September 18, 2009

Dear Colleagues,

We are writing to provide an update on and detailed description of the "Metabochip", a custom Illumina iSelect genotyping array designed to test, in a cost-effective manner, ~200,000 SNPs identified through genome-wide meta-analyses for metabolic and atherosclerotic / cardiovascular diseases and traits. In the attached document, we describe in detail the basic elements and design principles of the array. We encourage you to share this letter, and the accompanying documents, with members of your Consortia and others who might be interested in the array.

The Metabochip was designed by representatives of the following GWAS meta-analysis Consortia: CARDIoGRAM (coronary artery disease), DIAGRAM (type 2 diabetes), GIANT (height and weight), MAGIC (glycemic traits), Lipids (lipids), ICBP-GWAS (blood pressure), and QT-IGC (QT interval). It supports genotyping of SNPs selected according to five sets of criteria: (1) individual SNPs displaying evidence for association in GWA meta-analyses to diseases and traits relevant to metabolic and atherosclerotic-cardiovascular endpoints, (2) detailed fine mapping of loci validated at genome-wide significance from these meta-analyses, (3) all SNPs associated at genome-wide significance with any human trait, (4) "wildcards" selected by each meta-analysis Consortium for Consortium-specific purposes, and (5) other useful content, including SNPs that tag common CNPs, SNPs in the HLA region, SNPs marking the X and Y chromosomes and mtDNA, and for sample fingerprinting (common SNPs represented on major genome-wide array products from both Illumina and Affymetrix). Details of all these SNP sets can be found in the attached document.

After merging and pruning the lists (to remove redundant SNPs), a total of 217,697 SNPs representing 245,243 bead types was submitted to Illumina for manufacturing on August 19, 2009. The final chip is expected to genotype ~200,000 SNPs per sample. We have also compiled a supplemental text file that lists, for each SNP on the array: Chromosome, Base Position (Build 36), SNP ID (rsID or chr[chr#]:[bp] format), Illumina SNP ID, population alternative allele frequency (where known from 1000 Genomes Pilot 1 data), Illumina design score, number of bead types, functional annotation (nonsynonymous, essential splice site, stop codon mutations). That list also is attached.

We note that while the identities of the SNPs will be freely available to all users, the criteria that led to their selection (for example, the level of association and the associated trait) will not automatically be distributed by the Metabochip consortium nor have these been provided to Illumina. We believe that there would be considerable scientific merit for each contributing Consortium to divulge the provenance of the SNPs it posted to the array to others within the Metabochip Consortium. We recognize that disclosure of provenance needs to be a matter for each Consortium to consider from the perspective of their individual publication plans. We suggest that disclosure of provenance should be accompanied by an understanding from other groups that the Consortium concerned would, within a to-be-agreed time period, be given the opportunity to write the first paper reporting Metabochip-based association results for "their" SNPs with respect to "their" traits. This safe-haven option, which was described in early Metabochip documents, clearly needs further discussion to resolve details.

We are pleased to report that we reached our collective goal to submit purchase orders to Illumina for >100,000 samples, enabling us to successfully obtain the Consortium price ($39.00 per sample for consumables). On August 14, purchase orders for 113,976 samples were confirmed by Illumina, with new orders coming in each day. Manufacturing is underway.

A great deal of time and effort went into designing the Metabochip. We thank each of the participating meta-analysis Consortia for providing SNP and locus lists, much of it pre-publication, and thank the representatives of the Consortia for their contributions to developing the Metabochip design. We thank Illumina for their support for this project and the consistent responsiveness of their team members Tristan Orpin, Jennifer Stone, and Kimberly Gietzen. We thank Quang Le and Richard Durbin from the Sanger Institute and Jared Maguire and Mark Daly from the Broad Institute for providing August 1000 Genome genotype calls prior to their public release. Most of all, we thank a small set of young scientists whose creative thinking and long hours turned the Metabochip concept into a real product: in particular, Noël Burtt, Ben Voight, and Cameron Palmer at the Broad Institute and Hyun Min Kang, Jun Ding, and Yanming Li at the University of Michigan. As papers are written based on results from Metabochip genotype data, we encourage you to include these individuals as coauthors.

