Development of a custom SNP chip for dairy and beef cattle breeding, parentage and research

M.P. Mullen[1], M.C. McClure[2], J.F. Kearney[2], S.M. Waters[3], R. Weld[4], P. Flynn[4], C.J. Creevey[3], A.R. Cromie[2] and D.P. Berry[5].

Animal & Grassland Research and Innovation Centre, Teagasc, [1]Athenry, Co. Galway, [3]Grange, Co. Meath, [5]Moorepark, Fermoy, Co. Cork. [2]Irish Cattle Breeding Federation, Bandon, Co. Cork. [4]Weatherbys, Kildare, Co. Dublin.

Abstract

Genomics is currently being utilized for genetic evaluations, parentage verification and screening for lethal recessives, congenital disorders and other mutations with large effects on performance in cattle populations. However, many of these analyses are routinely undertaken independently and on different platforms. The objective of the current paper is to describe the development of a low cost custom genotyping panel to service all of these requirements for both dairy and beef cattle breeding industries. In total, 9,973 variants were successfully added to the commercially available Illumina low density genotyping platform (6,909 SNPs) and included 5,500 SNPs to aid imputation to high density genotypes, 2,176 SNPs to facilitate imputation to microsatellite genotypes, 424 variants for major gene effects and 1,873 variants of research interest. A total of 9,852 cattle were genotyped between March to August 2013 with a median animal call rate of 0.989 and < 5% of animals genotyped below 0.95. Illumina SNP call rates, i.e. present on the LD, 50K or HD panels, were high with over 99.4% of SNPs with call rates ≥ 0.95. Non Illumina SNPs, i.e. novel to an Illumina platform, had lower SNP call rates with 87.6% at ≥ 0.95. For parentage verification, for the 2,891 cattle with sires without SNP genotypes and requiring imputation to microsatellite genotypes the imputation accuracy was 96%. Carriers of lethal recessive conditions, brachyspina and complex vertebral malformation, were detected in the Holstein-Friesian population at 2% and 4%, respectively. In addition, congenital disorders citrullinaemia, osteopetrosis and syndactyly were identified at low frequencies (< 1%). Variants in the Myostatin gene, nt821, F94L and Q204X were segregating in Angus, Belgian Blue, Charolais, Limousine and Simmental populations. Development of the custom genotyping platform servicing these requirements will provide a valuable 'one-step' tool to service current and, due to its ongoing development, future needs of both dairy and beef cattle industries.

Key words: custom genotyping panel, cattle breeding, imputation, genomic selection

Introduction

Exploiting imputation of high density genotypes using low density SNP panels is an attractive cost saving approach for implementation of genomic selection in livestock. However, the imputation to higher density panels in certain populations, for example beef cattle in Ireland, is not sufficiently accurate using commercially available low density panels (Berry et al 2013).

Parentage verification in cattle and other species has been routinely carried out for several decades using microsatellites based analysis. Recently, genotyping panels exploiting single nucleotide polymorphism (SNP) markers both reduce cost and improve the reproducibility of parentage verification. However, a change in the technology used for parentage verification usually implies that back-pedigree would be required to be re-genotyped using the updated technology. For example, during the previous transition between blood typing to microsatellites based analysis for parentage, all animals had to be re-genotyped using microsatellites. This would suggest that the transition from microsatellites to SNPs would also incur a similar additional cost.

Furthermore, cattle breeding societies also routinely test for deleterious mutations, e.g. BLAD, Brachyspina, CVM, DUMPs, in addition to mutations with large effects on performance such muscle development and

milk composition, e.g. Myostatin, A1/A2 β casein.

The objective of this study was to develop a low-cost custom SNP genotyping platform suitable for international dairy and beef cattle populations that will 1) improve imputation or prediction of a larger number of SNPs for genomic selection (especially in beef cattle), 2) be able to accurately compare offspring genotyped using the SNP panel with parents genotyped using microsatellites, 3) screen for known lethal recessives and major genes and 4) provide a vehicle to genotype a large number of animals for SNPs of research interest. The custom panel, called the 'International Dairy and Beef' (IDB), will add these chosen variants to the commercially available Illumina low density (LD) genotyping panel (Illumina Inc., 2011a) which contains 6,909 SNPs and will be applicable for both beef and dairy cattle.

