



Novel Design of Imputation-Enabled SNP Arrays for Dual Hybridisation of Wheat and Barley

Gabriel Keeble-Gagnère

with colleagues from Agriculture Victoria and InterGrain

IWGSC/Illumina Webinar

15th April 2021

Disclaimer

The opinions expressed during this presentation are those of the speaker and may not represent the opinions of Illumina. Any uses of Illumina products described in this presentation may be uses that have not been cleared or approved by the FDA, or any other applicable regulatory body.



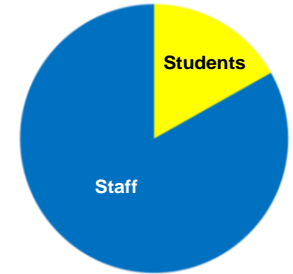
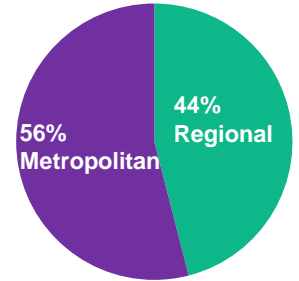
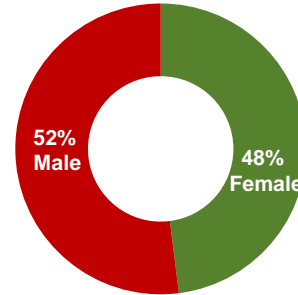
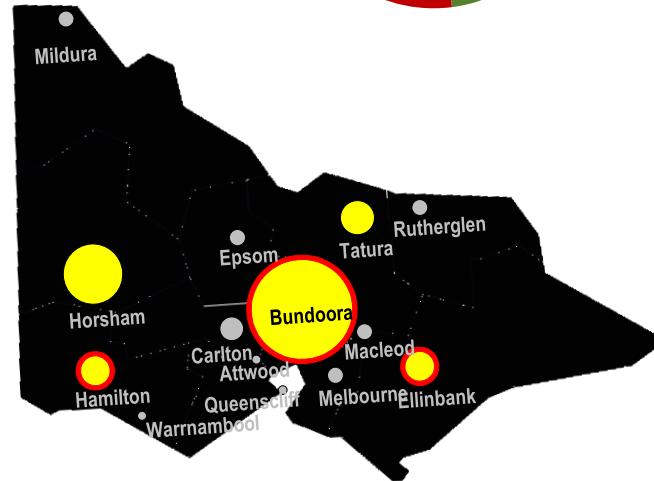
Agriculture Victoria Research

– Science Supporting Agriculture

Achieving **step change** improvements in agriculture through **innovation** for enduring profitability

Enhancing **response and management** of plant and animal **pest and disease outbreaks**

Enhancing the underpinning **innovation ecosystem**





InterGrain Introduction

Australian wheat, barley and oat breeding company

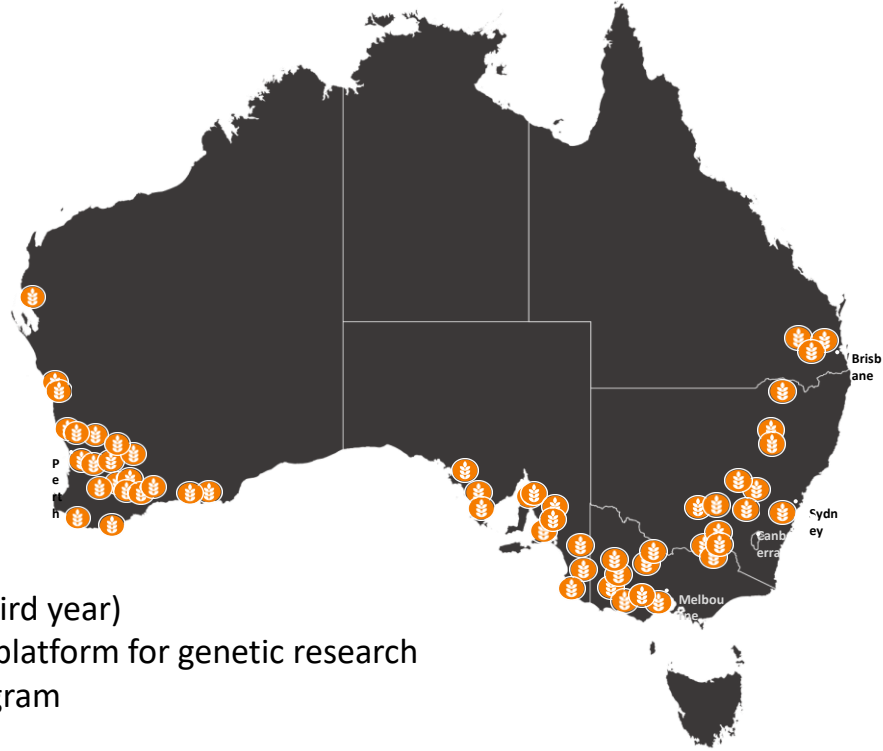
- National breeding programs
- >300,000 plots
- 63 field sites across Australia
- Staff of 55 (including 12 Breeder/Scientists)

InterGrain objectives for chip development:

- 1) Deployment in InterGrain breeding program at scale (now in third year)
- 2) Enhance global research relationships by providing a common platform for genetic research
- 3) Enable rapid translation of research outputs into breeding program

Business takes a collaborative breeding approach
Open to two-way global linkages

Key contact: Hannah Robinson hrobinson@intergrain.com



Variety improvement is applied genomics

High-quality assembled reference genomes



Article | [Open Access](#) | Published: 25 November 2020

Multiple wheat genomes reveal global variation in modern breeding

Sean Walkowiak, Liangliang Gao, [...] Curtis J. Pozniak

Nature 588, 277–283(2020) | [Cite this article](#)

Article | [Open Access](#) | Published: 25 November 2020

The barley pan-genome reveals the hidden legacy of mutation breeding

Murukarthick Jayakodi, Sudharsan Padmarasu, [...] Nils Stein

Nature 588, 284–289(2020) | [Cite this article](#)

Extensive genome-scale genomic resources

Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome

Fel He¹, Raj Pasam², Fan Shi³, Surya Kant⁴, Gabriel Keeble-Gagnere⁵, Pippa Kay⁶, Kerrie Forrest⁷, Allan Fritz⁸, Pierre Huel⁹, Krystalee Wiebe⁹, Ron Knox¹⁰, Richard Cuthbert¹¹, Curtis Pozniak¹², Aina Akhunova¹³, Peter L. Morrell¹⁴, John P. Daleka¹⁵, Steve R. Webb¹⁶, German Spangenberg¹⁷, Ben Hayes¹⁸, Hans Daetwyler¹⁹, Joaquin Tibbits²⁰, Matthew Hayden^{21,22} and Eduard Akhunov²³

Tracing the ancestry of modern bread wheats

Caroline Pont^{1,2}, Thibault Leroy^{1,2,3}, Michael Seidel^{4,5,6}, Alessandro Tondelli^{1,2,3}, Wandrille Duchemin^{1,2}, David Armerisen^{1,2}, Daniel Lang^{7,8,9}, Daniela Bustos-Korts^{10,11}, Nadia Gouel¹², François Balfourier¹³, Márta Molnár-Láng¹⁴, Jacob Lage¹⁵, Benjamin Kilian^{16,17}, Hakan Özkan¹⁸, Darren Walter¹⁹, Sarah Dyer²⁰, Thomas Leffler²¹, Michael Almaraz²², Wheat and Barley Legacy for Breeding Improvement (WHEALB) consortium²³, Joanne Russell²⁴, Beat Keller²⁵, Fred van Eeuwijk²⁶, Manuel Spannagl²⁷, Klaus F. X. Mayer^{28,29}, Robbie Waugh^{30,31}, Nils Stein³², Luigi Cattivi³³, Georg Haberer³⁴, Gilles Charret³⁵ and Jérôme Salse³⁶

Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation

Joanne Russell^{1,2}, Martin Mascher^{3,3,3,3}, Ian K Dawson⁴, Stylianos Kyriakidis⁵, Cristiane Calisto⁶, Fabian Freund⁷, Micha Bayer⁸, Iain Milne⁹, Tony Marshall-Groffins¹⁰, Shane Heinen¹¹, Anna Hofstätter¹², Rajiv Sharma¹³, Axel Himmelbach¹⁴, Manjula Kaur¹⁵, Maarten van Zonneveld¹⁶, John W. Brown¹⁷, Karl Schmid¹⁸, Benjamin Kilian¹⁹, Gary Maehbauer^{20,21}, Nils Stein²² & Robbie Waugh^{23,24}

Hybridisation-based target enrichment of phenology genes to dissect the genetic basis of yield and adaptation in barley

Camilla Beale Hill^{1,2}, Tefera Tolera Angessa^{3,4}, Lee-Anne McFaver⁵, Debbie Wong⁶, Joaquin Tibbits⁷, Xiao-Qi Zhang⁸, Kerrie Forrest⁹, David Moody¹⁰, Paul Telfer¹¹, Sharon Westcott¹², Dean Diepeveen¹³, Yanhao Xu¹⁴, Cong Tan¹⁵, Matthew Hayden^{16,17} and Chengtao Li^{18,24,24}

Genebank genomics highlights the diversity of a global barley collection

Sara G. Milner^{1,2}, Matthias Jost^{3,4,5}, Shin Taketa⁶, Elena Roy Mazzon⁷, Axel Himmelbach⁸, Markus Oppermann⁹, Stephan Weise¹⁰, Helmut Knipfner¹¹, Martin Basterrechea¹², Patrick König¹³, Danuta Schüller¹⁴, Rajiv Sharma^{15,16}, Raj K. Pasam¹⁷, Twan Rutten¹⁸, Ganggang Guo¹⁹, Dongdong Xu²⁰, Jing Zhang²¹, Gerhard Herren²², Thomas Müller²³, Simon G. Krattinger²⁴, Beat Keller²⁵, Yong Jiang²⁶, Maria Y. Gonzalez²⁷, Yisheng Zhao²⁸, Antje Habekuß²⁹, Sandra Färber³⁰, Frank Ordon³¹, Matthias Lang³², Andreas Börner³³, Andreas Gramer³⁴, Jochen C. Reif³⁵, Uwe Scholz³⁶, Martin Mascher³⁷ and Nils Stein³⁸

Multiple genotyping platforms

- Illumina iSelect
 - 90K wheat
 - 50K barley
- Axiom 35K, 820K
- Targeted GBS
- GBS
- Skim WGS
- WGS



How to choose?

How can researchers and breeders best choose a genotyping platform, considering they have different priorities?



Pre-breeding researcher

High density

200 per study

Detailed understanding of the biology, causal variants

Marker density

Number of samples

Key questions

Low density

20,000 per year

How to relate research outputs to my material



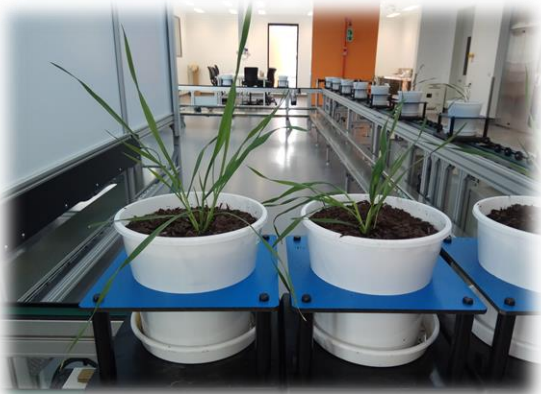
Breeder



Aim

To use the latest advances in genomic resources to design an efficient genotyping platform that provides a bridge between research and breeding.

In particular, a genotyping platform that enables **research**, **breeding** and **rapid deployment** *for impact*.