For those of you who have not yet ordered the Metabochip or have done so but would like to order more, we suggest you contact your local Illumina representative or Jennifer Stone (jstone@illumina.com). The current pricing level will be available while supplies last to all Cardio-MetaboChip Consortium (CMC^2) members. To join this consortium, all that is required is that you register your interest with Noël Burtt (burtt@broadinstitute.org). Membership in the consortium will then be communicated to Illumina by Ms. Burtt and will enable purchase of the Metabochip at the price set forth in this letter. In addition to the Consortium-specific price, the Broad Institute's Center for Genotyping and Analysis, through a grant (U54 RR020278) from the National Center for Research Resources, is now offering subsidies for genotyping on the Cardio-Metabochip. A complete description of the subsidy opportunity and the application process and deadline can be found at: http://www.broadinstitute.org/science/projects/broad/ncrr-center-genotyping-analysis/metabochip-genotyping-project-questionnaire

Finally, we note that each investigator has complete freedom to buy, run, and analyze its own data. Nonetheless, we hope and expect that groups interested in different sets of Metabochip traits will likely get together for joint analysis of the resulting data, in many cases nucleated by the Consortia that provided the results on which the Metabochip is based. We welcome new members to the Cardio-MetaboChip Consortium and are eager to commence genotyping and continued collaboration.

Best regards,

David Altshuler, Gonçalo Abecasis, Mark McCarthy, and Michael Boehnke
on behalf of the Cardio-MetaboChip Design Team

Cardio-MetaboChip Consortium (CMC^2) Design and SNP Content Review
September 18, 2009

Summary:
Our aim was to design a high-density custom array (~200,000 SNPs) that captures DNA variation at regions identified by well-powered genome-wide association study meta-analyses for diseases and traits relevant to metabolic and atherosclerotic-cardiovascular endpoints. The application for this array is the Illumina iSelect technology.

Cardio-MetaboChip Consortium Steering Committee Members from GWAS Meta-Analysis Consortia:

CONSORTIUM	Representatives	Content Submitters
DIAGRAM	David Altshuler Michael Boehnke Mark McCarthy	Andrew Morris Ben Voight
CARDIoGRAM	Heribert Schunkert Nilesh Samani Panos Deloukas Sekar Kathiresan	Jeanette Erdmann Panos Deloukas
Lipids	Sekar Kathiresan Leena Peltonen Michael Boehnke	Tanya Teslovich Sekar Kathiresan
GIANT	Joel Hirschhorn Mark McCarthy	Joshua Randall Liz Speliotes Ruth Loos Anne Jackson Iris Heid
MAGIC	Inês Barroso Gonçalo Abecasis	Inga Prokopenko Richa Saxena Eleanor Wheeler Gonçalo Abecasis
ICBP_GWAS	Christopher Newton-Cheh Patricia Munroe	Toby Johnson
QT-IGC	Christopher Newton-Cheh Arne Pfeufer	Christopher Newton-Cheh

Design Team Members:
Noël Burtt, Ben Voight, and Cameron Palmer at the Broad Institute
Hyun Min Kang, Jun Ding, and Yanming Li at the University of Michigan

Consortium Leaders:
David Altshuler, Gonçalo Abecasis, Mark McCarthy, and Michael Boehnke

Consortium Contact:
Noël Burtt (burtt@broadinstitute.org)

Core Features of the Array:

A. Traits:
Tier 1: Primary metabolic and atherosclerotic-cardiovascular traits (12 traits, 5000 SNPs/trait)
Type 2 diabetes (T2D), fasting glucose (fastGLU), myocardial infarction (MI), coronary artery disease (CAD), LDL-cholesterol, HDL-cholesterol, triglycerides (TG), body mass index (BMI), waist-to-hip ratio adjusted for body mass index (WHR), systolic blood pressure (SBP), diastolic blood pressure (DBP), and QT interval (QT).

Tier 2: Secondary metabolic and atherosclerotic-cardiovascular traits (12 traits, 1000 SNPs/trait)

Fasting insulin (fastINS), 2hr glucose (2hGLU), HbA1c, T2D age of diagnosis in T2D cases (T2DAOD), early onset T2D (cases < 45 years vs controls) (T2DEarlyonset), waist circumference adjusted for BMI (WC), height, percent fat mass (FATPCT), total cholesterol (TC), platelet counts (PLT), mean platelet volume (MPV), and white blood cell counts (WBC).

Tier 3: All human traits
This tier includes the index SNP for any validated (genome-wide significant) association to any human trait along with a proxy.