Materials and Methods

Imputation to HD genotypes

Illumina high density (HD) *g*enotypes (777,962 SNPs) were available on 3,122 dairy and beef animals including Aberdeen Angus (n=269), Belgian Blue (196), Charolais (710), Hereford (234), Holstein-Friesian (719), Limousin (730), and Simmental (264) bulls. A total of 5,500 SNPs was selected to be added to the LD panel with the threshold of additional SNPs per chromosome proportional to chromosome length. The most informative SNPs were sequentially chosen, within chromosome, based on a dual objective of maximising the minor allele frequency weighted across breeds, and minimising the maximum level of linkage disequilibrium between each candidate SNP and the SNPs already chosen (including those on the Illumina low density panel).

Animals were partitioned into either a reference or a validation population to test the accuracy of imputation to the Illumina HD platform. All animals, irrespective of breed, born after 2005 (n= 698) were assumed to represent the validation bulls; all other bulls were included in the reference population. Imputation from the original LD genotyping platform, the IDB, and the commercially available Illumina Bovine50 Beadchip (54,001 SNPs) to the Illumina HD platform was undertaken in dairy and beef cattle using BEAGLE (Browning and Browning, 2007, Ilumina Inc., 2010, Illumina Inc., 2011b).

Imputation to microsatellite genotypes

For imputation of the 12 ISAG recommended bovine microsatellite (MS) parentage panel (www.isag.us/Docs/CattleMMPTest_CT.pdf) initially the reference panel from McClure et al. (2013) was used. This panel consists of 7,130 individuals representing 36 *Bos taurus* pure breeds and 29 *B. taurus* crossbred animals with both Illumina HD SNP and microsatellite genotypes. The identification of haplotypes for MS imputation as described in McClure et al. (2013) is described briefly below.

BEAGLE input files for the reference population were created for each MS marker and flanking Illumina SNP within 500kb. Animals were filtered on their MS genotypes so that for each MS the BEAGLE file contained only individuals with a MS genotype. All reference individuals were phased together using BEAGLE with 100 iterations (Browning and Browning, 2007). Optimal haplotype size for MS imputation was determined by analysing phased haplotypes, centred on the MS, using sliding windows that increased in size (10-20 flanking SNP increments). The number of unique reference population haplotypes that were linked to 1 MS allele 100 % of the time and the number of haplotypes that were linked to > 1 MS alleles but matched 1 MS allele ≥ 90% of the time were tallied. The optimal haplotype size was determined when either of the following criteria was met:

1) The maximum number of unique haplotypes appearing ≥ 4 times and linked to only 1 MS allele 100% of the time or linked to 1 MS allele ≥ 90% of the time across all breeds was obtained.
2) Increasing the haplotype size by 10 SNP resulted in ≤ 1% increase in the total number of tallied haplotypes.

Since we adopted the process of MS imputation for parentage verification an additional 446 animals with SNP MS genotypes have been added to the base reference population. Each time the reference

population is increased it is fully rephased using BEAGLE and the original haplotype sizes are retained.

For animals genotyped on the IDB beadchip their sire verification is initially checked using the core 116 parentage SNP (Heaton et al. 2002). MS are imputed for any animal whose reported sire failed SNP verification, whose sire are not SNP genotyped, or the animal was generated by embryo transfer. These imputed MS are then checked against the reported parent MS profile for parentage verification. Any animal with > 1 mismatches between its imputed MS profile and reported parent's MS profile is MS genotyped and eventually is added to the reference panel using the MS genotype data.

Causative mutations

Known causative variants in 53 genes were selected for the custom panel including four lethal recessives (CVM, BLAD, DUMPS, Brachyspina); 33 congenital disorders / undesirable traits including Arthrogryposis, Hypotrichosis, Polledness; and 16 genes with large effects on quantitative traits including Myostatin and DGAT1 (Appendix 1). Causative variants were added in triplicate to the panel.

