 lines screened in 2 nurseries, 1 rep.

All 384 lines have been screened in 5 or more FHB environments

Scab Screening

Thresh 30 spikes or ~0.5 m of row

Kornservice seed cleaner

Micro
Test weight

30 spike wt.
VSK
DON

Resistance Traits Assessed

Incidence = Type I

Severity = Type II

30 Head Weight

Test Weight = Grain volume weight

Visually scabby kernels (VSK)

DON

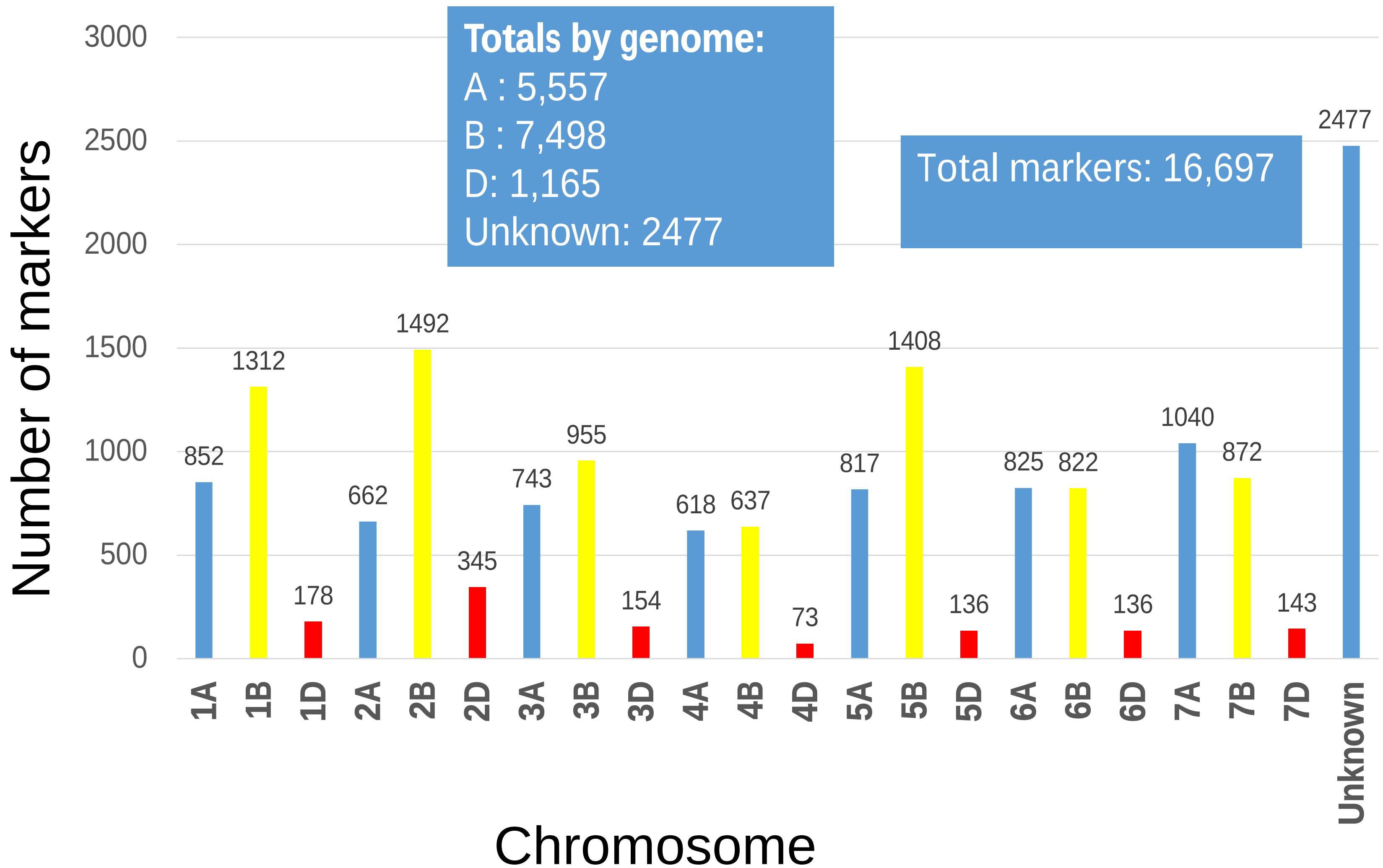
Heading date

Height

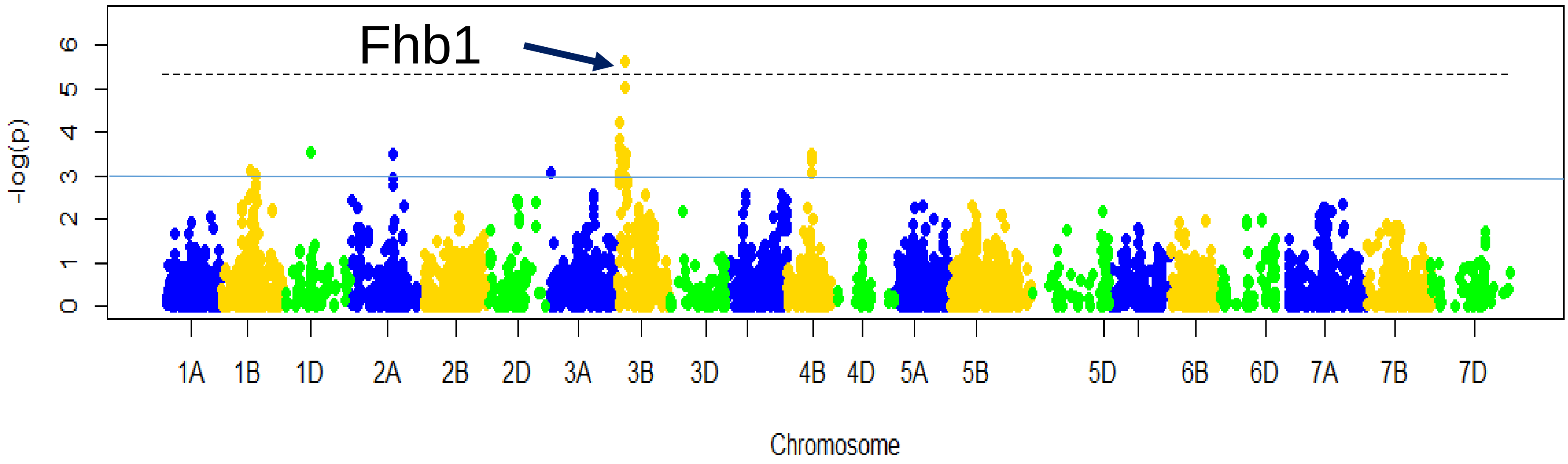
GWAS analysis

- Identify significant QTLs in the panel that could be used to improve prediction model accuracy
- All 384 lines genotyped on 90K Infinium iSelect Assay