B. Design Elements: 3 classes
1. Replication: Deep replication of signals arising from consortium-wide GWAs meta-analyses.
2. Fine mapping: Fine mapping of validated signals derived from GWAs with SNP content selected from HapMap and 1000 Genomes.
3. Wildcard and miscellaneous: SNPs from Consortium-specific hypotheses and interests (eg. from a particular pathway or deep sequencing efforts); trait-independent SNPs including common CNV-tags, MHC SNPs, fingerprint SNPs from all GWA array products, mtSNPs, and chromosome X and Y markers.

SNP Selection Criteria:

A. Replication:

A.1. Tier 1 and Tier 2 Replication:

66,117 SNPs utilizing **71,055 bead types** were promoted to the array based on the union of replication lists provided by each of the 12 Tier-1 and 12 Tier-2 trait groups listed above. The approach for selecting SNP content required two tasks: LD-pruning of replication lists and selection of SNPs from LD-grouped sets.

Task One: Linkage disequilibrium (LD) pruning

Phenotype consortia provided lists of their 50,000 (Tier-1) or 10,000 (Tier-2) most significant SNPs for each phenotype from their meta-analyses (with the exception of platelet counts, mean platelet volume, and white blood cell count, for which results with $p < 1 \times 10^{-3}$ were provided). For each phenotype, SNP results were "clumped" (grouped) together, starting from the most significant SNP and moving to least significant, indexed by the lead SNP (the most significantly associated in the clump), based on LD patterns in the phase 2 HapMap CEU. The clumping strategy was similar to that described in the PLINK software package (http://pngu.mgh.harvard.edu/~purcell/plink/clump.shtml). We used a clumping threshold of r2>= 0.2, meaning that index SNPs should not be in LD with any other index SNP at r2>= 0.2. SNPs not found in HapMap were assumed to be index SNPs separate from the others. In fact, many of these SNPs are in LD with other chosen SNPs, but we made no attempt to filter those out here.

Task Two: SNP Selection

After LD pruning, each SNP clump was rank ordered by the p-value of the index SNP. Within each clump of SNPs, SNPs were ranked by a priority score which incorporated the following information:
1. P-value rank of the SNP within the group, ordered by most significant to least significant (measured by the rank_in_group/total_snps_in_group);
2. Number of bead type used (=1 if two bead types, = 0 if one bead type); and
3. Design score of the proposed assay (calculated as 1.1 minus the Illumina design score).

The priority score was calculated as the sum of the products of the measures for each piece of information by its respective weight. Weights for measures 1 and 2 were set to 0.2 and weight for measure 3 was set at 0.6, so that stronger preference was given to SNPs with the best design scores. Under this weighting scheme, SNPs with the most significant p-value within the clump, that use only one bead type, and that have the highest possible Illumina design score (1.1) have the smallest (that is, best) priority score.

Once SNPs within each clump were ordered by priority score, the top SNP for each index was promoted to the array; SNPs representing the top 5000 (Tier-1) or top 1000 (Tier-2) SNP clumps were included. In addition, for the top 300 (Tier-1) or

100 (Tier-2) top signals for each phenotype, an additional SNP was included which had the next highest priority within the clump.

No specific minimum Illumina design score filter was applied to this list, although only ~850 assays have Illumina design score < 0.5.

Table 1: Tier 1 and Tier 2 Replication Summary:

CONSORTIUM	TRAIT NAME	TRAIT SHORT NAME	No. of SNPS	No. of Beads
DIAGRAM	Type 2 diabetes	T2D	5,270	5,639
	T2D Age Diagnosed	T2DAOD	1,081	1,147
	T2D Early Onset	T2DEarlyOnset	1,089	1,155
CARDIoGRAM	MI & CAD	MICAD*	6,756	7,229
	Mean Platelet Volume	MPV	697	769
	Platelet Count	PLT	598	648
	White Blood Cell	WBC	619	673
Lipids	HDL Cholesterol	HDL	5,249	5,658
	LDL Cholesterol	LDL	5,250	5,638
	Triglycerides	TG	5,256	5,657
	Total Cholesterol	TC	971	1,048
GIANT	Body Mass Index	BMI	5,276	5,697
	Waist Hip Ratio (BMI adjusted)	WHR	5,268	5,677
	Waist Circumference (BMI adjusted)	WC	1,093	1,178
	Height	HEIGHT	1,098	1,122
	Percent Fat Mass	FATPCT	1,076	1,158
MAGIC	Fasting Glucose	fastGLU	5,267	5,637
	Fasting Insulin	fastINS	1,086	1,159
	2hr Glucose	2hGLU	1,081	1,162
	HbA1c	HbA1c	1,082	1,182
ICBP_GWAS	Systolic Blood Pressure	SBP	5,269	5,618
	Diastolic Blood Pressure	DBP	5,267	5,664
QT-IGC	QT Interval	QT	5,244	5,623
TOTAL	Total with redundancy		70,943	76,138
	Number of unique SNPs		66,117	71,055