© Reinhold Möller



AGRICULTURE VICTORIA

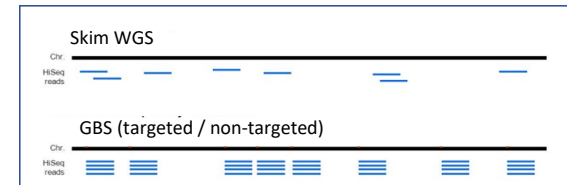
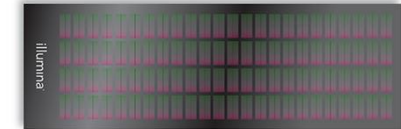
Why Illumina Infinium™ XT technology?

SNP chips vs genotyping-by-sequencing

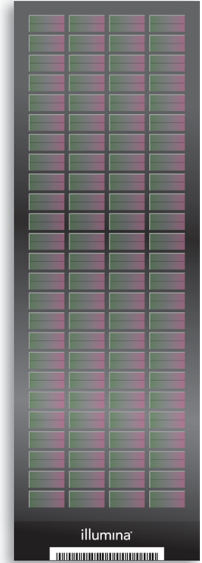
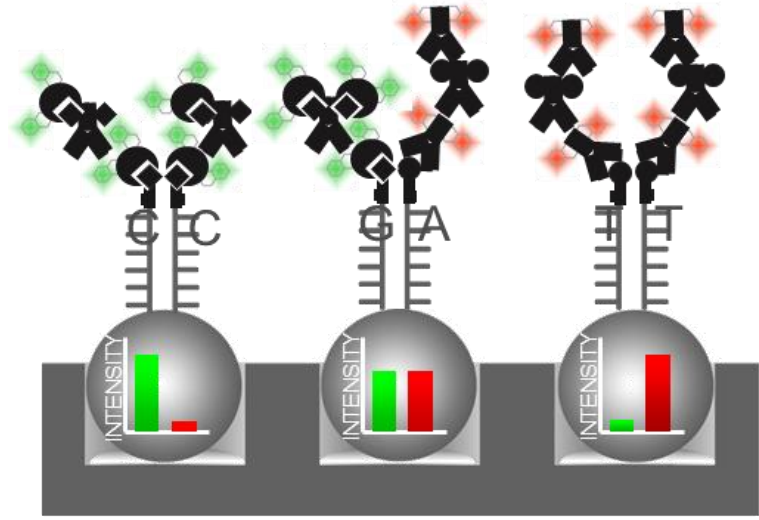
- faster sample turnaround
- higher genotype call rates, typically >90%
- more accurate HET calling
- better suited for large genomes
- easier to deploy at scale

Limitations of existing SNP chips

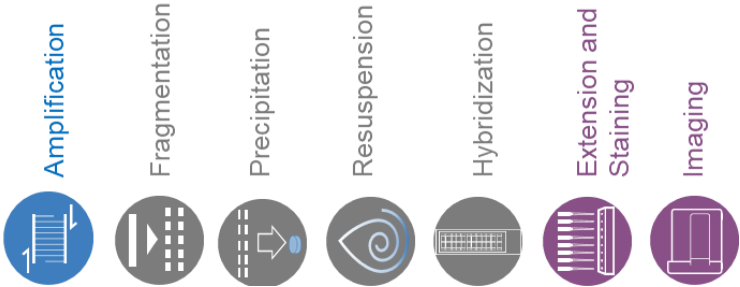
- potential ascertainment bias
- based on incomplete genome assemblies, limited diversity
- markers not always single dose
- SNP density often higher than needed
- designed for specific applications; i.e. not ideal for both research and breeding
- not designed for imputation
- cost



Infinium™ XT Technology



- Array**
 - 96 samples per Array
 - For cost effective, efficient processing
- Content**
 - iSelect up to 50k markers
 - 95%+ conversion rate on custom content
 - Multi-species supported (up to 8 species/array)
- Workflow**
 - Workflow options to meet demand
 - Choose from standard and high-throughput workflow/kits depending on your needs



For Research Use Only. Not for use in diagnostic procedures.

Image Source: Illumina



From theory to practice



Infinium™ Wheat Barley 40K design objectives

To design a genotyping platform that enables *research, breeding and rapid deployment for impact.*

- Design **single locus SNP targets** with **no nucleotide variation underlying assay probes** to maximise genotype calling accuracy
- Use **maximally diverse and representative worldwide germplasm** to minimise ascertainment bias
- Design SNP assays to **tag LD blocks** in worldwide germplasm to **enable accurate imputation**
- Minimise cross-species probe hybridisation to **enable dual sample hybridisation**
- Include specific SNP content to **drive utility in both pre-breeding research and breeding**

Requires good capture of diversity in reference set

Relies on high quality genome assemblies

Literature, in-house research, etc.

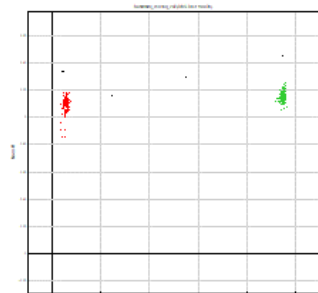


Importance of single locus SNP targets

SNP cluster compression affects genotype calling accuracy especially for heterozygous loci

Each SNP cluster position represents the sum of all loci assayed

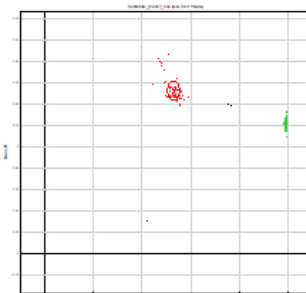
- Dependent on ratio of nucleotides (A or T vs. C or G) and SNP locus copy number
- y axis [R = normalized X + normalized Y] vs. x axis [Theta = $2/\pi \tan^{-1}(C_{y5}/C_{y3})$]



	Red Cluster	Green Cluster
A-genome	T	C
B-genome	-	-
D-genome	-	-

Diploid

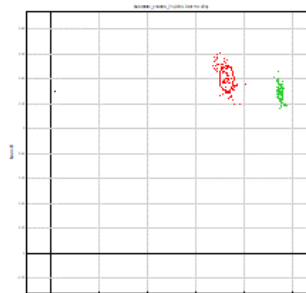
Theoretical Theta value
0.0 and 1.0



	Red Cluster	Green Cluster
A-genome	T	C
B-genome	C	C
D-genome	-	-

Tetraploid

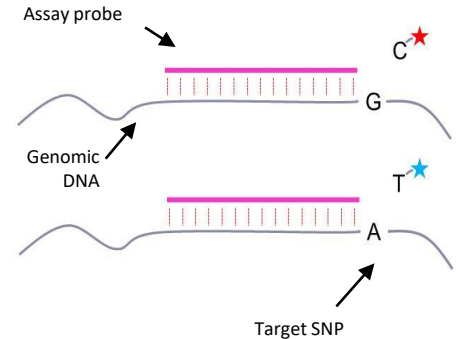
Theoretical Theta values
0.0 and 0.5
0.5 and 1



	Red Cluster	Green Cluster
A-genome	T	C
B-genome	C	C
D-genome	C	C

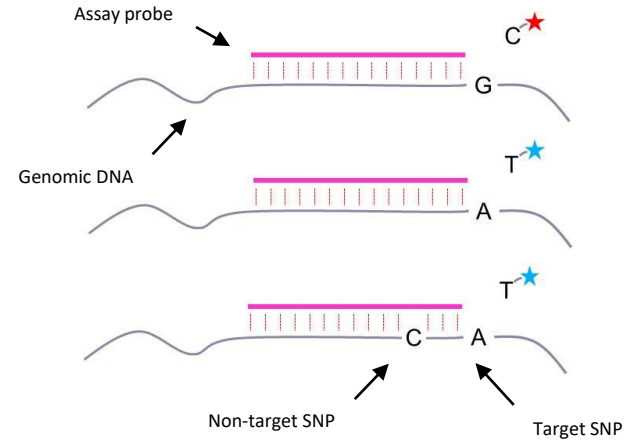
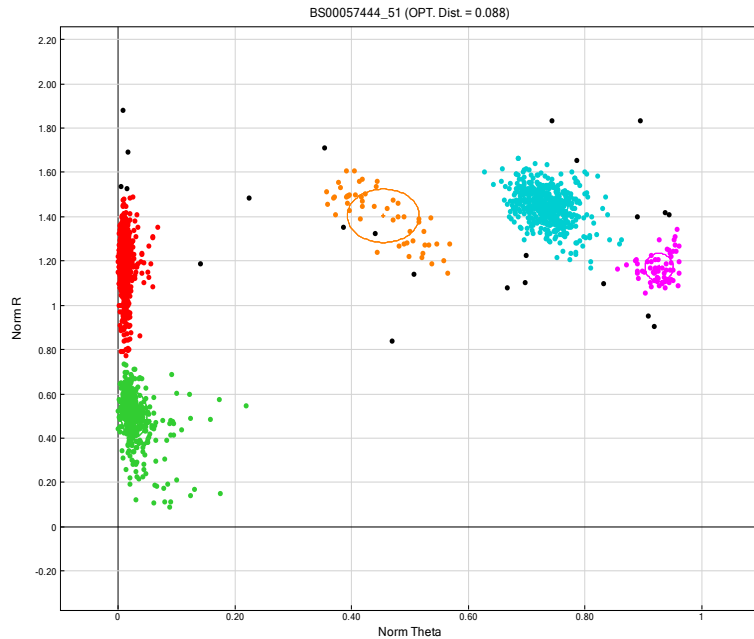
Hexaploid

Theoretical Theta values
0.0 and 0.33
0.67 and 1.0



Importance of probe design

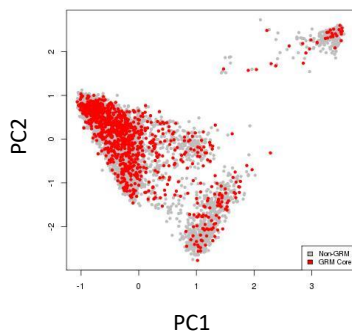
Segregation of nucleotide variants underlying a SNP probe affects genotype calling accuracy



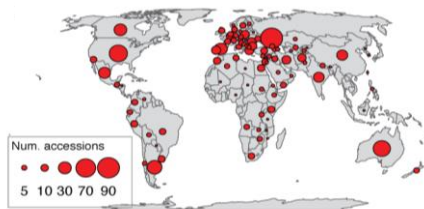
Wheat germplasm

1,200 accessions selected to maximally capture genetic diversity among 6,700 globally diverse wheat accessions comprising landraces, released varieties, synthetic derivatives, and novel trait donor and historical breeding lines

Principal component analysis of world-wide wheat collection based on 90K SNP chip (red circles, exome-sequenced lines)



Geographical distribution of wheat accessions used in He *et al.* 2019 according to country of origin



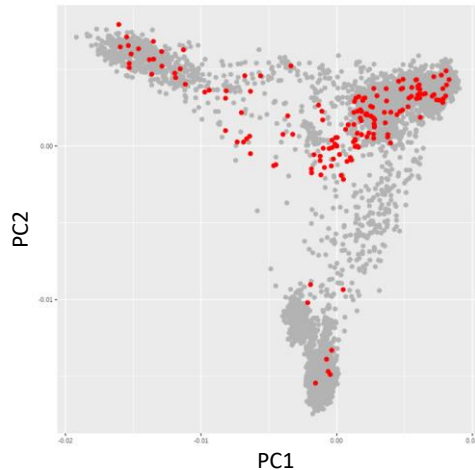
Dataset	Type	# lines	Reference
He <i>et al.</i>	Exome	890	Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nat Genet. 2019 51:896-904
AVR	Exome	310	Globally diverse accessions; unpublished
InterGrain	Exome	102	Historical breeding lines; unpublished