Research

A total of 1,873 variants were selected which are part of on-going research in Teagasc into the role of mutations in the genes of the somatotrophic axis on performance in cattle (Mullen et al. 2010, 2011, 2012a, 2012b).

Results and Discussion

IDB genotyping in Ireland

Since its release in March 2013 and up until the end of July 2013 approximately 9852 cattle in Ireland across both beef and dairy have been genotyped using the IDB (Table 1).

Table 1. Breed composition of the 9852 cattle genotyped using the IDB between March to August 2013.

Breed composition			
Breed	#	Breed	#
Holstein Friesian	5714	Limousine	1478
Aberdeen Angus	49	Simmental	296
Belgian Blue	59	Hereford	394
Charolais	1773	Other	89

Animal call rates

The median animal call rate was 0.988 (ranged between 0.3313 – 0.9993), with less than 6% of the animals genotyped having call rates below 0.95. There were no significant differences in mean animal call rates between breeds.

SNP call rates and GT scores- Illumina SNPs

Of the 14,585 Illumina SNPs 99.4% had call rates of ≥ 0.95 with a median GT score of 0.86 (Figure 1).

SNP call rates and GT scores- non Illumina

Of the 2,297 non Illumina SNPs 87.6% had call rates of ≥ 0.95, with a median GT score of 0.815 (Figure 2). 258 SNPs had a call rate of < 90%.

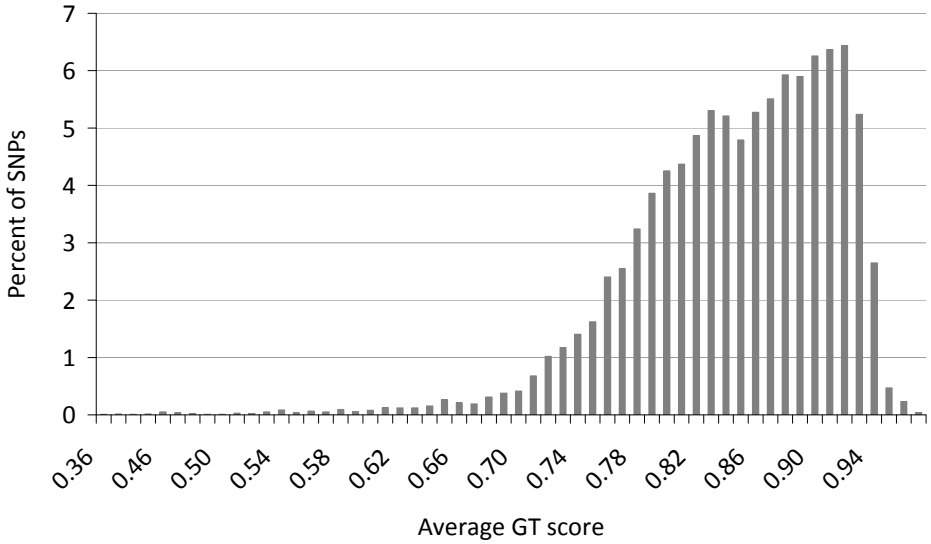

Figure 1. Average GT scores of the 14,585 Illumina SNPs on the IDB.