*Due to a misunderstanding in the file submission process, an additional 1,516 SNPs were added to the CAD/MI list, most of which represent additional proxies for signals on the list.

A.2. Tier 3 Replication: Associated SNP and Proxies

2,207 unique SNPs utilizing **2,517** bead types (all with Illumina design score > 0.5) were selected from SNPs identified as associated with any trait at a genome-wide significant threshold of $P < 5 \times 10^{-8}$. This list was obtained from the catalog of published association results at the NHGRI website (http://www.genome.gov/gwastudies/, obtained on August 1, 2009). We augmented this list of SNPs with one proxy per SNP from the CEU HapMap2 data to guard against genotyping failure (r2>0.9 with original SNP). This list was then tagged in the YRI HapMap2 with up to 4 additional SNPs (r2>0.5 with SNP, 77% final coverage) to capture different LD patterns around the signal in the YRI sample.

Citation: Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, and Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. [May 27, 2009.]

B. FINE MAPPING:

139,879 unique SNPs utilizing **160,993** beadtypes were promoted to the array for fine mapping of a total of 257 loci nominated by the GWA meta-analysis Consortia. For 214 loci submitted for signal fine mapping (SFM), the relevant Consortium provided the index SNP representing the strongest association signal at that locus. For 43 loci submitted for locus fine mapping (LFM), genomic intervals were specified by the Consortium. Priority for each locus was specified by the submitting Consortium as GOLD (highest), SILVER, or BRONZE (lowest).

1000 Genomes Pilot 1 SNP calls obtained from Sanger Institute (August 12, 2009) and the Broad Institute (August 11, 2009) were combined and used to determine intervals for SFM and to select SNPs within the fine mapping intervals. We also incorporated existing HapMap phase 2 data to complement LD estimates given sparse coverage of 1000 Genomes data in some regions.

Task One: SFM boundary selection.

Intervals for signal fine mapping were determined using LD information from the 1000 Genomes (Sanger calls with LD information) and HapMap phase 2 CEU data. Initial boundaries were determined by identifying all SNPs with r2>=0.5 with the index SNP, and then expanding the initial boundaries by 0.02cM in either direction using the HapMap-based genetic map (The International HapMap Consortium (2007) *Nature* **449**: 851-861). Practically, this approach extends the interval in each direction to the nearest flanking hotspot, but stops if there is no hotspot nearby. To avoid loss of LD information due to low coverage in the 1000 Genomes data, we used HapMap data to identify perfect proxies of the index SNPs within a 20kb window, and estimated boundaries based on the maximum r2 with any of those perfect proxies. All boundaries were manually screened with visualized LD plots and further trimmed or expanded if deemed appropriate. This occurred for example if the interval spanned >500kb or the r2>=0.5 boundary was ambiguous due to low MAF or a single distant SNP in high LD with the index SNP; 12 of the 214 SFM intervals were manually modified in this way. An additional 16 loci had their boundaries trimmed or expanded based on the specific requests from the relevant Consortium representatives after initial boundary review. Before considering overlap, the total region size was 64.97Mb; taking overlap into account reduced the total region size to 45.52 Mb. Regions present for m different traits were weighted by 1/m in the interval sizes listed in Table 2.

