Mining the wheat genome for useful polymorphisms

Gary Barker, Sacha Allen
Jane Coghill, Keith Edwards

University of Bristol, UK

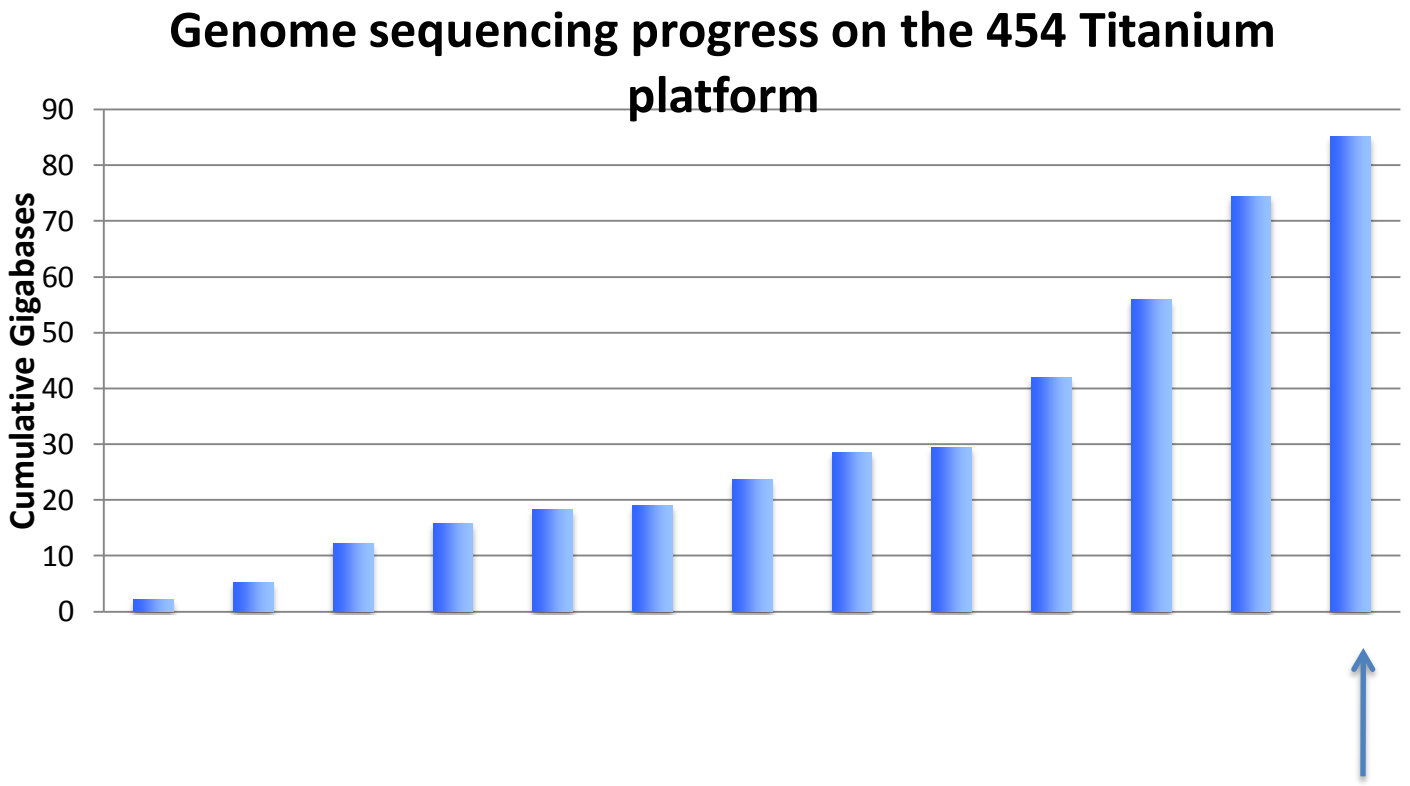
Aims (1)

- Sequence Chinese spring to 5x using 454
- Re-sequence four UK wheat varieties to 20x on SOLiD platform and map on to CS reference
- Call UK varietal SNPs
- Validate 1000 SNPs on 20 UK varieties using KASPAR

Aims (2)

- Generate SNPs from UK varieties via Illumina sequencing of normalized cDNA
- Map SNPs onto Kansas deletion lines and selected nullisomic lines.
- Map SNPs to tetraploid or diploid progenitors as part of a study of synthetic hybridization