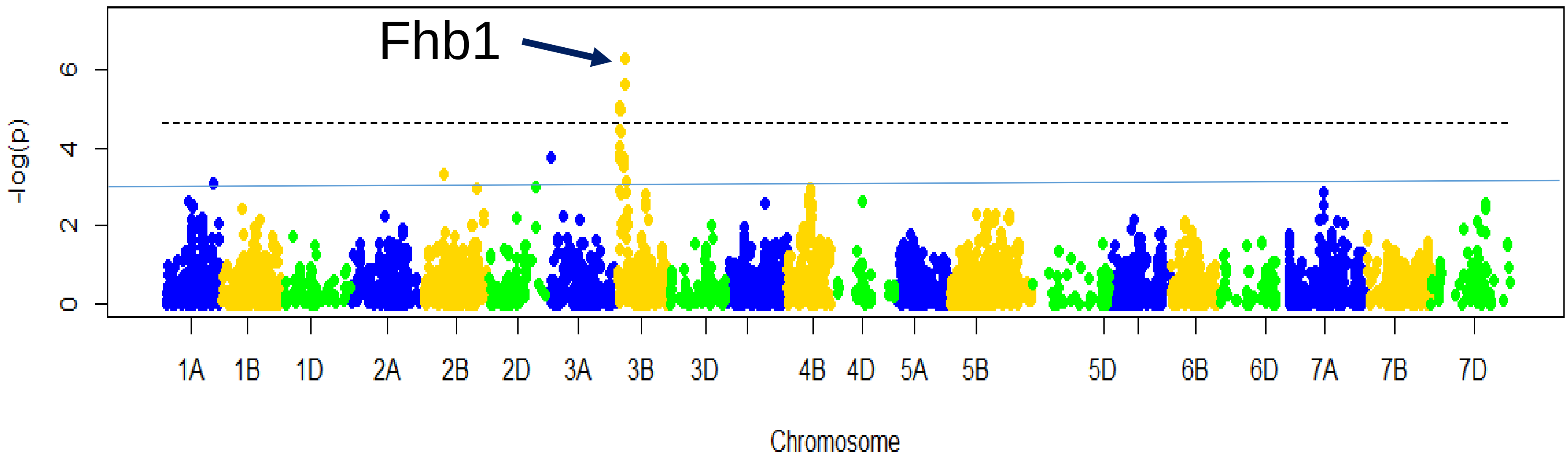
Number SNPs by chromosome

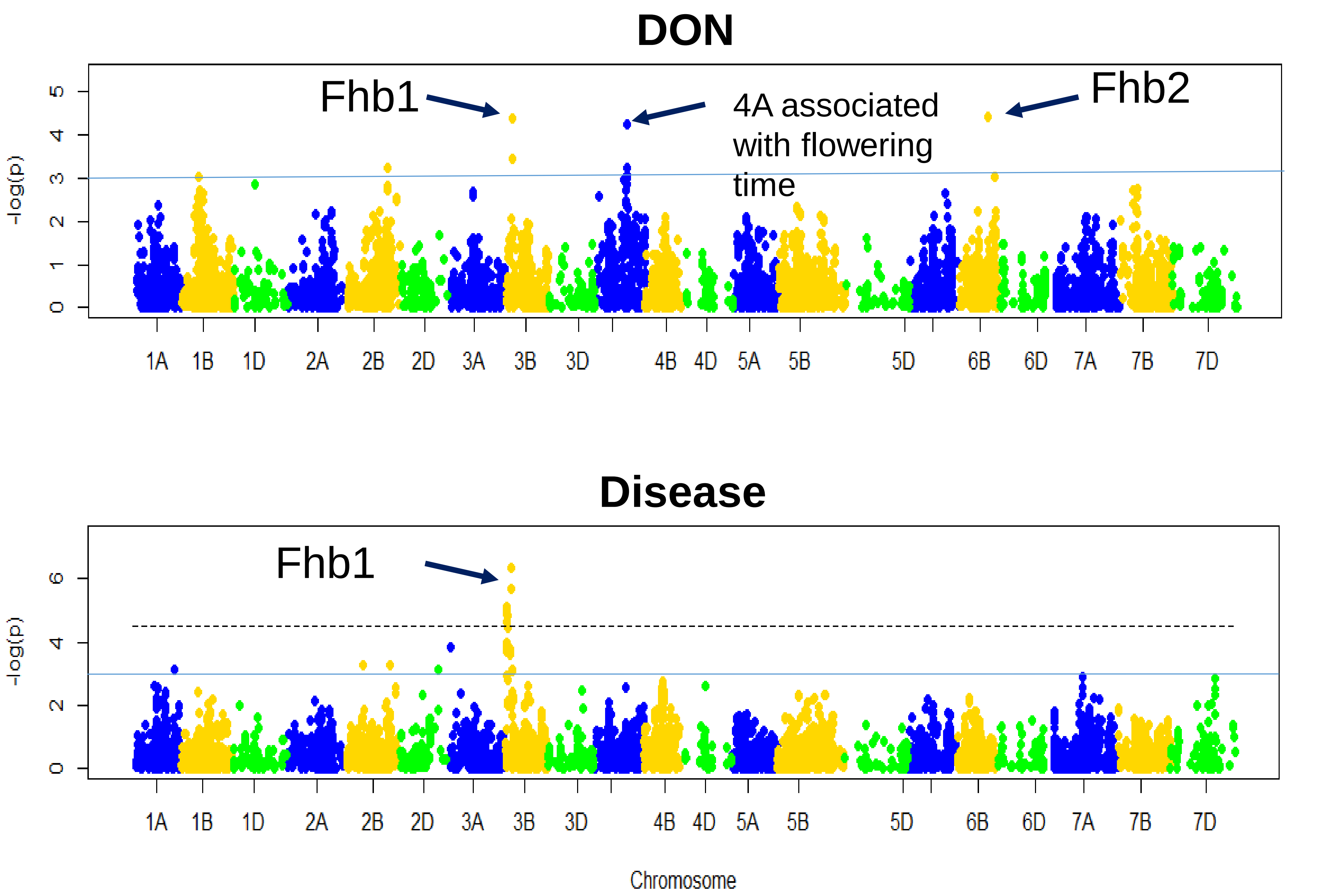


VSK



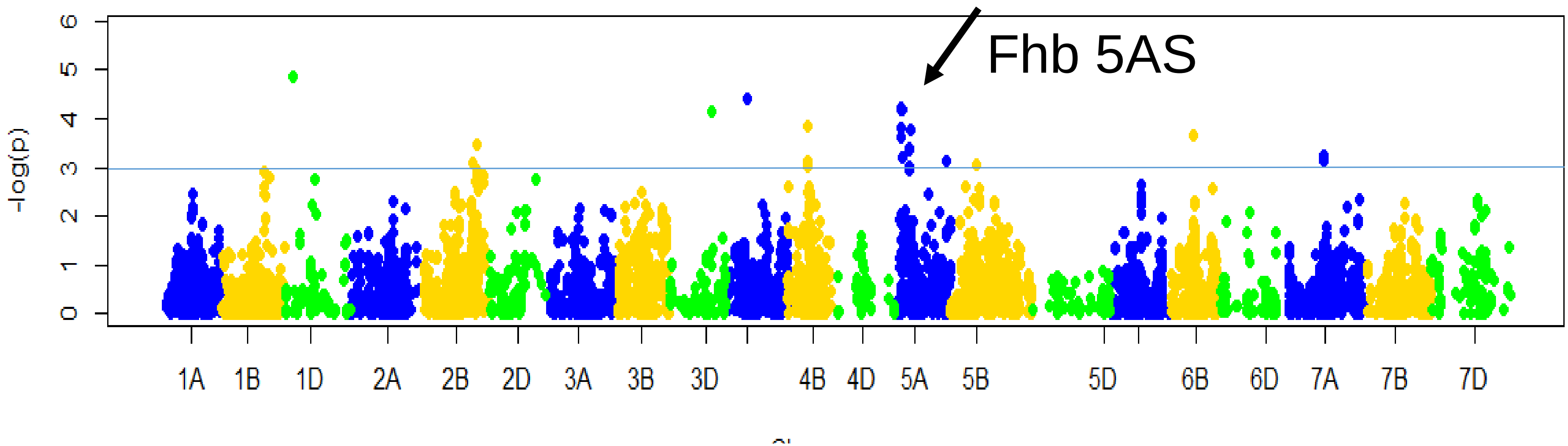
Severity





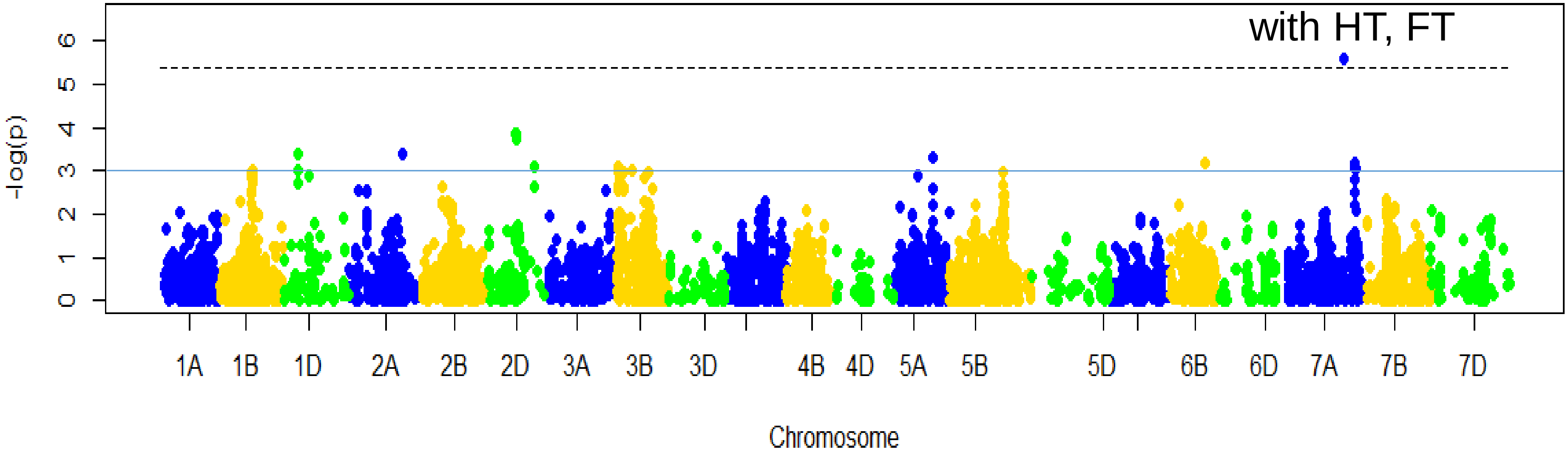
DON
Fhb1
4A associated with flowering time
Fhb2
-log(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4B 4D 5A 5B 5D 6B 6D 7A 7B 7D
Disease
Fhb1
-log(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4B 4D 5A 5B 5D 6B 6D 7A 7B 7D
Chromosome

FHB Incidence

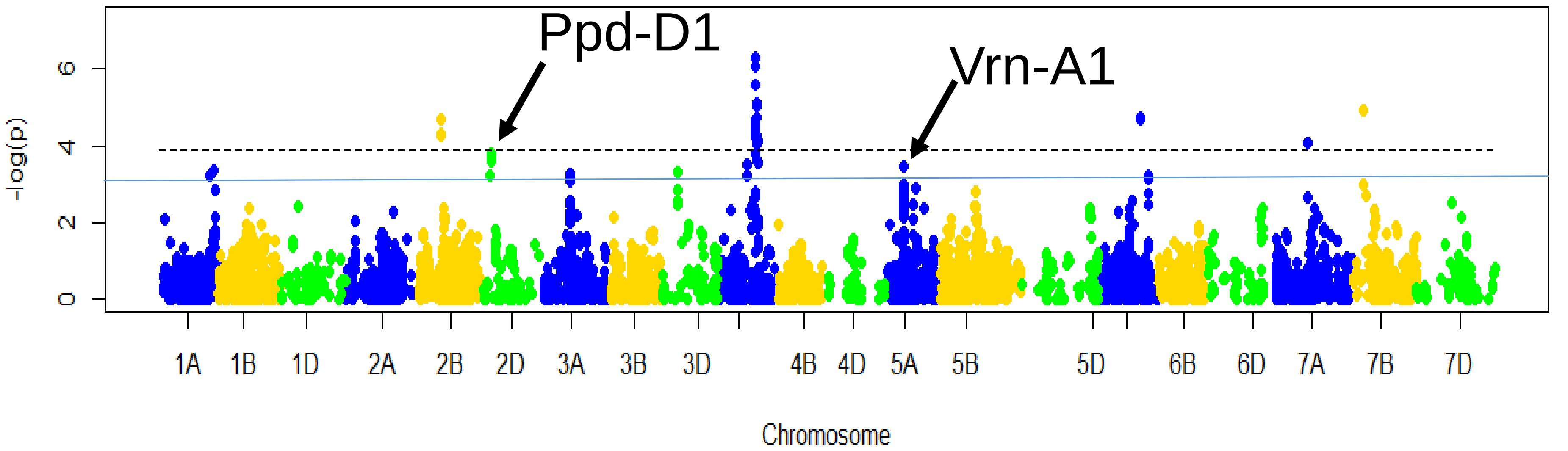


Test Weight

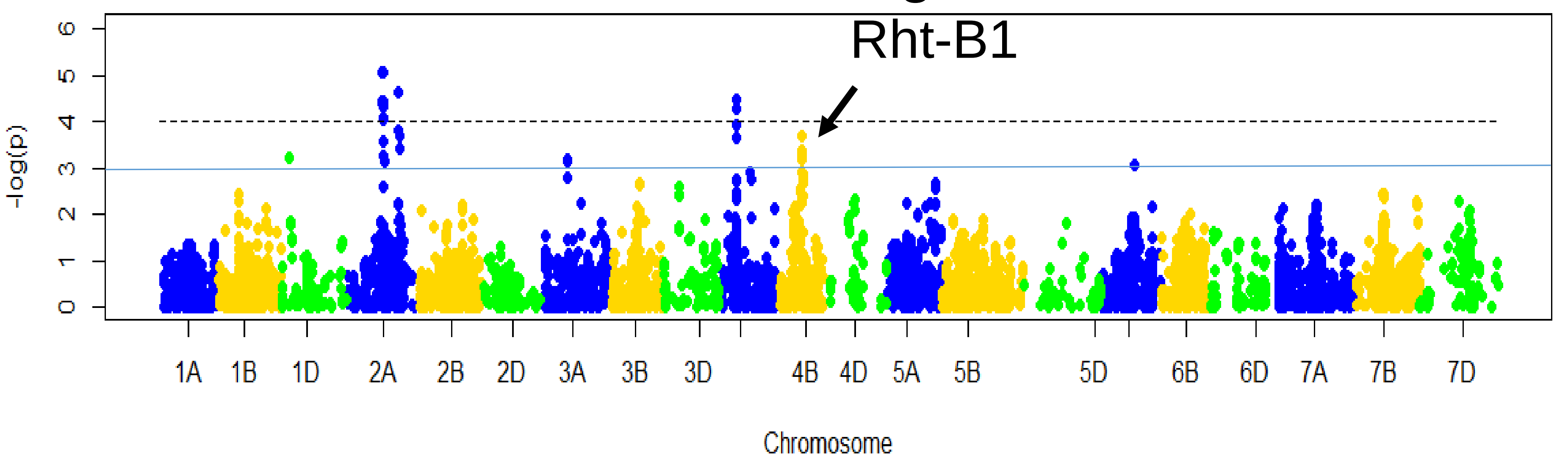
7A reported to have association with HT, FT