Barley germplasm

267 exome-sequenced *H. vulgare* landraces and *H. spontaneum* accessions selected to capture geographical diversity in barley, as well as 120 whole genome sequenced historical breeding lines and targeted sequence capture data from 1,000 globally diverse *H. vulgare* accessions

Principal component analysis of ~6000 accessions from the 20K IPK genebank collection based on GBS markers (red circles, exome-sequenced landraces)



Dataset	Type	# lines	Reference
Russell <i>et al.</i>	Exome	267	Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. Nat Genet. 2016 48:1024-30
Milner <i>et al.</i>	GBS	19,778	Genebank genomics highlights the diversity of a global barley collection. Nat Genet. 2019 51:319-326
Hill <i>et al.</i>	Capture	1000	Hybridisation-based target enrichment of phenology genes to dissect the genetic basis of yield and adaptation in barley. Plant Biotech. J. 2019 17:932-944
InterGrain	WGS	120	Historical breeding lines; unpublished

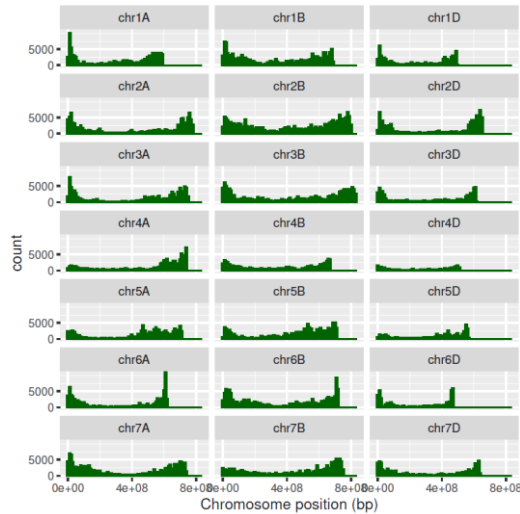


SNP discovery

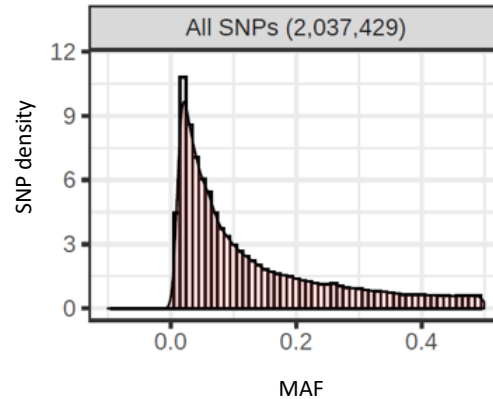
Wheat

2,037,429 SNPs

Against IWGSC RefSeq v1.0



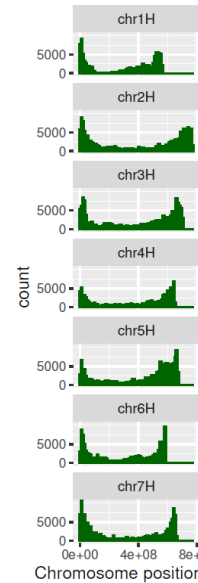
MAF distribution of wheat SNP



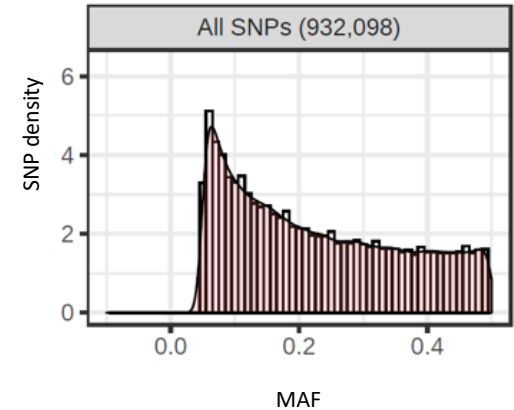
Barley

932,099 SNPs

Against Morex v1.0



MAF distribution of barley SNP



Filters Applied: missing < 60%, MAF > 1% in wheat (MAF > 5% in barley), < 15% HET



Linkage Disequilibrium (LD)

Linkage disequilibrium refers to the non-random association of alleles at two or more loci in a general population. When alleles are in **linkage disequilibrium**, haplotypes **do not** occur at the expected frequencies

r^2 is a common measure of linkage disequilibrium,

$$r = \frac{D}{\sqrt{p_A(1-p_A)p_B(1-p_B)}}.$$

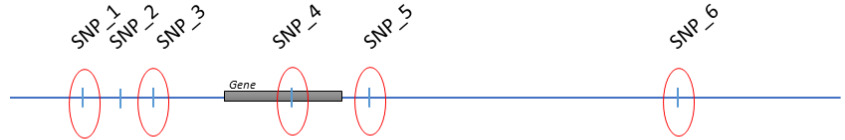
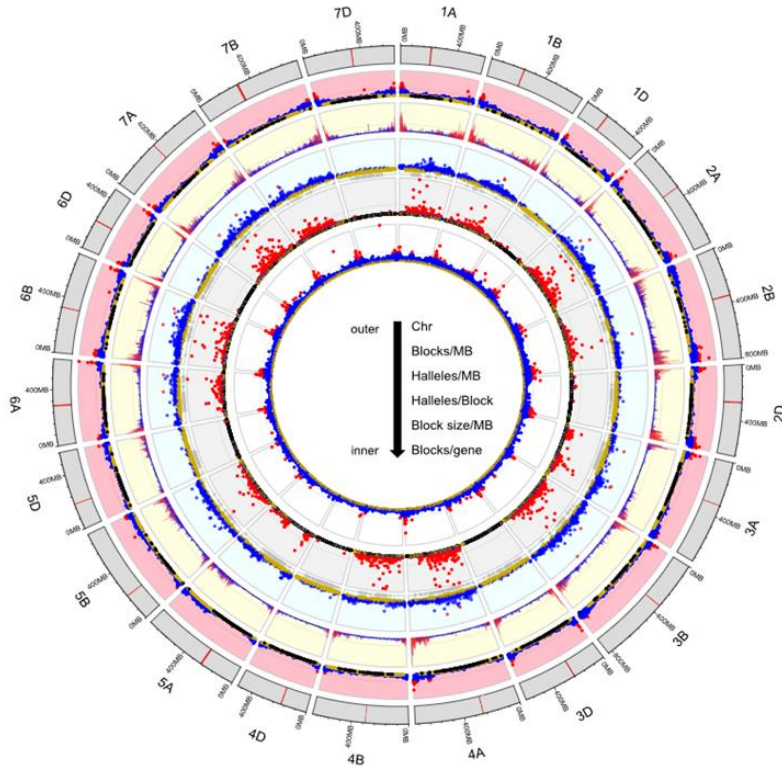
Importantly, when alleles are in LD, they are not independent.

This means that information about one locus is informative about other locus.

Goode E.L. (2011) Linkage Disequilibrium.
In: Schwab M. (eds) Encyclopedia of Cancer. Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-642-16483-5_3368



Defining relationship between SNP using linkage disequilibrium



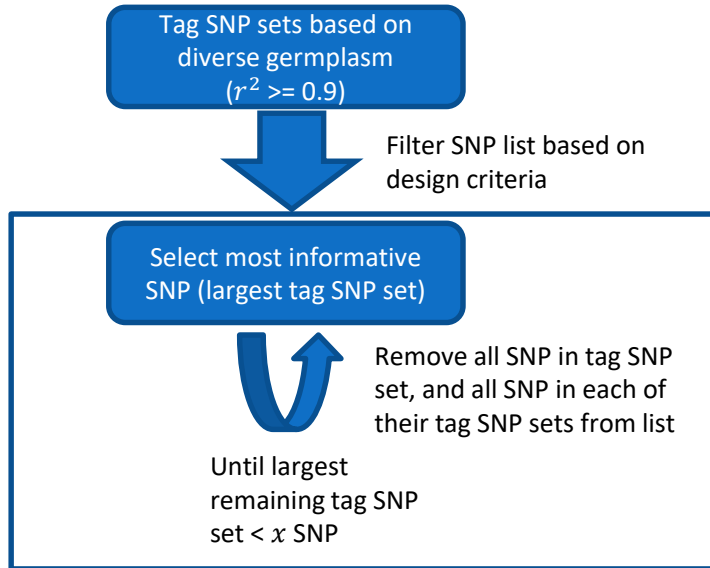
	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	SNP_6
SNP_1	0.960541	0.22	0.966894	0.973496	0.990057	0.960252
SNP_2		0.2	0.986795	0.953663	0.963772	0.973496
SNP_3			0.13	0.23	0.2	0.2
SNP_4				0.952943	0.963518	0.972976
SNP_5					0.976748	0.952773
SNP_6						0.956864

r^2

SNP_1, SNP_3, SNP_4, SNP_5, SNP_6 form a **tag SNP set**. Selecting one of these SNP, eg: SNP_1, the remaining 4 SNP are referred to as tSNP



SNP selection algorithm

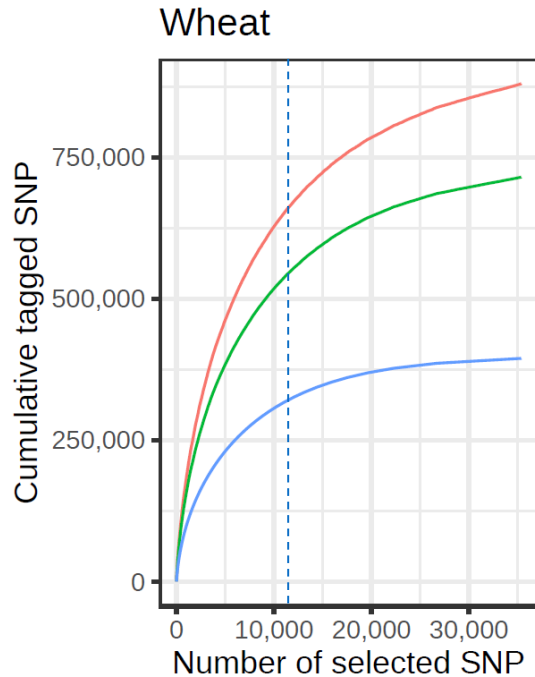


Repeat process with $x = 10, 20, 50$

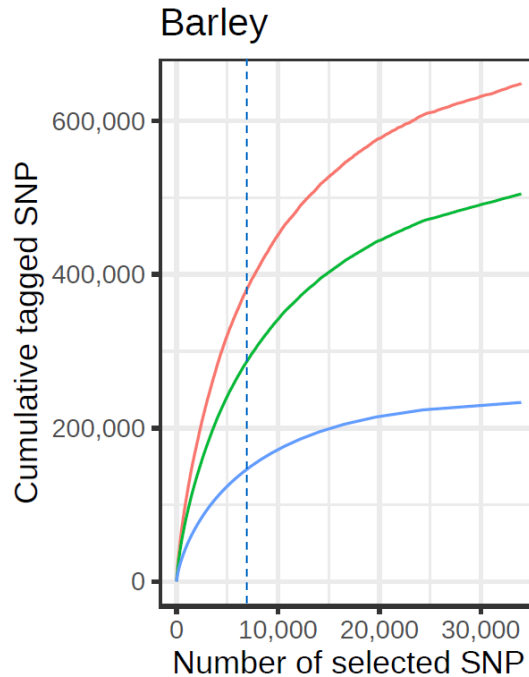
- Considers SNP **equivalent** allowing further filtering
- **Not biased** by pre-existing biological knowledge
- **Minimises** SNP number to **maximise** LD capture
- Repeat process builds in **redundancy** for large tag SNP sets, as well as **increases confidence** for NULL calls