Chinese Spring sequencing



Target 5x coverage reached August 2010

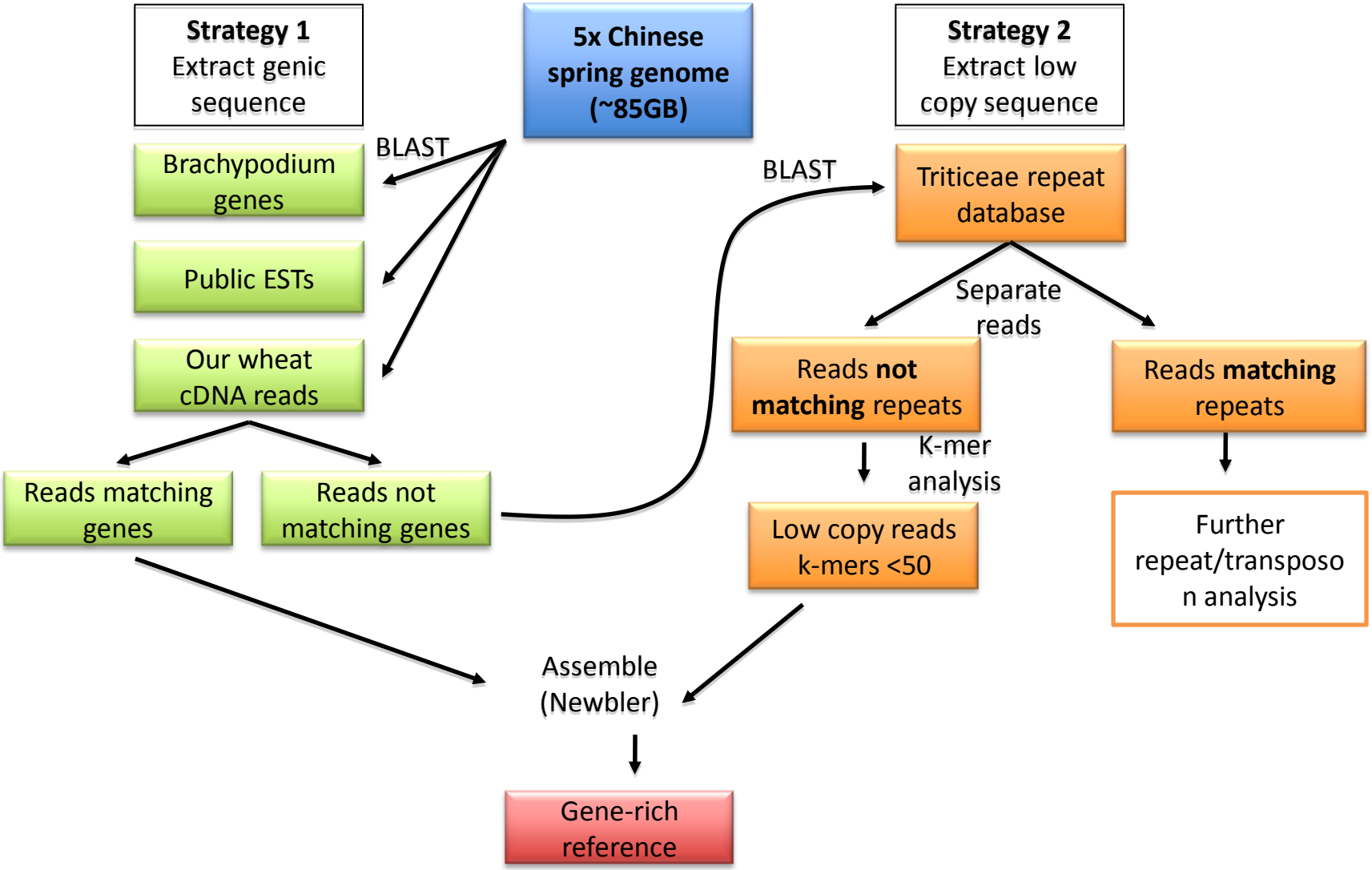
5x Genome interest:

>10,000 BLAST searches

159 email requests for a ftp download details

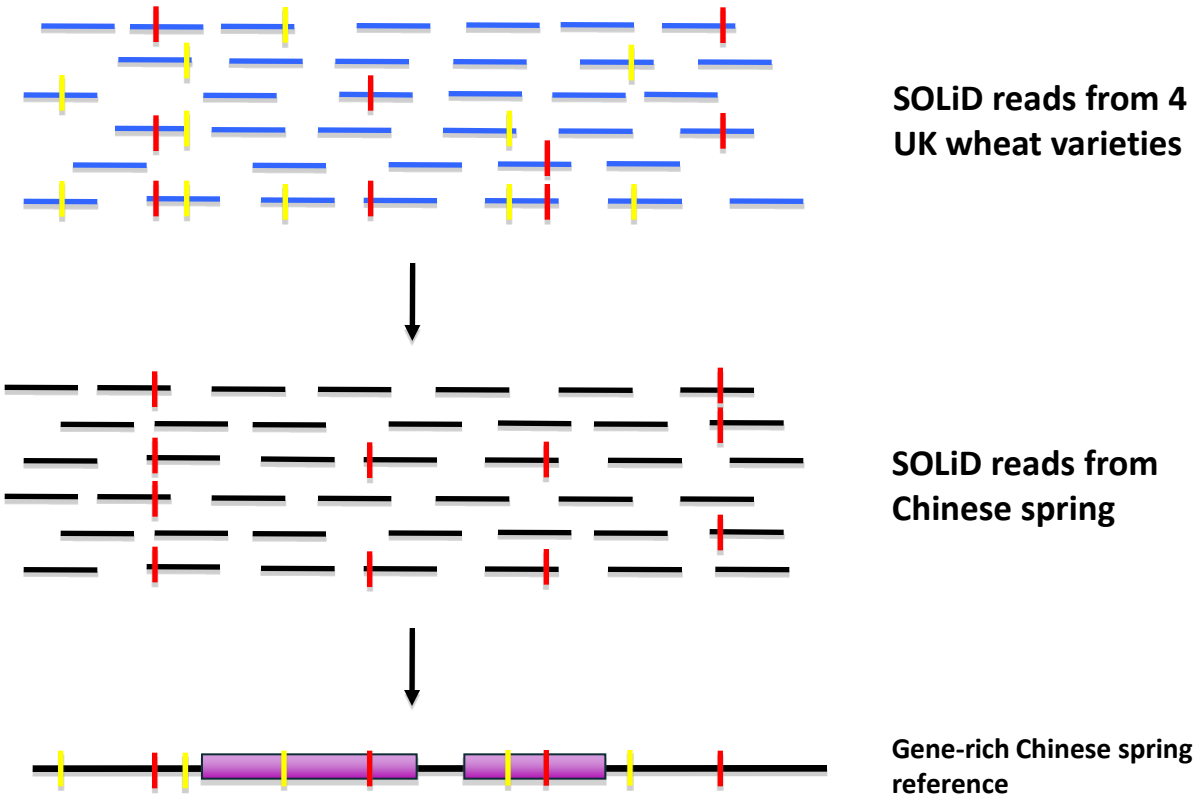
Good publicity for UK wheat research!

Utilizing the CS reference



Sequence four U.K. cultivar genomes plus Chinese Spring
To 20x coverage on the SOLiD platform

Map the SOLiD reads to the gene-rich reference sequence, taking into account those SNPs known to be homeologous



Illumina sequencing for SNP discovery and mapping

- Began before the Genome sequencing effort
- Sequenced normalized cDNA from UK varieties, deletion lines and landraces.
- UK: Avalon, Cadenza, Rialto Savannah, Robigus
- Deletion lines: CS Nullisomic 5A, 5B, 5D plus 42 Kansas lines deleted for partial chromosome arms.
- Landraces: Four from Iraq/Iran

Illumina sequencing for SNP discovery and mapping

- Sequenced UK varieties, deletion lines and landraces.

	Total reads	Mappable reads	Gbases
CS deletion lines	974,015,746	348,313,554	146
CS Nullisomics	141,942,170	73,961,422	21
UK varieties	206,999,667	105,981,163	31
Landraces	136,879,620	55,508,149	21
Synthetics + parents	117,127,199	58,085,906	18
Total	**1,576,964,402**	**641,850,194**	**237**

(Total bases sequenced = 236 billion)

Illumina sequencing for SNP discovery and mapping

- Sequenced UK varieties, deletion lines and landraces.