Heading Date



Plant Height

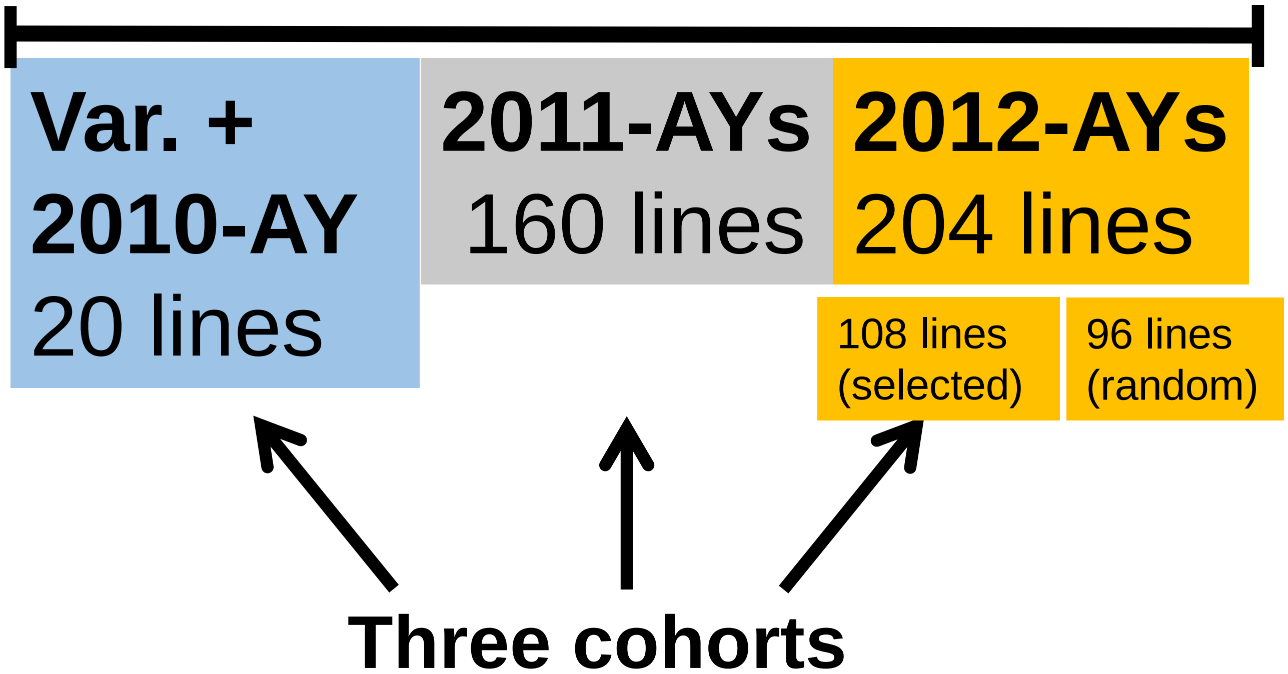


Significant regions detected:

Trait	Chr	Pos	#SNPs	P-value	Freq. of favorable allele	Trait	Chr	Pos	#SNPs	P-value	Freq. of favorable allele
SEV	1A	139.7	1	3.10	0.48	DIS	3B	10.9-14.1	15	5.14	0.64
DIS	1A	139.7	1	3.16	0.48	TWT	3B	11.6	1	3.08	0.70
HD	1A	139.7-149.8	2	3.38	(late) 0.69	VSK	3B	19.3-30.3	14	5.63	0.53
DON	1B	60.6	1	3.03	0.23	SEV	3B	19.3-30.3	14	6.54	0.53
TWT	1B	90.6	1	3.01	0.84	DIS	3B	19.3-30.3	14	6.37	0.53
VSK	1B	81.1-96.3	7	3.11	0.93	DON	3B	25.1-25.4	2	4.40	0.53
HT	1D	21.8	1	3.23	(short) 0.53	TWT	3B	45.9	1	2.99	0.05
INC	1D	21.8	1	4.86	0.10	HD	3D	40.5	1	3.30	0.42
TWT	1D	39.5	6	3.39	0.16	INC	3D	113.1	1	4.15	0.57
VSK	1D	67.7	1	3.53	0.90	HT	4A	47.5-48.5	7	4.47	0.41
HT	2A	98.4-104.0	29	5.06	0.19	INC	4A	58.4	1	4.41	0.07
VSK	2A	116.2	1	3.50	0.13	HD	4A	80.1-109.3	31	6.30	0.57
TWT	2A	143.2	1	3.38	0.31	DON	4A	106.5-107.6	8	4.26	0.54
HT	2A	142.6-144.2	4	4.62	0.27	INC	4B	55.5-56.2	4	3.83	0.16
SEV	2B	67.1	1	3.24	0.20	HT	4B	59.5	6	3.68	0.84
DIS	2B	67.1	1	3.29	0.20	VSK	4B	66.3-66.8	5	3.51	0.27
HD	2B	73.7-74.9	6	4.67	0.13	INC	5A	15.6-19.9	7	4.18	0.08
DON	2B	145.6	1	3.23	0.87	INC	5A	38.7-43.3	4	3.78	0.09
INC	2B	146	1	3.10	0.14	HD	5A	50.4	1	3.47	0.33
SEV	2B	157.2	2	3.03	0.28	TWT	5A	106.0	1	3.30	0.60
DIS	2B	157.2	2	3.27	0.28	INC	5A	139.8	1	3.11	0.08
INC	2B	157.2	2	3.48	0.28	INC	5B	71.6	1	3.04	0.56
HD	2D	22.5-26.0	4	3.77	0.28	HT	6A	91.9	1	3.06	0.60
TWT	2D	76.6-80.5	3	3.87	0.31	HD	6A	112.6	3	4.73	0.16
TWT	2D	129	1	3.11	0.19	HD	6A	135.8	2	3.20	0.93
DIS	2D	129	1	3.16	0.19	INC	6B	61.8	1	3.66	0.08
VSK	3A	15.1	1	3.07	0.63	TWT	6B	95.8	1	3.19	0.31
SEV	3A	15.1	1	3.99	0.63	DON	6B	95.8-116.2	2	4.41	0.31
DIS	3A	15.1	1	3.82	0.63	HD	7A	119.2	1	4.08	0.54
HT	3A	73.2	3	3.20	0.52	INC	7A	126.4-126.8	5	3.24	0.80
HD	3A	86.2	5	3.25	0.40	TWT	7A	181.4	2	5.58	0.94
VSK	3B	10.9-14.1	11	4.24	0.64	TWT	7A	212.7-216.4	5	3.19	0.57
SEV	3B	10.9-14.1	17	5.34	0.70	HD	7B	29.5	1	4.90	0.48

1A: Fhb severity and disease, overlaps with heading
1B: DON, VSK and test weight
1D: Fhb incidence (overlaps with height); test weight, VSK
2A: Plant Height (2 regions), VSK; test weight (overlaps with height)
2B: Fhb severity and disease (2 regions), DON, incidence, heading
2D: Test weight, disease, heading
3A: VSK, severity, disease, height and heading (non-overlapping)
3B: VSK, severity, disease test weight (2 regions), DON
3D: Incidence, heading (non-overlapping)
4A: Height, incidence, DON (overlaps with heading)
4B: Incidence, VSK, height
5A: Incidence, test weight, heading
5B: Incidence
6A: Heading, height
6B: Incidence, test weight, DON
7A: Incidence, test weight, heading
7B: Heading