Table 2: Number of fine mapping loci and size of intervals

CONSORTIUM	TRAIT NAME	Number of regions				Interval size in Mb*			
		SFM GOLD	SFM Total	LFM Total	Grand Total	SFM GOLD	SFM Total	LFM Total	Grand Total
DIAGRAM	T2D	16	29	5	34	1.39	3.65	1.88	5.53
CARDIoGRAM	MICAD	16	24	6	30	2.99	3.98	1.73	5.71
Lipids	LDL	9	17	4	21	1.07	2.55	1.00	3.55
	HDL	8	20	3	23	0.49	3.35	0.53	3.88
	TG	6	17	3	20	0.52	1.71	1.78	3.49
GIANT	BMI	12	21	3	24	3.03	5.45	1.10	6.55
	WHR	6	14	1	15	0.62	1.68	0.18	1.86
	WC	1	2	0	2	0.17	0.33	0.00	0.33
MAGIC	fastGLU	9	14	5	19	0.61	1.47	1.86	3.33
	fastINS	1	1	1	2	0.03	0.03	0.56	0.59
	2hrGLU	1	2	1	3	0.02	0.17	0.25	0.42
	HbA1c	1	5	0	5	0.26	0.37	0.00	0.37
ICBP_GWAS	SBP	8	17	4	21	0.97	2.20	0.55	2.75
	DBP	7	16	4	20	0.74	1.77	1.36	3.13
QT_IGC	QT	10	15	3	18	1.26	2.56	1.47	4.03
TOTAL		111	214	43	257	14.17	31.27	14.25	45.52

* Overlapping interval sizes are corrected (dividing by the number of overlaps)

Task Two: SNP selection

Within a specified interval, 1000 Genomes SNPs (in the Sanger or Broad call sets) were considered as potential fine mapping SNPs unless SNP minor allele frequency (MAF) was <0.01 in all three HapMap samples. Since the available Broad SNP calls lacked LD information, for Broad-only SNPs, we used a MAF cutoff of 0.02. To reduce the risk of genotyping failure, SNPs were eliminated from consideration if (a) the Illumina design score was < 0.5 or (b) there were SNPs within 15bp in both directions of the SNP of interest with CEU MAF > 0.02. As an exception to this filtering strategy, SNPs annotated as nonsynonymous, essential splice site, or stop codons (henceforth "likely functional") were included regardless of MAF, design score, or nearby SNPs in the primer. In total, 73,070 of 242,301 SNPs (30.1%) failed one or more of our filtering criteria and were removed from consideration.

For the SFM GOLD loci, all candidate SNPs passing the above QC filtering criteria were promoted to the Metabochip. For LFM GOLD and all SILVER loci, SNPs were selected based on r2=1.0 tagging across all three continental groups, with redundant proxies added at 30kb intervals to increase the coverage of highly represented SNPs. Moreover, SNPs in moderate to high LD with the index SNPs (r2>=0.3) and all likely functional SNPs were unconditionally included before starting the tagging procedure. SNPs specific to the Broad 1000 Genomes calls and SNPs already included based on other categories of the Metabochip design were also unconditionally included prior to tagging. This tagging procedure resulted in inclusion of 79.7% of all available SNPs. For BRONZE loci, we tagged using exactly the same SNP selection strategy but with a lower r2 threshold of 0.8, resulting in inclusion of 71.6% of all BRONZE candidate SNPs.

Table 3: Number of SNPs selected for the fine-mapping:

CONSORTIUM	TRAIT NAME	Number of SNPs included in the Metabochip*			
		SFM GOLD	SFM Total	LFM Total	Grand Total
DIAGRAM	T2D	5,392	11,069	6,483	17,553
CARDIoGRAM	MICAD	9,442	13,287	5,545	18,832
Lipids	LDL	3,578	7,155	3,310	10,466
	HDL	2,324	8,692	2,329	11,021
	TG	1,776	4,950	3,962	8,912
GIANT	BMI	9,102	15,322	4,020	19,342
	WHR	2,250	5,044	782	5,826
	WC	594	1,068	0	1,068
MAGIC	fastGLU	2,708	4,740	7,045	11,784
	fastINS	143	143	1,326	1,469
	2hrGLU	50	544	638	1,182
	HbA1c	1,538	2,006	0	2,006
ICBP_GWAS	SBP	3,272	6,500	1,726	8,226
	DBP	2,590	5,626	3,825	9,461
QT_IGC	QT	5,208	8,026	4,706	12,732
TOTAL		49,967	94,710	45,709	139,879
SNP density (#SNPs/kb)		3.53	3.03	3.21	3.08
Saved by tagging (SILVER/BRONZE)		0%	12.7%	20.1%	15.3%

* Overlapping SNPs are corrected by dividing by the number of overlaps.