	Total reads	Mappable reads	**SNPs**	Gbases
CS deletion lines	974,015,746	348,313,554	**13,205**	146
CS Nullisomics	141,942,170	73,961,422	**7,131**	21
UK varieties	206,999,667	105,981,163	**13,239**	31
Landraces	136,879,620	55,508,149	**-**	21
Synthetics + parents	117,127,199	58,085,906	**60,293**	18
Total	**1,576,964,402**	**641,850,194**	**93,868**	**237**

SNPs so far = tip of the iceberg: we haven't started mining the SOLiD data at all yet!

SNP discovery: Landraces and UK varieties

TaContig152041_152		
	Base content	
Landrace 1190304	C(20)	
Landrace 1190306	C(7)	T(26)
Landrace 1190311	C(16)	
Landrace 1190326	C(6)	T(27)
Avalon	C(71)	T(135)
Cadenza	C(64)	T(151)
Rialto	C(90)	T(155)
Robigus	C(86)	T(267)
Savannah	C(62)	

SNPs are called when a base is present in one or more varieties and absent in at least one.

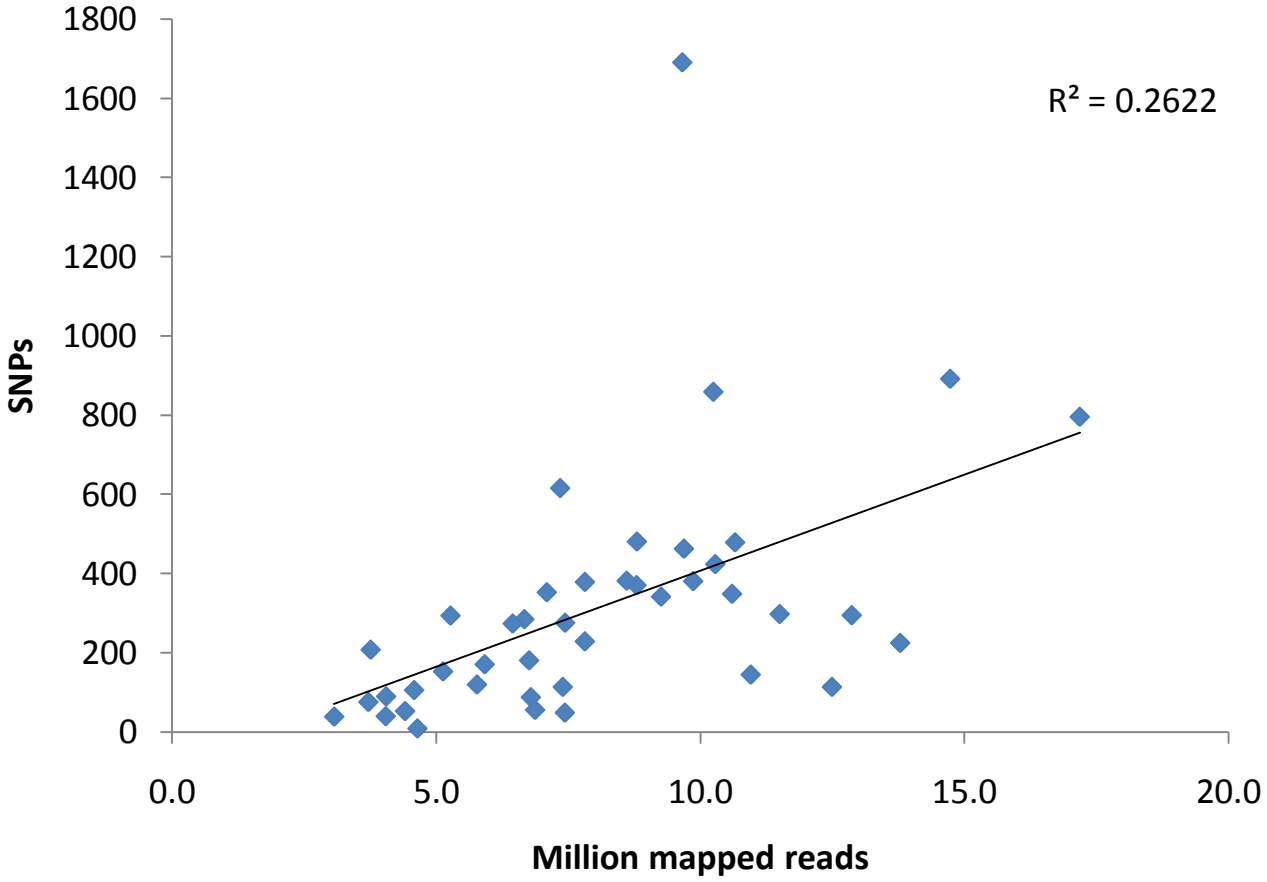
The expected count for the missing base must exceed a threshold determined by randomization tests. FDR < 5%