Tagging LD blocks in globally diverse germplasm

Result of applying selection algorithm to wheat and barley. Dotted line shows end point of first pass



r^2
— 0.5
— 0.7
— 0.9

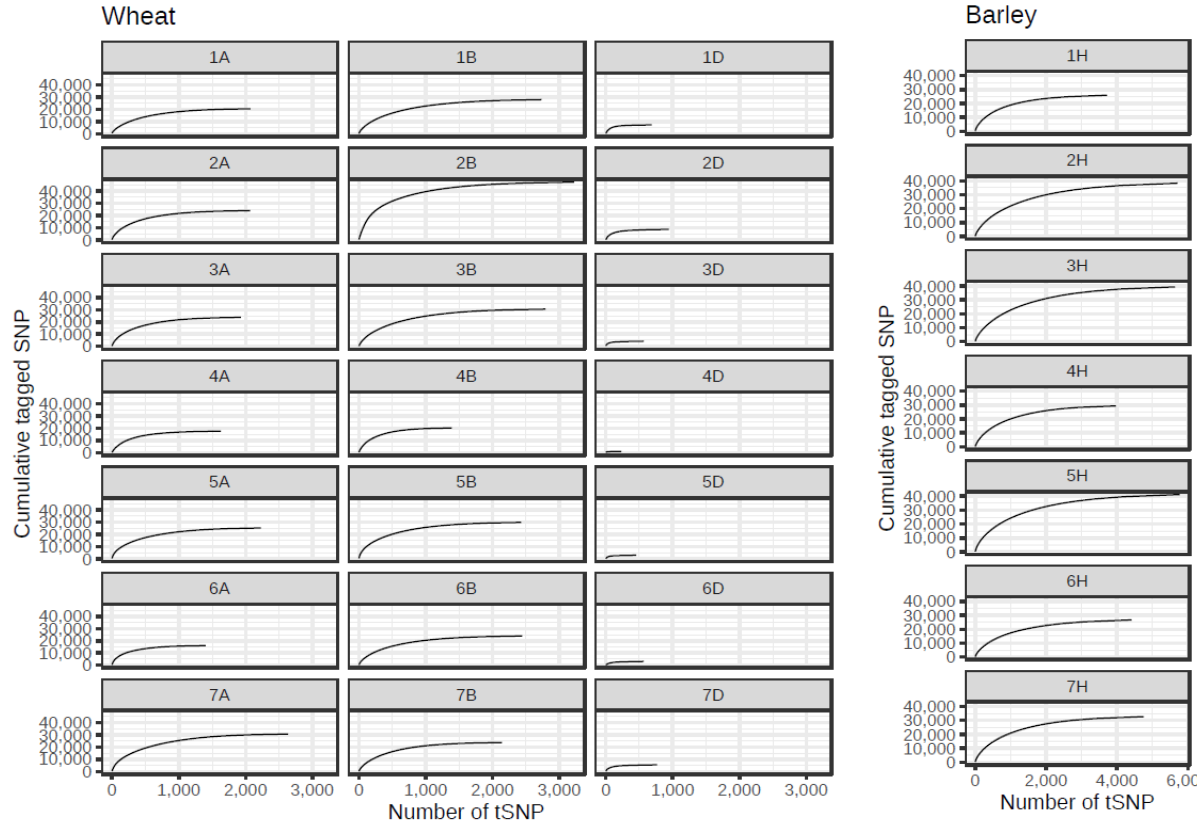


r^2
— 0.5
— 0.7
— 0.9



Tagging LD blocks in globally diverse germplasm

Result of applying selection algorithm to wheat and barley, per chromosome

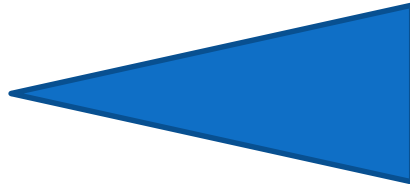


Importance of minor allele frequency (MAF) distribution

Need to balance between a range of application-specific needs: from bi-parental mapping populations (where any polymorphism between parents is sufficient) to global diversity (where rare subpopulation-specific variants are important to differentiate divergent samples, important when introgressing novel genetic variation, for example).

Low MAF

- Important to differentiate varieties from each other (reduce ascertainment bias)
- Chance SNP appears as monomorphic SNP when population diversity is limited

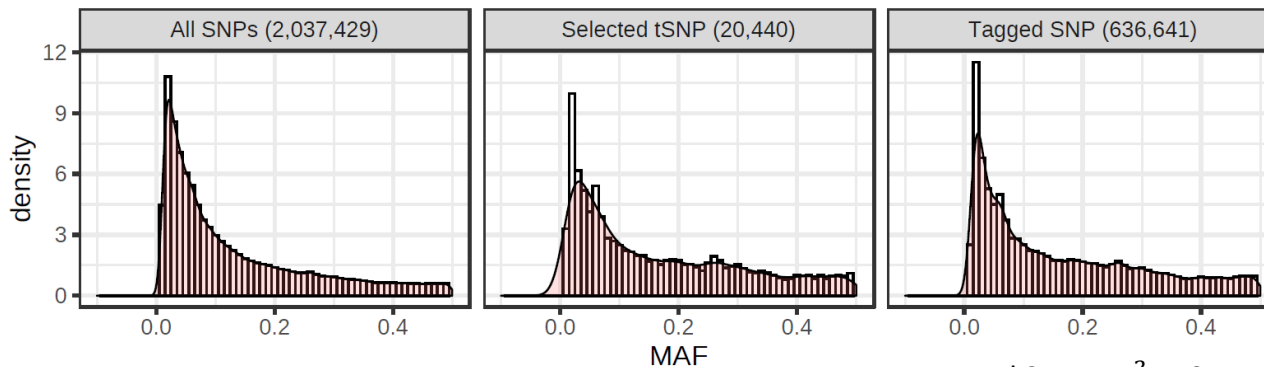


High MAF

- Important for background variance and LD estimates
- Lower MAF variants will regress onto high MAF SNP

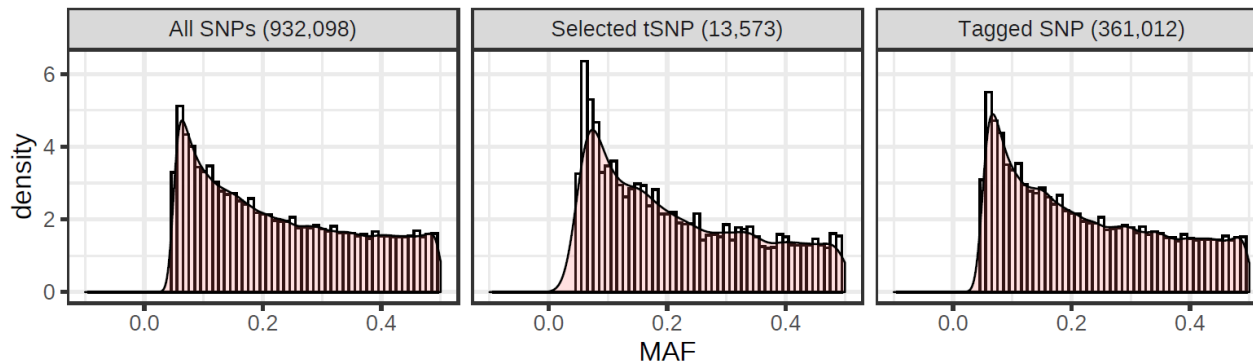
MAF distribution

Wheat



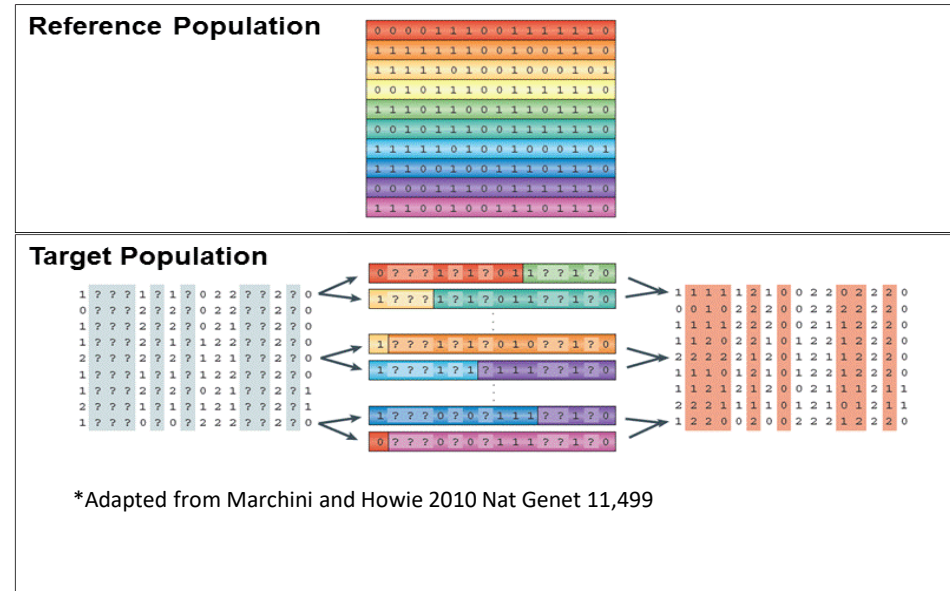
Tagged SNP at $r^2 \geq 0.7$

Barley

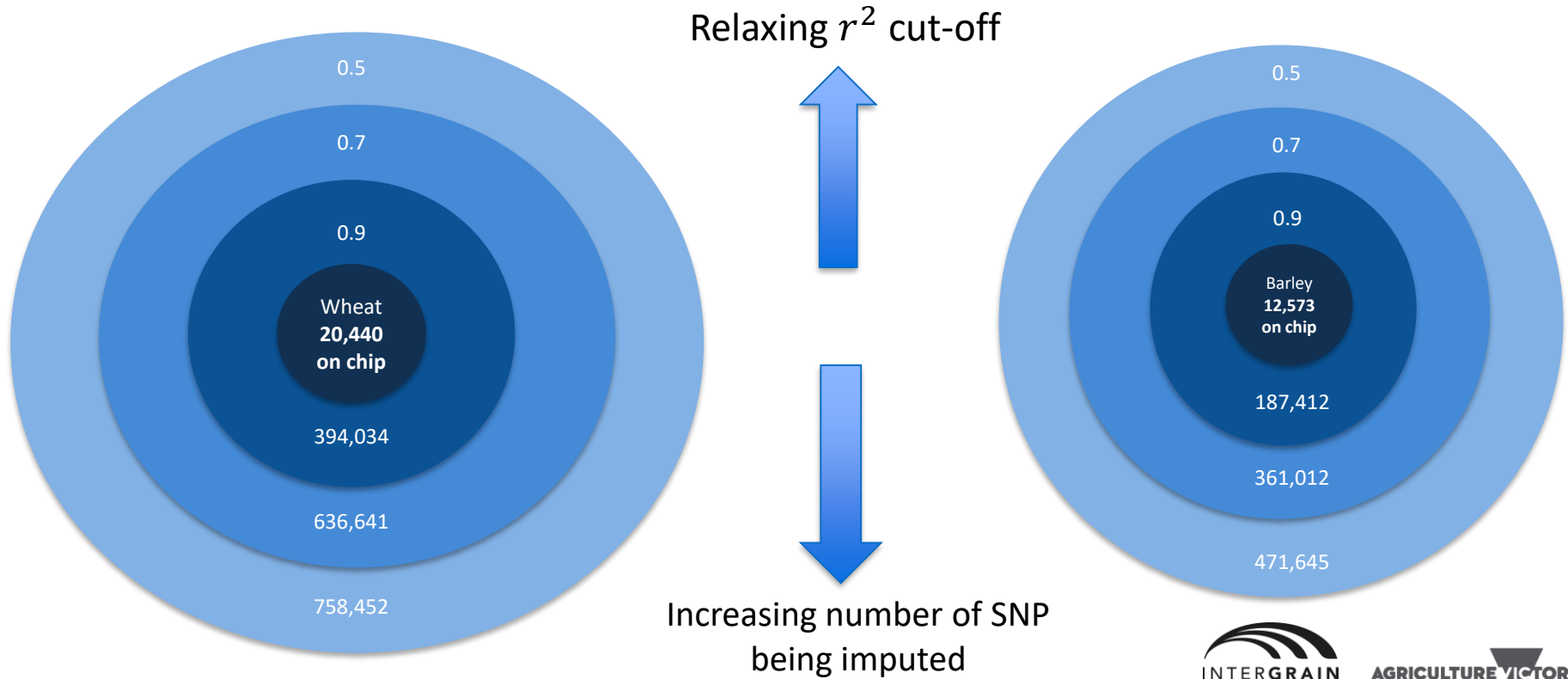


Using imputation to recover statistical power

- With the SNP selection algorithm we have defined a method to maximise informativeness of a small set of SNP.
- This is important because in order to reduce cost per sample, number of SNP on chip needs to be as low as possible.
- The downside is that with fewer SNP, we have less statistical power relative to higher density assays, eg: exome SNP discovery.
- Imputation is a statistical method to infer missing genotypes using a set of reference haplotypes
 - The selected SNP maximise number of tSNP
 - Using imputation we can recover a large proportion of the SNP we do not genotype and *increase statistical power*



