Supplemental Information for

Pervasive haplotypic variation of the spliceo-transcriptome of the human Major Histocompatibility Complex

Claire Vandiedonck[1,2,3±], Martin S Taylor[1,4], Helen E Lockstone[1], Katharine Plant[1], Jennifer M Taylor[1], Caroline Durrant[1], John Broxholme[1], Benjamin P. Fairfax[1] and Julian C Knight[1±]

[1] *Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, OX3 7BN, UK.*

[2] *INSERM UMRS-958, 10 avenue de Verdun, 75010 PARIS, France*

[3] *Université Paris 7 Denis-Diderot, Paris, France*

[4] *MRC Human Genetics Unit, Edinburgh, EH4 2XU, UK*

[±] to whom correspondence should be addressed :
claire.vandiedonck@inserm.fr and julian@well.ox.ac.uk

Supplemental Material includes:

Supplemental Methods
Supplemental Figures 1 to 8
Supplemental Tables 1 to 8
Supplemental References

TABLE OF CONTENTS:

1. SUPPLEMENTAL METHODS

1.1. The MHC Array

The MHC array was manufactured by Affymetrix using their 49-7875 format with 11 μm features and all probes were 25 nucleotides in length. The design incorporated tile probes and splice junction probes as well as alternate allele versions in both the tile and splice junction sets. For every Watson strand probe the corresponding Crick probe was also included on the array, for both tile and junction probes. For the junction set, we spotted each probe in four technical replicates on the array.

1.1.A. Target regions and annotation

The reference sequence for the MHC array design was based on the NCBI:35 (aka. UCSC:hg17, May 2004) reference human genome assembly. Except where otherwise stated, the annotation used in the array design was from the hg17 UCSC Genome Browser annotation database (September 2006). Regions targeted for tile probe design correspond to chr6:29748239-33231091. We were able to construct tile paths at an average resolution of one (Watson strand) probe per 18 nucleotides (nt) across 86% of this target region. The sequence of the remaining 14% was not sufficiently unique in the genome for discriminatory probes to be designed. A more tightly defined region of chr6:31593635-32482819 was additionally targeted for the design of splice junction probes. The complete extent of the hg17 reference assembly alternate HLA_HAP1 and HLA_HAP2 haplotype assemblies were also targeted for tile probe design.

Subsequently, we remapped all the probes using the hg18 reference sequence and analyzed and presented the data using this build. Importantly, the MHC reference sequence (that is from PGF) has remained unchanged between hg17, hg18 and the current build, hg19. The latter two builds differ only by a constant shift in the numbering that is of 107,079 nt on the reference sequence, of 63,673 nt on COX sequence and of 46,053 nt on QBL sequence.

1.1.B. Probe types

Tile probes. For the targeted genomic intervals, every possible tile probe was evaluated (*n*-25+1 Watson strand probes, where *n* is the length of the target interval in nt). Our

target density was one probe per 18 nucleotides of target sequence, resulting in a tile path of overlapping probes of unprecedented resolution.

Splice junction probes. A genome wide, database of splice junction sequences was produced, comprising 850,552 unique junctions. This was based on annotation of splice donors and acceptors derived from pre-computed alignments of UCSC known genes, RefSeq transcripts, VEGA annotation, public mRNA sequences and Acembly clustered ESTs on the reference genome. All alignment data were obtained from UCSC hg17 database tables. Splice junction sequences used for the design typically comprised 24 nt exonic splice donor concatenated with 24 nt exonic splice acceptor, though exonic sequences were truncated at points of overlap with other annotated splice sites. Splice junctions mapping to the hg17 reference sequence coordinates chr6:31593635-32482819 were targeted for the design of splice junction traversing probes. Our aim was six overlapping, Watson strand, splice junction traversing probes per targeted splice junction.

Alternate allele probes. Where a candidate probe overlapped a known single nucleotide (substitution) polymorphism, an alternate allele version of that probe was also designed. In the cases where multiple known polymorphisms overlap the probe position, alternate allele probes were designed that represented every permutation of SNP phasing.