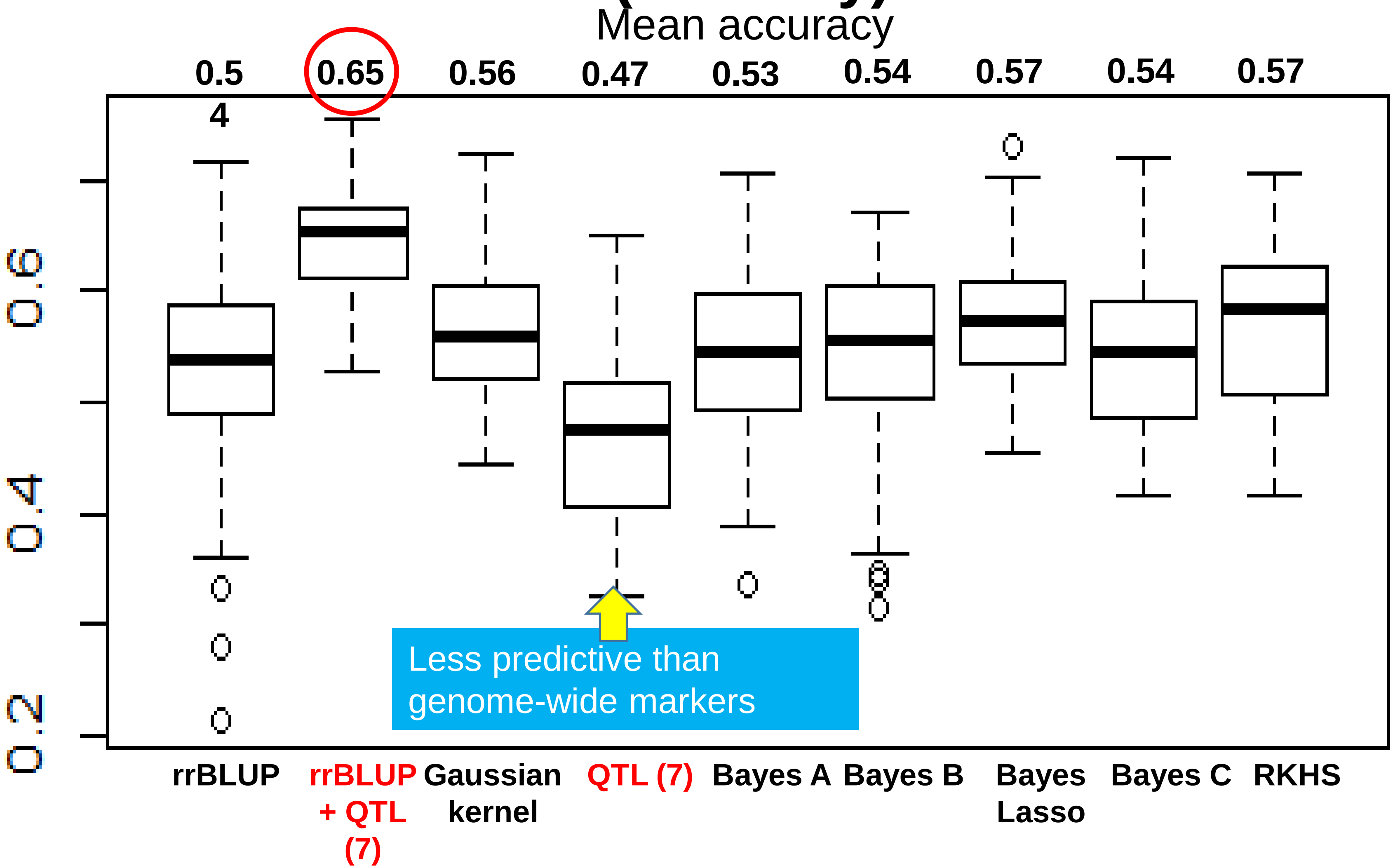
Training and Validation



1. Cross Validation: 4/5 used to predict 1/5 (100 permutations)

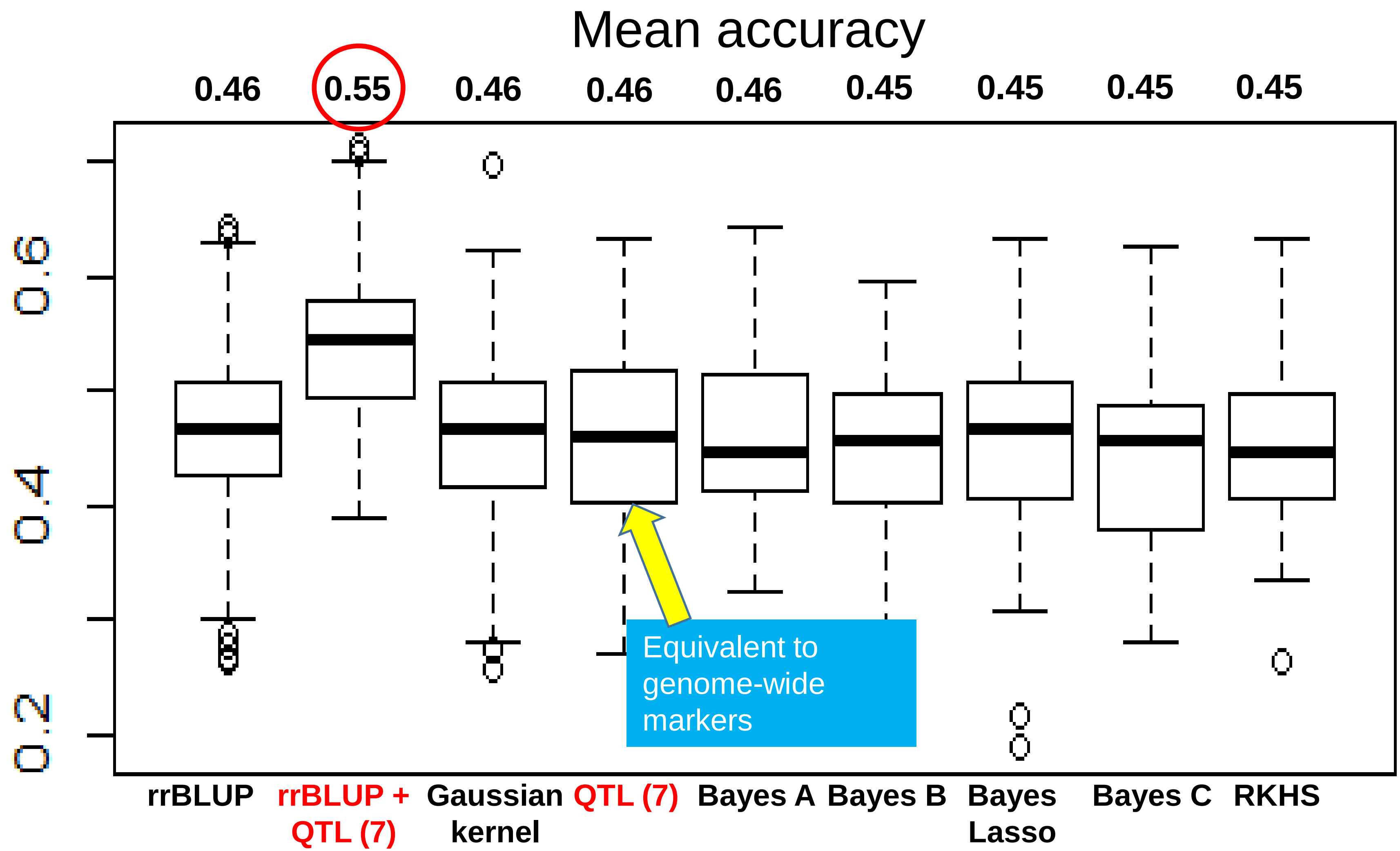
2. Interset Validation: 288 used to predict 96 (“random”)

Cross-Validation Prediction Accuracy (Severity)

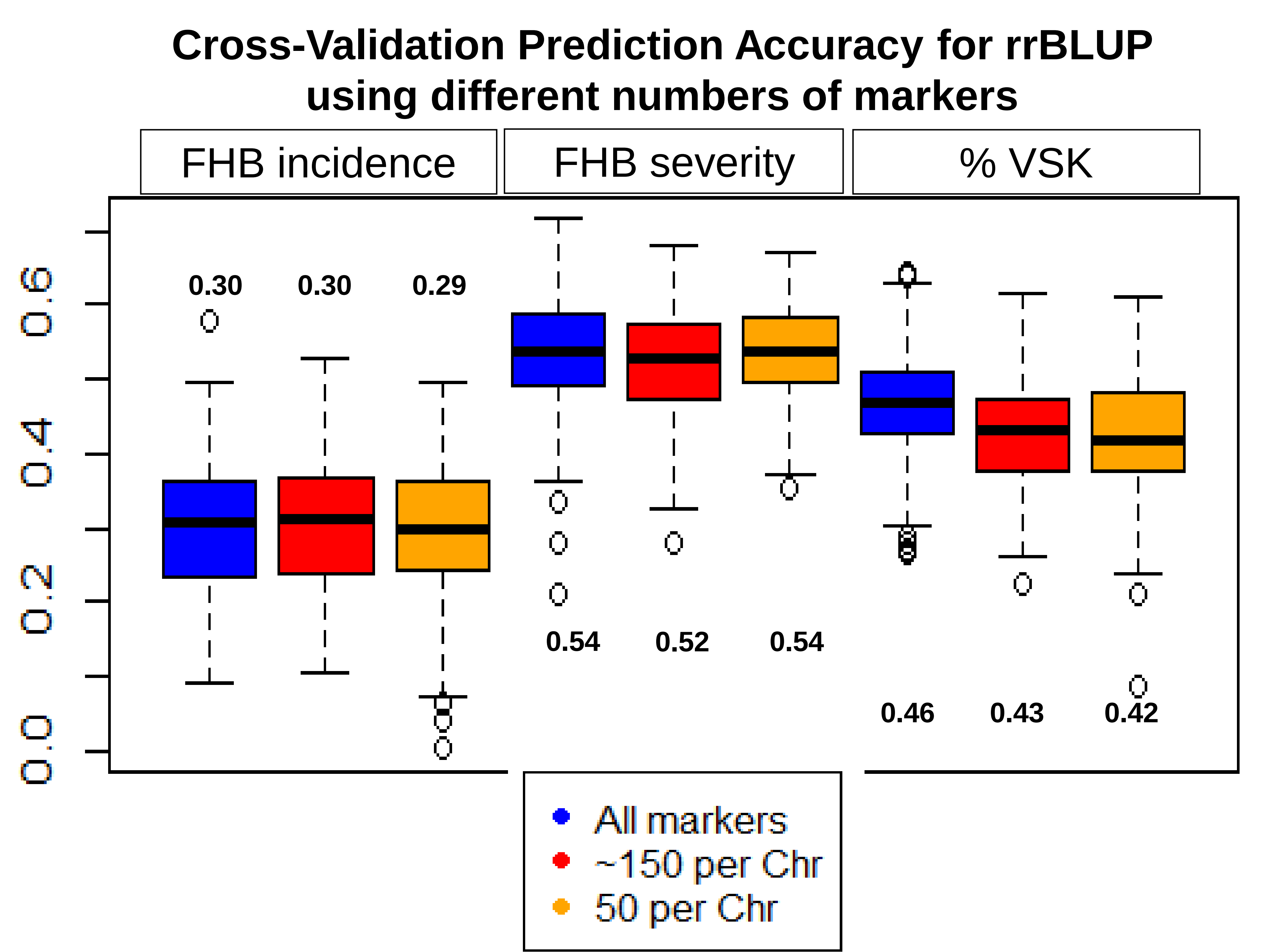


- Highest prediction accuracy using rrBLUP+QTL
- Other models mostly equivalent, QTL only less accurate

Cross-Validation Prediction Accuracy (VSK)

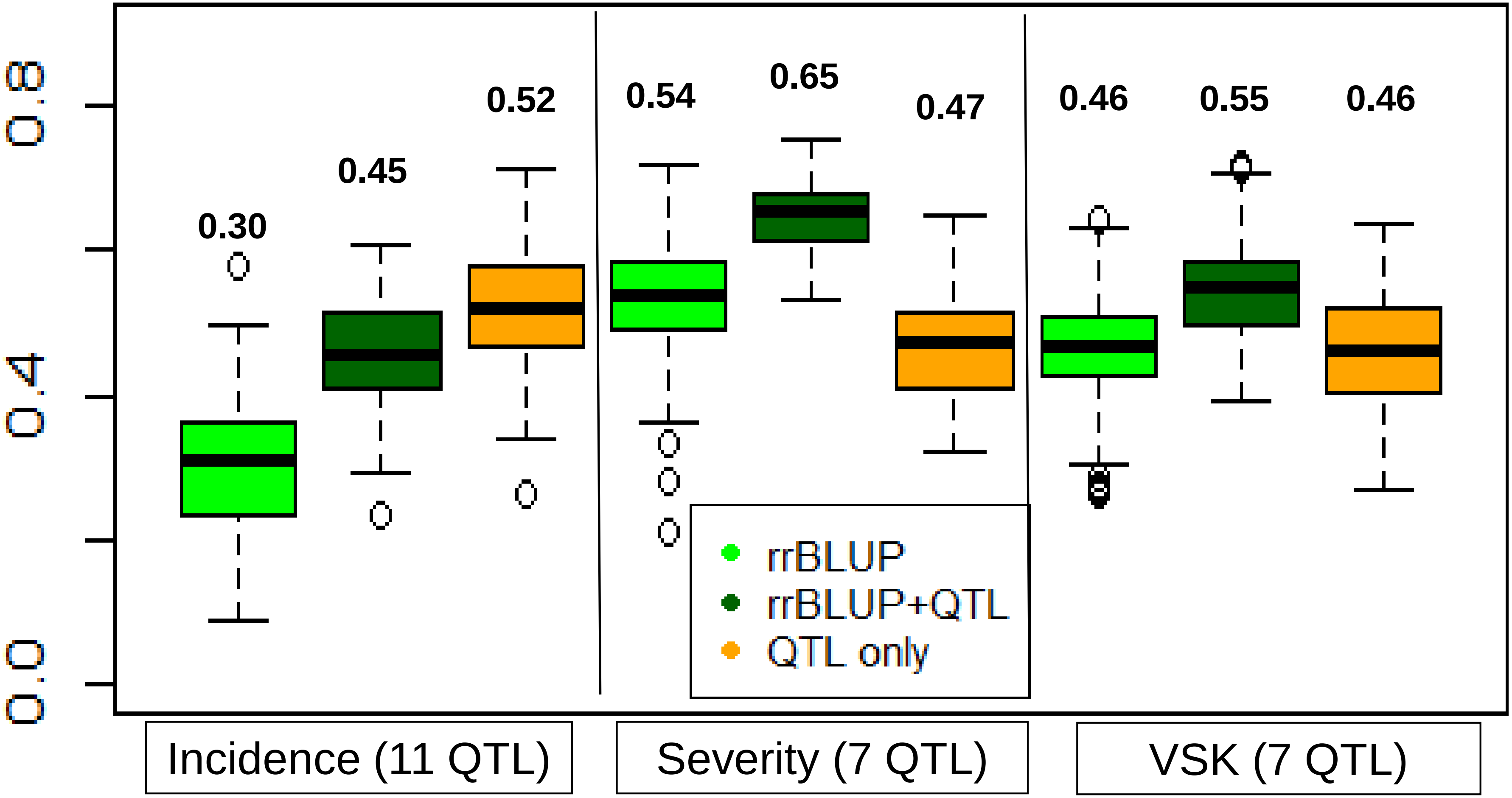


- Highest prediction accuracy using rrBLUP+QTL
- All other models equivalent, including QTL only