Defining the imputation target set



Imputation accuracy in wheat

Based on 100-fold in-silico cross-validation

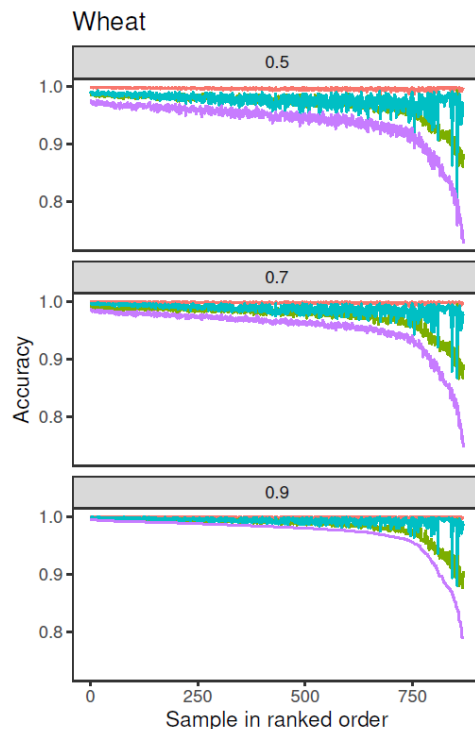
XT SNP imputed to tSNP set for:

$r^2 \geq 0.5$ (758,452 SNP)

$r^2 \geq 0.7$ (636,641 SNP)

$r^2 \geq 0.9$ (394,034 SNP)

Strictest measure of accuracy (r^2 including heterozygous positions) remains well above 90% for majority of samples



Exome genotypes

Sporadic missing data filled in with Beagle, phased with Eagle

Filled-in exome genotypes

Extract random set of 100 samples, selected SNP only

XT genotypes for 100 lines

Exome genotypes for remaining samples

Imputation with Minimac 3 *

Imputed genotypes

Minimac: Das et al. (2016) Nature Genetics doi:10.1038/ng.3656

Beagle: Browning and Browning (2016). Am J Hum Genet. doi:10.1016/j.ajhg.2015.11.020

Eagle: Loh et al. (2016) Nature Genetics doi: 10.1038/ng.3571



Imputation accuracy in barley

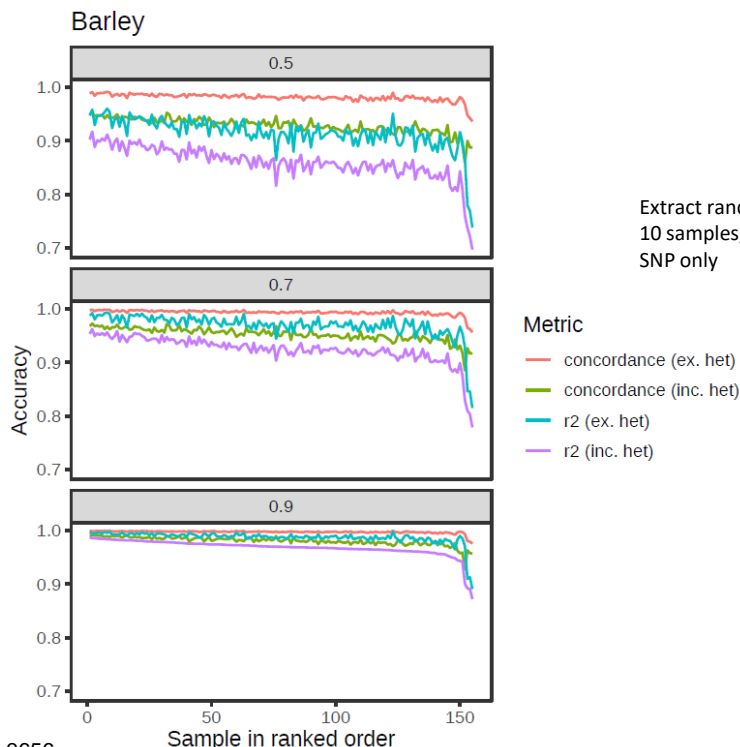
Based on 100-fold in-silico cross-validation

XT SNP imputed to tSNP set for:

$r^2 \geq 0.5$ (471,645 SNP)

$r^2 \geq 0.7$ (361,012 SNP)

$r^2 \geq 0.9$ (187,412 SNP)



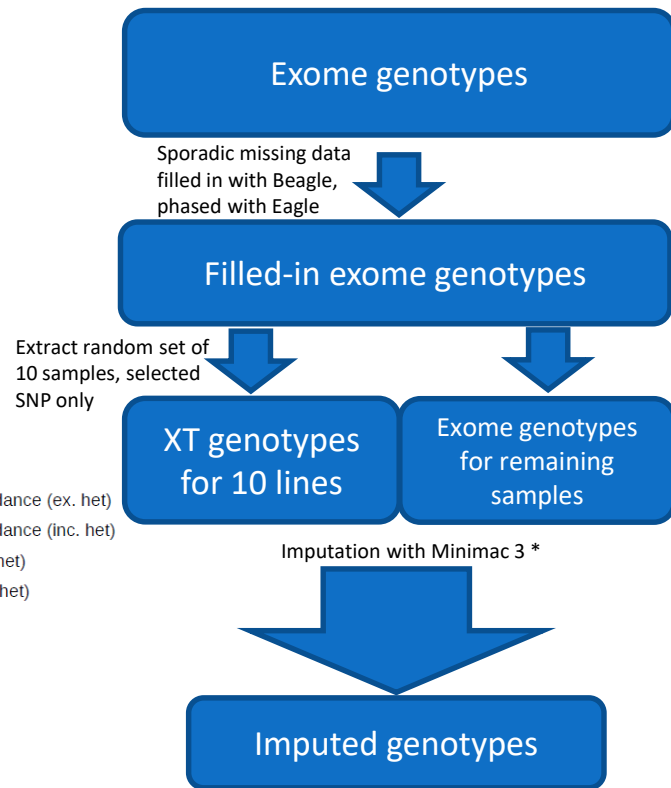
Reduced accuracy relative to wheat, though strictest measure of accuracy (r^2 including heterozygous positions) remains above 90% for majority of samples in $r^2 \geq 0.7$ target.

Expect improved accuracies as reference population size is increased.

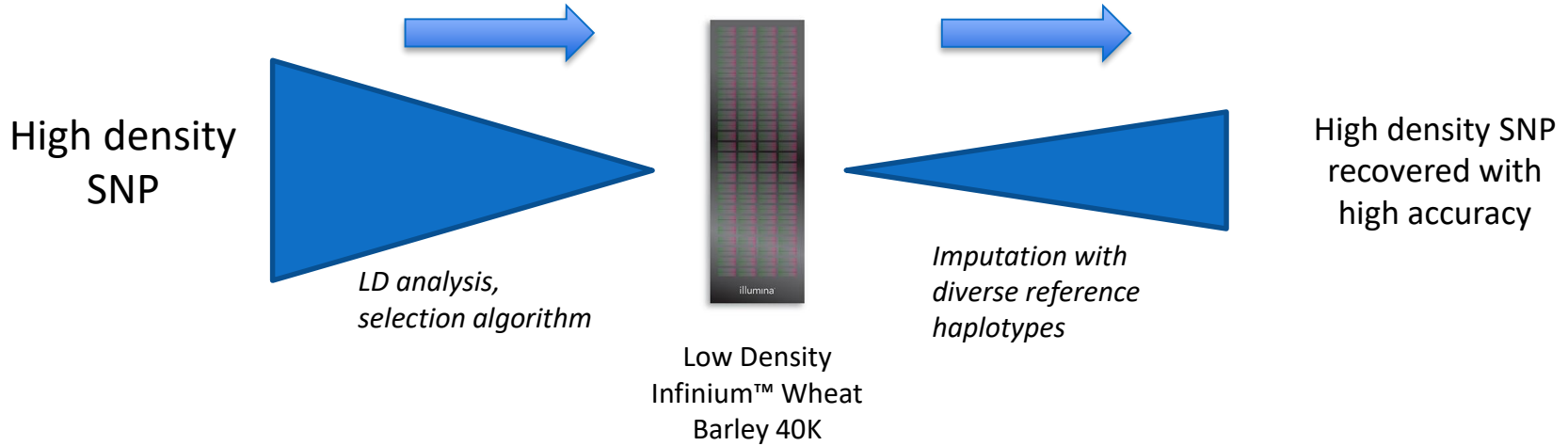
Minimac: Das et al. (2016) Nature Genetics doi:10.1038/ng.3656

Beagle: Browning and Browning (2016). Am J Hum Genet. doi:10.1016/j.ajhg.2015.11.020

Eagle: Loh et al. (2016) Nature Genetics doi: 10.1038/ng.3571



Imputation Summary



In addition to increasing statistical power relative to raw genotypes, we can use imputation to *integrate different genotyping platforms*.