Affymetrix exon probes. To enable direct comparison with gene expression measured by Affymetrix human Exon array 1.0 ST (HuEx1), we incorporated all probes from HuEx1 that mapped into our tiling array target regions, regardless of their quality score or uniqueness. Many of these HuEx1 probes are not genome-wide unique, only unique probes were included in the tile-path analyses reported in this work.

Control probes. A total of 19,184 control perfect match (PM) probes were selected from the Affymetrix arrays: 99 for spikes (antisense), 640 for housekeeping genes (both strands), 138 for *Alu* (both strands) and 264 for polyA (both strands) from the Human Genome U133 Plus 2.0 array, as well as 1,100 probes corresponding to introns and exons of housekeeping genes and 16,943 antigenomic with different content of GC from the Exon 1.0 ST array. We also included 10,973 probes, only PM and in sense orientation, corresponding to 994 probesets of 371 non-MHC genes involved in alternative splicing, immune response, cell cycle, signaling, as well as human tissue specific genes. In addition, 17 genomic regions (range 212 nt to 19.5 kb), considered as positive and

negative controls for future array applications, were incorporated into the tiling and/or splicing design.

1.1.C. Probe selection

Each candidate probe was scored for its genome-wide uniqueness, sequence properties (probe quality), and overlap with polymorphic sites. For each of these measures we defined an optimum (opt) value as well as thresholds for acceptable minima (min) and maxima (max).

Uniqueness. We considered two measures of uniqueness for candidate probe sequences. First, the number of nucleotide identities between the candidate probe and it's best non-self match in the genome or whole-genome splice junction dataset (opt=0, min=0, max=24; i.e. we required the candidate probe to be at least one substitution away from any other sequence). The second uniqueness measure was the number of highly similar (up to three substitutions from the candidate probe) sequences in the genome (opt=0, min=0, max=9). For tile probes, alignments were calculated using a customized version of Olly (Jim Kent, UCSC, Unpublished; customization to stop searching for matches after the tenth match is found) that finds all nearly-identical matches, up to a specified substitution distance, in ungapped alignment between a query sequence and a target genome. This allowed the uniqueness measures to be calculated for non-repeat masked genome, considered desirable as diverged repetitive sequences often contain genome-wide unique 25 mers, adding to the target sequence coverage, particularly in regions that are otherwise difficult to assay.

For the candidate junction probes, alignments to the reference genome and the genome wide splice junction dataset were performed using BLASTN (NCBI blastall with options -FF -w7) which allows for alignment gaps as well as substitutions. Self matches and partial self matches (where splice junctions share either a splice donor or acceptor exonic sequence) were filtered out. For uniqueness scoring, we considered alignment identity as with Olly based alignments, and ignored alignment gaps.

Probe quality score. The Affymetrix probe score as detailed in (Mei et al. 2003) was applied to candidate probes and averaged over the Watson and Crick strands (opt=0.8, min=0.08, max=0.8). This score incorporates measures of sequence composition, secondary structure and hybridization thermodynamics.

Overlapping polymorphisms. To minimize the number of permutations of alternate allele probes and to simplify the interpretation of hybridization results between haplotypes we sought to minimize the number of known polymorphisms (dbSNP build 126) overlapping a probe (opt=0, min=0, max=2).

Probe tiling/spacing. Tracts of target genome sequence (> 24 nt) for which no unique probes could be designed, and Affymetrix HuEx1 derived probes already incorporated into the design (see above) provided a natural punctuation with which to constrain the choice of the probe tiling path. Between these fixed positions, we calculated an optimal tiling path of probes using a scoring system that incorporated the uniqueness, probe quality and polymorphism overlap measures described above as well as relative spacing measured as the distance between the midpoint of adjacent probes (opt=18, min=12, max=24), all transformed onto a unified scale (*D*) and equally weighted using equation 1:

$$D = \frac{\sum_{i=1}^{N} \left(\frac{(Obs_i - Opt_i)}{(Opt_i - (Min_i | Max_i) + p)} \right)^2}{N}$$

Where Obs, Opt, Min and Max were respectively: observed values, optimal values and minimal and maximal allowed values. Min_i is used where $Obs_i < Opt_i$, otherwise Max_i was used. *p* is a small constant (0.0001) that overcomes the problem of divisions by zero. This rescaled score was averaged (sum divided by the number *N* of measures *i*) over each of the probe measures (*i* represents the probe measures: best non-self match, number of non-self matches, probe quality score, overlapping polymorphisms and probe spacing). As the probe spacing parameter was dependent on adjacent probes we applied Dijkstra's dynamic programming algorithm (Aho et al. 1983) to select an optimal tile path.