SNP discovery and mapping

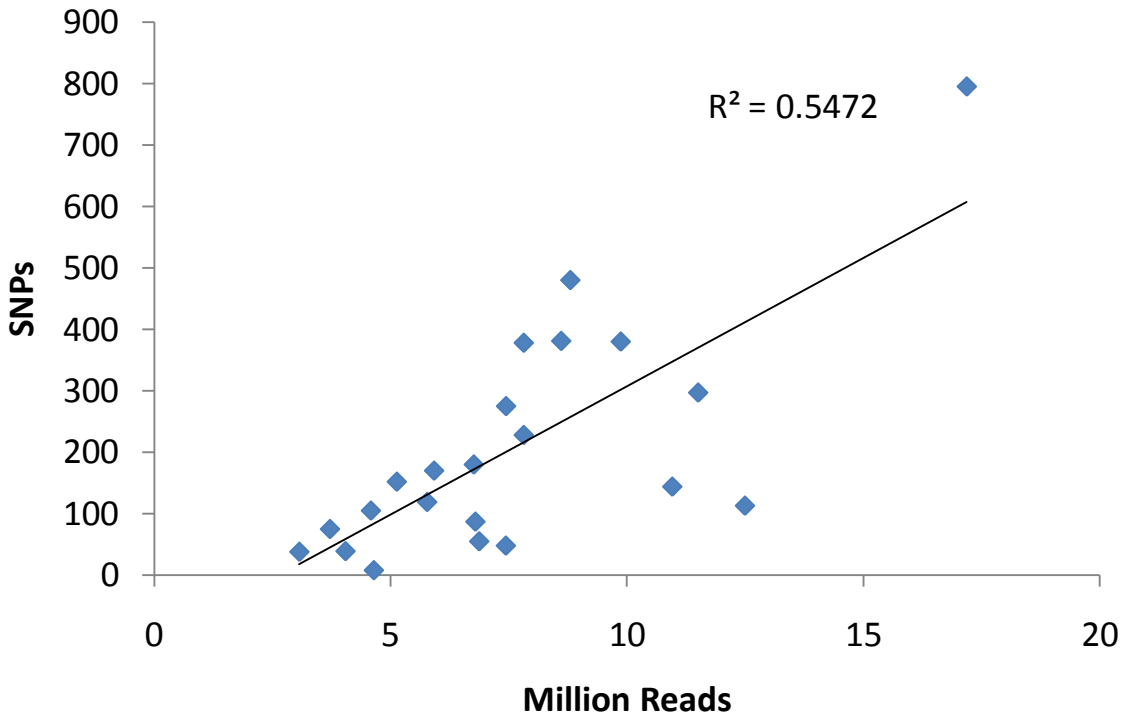
TaContig121741_236

Landrace 1190304	A(13)	G(30)
Landrace 1190306		G(45)
Landrace 1190311	A(16)	G(27)
Landrace 1190326	A(7)	G(43)
Avalon	A(60)	G(77)
Cadenza		G(69)
Rialto	A(29)	G(43)
Robigus	A(48)	G(60)
Savannah	A(20)	G(38)
CS	A(13)	G(34)
CS Nullisomic 5D		G(41)
CS deletion 5DL-1		G(55)