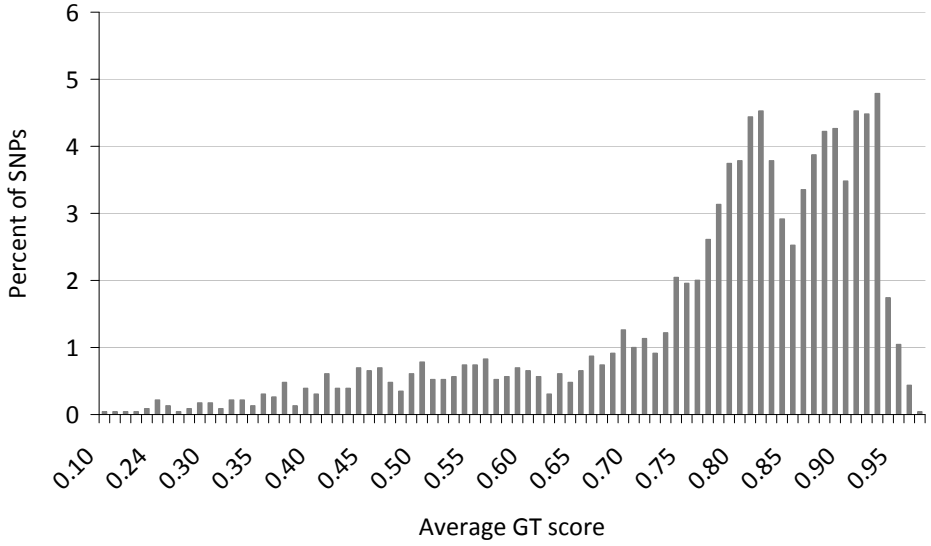

Figure 2. Average GT scores of the 2,297 non Illumina SNPs on the IDB.

Illumina missing SNPs

Unexpectedly, 832 Illumina SNPs failed to make it into the final panel. Illumina SNPs originally chosen to be on the IDB had to have call rates greater than 95%, GT scores greater than 0.55 and be in Hardy Weinberg equilibrium.

Imputation to HD genotypes

Figure 3 shows the accuracy of imputation to high density genotypes. The accuracy of imputation from the Illumina low density panel (LD) to HD genotypes improved with the addition of the extra selected SNP markers, especially when the sire was not genotyped on the HD panel. The mean accuracy of imputation to high density from the commercial available LD, the custom SNP chip described here (International Dairy and Beef, IDB), and the commercially available Bovine50 beadchip was 0.95, 0.97 and 0.99, respectively; moreover if the sire was genotyped on the higher density genotype platform then the respective accuracy of imputation was 0.97, 0.98 and 0.99.

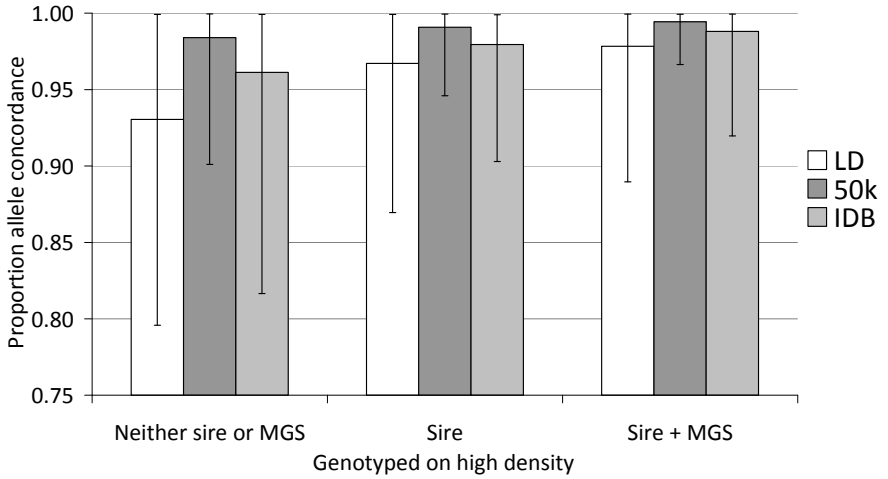

Figure 3. Mean and range (error bars) in the accuracy of imputation of HD genotypes from low, medium and custom panel (IDB) density genotypes where neither the sire nor the maternal-grandsire (MGS), only the sire or both the sire and MGS were genotyped on the HD platform.

Imputation to microsatellite genotypes

The accuracy of the microsatellite imputation in the validation animals, determined by comparing the imputed microsatellite allele to the parental microsatellite alleles and allowing 1 MS marker conflict was 95.6%. As a comparison to McClure et al. (2013) when also allowing 1 MS marker conflict, 90.1% of the validation animals with imputed MS alleles and 95.9% of the reference animals with genotyped MS alleles was correct when compared to their parent's genotyped MS alleles. The 1 MS marker conflict was allowed as microsatellite genotyping errors are known to exist and are thought to be between 1% and 5% (Bonin et al., 2004; Weller et al., 2004). These known errors are why it has been suggested that 2 marker conflicts must exist for an animal to be excluded from parentage verification (Bonin et al. 2004; Weller et al. 2004; Baruch et al. 2008).