Segmental duplications. There is a known, high identity segmental duplication within the targeted MHC region, defined as hg17 coordinates chr6:32056288-32089047 and chr6:32089084-32121844. Specifically for these regions, uniqueness criteria were adjusted so that a probe match to either region was considered a self-match, allowing tile-paths to be designed through these regions.

Alternate haplotypes. Initial probe selection (including alternate allele probe design) was based on the PGF haplotype in the main reference assembly. After mapping designed probes (including alternate allele versions) into the alternate COX and QBL haplotypes using a hash table, remaining gaps in the tile path were closed by selecting an optimal tile

path as described above. As with the segmentally duplicated regions, uniqueness criteria were adjusted over the targeted alternate haplotype intervals, so that matches in any of the haplotypes were considered self-matches.

Junction probe specific parameters. Splice junction traversing probes were selected using the same set of parameters as for tile probes, but the probe spacing was adjusted to (opt=3, min=1, max=24). We also constrained selection to only consider probes that had nucleotides corresponding to at least 4 nucleotides from both exons of the splice junction.

1.2. Array signal processing

1.2.A. Custom MHC array processing. Custom array signals were processed using an in-house pipeline under the R and Bioconductor environment, and using Perl scripts.

Preprocessing. Affymetrix data intensities (.CEL files) were read using the "affy" Bioconductor software package. Signal intensity was first background corrected for each array by subtracting the median intensity of 2,524 blanks. Using control antigenomic GC bins, we corrected probe intensities by subtracting the corresponding GC bin median intensity. Experimental probe GC content ranged from 3 to 21. Between array normalization was done using vsn (Huber et al. 2002), which also stabilizes the variance by transforming intensities to a generalized log scale to base 2. Finally, probes showing excessive dispersion between probe replicates first, and then between biological replicates, were filtered out. In all cases, they correspond to the first percentile of the distribution of the ratio standard deviation/mean.

Tiling path analysis. The smoothing process involved two steps (Sabo et al. 2006). First, using a moving window of 100 bases, we averaged the signal of the probes that were above the background set at the fifth percentile. Then the signal was weighted for distance from the centre of the window using a Gaussian function whose σ was defined at 30 as being the expected standard deviation of the size of the fragmented hybridized samples, ranging from 40 to 70 bases. Transcriptionally active regions (TARs) were determined by identifying windows where the signal exceeded a threshold determined by permutation of probe signal intensities. We considered 51-base (25 bases on each side of the tested position) windows with at least 3 probes (corresponding to the 18 nt average resolution of the tiling) above background. The median intensities across all probes above

background in each window were computed. To determine thresholds, we generated 1,000 random datasets of 1,000 consecutive probes with a randomly attributed intensity, similarly considered 51-base windows and used the 99th and 95th percentiles of the generated median distributions. Thresholds corresponding to 1% or 5% false discovery rate (FDR) were used depending on purpose. Eight tiling paths were considered, including the "shared paths" on each strand, corresponding to probes mapping uniquely in all three haplotypes, and the "alternate paths" on each strand and also including probes specific to each haplotype. Overlapping TARs were merged into transcribed blocks.

Gene and exon level computations were performed on each of the haplotypic "alternate paths". To this end, we grouped probes into metaprobesets corresponding to Vega annotations per haplotypic path. At the exon level, they comprise all probes associated with each exon of each transcript. At the gene level, they include all unique probes corresponding to exons of each gene, such that if an exon is present in different transcripts of a given gene, the corresponding probes are only counted once. Probes corresponding to introns were ignored. Within a gene or exon, the hybridization to each probe can differ depending on the affinities and transcript diversity. To account for these variations, we computed the median rather than the mean intensities of all probes in the given metaprobeset. For each cell line, this was done using its corresponding "alternate path" metaprobeset. This provided the most complete and robust information on gene and exon expression in each cell line.