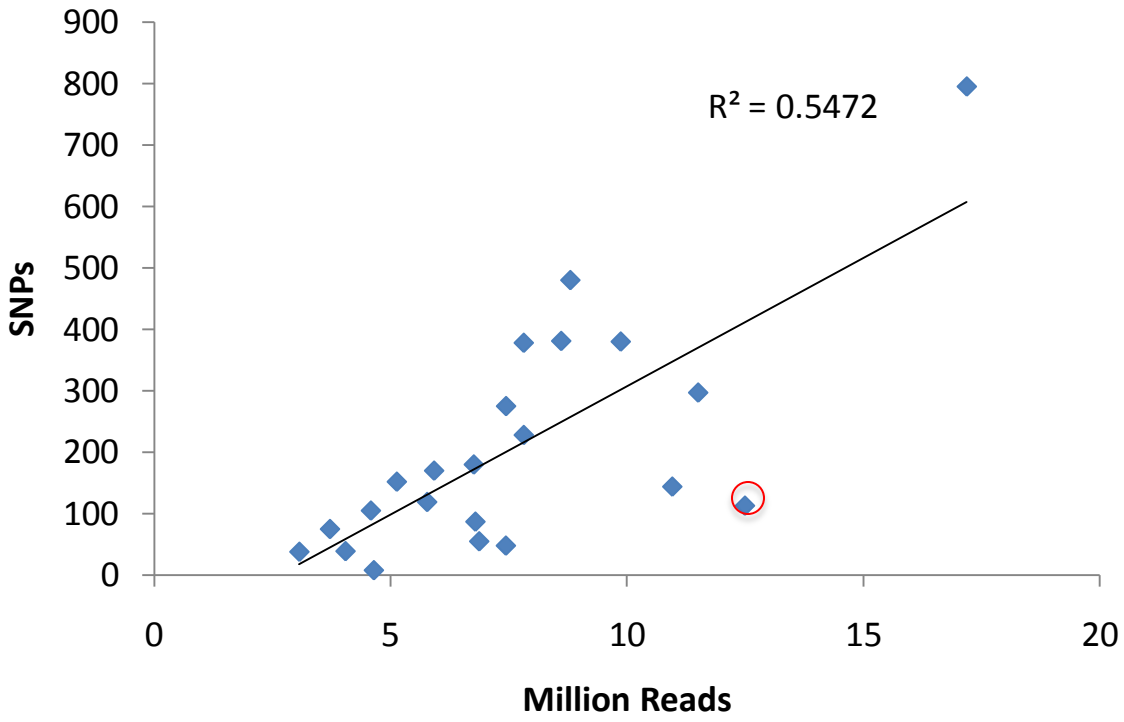
SNP discovery per sequencing effort – all data