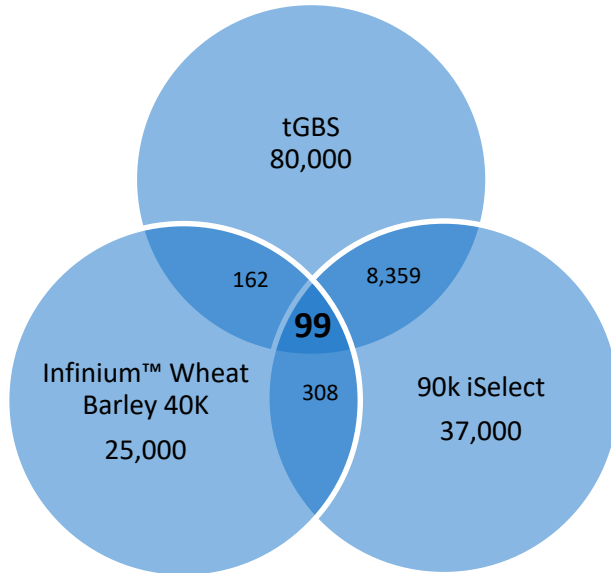
Integrating across platforms – before imputation

Wheat

90K iSelect
~37,000 SNPs
(anchored to genome)

tGBS
~80,000 SNPs

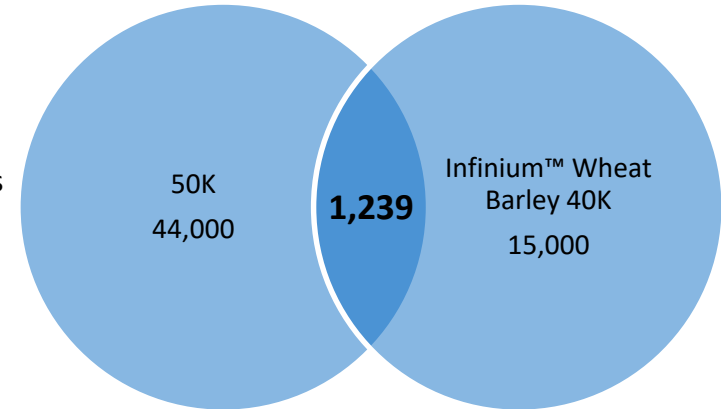
Infinium™ Wheat
Barley 40K
~25,000 SNPs



Barley

50K iSelect
~44,000 SNPs

Infinium™ Wheat
Barley 40K
~15,000 SNPs



Limited overlap between SNP content on different platforms – especially when the intersection of multiple platforms is considered



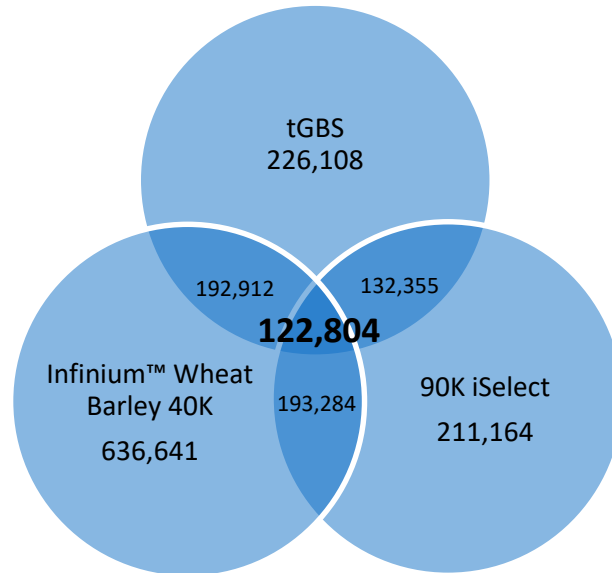
Infinium™ Wheat Barley 40K provides a step-change for imputing to high density

Imputation to high density allows **legacy** genotypic datasets generated from germplasm genotyped using existing technologies to be integrated for downstream analyses such as genomic selection and GWAS

Population 1
(Illumina 90K wheat SNP array)

Population 2
(tGBS)

Population 3
(Infinium™ Wheat Barley 40K)



Combined dataset with
122,804 SNP imputed
with high accuracy

Numbers based on current data

Based on $r^2 \geq 0.7$ imputation target
Note: Relies on SNP overlap with exome



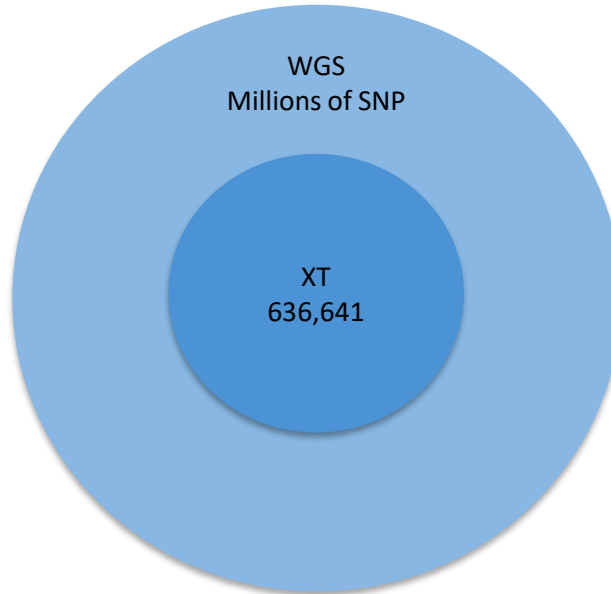
Infinium™ Wheat Barley 40K provides a step-change for imputing to high density

Future-oriented design creates opportunity to achieve much higher SNP density for integration of future genotypic datasets, as well as **a process to continually update chip content**

Population 1
(Infinium™ Wheat Barley 40K)

Population 2
(WGS)

Population 3
(skim WGS)



Whole genome density

Three things deliver this:

1. WGS of reference population increase number of SNP in LD
2. Increased SNP density for other germplasm provides more overlap to give larger common imputed SNP set
3. Future iterations of array content will improve tag SNP set size

As datasets grow, Infinium™ Wheat Barley 40K chip performance can evolve in parallel



Enhancing utility of Infinium™ Wheat Barley 40K for research and breeding

Provides a bridge between marker-assisted breeding (MAS) and Genomic Selection (GS)

SNP Content	Barley	Wheat
Published trait linked markers	184	483
Published SNP tagging GWAS signals for key breeding traits	0	467
Markers providing direct link to GBS genotyped 20K barley accessions from IPK genebank (Milner et al. 2019 Nat Genet. 51:319-326)	685	n/a

Wheat traits: metal ion tolerances, fungal and nematode resistances, frost tolerance, phenology, glutenin sub-units, LMA, PHS, grain end-use quality, plant height, herbicide tolerance

Barley traits: metal ion tolerances, fungal and nematode resistances, frost tolerance, phenology, malting quality, grain characteristics, row number, herbicide tolerance

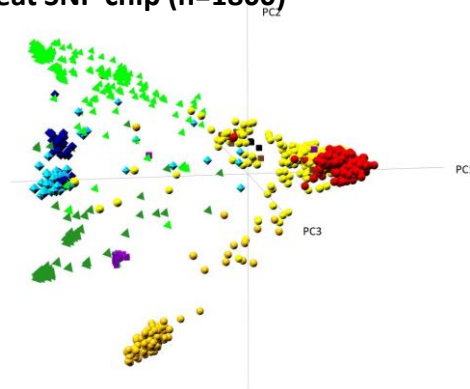


Enhancing utility of Infinium™ Wheat Barley 40K for research and breeding

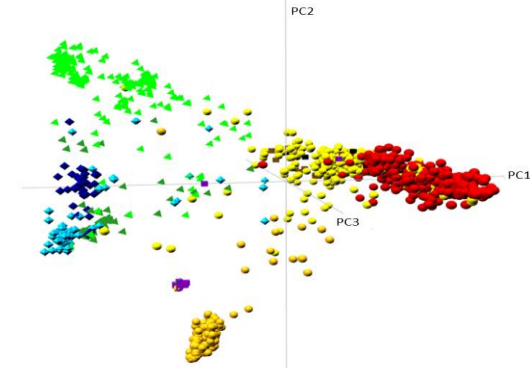
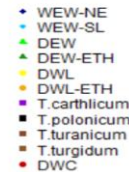
2,615 iSelect 90K SNP chip markers tagging tetraploid (A- and B-genome) diversity

- comprises SNP private to wild and domesticated emmer, durum landraces and *T. turgidum* subspecies, identified from genotyping 2,000 accessions capturing tetraploid diversity (Maccaferri *et al.* 2019 Nat. Genet. 51, 885-895)

PCA showing various groups of tetraploid wheat accessions genotyped using 90K wheat SNP chip (n=1800)



PCA showing various groups of tetraploid wheat accessions genotyped using Infinium™ Wheat Barley 40K chip (n=1800)



Same set of accessions

WEW-NE: Wild emmer wheat from North Eastern Fertile Crescent; WEW-SL: Wild emmer wheat from Southern Levant Fertile Crescent; DEW: Domesticated emmer wheat; DEW-ETH: Domesticated emmer wheat from Ethiopia; DWL: Durum wheat landraces; DWC: Durum wheat cultivars

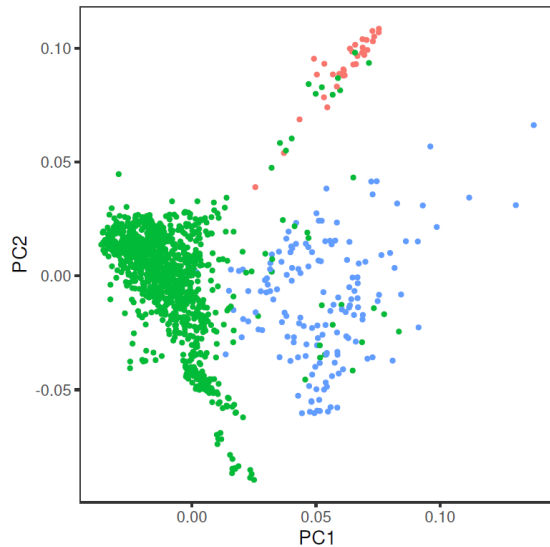


Enhancing utility of Infinium™ Wheat Barley 40K for research and breeding

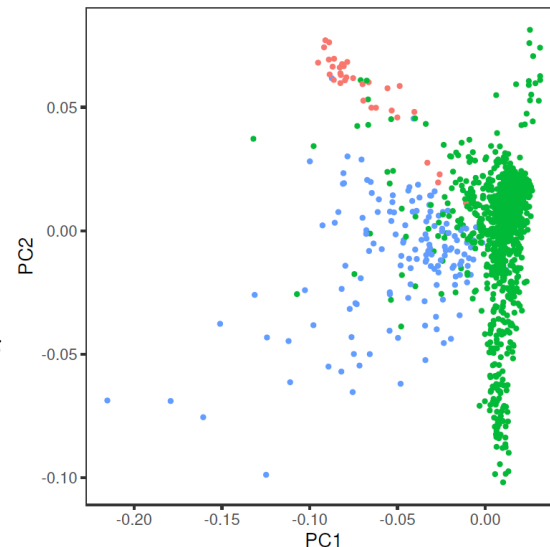
1,226 exome SNPs tagging *Ae. tauschii* (D-genome) diversity

- comprises *tauschii*-specific SNP selected for high D-genome coverage, MAF ≥ 0.10 in synthetic derivatives (derived from crosses involving 100 *tauschii* accessions) and enriched density in highly recombining chromosomal regions

PCA showing bread wheat accessions genotyped using 90K wheat SNP chip



PCA showing bread wheat accessions genotyped using Infinium™ Wheat Barley 40K chip



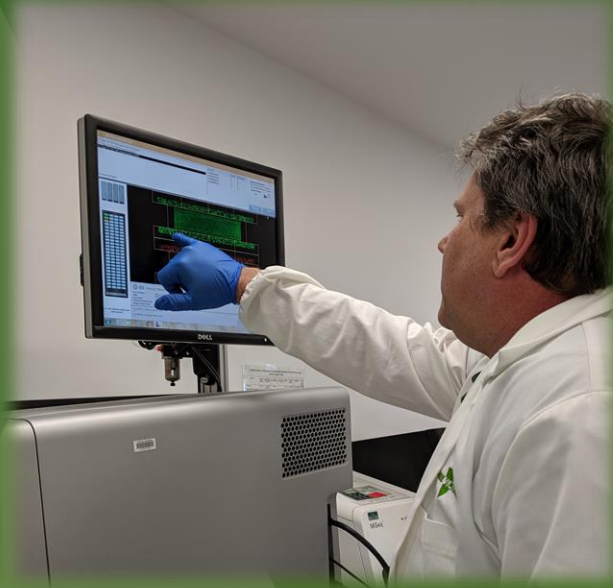
Bread wheat
Synthetic derivatives
Durum-like bread wheat

Same set of accessions

Summary of content on Infinium™ Wheat Barley 40K

Version 1.0	Wheat	Barley	Total
Tagging SNP for imputation	20,665	13,469	34,134
Trait associated SNP	457	178	635
SNP linking germplasm resources	4,271	614	4,885
Total number of SNP	25,393	14,261	39,654





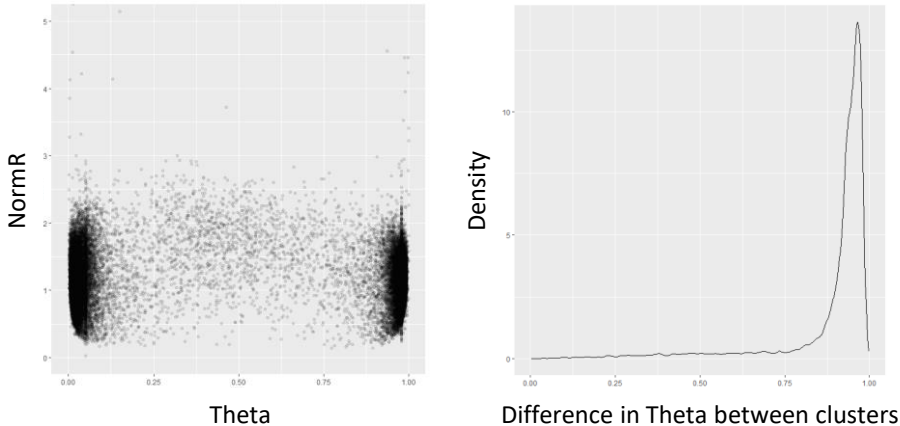
So, how does it actually perform?