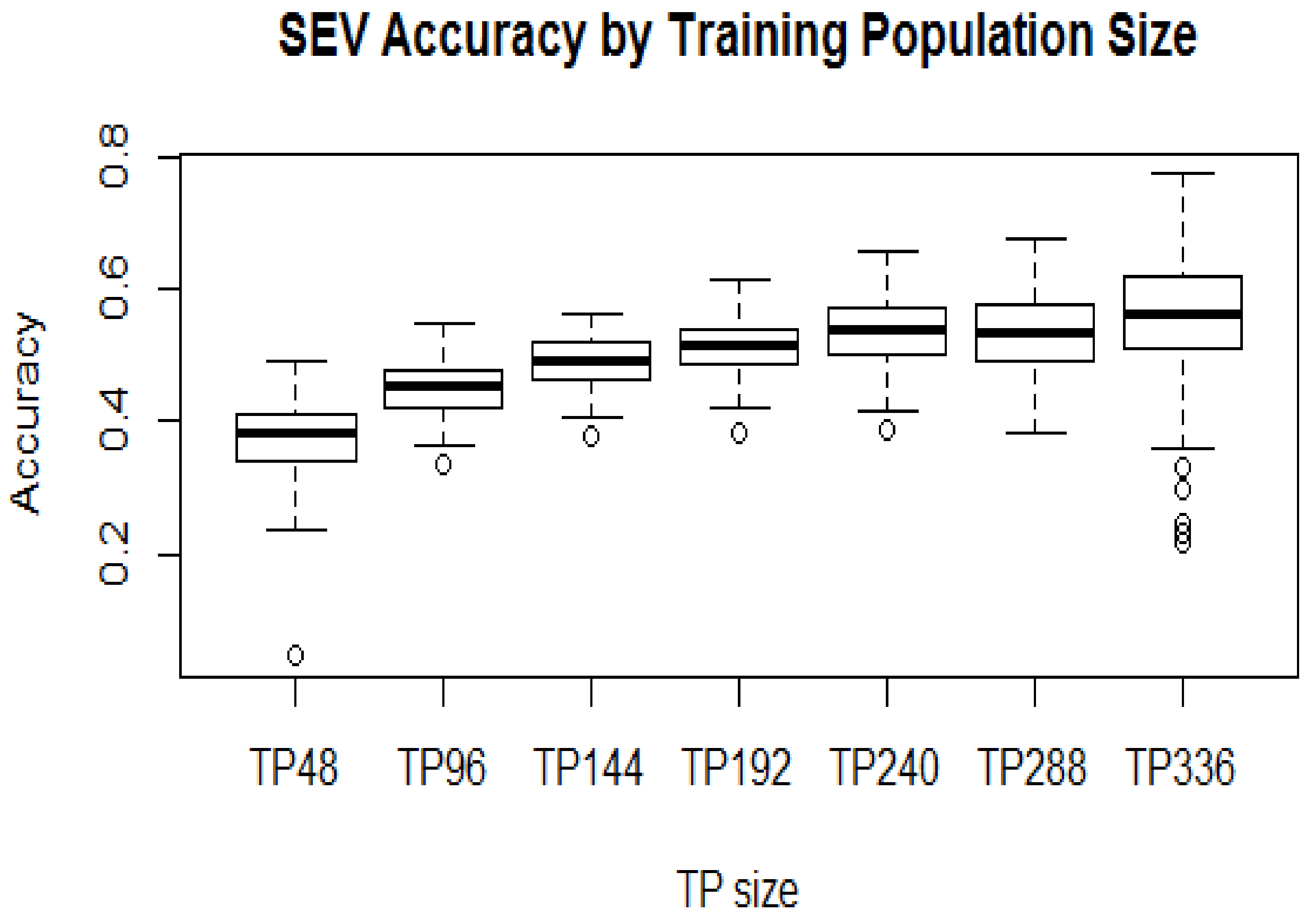


Cross-Validation Prediction Accuracy for rrBLUP
using different numbers of markers
FHB incidence
FHB severity
% VSK
0.0
0.2
0.4
0.6
0.30
0.30
0.29
0.54
0.52
0.54
0.46
0.43
0.42
All markers
~150 per Chr
50 per Chr

Cross-Validation using rrBLUP, rrBLUP+QTL, QTL only

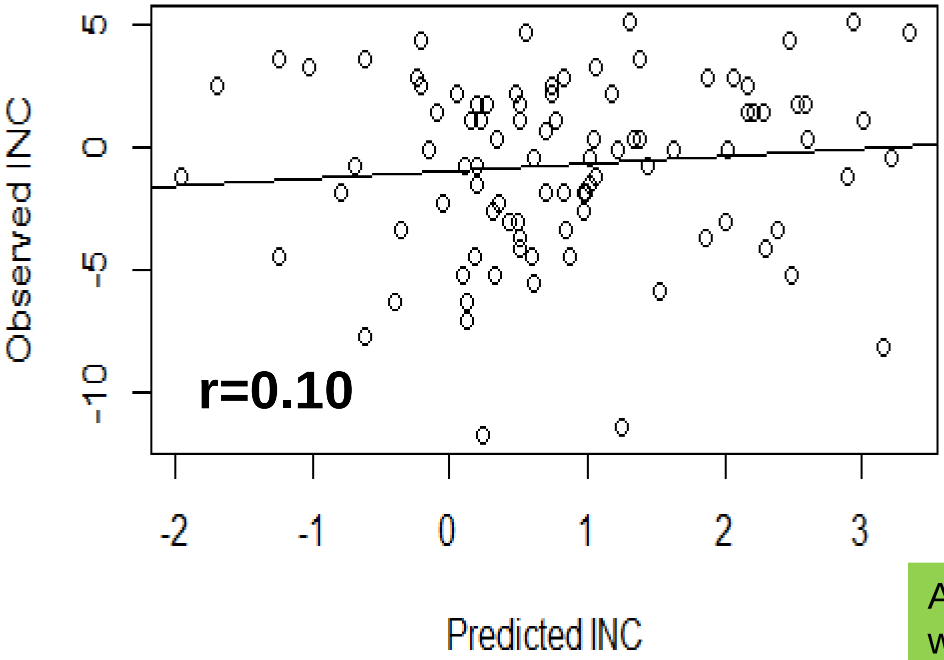


QTL only most accurate for incidence, least accurate for severity, and equivalent to rrBLUP for VSK

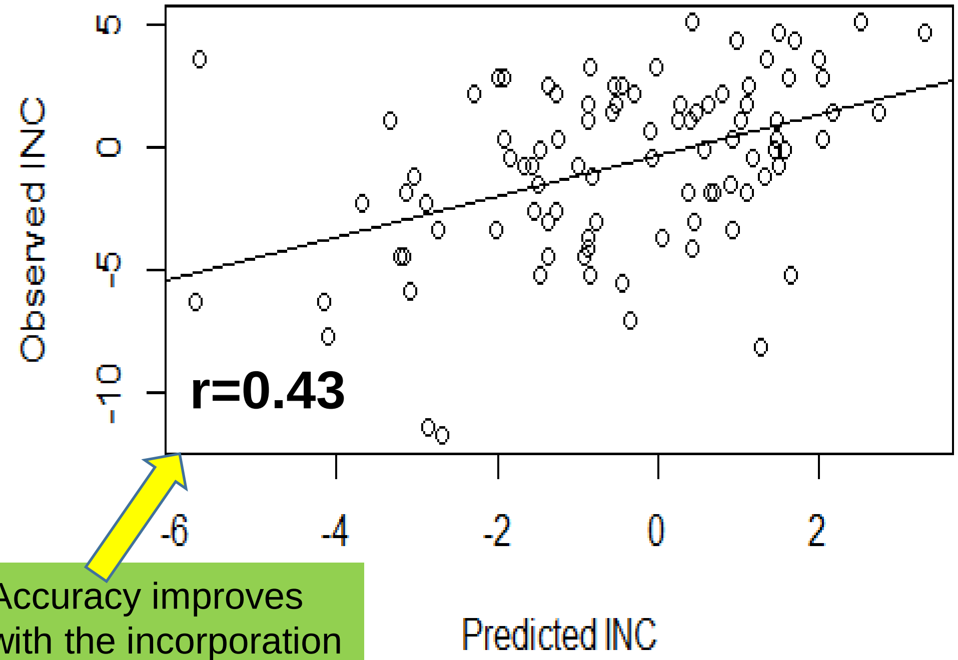

SEV Accuracy by Training Population Size
Accuracy
0.2
0.4
0.6
0.8
TP48
TP96
TP144
TP192
TP240
TP288
TP336
TP size

Interset Prediction (288 to predict 96) Incidence

rrBLUP



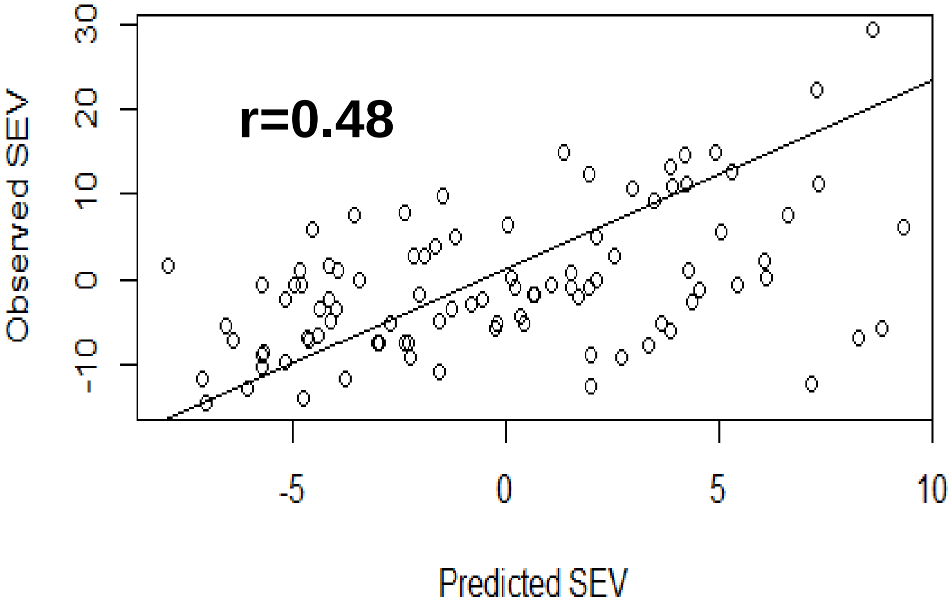
rrBLUP + QTL (11)



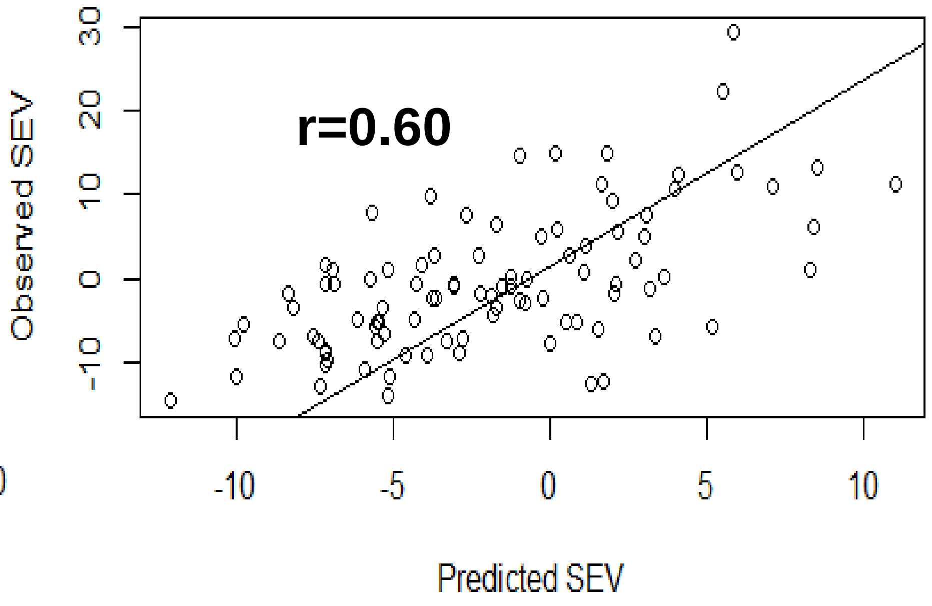
Accuracy improves with the incorporation of significant QTL in the model