Splicing computation was performed on each of the "alternate paths" per gene. We first considered every exon with reference to their transcript and generated normalized intensities (NI) (defined as the log2 ratio of exon level intensities and gene intensities). A null value indicates the exon is not alternatively spliced in the gene. An exon can be represented with different names in different transcripts of a given gene, but the underlying probes are the same and therefore the NI values are identical. To better estimate the transcript abundance, we considered a second complementary level by computing junction normalized intensities (JNI). To this end, we generated junction metaprobesets per "alternate path" by grouping all junction probes targeting a common splice junction of a transcript. Note that if a splice junction is shared between several transcripts or even genes, a metaprobeset exists for each transcript. Junction probes were

present in four technical replicates on the array. Their signal intensities were averaged to obtain a more accurate estimate of intensities. Junction intensities were next computed as the median of all probes in the junction metaprobeset and normalized against the gene level intensity.

1.2.B. Exon array processing. Exon array signals were processed using Affymetrix Power Tools (APT) and R.

Cross-platform correlation with the MHC array:

Probe level data for the entire exon array were first extracted using apt-cel-extract in Affymetrix Power Tools (apt-1.8.0), with a GC-based background correction. The 9 unstimulated samples were then normalized using vsn under the R/Bioconductor environment, so processing was the same as for MHC array data. A set of 10,572 probes from the Affymetrix Human Exon 1.0 ST array were included in the MHC array, each targeting a unique genomic location. An expression data matrix (10,572 x 9) for each platform was created and the correlation across all probes for the same sample on the two different platforms was measured (positive Pearson correlation test). Consistency between MHC and exon array data was also assessed by looking at differences between haplotypes. We computed fold changes between each pair of haplotypes for the 2,129 probes varying between haplotypes (defined as standard deviation across all samples > 0.5 on both platforms) and performed positive Pearson correlation tests, on all probes or after filtering out low intensity probes.

Gene, exon and splicing computation:

We used probeset definitions based on custom CDF libraries from the Microarray Lab (University of Michigan) (Dai et al. 2005) corresponding to the Vega and Ensembl gene and exon annotations (version 11 - November 12, 2008 which is based on the hg18 build). Probe level data from PGF, COX and QBL unstimulated samples were extracted, background corrected, normalized and summarized at the probeset level (exon or gene) using Robust Multichip Average (RMA) with the apt -probeset-summarize -a rma-sketch command and the appropriate CDF in APT (apt-1.8.0). Exon normalized intensities (NI) were computed for PGF by subtracting the log2 gene intensity from the log2 exon intensity (each first summarized as the median across the 3 replicates).

1.2.C. Quality controls.

We first assessed hybridization efficiency using spikes and polyA signals provided by Affymetrix.

The specificity of the strandedness was estimated by computing the ratio of the difference between the sense and antisense probeset signals to the sense probeset signal using 5 known expressed housekeeping genes (*ACTB*, *GAPDH*, *IRF9*, *RN28S1* and *RN18S1*). We provide the mean of all these ratios with its confidence interval at 95%.

The coverage of full length transcript was estimated by comparing signal intensities from the same housekeeping genes for the 5', middle located and 3' metaprobesets. We computed the coefficient of variation (standard deviation/mean).

Similarly, we computed the coefficient of variation of the 4 replicates of the junction probes spotted on the array. This was 0.0373 (95% confidence interval, 0.0362-0.0384).

We computed the Pearson correlation coefficient for all probes between each pair of samples and performed a hierarchical clustering on these coefficients (shown as a heatmap in Supplemental Fig. 3).

The sensitivity of junction probes was checked by comparing array data and quantitative real-time PCR data for the two main transcript isoforms of *CD79A* and *CD79B* genes.