As more animals are added to the reference population the accuracy of the imputation process improves as seen in Figure 4, where the currently used reference panel contains 446 more animals than the initial reference panel from McClure et al. (2012). The addition of animals to the reference population will increase the MS imputation accuracy and will also allow for the capture of rare MS-SNP haplotypes.

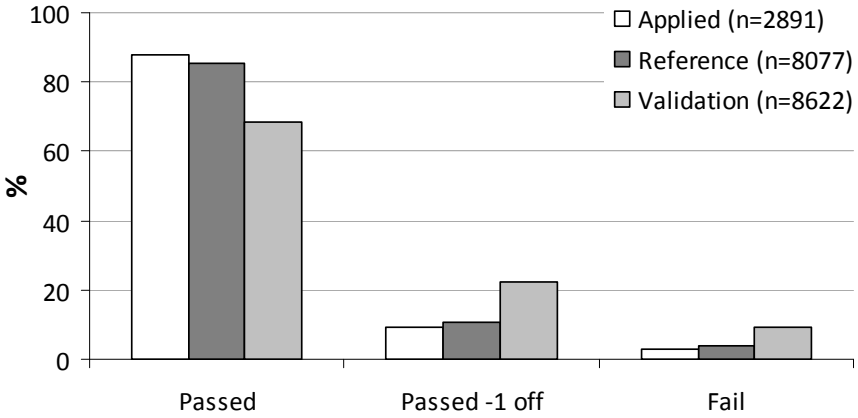

Figure 4. Accuracy of sire assignment using traditional microsatellites (reference; McClure et al. 2013), imputation to microsatellite genotypes (validation; McClure et al. 2013) and IDB dataset (applied). Passed indicates that 100% of an animal's MS profile matched its reported parent's MS

profile. Passed -1 off indicates that 1 marker conflict occurred between the animal's and its parent's MS profile. Fail indicates that 2 or more marker conflicts exist.

Frequency of lethal recessives, congenital disorders and major genes

Carriers (both male and female) of lethal recessive mutations, brachyspina and complex vertebral malformation (CVM), were detected in the Holstein-Friesian sample at 2% and 4%, respectively. Congenital disorders citrullinaemia, osteopetrosis, syndactyly were also present in the HF sample at low frequencies (< 1%). 33%, 12% and 15% of the HF cattle were homozygous carriers for A2 β-casein, K- casein and DGAT1, respectively (Table 2).

Variants in the Myostatin gene (*GDF8)*, nt821, FL94 and Q204X, were segregating in five of the (except Hereford) six main beef breeds in Ireland (Table 3). In addition, a mutation in the *KRT71* gene resulting in hypotrichosis (hairlessness) was identified in a small proportion (2%) of the Hereford cattle genotyped.

Table 2. Genotype frequencies (%) of lethal recessives, congenital disorders and mutations with effects on milk yield and composition identified in 5714 Holstein Friesian cattle.

Mutation		% carriers[a]
		HF
Brachyspina		2 (0)
CVM		4 (0)
Citrullinaemia		< 1 (0)
Osteoporosis		
Syndactyly		
β casein	A1/A1	21
	A1/A2	46
	A2/A2	33
DGAT1	A/A	40
	A/K	45
	K/K	15
Kappa casein	A/A	43
	A/B	45
	B/B	12

[a]: % homozygous carriers in parenthesis.

Table 3. Genotype frequencies (%) of mutations in *GDF8* (Myostatin) across six beef breeds sampled between March to August 2013.

Mutation		% carriers[a]					
		AA	BB	CH	HE	LM	SI
Myostatin	nt821	2.5 (0)	33 (18)	< 1 (0)	0	6 (0)	0
	F94L	2.5 (0)	6 (0)	27 (2)	0	98 (83)	1 (0)
	Q204X	0	0	27 (1)	0	7 (0)	0

[a]: % homozygous carriers in parenthesis.