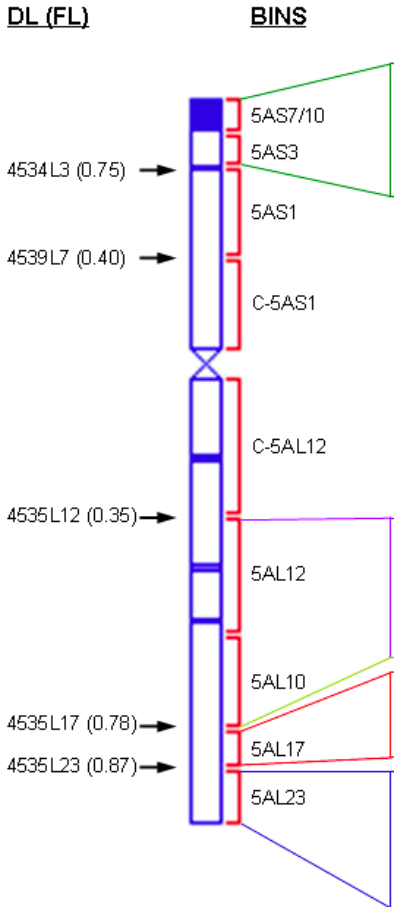


SNP discovery per sequencing effort – short arm data



SNP discovery per sequencing effort – short arm data

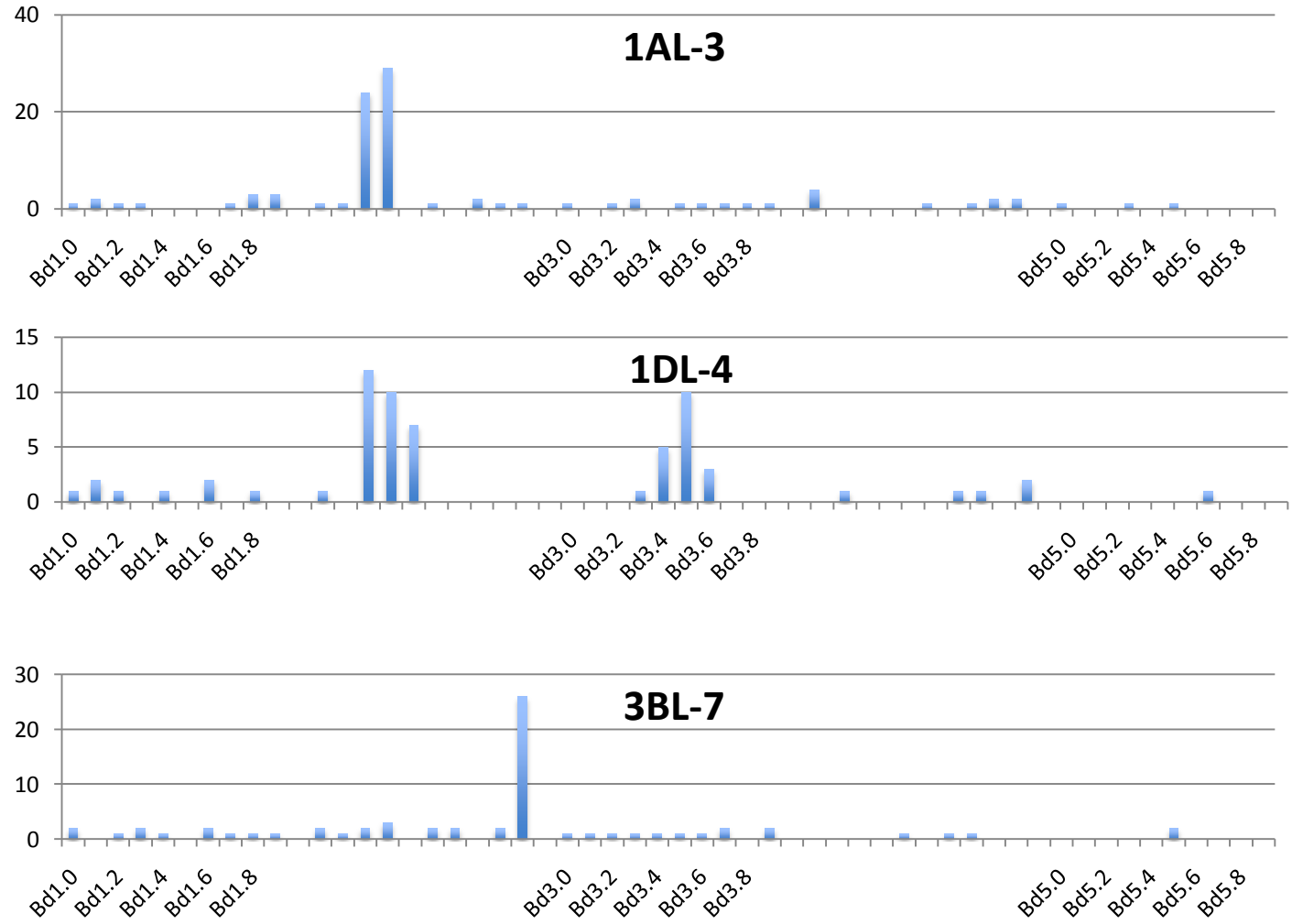




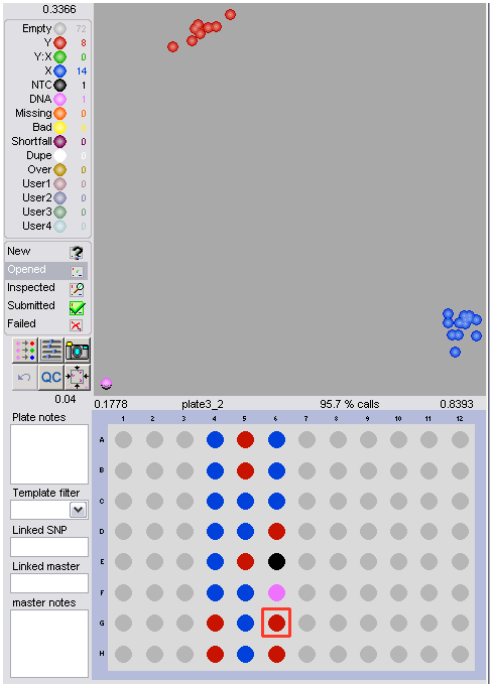
Validation by Independent chromosme assignment of SNP to CS Nullisomic 5 and Kansas Chr 5 deletions.