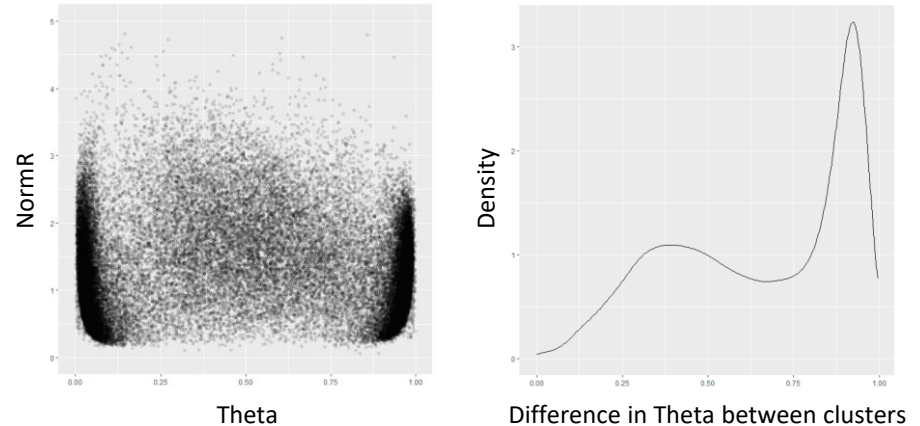
SNP assay specificity – Single sample hybridisation

Majority of SNP assays interrogate a single locus in barley and wheat

Barley SNP cluster positions for 14,168 assays revealing polymorphism



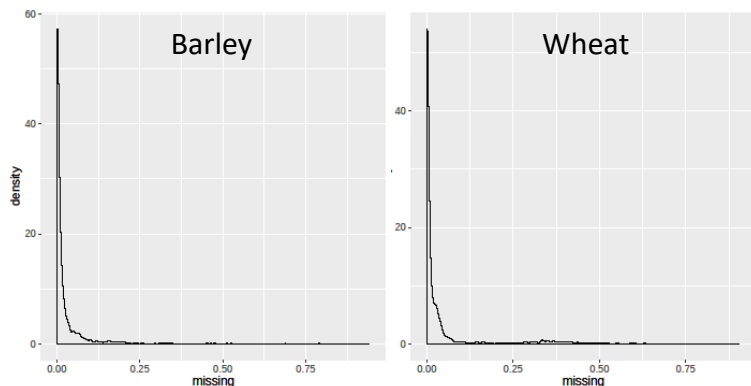
Wheat SNP cluster positions for 24,100 assays revealing polymorphism



Genotype calling – Single sample hybridisation

High genotype call rate per sample; very high concordance between called and true genotype

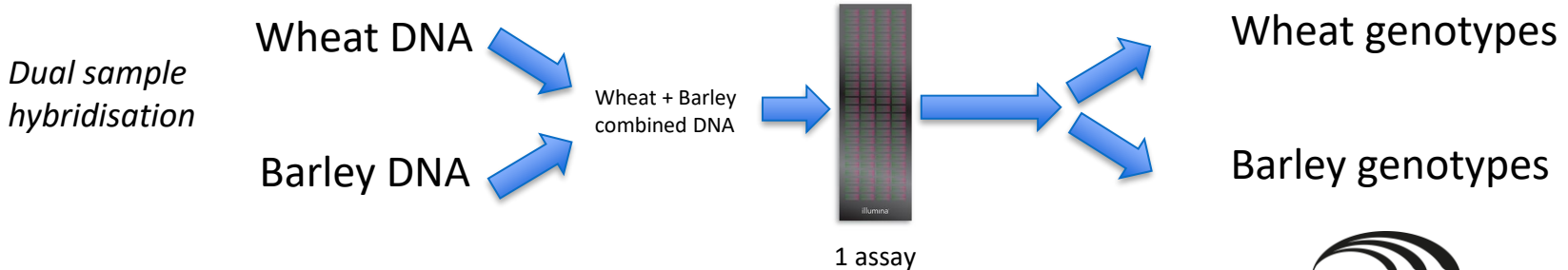
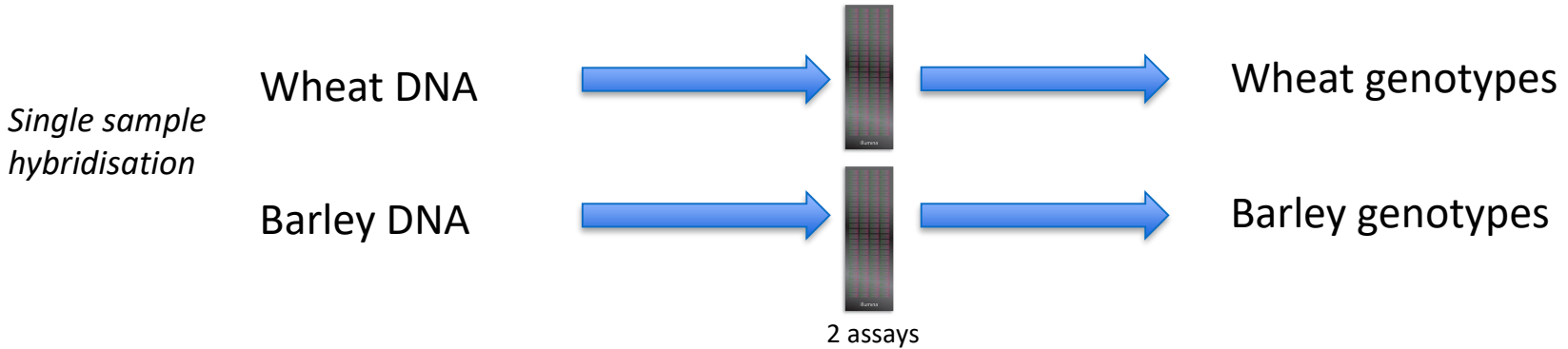
Missing data rate (%)



Average (%) correlation and concordance between Infinium™ Wheat Barley 40K (single sample hybridisation) and exome-sequenced genotypes

	Including HET calls		Excluding HET calls	
	Correlation	Concordance	Correlation	Concordance
Barley (90 samples)	97.2	98.2	99.2	99.8
Wheat (423 samples)	95.7	97.6	98.1	99.5

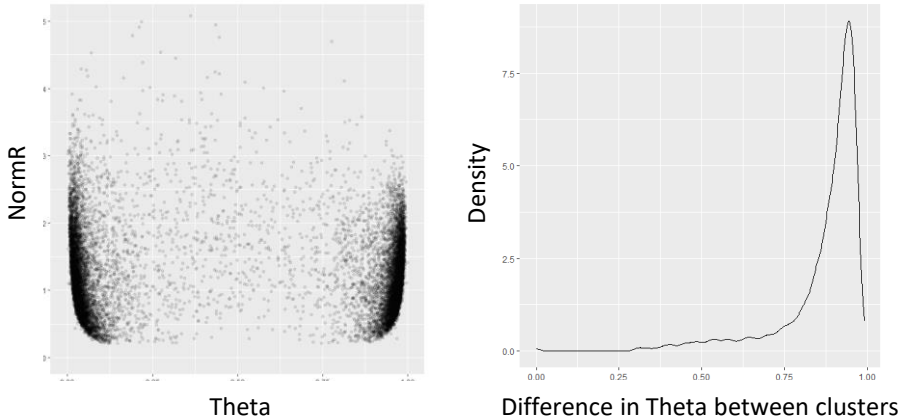
Single sample vs Dual sample hybridisation



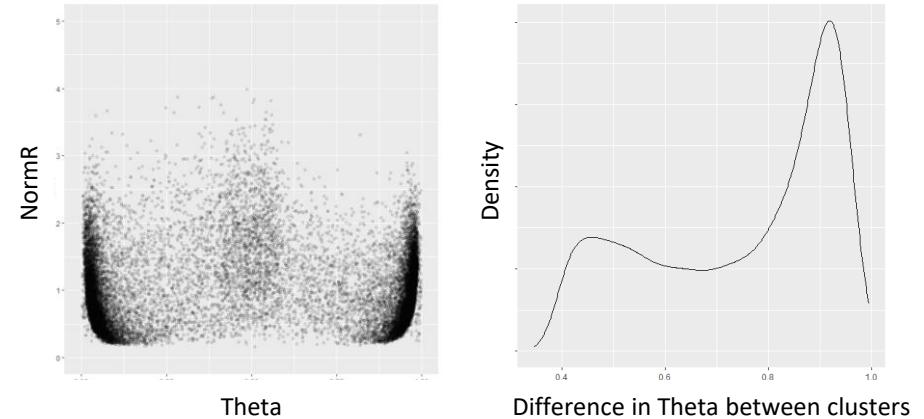
SNP assay specificity – Dual sample hybridisation

70% and 40% of SNP assays show good cluster separation in barley and wheat

**Barley SNP cluster positions for
9,826 assays revealing polymorphism**



**Wheat SNP cluster positions for
9,118 assays revealing polymorphism**



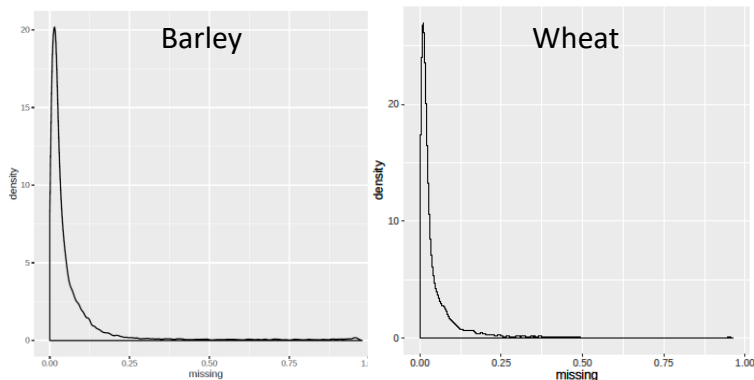
⇒ Very strict criteria was applied to choose which SNP were used for genotype calling; in hindsight this criteria could be significantly relaxed



Genotype calling accuracy – Dual sample hybridisation

High genotype call rate per sample; very high concordance between genotype calls from single and dual hybridisation assays of same samples

Missing data rate (%)



Average (%) correlation and concordance between Infinium™ Wheat Barley 40K single and dual sample hybridisation genotypes