The workflow

DNA extraction and genotyping are carried out using a commercial service provider (Weatherbys, Ireland) and genotypes uploaded to the Irish Cattle Breeding Federation (ICBF). Imputation of microsatellites for sire verification, generation of dairy genomic EBI's, screening for causative mutations including final report generation are all carried out by the ICBF and communicated back to the farmer (Table 4). The information on all causative mutations which incur no royalty fees (unless ordered) is provided to the farmer at no extra cost. Any royalty bearing genotypes can be retrospectively requested at a cost.

Table 4. Timeline of IDB genotyping and data analysis.

Data		Timeframe (working days)
gEBI (dairy)		15
Sire verification	SNP genotypes available on sire	10
	Imputation to microsatellite genotypes	20
Genes with large effects		10

Conclusions

This custom genotyping panel, known as the International Dairy & Beef (IDB), contains 16,882 variants, to increase the accuracy of imputation to greater SNP density for genomic selection, enables microsatellite imputation for parentage verification, screens for mutations relevant to the cattle industry such as lethal recessive mutations, genetic defects and major genes as well as providing an avenue to generate genotype information on mutations of research interest. The list of variants included in the panel is freely available. The custom panel is under annual development with version two planned for release in January 2014 with additional congenital disorders/undesirable traits and bovine breed assignment and is open for research collaboration.

Acknowledgements

We would like to acknowledge the contributions of the breed societies and AI stations in addition to Paul VanRadan (USDA) and Jon Beever (Illinois University). Financial support from the Irish Department of Agriculture, Food and Marine through the Research Stimulus Fund (RSF-06-353; 11/S/112), Science Foundation Ireland (07/SRC/B1156), breed societies and AI stations is acknowledged.

References

Baruch, E. and Weller, J. I. 2008. Estimation of the number of SNP genetic markers required for parentage verification. *Anim. Genet.* 39(5), 474-9.

Berry, D.P., Mullen, M.P. & Cromnie, A.R. (2013). Across breed, beef and dairy cattle genotype imputation to high-density genotypes. *Proc. Ir. Agr. Res. Forum*, p70.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C. and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. *Mol. Ecol.* 13(11), 3261-73.

Browning, S. R. and Browning, B. L. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084-1097.

Heaton, M. P., Harhay, G. P., Bennett, G. L., Stone, R. T., Grosse, W. M., Casas, E., Keele, J. W., Smith, T. P., Chitko-McKown, C. G. and Laegreid, W. W. 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian genome,* 13(5), 272-81.

Illumina Inc. 2011a. BovineLD Genotyping BeadChip. *Accessed Aug. 31, 2012.*

http://www.illumina.com/Documents/products/datasheets/datasheet_bovineLD.pdf.

Illumina Inc. 2010. BovineHD Genotyping BeadChip. *Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf.*

Illumina Inc. 2011b. BovineSNP50 Genotyping BeadChip. *Accessed Aug. 31, 2012.http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf.*

McClure, M., Sonstegard, T., Wiggans, G. and Van Tassell, C. P. 2012. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. *Front. Genet.* 3, 140.

McClure, M., Sonstegard, T., Wiggans, G., Van Eenennaam, A., Weber, K., Penedo, M.C., Berry, D.P., Flynn, J., Garcia, J.F., Santana do Carmo, A., Regitano, L.C.A., Albuquerque, M., Silva, M.V., Coffey, M., Moore, K., Boscher, M.Y., Genestout, L., Mazza, R., Taylor, J., Schnabel, R., Machado, M.A., Simpson, B., McEwan, J.C. Cromie, A., Coutinho, L.L., Kuehn, L.A., Keele, J.W., Piper, E.K., Cook, J., Marques, E. and Van Tassell. C.P. Imputation of Microsatellite Alleles from Dense SNP Genotypes for Parentage Verification Across Multiple *Bos taurus* and *Bos indicus* breeds. *Front.Genet.,* In Review 08/2013.