Interset Prediction Severity

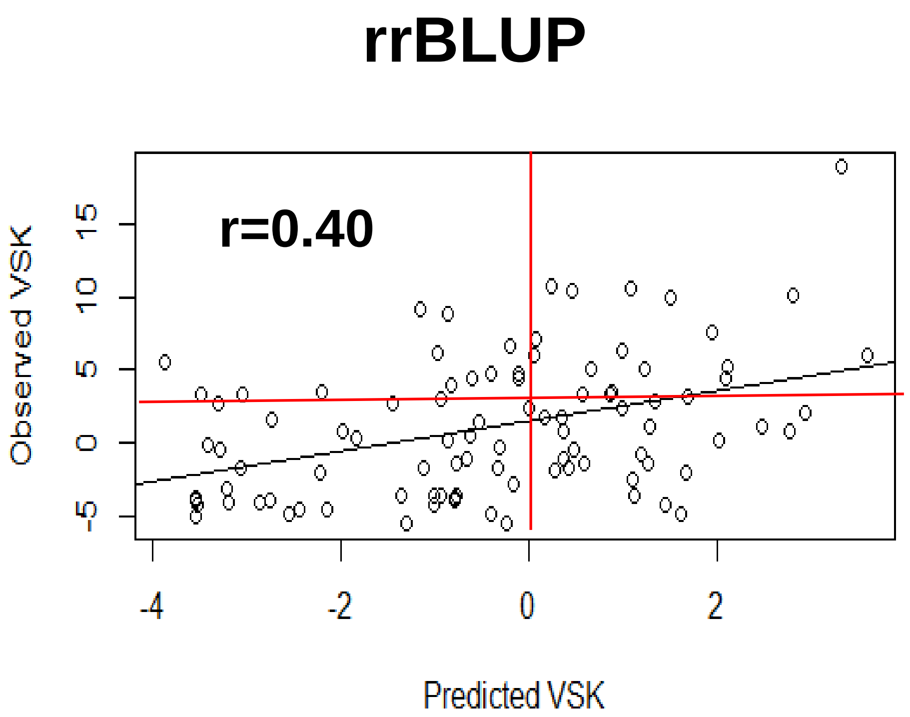
rrBLUP

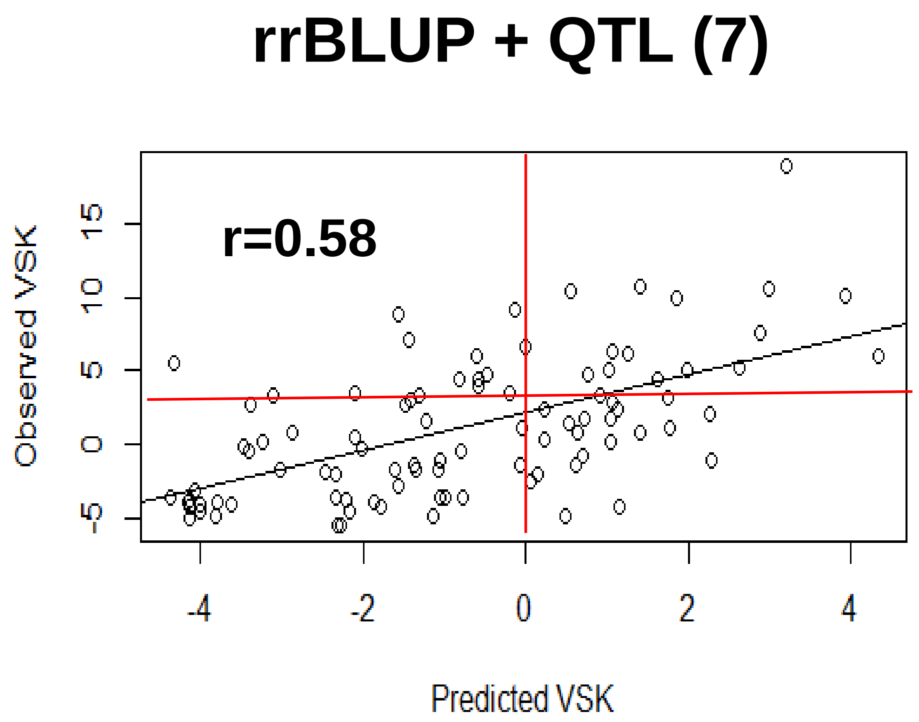


rrBLUP + QTL (7)



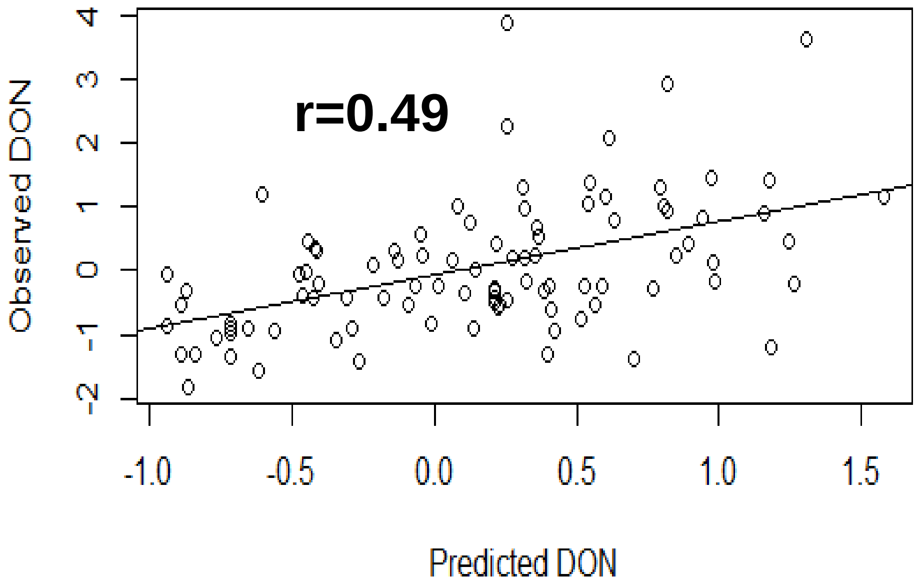
Interset prediction
VSK



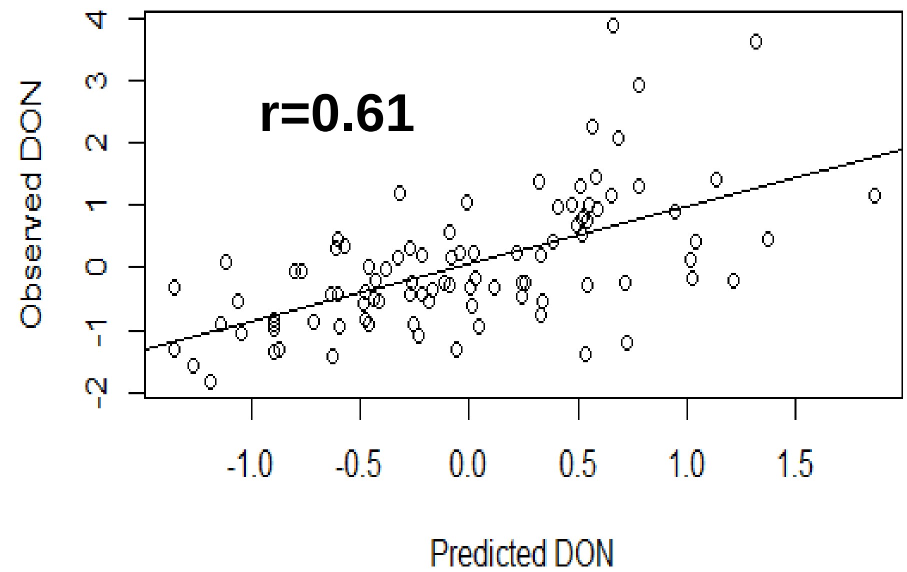


Interset Prediction
DON

rrBLUP



rrBLUP + QTL (5)



Conclusions

- Genomic selection models perform similarly
- Reducing the number of markers to 50 per chromosome does not greatly reduce prediction accuracy
- Incorporating known significant QTL as fixed effects improves prediction accuracy
- Next Steps:
    1. Predict F_5's
    2. Predict 384 lines from NDSU and SDSU

Predicting F_5's using 384 line Training Pop.

Why low correlations?

1. Different crosses
2. Different Marker platforms (90K vs GBS)
3. Different Environments

Genomic Selection v.2

- Concern that a breeding program-wide model will not predict future lines

- A model with more closely related individuals in same environment(s) should perform better.
  - Use subset of F_5's to predict all F_5's
  - Opportunity to optimize training pop.

Acknowledgements

The Wheat Project:
Susan Reynolds
Roger Caspers
Dr. Xiaofei Zhang
Prabin Bajgain
Kathryn Turner
Kayla Altendorf
Jen Flor
Dr. Katherine Frels

U.S. Wheat & Barley Scab Initiative

Crookston scab nursery:
Galen Thompson
Madeleine Smtih
Dept. of Plant Pathology:
Dr. Ruth Dill-Macky

USDA-ARS Small Grain Genotyping Center, Fargo, ND:
Dr. Shiaoman Chao
Mary Osenga

U of MN Small Grains Initiative

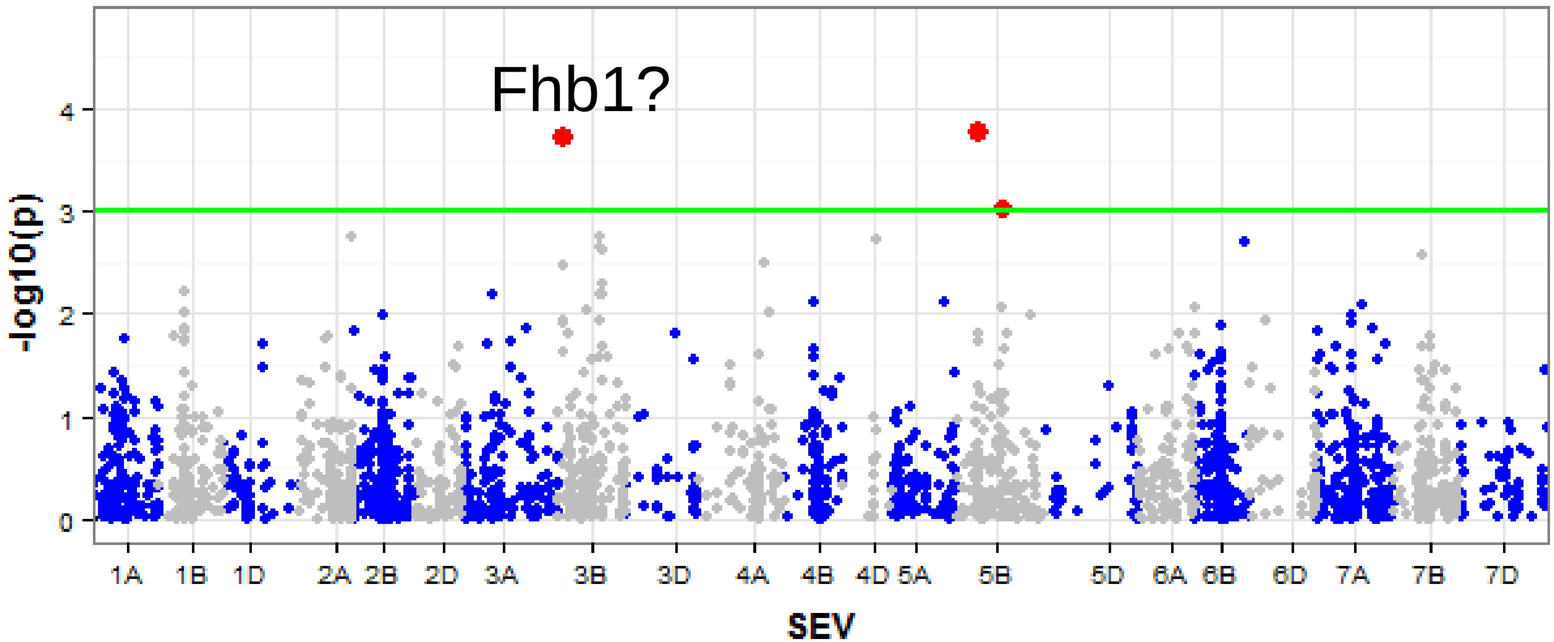
MN Wheat Research & Promotion Council

The Barley Project:
Ahmad Sallam
Celeste Falcon
Karen Beaubien
Tyler Tiede
Mohsen Mohammadi