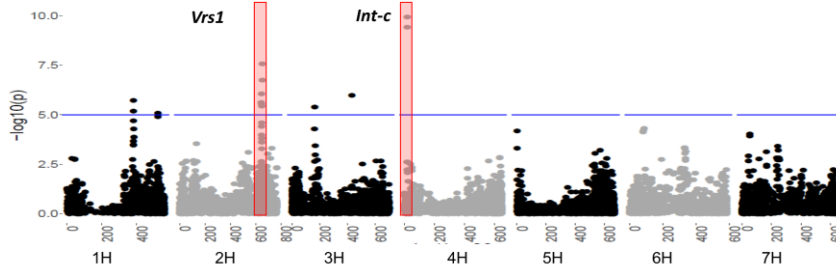
	Including HET calls		Excluding HET calls	
	Correlation	Concordance	Correlation	Concordance
Barley (547 samples)	99.2	99.4	99.8	99.9
Wheat (63 samples)	95.3	97.8	97.0	99.3

⇒ SNP attrition in dual hybridisation assays can be accurately recovered in *breeding germplasm* via imputation



GWAS with Infinium™ Wheat Barley 40K

- Germplasm* from the IPK exome-sequenced landraces (Russell *et al.*), genotyped on the Infinium™ Wheat Barley 40K



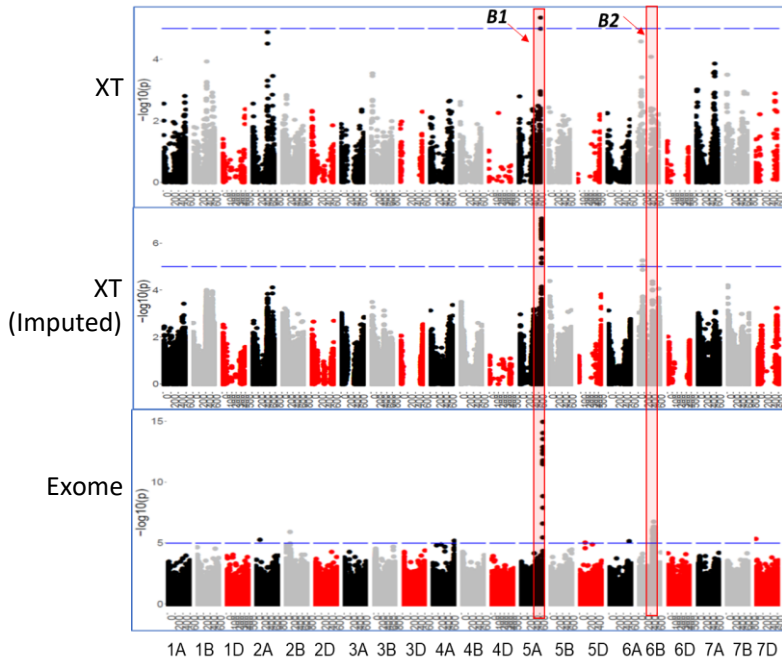
Barley
Row type

* Germplasm provided by Nils Stein, IPK

<https://www.homebrewersassociation.org/zymurgy/zymurgy-extra-2-row-vs-6-row-barley/>



GWAS with Infinium™ Wheat Barley 40K



Wheat
Awned trait

- Germplasm from He *et al.* genotyped on the Infinium™ Wheat Barley 40K



Awned



Awnless

Image Source: AVR



Genomic Prediction with Infinium™ Wheat Barley 40K

Wheat	Training Population Size	Obs.	Trials	BLUE Accuracy	h^2	acc/sqrt(h^2)
Yield	16,281	74,114	593	0.39	0.21	0.85
Stem Rust	8,271	17,772	74	0.35	0.42	0.54
B* colour	1,048	4,834	185	0.58	0.68	0.69

Barley	Training Population Size	Obs.	Trials	BLUE Accuracy	h^2	acc/sqrt(h^2)
Yield	3,952	40,345	264	0.31	0.35	0.52
Spot Form Net Blotch	3,522	6,009	-	0.56	0.56	0.75
Diastase	2,695	6,570	219	0.57	0.58	0.75



Image Source: InterGrain



Updating the Infinium™ Wheat Barley 40K

Version 1.0	Wheat	Barley	Total
Tagging SNP for imputation	20,665	13,469	34,134
Trait associated SNP	457	178	635
SNP linking germplasm resources	4,271	614	4,885
Total number of SNP	25,393	14,261	39,654

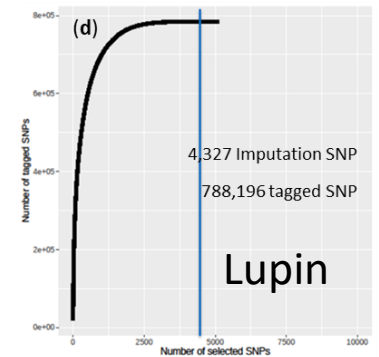
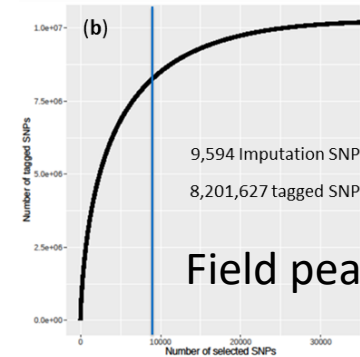
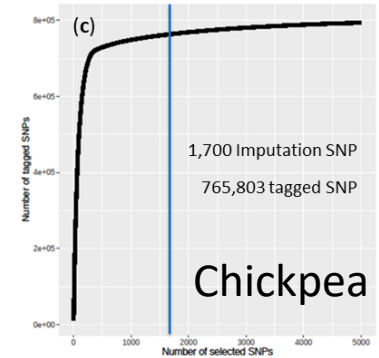
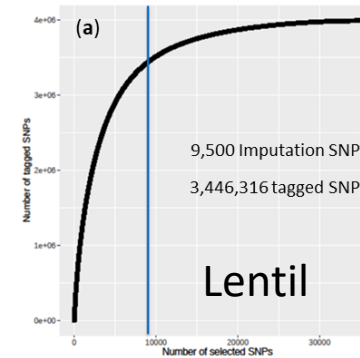
Version 1.1 Additions	Wheat	Barley	Total
Extra tagging SNP for imputation	1,282	673	1,955
Filling in sparsely covered regions of genome	1,378	7	1,385
Trait associated SNP	1,878	69	1,947
Total number of SNP	4,538	749	5,289



Multispecies pulse 30K SNP chip

Species	Probes	Probe categories	
Lentil	10,528	Imputation	9,500
		Canadian linking	1,000
		Trait linked	28
Chickpea	4,449	Imputation	1,700
		Top 1k wild	1,000
		ICRISAT	1,000
		Wild spaced	735
		Trait linked	14
Fieldpea	9,597	Imputation	9,594
		Trait linked	3
Lupin	5,425	Imputation	4,327
		Wild	1,038
		Trait linked	60
Total	29,999		

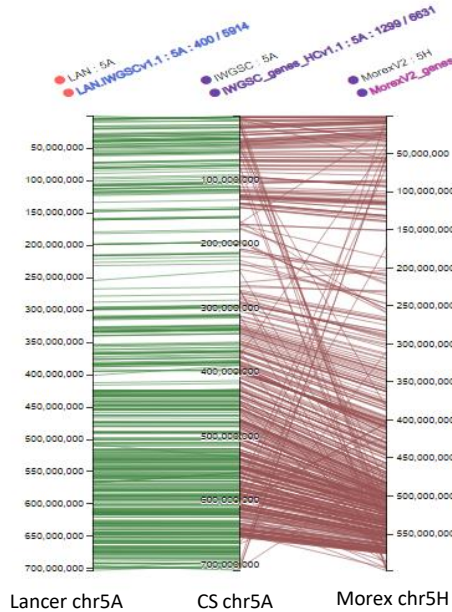
⇒ Multi-samples hybridisation also demonstrated with pulse 30K SNP chip, with ability to co-hybridise chickpea, lentil, lupin and field pea



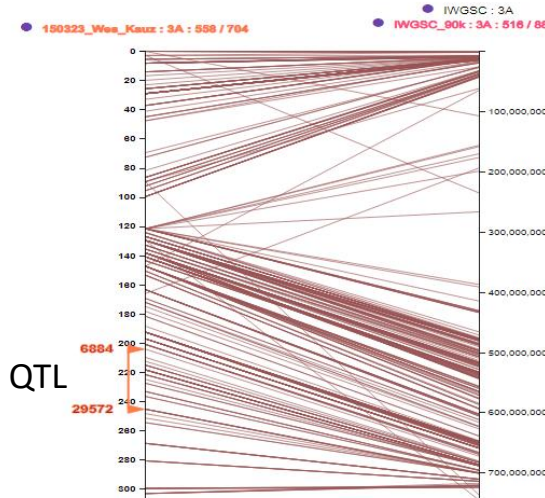
A web tool to enhance utility of Infinium™ Wheat Barley 40K in research and breeding

<https://plantinformatics.io>

<https://github.com/plantinformatics/pretzel>

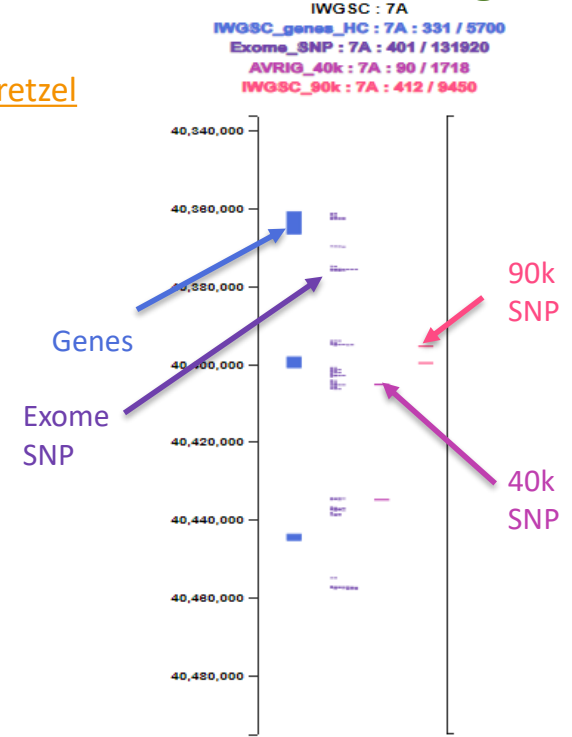


Chromosome vs chromosome alignment



Genetic map to physical alignment

Chromosome scale



Base pair scale



Previously funded through GRDC Project DAV00127



Conclusion

We have designed and tested a new Infinium™ Wheat Barley 40K genotyping platform that bridges research and breeding applications:

- Delivers high quality, highly reproducible genotypes in wheat and barley;
- Enables imputation to high density to deliver increased statistical power and ability to integrate different platforms;
- Can dual hybridise wheat and barley samples, halving costs in high volume applications;
- Is being effectively deployed in InterGrain breeding program;
- Is being made available for use by the international community through Illumina.



Image Source: InterGrain



Acknowledgements

Agriculture Victoria

Raj Pasam *Lab staff*
Debbie Wong Ashley Robertson
Don Isdale Sidhra Jani
Kerrie Forrest Thanuja De Alwis
Adam Dimech Yasodha De Silva
Sukhjiwan Kaur Sanka Perera
Zibei Lin
Shiva Azizinia
Yongjun Li
Hans Daetwyler
Josquin Tibbits
Matthew Hayden

InterGrain

David Moody
Allan Rattey
Daniel Mullan
Jayfred Godoy
Hannah Robinson
Tress Walmsley



PRETZEL
plantinformatics.io

github.com/plantinformatics/pretzel
Pretzel source code

<https://github.com/plantinformatics/pretzel-input-generator>
A pipeline for generating inputs to Pretzel. Includes pre-computed input data for a set of publicly available genomes and syntenic relationships

<https://hub.docker.com/r/plantinformatics/collaboration/>
Automated Docker builds for simple, fast deployment

* Infinium™ Wheat Barley 40K chip is available for use in public pre-breeding research and breeding, and is distributed with a pre-trained SNP cluster genotype calling file

Contact matthew.hayden@agriculture.vic.gov.au or HRobinson@intergrain.com