Mullen, M., Creevey, C., Berry, D., McCabe, M., Magee, D., Howard, D., Killeen, A., Park, S., McGettigan, P., Lucy, M., MacHugh, D. and Waters, S. 2012a. Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples. *BMC Genomics* 13(1), 16.

Mullen, M. P., Berry, D. P., Creevey, C. J., McCabe, M. S., Howard, D. J., Magee, D. A., Lucy, M. C., MacHugh, D. E. and Waters, S. M. 2012b. Characterization of genetic variation within the somatotrophic axis in DNA pools of beef and dairy cattle divergent for milk production, size, fertility and immune response. *J. Anim. Sci.* Vol. 90, Suppl. 3, LB2i.

Mullen, M. P., Berry, D. P., Howard, D. J., Diskin, M. G., Lynch, C. O., Giblin, L., Kenny, D. A., Magee, D. A., Meade, K. G. and Waters, S. M. 2011. Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF-1) gene are associated with performance in Holstein-Friesian dairy cattle. *Front. Genet.* 2:3.

Mullen, M. P., Berry, D. P., Howard, D. J., Diskin, M. G., Lynch, C. O., Berkowicz, E. W., Magee, D. A., MacHugh, D. E. and Waters, S. M. 2010. Associations between novel single nucleotide polymorphisms in the Bos taurus growth hormone gene and performance traits in Holstein-Friesian dairy cattle. *J. Dairy Sci.* 93(12), 5959-69.

Weller, J. I., Feldmesser, E., Golik, M., Tager-Cohen, I., Domochovsky, R., Alus, O., Ezra, E. and Ron, M. 2004. Factors affecting incorrect paternity assignment in the Israeli Holstein population. *J. Dairy Sci.* 87(8), 2627-40.

Appendix:

Appendix 1: Content of the IDB v1

	Lethal recessives		Congenital disorders (cont.)
1	CVM (Complex Vertebral Malformation) [a]	23	Citrullinemia
2	BLAD	24	CMDI: Congenital muscular dystonia I
3	DUMPS	25	CMDI: Congenital muscular dystonia II
4	Brachyspina [a]	26	Crooked Tail Syndrome [a]
		27	Factor XI
	Congenital disorders	28	Heterochromia Irides (White Eye)
1	Arthrogryposis (Curly Calf) [a]	29	SDM
2	Fawn Calf Syndrome or Contractural Arachnodactylyl [a]	30	Idiopathic Epilepsy [a]
3	Hypotrichosis_PMel17	31	Pulmonary Hypoplasia [a]
4	Hypotrichosis in Belted Galloway, HEPHL1 SNP	32	Weaver
5	Hypotrichosis_KRT71 [a]	33	Neuropathic hydrocephalus (water head syndrome) [a]
6	Spiderleg- MOCS1 gene- Simmental		
7	Spiderleg- SOUX gene- Brown Swiss		**Major genes**
8	Polledness	1	DGAT1
9	Mule Foot	2	MSTN (GDF8) Double Muscling [a]
10	Tibial Hemimelia [a]	3	A1/A2 beta casein + [a]
11	Black/Red Coat Color/Red Factor	4-7	Fertility Haplotypes (HH1, HH2, HH3, JH1)
12	Red Recessive coat colour (Different to red factor)	8	Kappa Casein I
13	Silver Color Dilutor	9	Kappa Casein II
14	Dun Color	10	ABCG2
15	RNF11 (affects growth and stature)	11	GH2141 and GH2291 (Marbling and growth rate) [a]
16	Osteopetrosis (Marble Bone Disease)	12	IGF1-AF017143
17	Pink Eye (Infectious Bovine Keratoconjunctivitis)	13	STAT1 [a]
18	Protoporphyria (Photosensitization)	14	STAT3 [a]
19	SMA (Spinal Muscular Atrophy)	15	STAT5 [a]
20	Beta Lactoglobulin	16	Calpain (Tenderness) loci
21	Beta Mannosidosis		
22	Alpha Mannosidosis		

[a]: royalties may apply