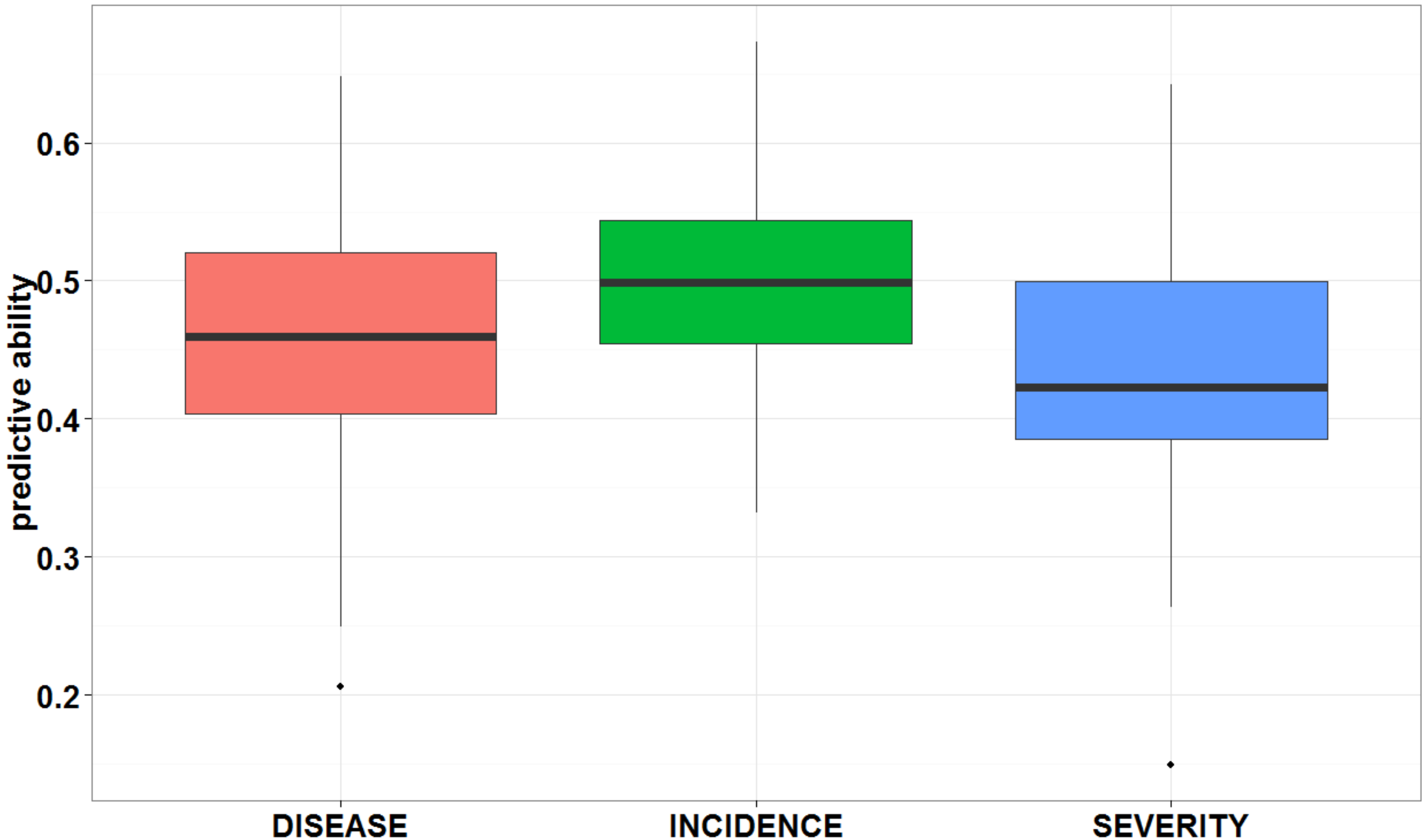
Overview of F_5 Genomic Selection predictions

- 380 F_5 lines, phenotyped for FHB Incidence, Severity Disease (VSK and DON in progress) in 3-5 environments
- F_5's genotyped by GBS. > 4,000 high quality GBS SNP markers were obtained
- Cross validation: F_5's to predict F_5's
- Exploratory Genomic Selection: using advanced lines genotyped with 90K SNP arrays to predict F_5's genotyped using GBS

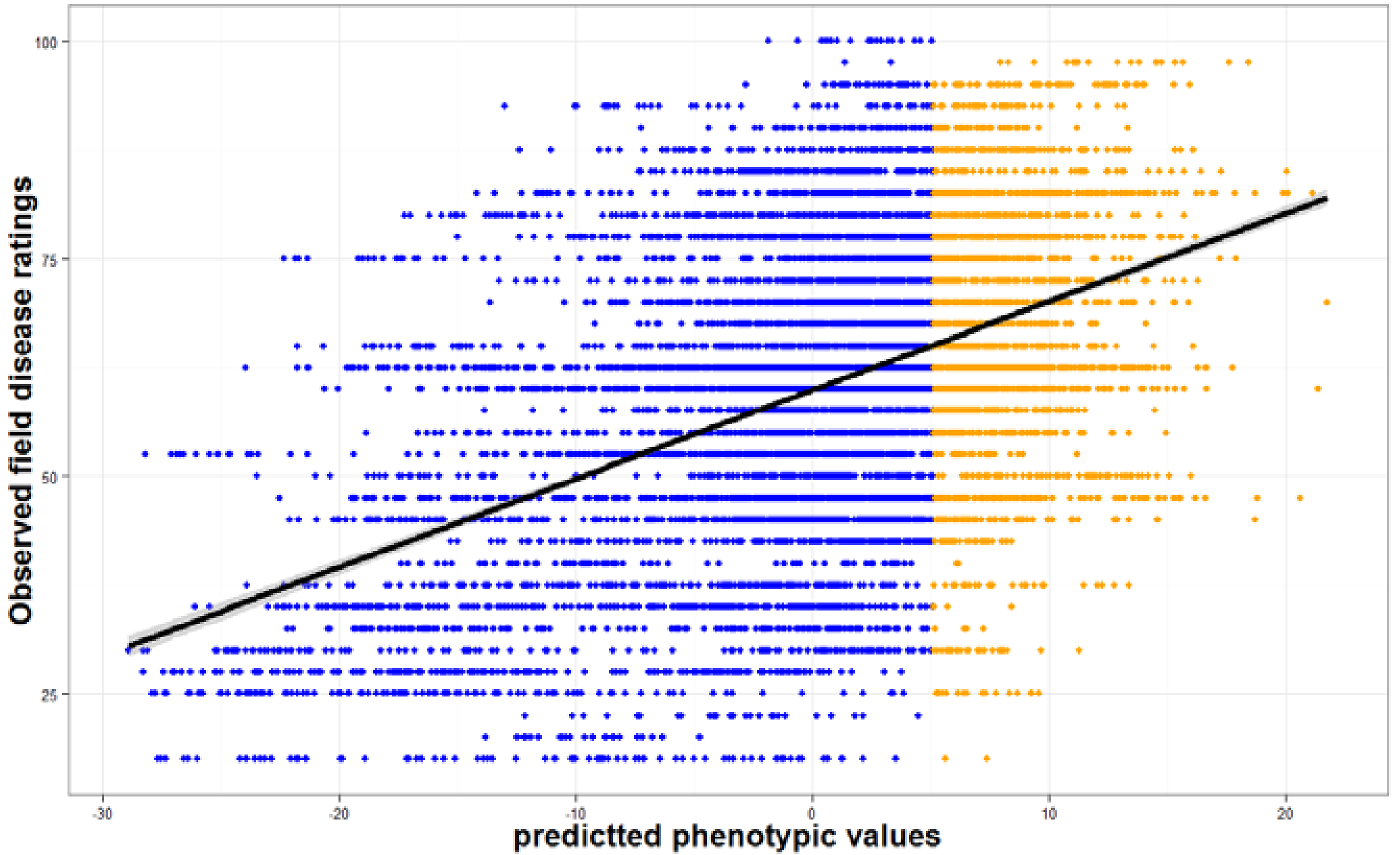
Preliminary GWAS study on FHB resistance using F_5s and Genotyping by Sequencing (GBS)



- >4,000 GBS SNP markers identified; >2,500 GBS markers were assigned to approximate positions based on blast hit information
- Preliminary GWAS study was conducted using the 380 F_5's