C. WILDCARDS AND MISCELLANEOUS:

C.1. Wildcard SNPs:

5,314 SNPs utilizing **6,104** bead types were placed on the array, regardless of Illumina design Score. These SNPs were submitted by each Consortium and are based on Consortium-specific criteria.

C.2. CNV Tag SNPs:

6,888 SNPs utilizing **7,671** bead types were promoted to the array based on the union of SNPs tagging (based on LD) copy number polymorphisms (CNPs) detected from three sources: HapMap 3 based CNP tags from Jim Nemesh [nemesh@broadinstitute.org] and Steve McCarroll [smccarro@broadinstitute.org]; HapMap 2 based CNP tags from Matt Hurles [meh@sanger.ac.uk], and *WTCCC+* based CNP tags from Richard Pearson [richard.pearson@well.ox.ac.uk]. All CNP tagging SNPs promoted to the array had Illumina design scores > 0.5.

C.3. MHC SNPs:

Three of the signal fine mapping loci were handled differently because they reside in the MHC region on chromosome 6. Due to the extremely high SNP density and low recombination rate in the MHC, applying the same SNP selection strategy used for other genomic regions would have included tens of thousands of SNPs each. Instead, we included **3,203** SNPs utilizing **3,314** bead types encompassing the extended MHC region (de Bakker PI et al. (2006) *Nat Genet*. **38**:1166-72) utilizing the information from two existing commercial Illumina chips (660W-Quad and Cyto SNP-12). Although the MHC regions were not the main target of the Metabochip, including these MHC SNPs should help explore the direct and indirect effect of the MHC variations on these metabolic traits.

C.4. X and Y Chromosome SNPs (GENDER):

112 SNPs utilizing **112** bead types, 90 on the X chromosome and 22 on the Y chromosome, were promoted to the array to provide a genetic assessment of gender. Promoted SNPs met the following criteria:

1. Illumina design score of 1.1 indicating a previously validated SNP assay
2. Heterozygote frequency > 0.2 for all three HapMap 2 samples
3. For X chromosome markers, MAF>35% in the CEU HapMap samples
4. Clear genotyping plots (no compressions, good intensity, tight clusters, etc.)
5. Used only a single bead type
6. No SNP was within 200kb of another.

C.5. Fingerprint SNPs:

46 SNPs using **50** beadtypes were placed on the array. These markers cover all 22 autosomes and have MAF>=15% in HapMap CEU, CHB, JPT, and YRI samples. These makers are validated assays on the Genome-Wide Human SNP Array 6.0 and the Infinium® HD BeadChips**.** These SNPs were included since some genotyping centers routinely use these markers as a genotypic barcode during the QC process to assess sample concordance across genotyping platforms and array types. All SNPs for this selection had Illumina design scores >0.5.

C.6. MtDNA Tag SNPs:

144 mtDNA SNPs requiring **147** beadtypes were selected for the array. These markers capture all variation to 5% in Asians and Africans, and to 1% in Europeans. We included redundant markers for the 5% SNPs. Finally, the rare MIDD mutation and a putative T2D SNP were also added. All SNPs had Illumina design score >=0.5.

Citation: Saxena R, de Bakker PI, Singer K, Mootha V, Burtt N, Hirschhorn JN, Gaudet D, Isomaa B, Daly MJ, Groop L, Ardlie KG, Altshuler D. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. Am J Hum Genet. 2006 79:54-61 and Saxena *et. al.,* unpublished data, 2009.

Table 4: Review of SNP Territory Allocation by Design Element:

Design class	Allocation breakdown	Total SNPs	Total BeadTypes
REPLICATION	Tier 1: 5000 X 12 traits≈60,000 *Proxies for top 300 SNPs=300X 12 trait≈3600* Tier 2: 1000 X 12traits≈12,000 *Proxies for top 100 SNPs 100 X 12 trait≈1200*	66,117	71,055
	Tier 3: Index SNP and Proxy	2,207	2,517
FINE-MAPPING	Signal FM and Locus FM	139,879	160,993
WILDCARD	Up to 1000 SNPs per consortium	5,314	6,104
MISCELLANEOUS	CNV tagging SNPs	6,888	7,672
	MHC SNPs	3,203	3,314
	X and Y chromosome SNPs	112	112
	Fingerprint SNPs	46	50
	mtDNA tag SNPs	144	147
TOTAL	**All lists merged and redundant markers excluded**	217,697	245,243