	Chinese spring **5A**	nullisomic **5D**
Kansas 5AL	704	5
Kansas 5AS	7	1
Kansas 5BL	1	4
Kansas 5BS	0	3
Kansas 5DL	2	377
Kansas 5DS	0	0
Mapped kansas Other	10	17
Total Mapped	722	406
Total On target	711	377
Perent on target	**98**	**92**
Total mapped to kansas 5	714	390
Percent mapped to Kansas 5	**98**	**96**

Validation by mapping reference contigs to *Brachypodium*

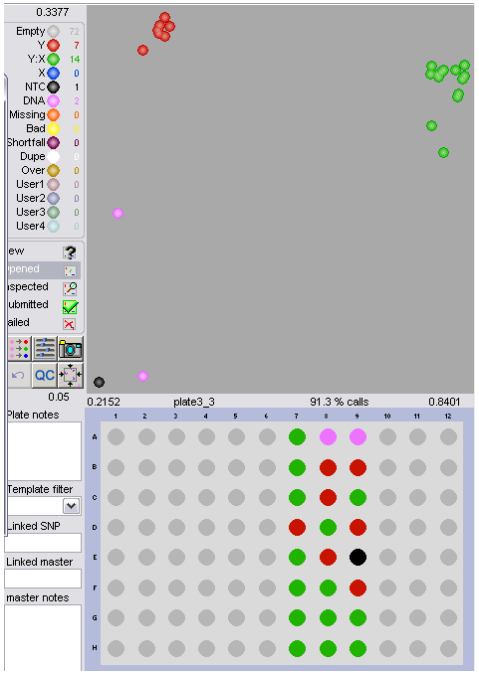


Probe BS00009871



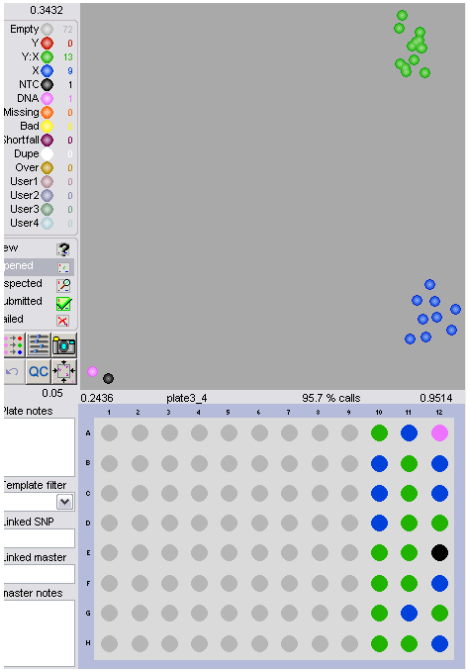
C or G

Probe BS00009861



T or TC

Probe BS00009848



C or GC

1. Alchemy
2. Alcedo
3. Avalon
4. Bacannora
5. Brompton
6. Cadenza
7. Chinese spring
8. Claire
9. Hereward
10. Opata
11. Recital
12. Renan
13. Rialto
14. Robigus
15. Savannah
16. Shamrock
17. Soissons
18. Synthetic
19. Weebil
20. Xi19
21. Water
22. Diploid
23. Tetraploid
24. Synthetic

“Wet” validation:

2000 kaspar primer sets being validated for UK varietal SNPs in Bristol

~1700 additional SNPs being validated in collaboration with Limagrain

~3000 additional SNPs being validated in collaboration with Syngenta

People

Illumina sequencing:

Keith Edwards	Normalised libraries
Gary Barker	Bioinformatics
Sacha Allen	Plant growth, normalised libraries
Jane Coghill	Sequencing

454 + SOLiD Sequencing:

Liverpool- Neil Hall, Anthony Hall, Rachel Brenchley, Linda D'Amore

John Innes Centre - Mike Bevan, Darren Waite

Bristol – Keith Edwards, Gary Barker, Sacha Allen

Addition SOLiD sequencing

TGAC, Norwich, UK – Jane Rogers, Mario Caccamo