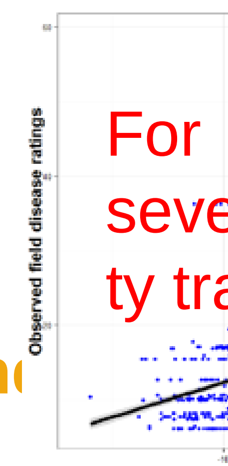


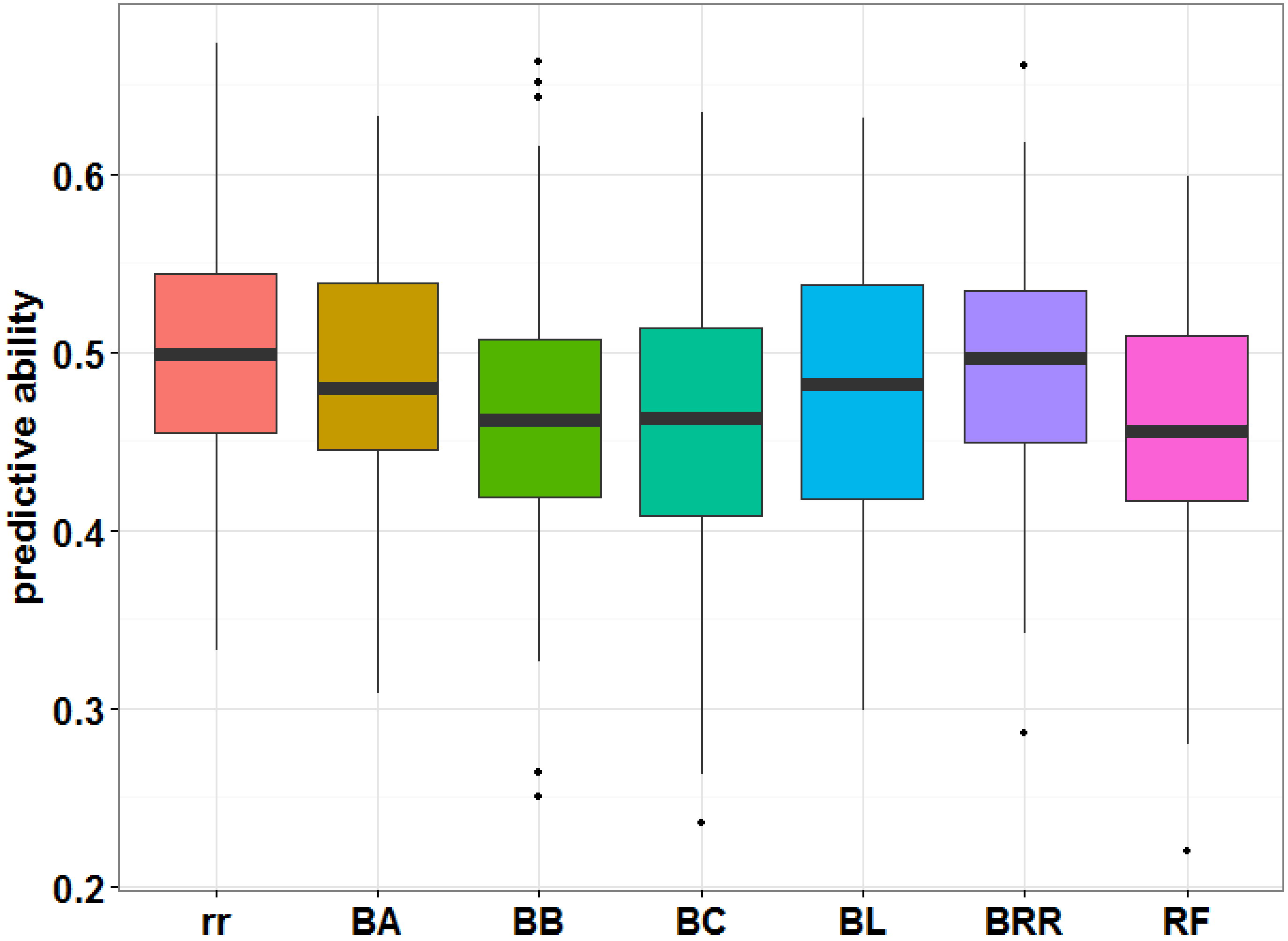
Predictability for F_5 FHB Incidence (0.5) is higher than severity and disease (for both 2014 and 2015 data).



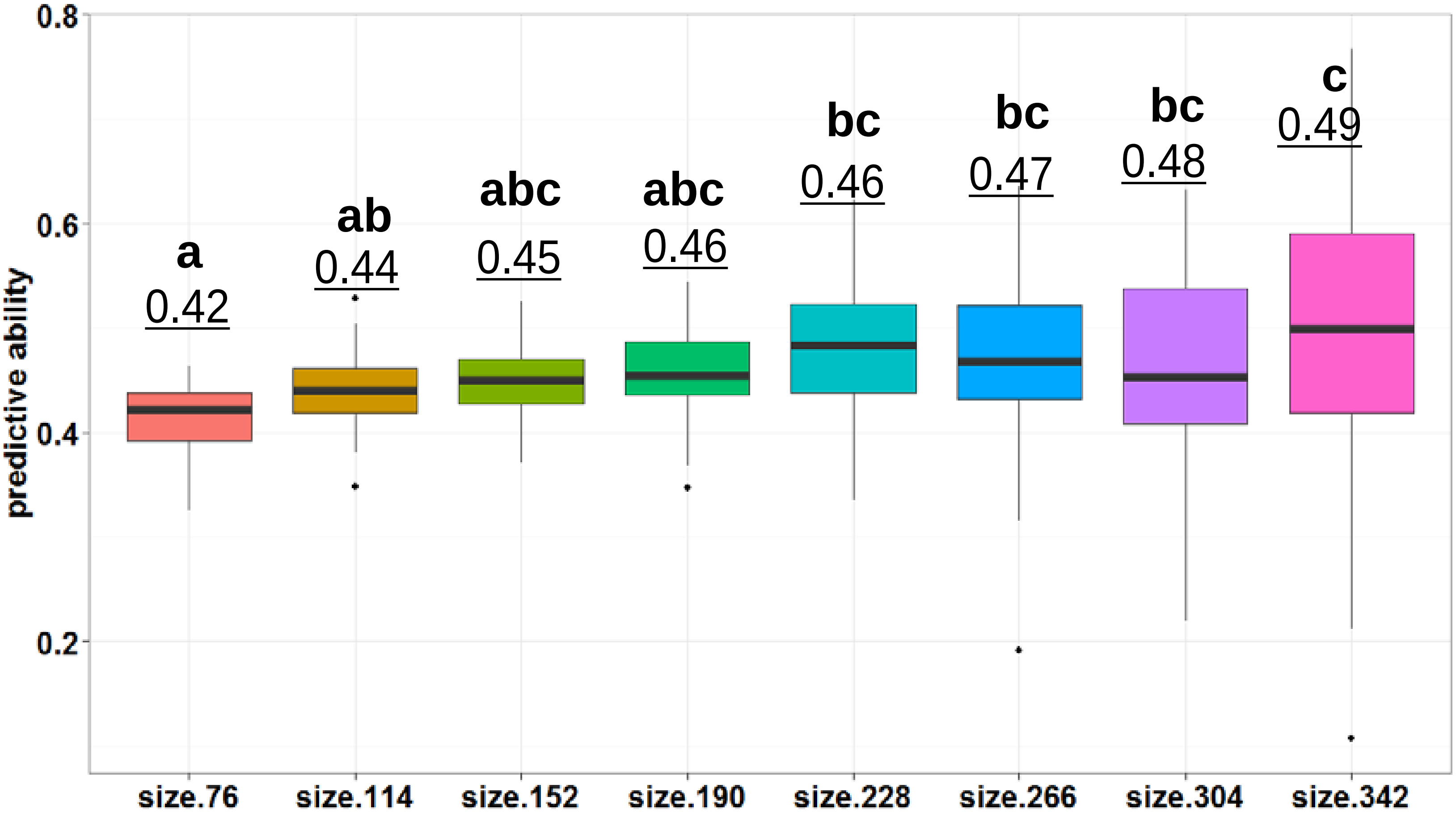
For incidence traits: **orange highlighted are cand**

Scatterplot of predicted vs. observed for **50 cycles** of random sampling, using half (190 F5s) as training set and half (190 F5s) as validation set





Genomic Selection Models for F5 FHB trait do not differ significantly (Tested models include: rrBLUP; BayesA; BayesB; BayesC; BayesBL; BayesBRR; RandomForest)

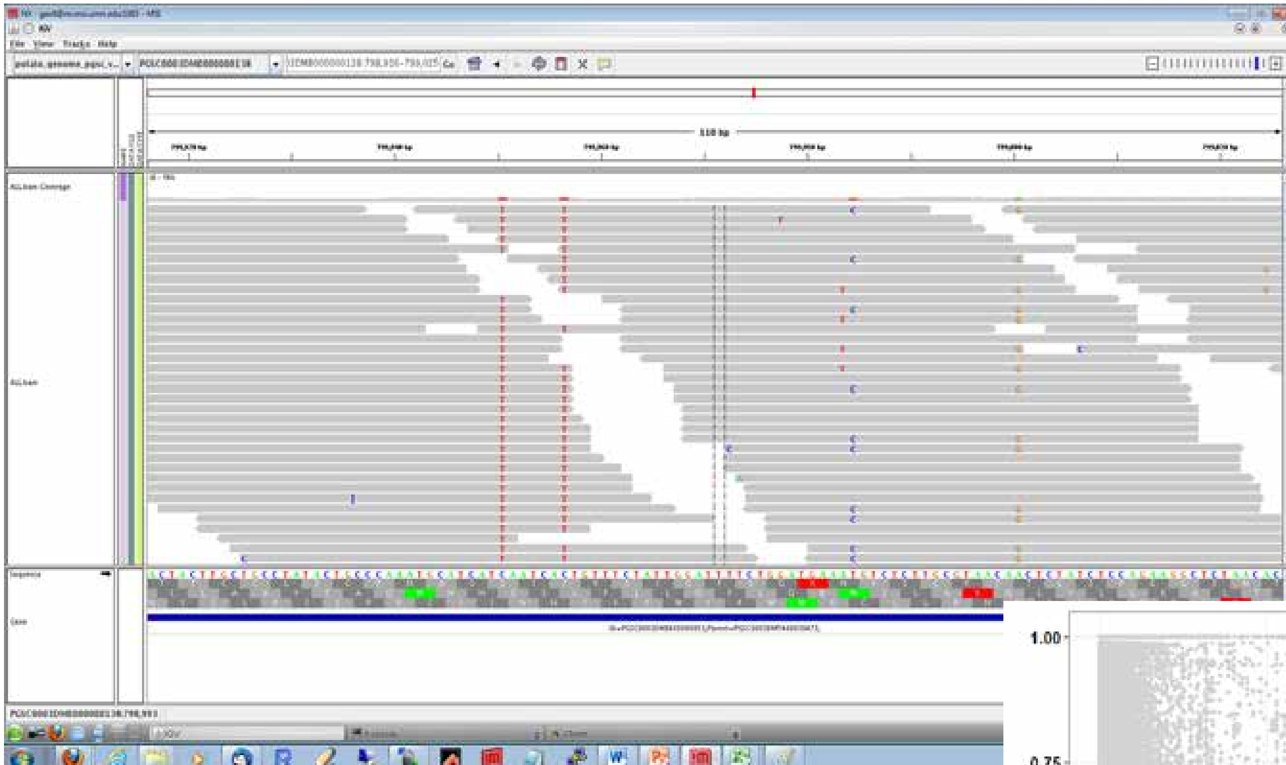


Predictive ability remains statistically unchanged:

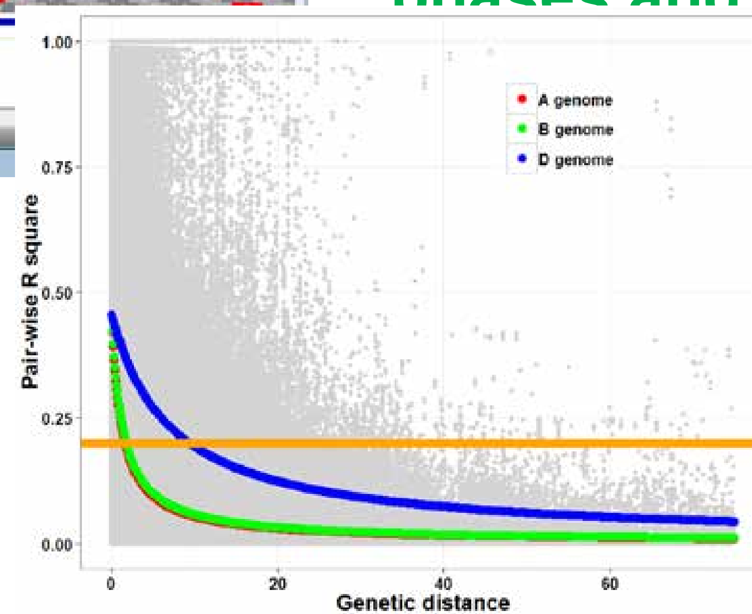
from 228 to 304 F_5's as training individuals

Decreasing TP size might be OK, TP optimizations are being explored

LD based method for finding equivalent markers between GBS and 90K



SNPs mapped to the same positions might have very different phases and



Thus, a common set of individuals (96 or more) may be needed to assign common markers between 90K and GBS, based on LD