The perfect match design specificity was estimated first by a global analysis comparing the signal of PGF ccDNA on PGF-, COX- and QBL-specific probes. The analysis was restricted to probes included in TARs to filter out background signal. Second, we carried a probe-wise comparison of the signal produced by the PGF samples on the 123 perfect match probes paired with probes carrying one mismatch corresponding to the COX path. Again, only probes included in TARs were considered. We used analysis of variance with repeated measures to account for the three biological replicates.

1.3. Statistical analyses

They were performed using R, Perl and PLINK.

1.3.A. TARs distribution:

Localization of TARs was studied with reference to the strand orientation. If a TAR was present on the sense path, it was expected to correspond to a transcript on the forward strand and vice versa. Hence, TARs tracks available at GEO (GSE22455) are given with respect to the transcript orientation for the strand column, although their names correspond to the path considered. We searched for overlap of at least one base inside gene boundaries obtained from Vega gene or pseudogene annotations. Percentage overlap, overlapped gene name and distance to neighboring genes relative to strand are provided in Supplemental Table 2. Intergenic TARs were defined by the absence of gene overlap on both strands. Median distance between Vega genes is 18.09 kb. Distal intergenic TARs were defined as being at least 10 kb away from the nearest gene. Their distribution was compared to that of genomic features extracted from UCSC hg18, including repeat elements from RepeatMasker (rmsk table in UCSC), CpG islands (cpgIslandEx table), open chromatin (EncodeDukeDNaseSeqPeaksGm12891V2 table where GM12891 carries *HLA-DR3/DR2* and is therefore heterozygous for COX/QBL and PGF alleles) and conservation (phastConsElements17way table). Colocalization was defined by an overlap of at least one base.

1.3.B. Differential expression:

Probe level comparisons were performed using the "shared path" probes only (269,678 of 381,916 tiling probes). Among the 230 genes and pseudogenes present in the MHC region, 206 are found on the three haplotype sequence annotations and were considered for gene level comparison using their respective "alternate path". Similarly, after filtering for redundancy between transcripts, we considered the normalized values for 2,198 exons and 591 junctions present in all three sequence annotations, thus directly comparable using their respective "alternate path".

All differential expression analyses were performed with the Bioconductor software package limma (Smyth 2004) which uses linear models and empirical Bayesian methods. We used a group-means parameterization with the model formula ~0+groups, where

groups was a factor describing the 18 samples in terms of the 6 haplotype-condition combinations. Contrasts were used to extract the comparisons of interest and Benjamini-Hochberg's adjustment was used to control the false discovery rate. Adjusted p-values below 0.05 were considered significant. First, a brief assessment of overall stimulation, haplotype and interaction effects was performed on the processed probe-level data: of the 269,678 probes, 9,232 probes showed a significant response to stimulation in at least one haplotype; 16,250 probes differed significantly between haplotypes in either stimulated or unstimulated conditions, and just 24 probes showed a significant interaction effect (i.e. the response to stimulation differed between haplotypes). The detailed analyses of haplotypic differences at the gene level, exon level NI values and junction JNI values presented in the current manuscript are based on the unstimulated samples. The same approach was used for the gene level analyses using data from the Affymetrix Exon 1.0 ST array.

1.3.C. Correlation of differentially expressed (DE) probes with SNP distribution:

The lists of polymorphic SNPs between the three haplotypes or haplotype pairs were generated using hg18 snp129 table at UCSC. Altogether there were 23,556 polymorphic SNPs: 16,787 between PGF and COX, 15,993 between PGF and QBL, and 13,734 between COX and QBL. They were compared to the lists of differentially expressed probes obtained with limma using the "shared path". To this end, we took the MHC reference sequence and removed genomic segments corresponding to contig gaps in the QBL sequence. The MHC region was then segmented into two series of 10 kb windows shifted by 5 kb. The counts of DE probes per window were normalized by the number of probes designed in each window. Correlation between the normalized counts of DE probes and counts of polymorphic SNPs across windows was tested using the Pearson test (after log-transformation) and nonparametric Spearman tests. This was done either using all the windows ("_poly.all" sheet of Supplemental Table 5) or after filtering out windows with low SNP counts in the lower quartile of the SNP count distribution ("_poly.quartile" sheet).

To assess the relevance of the correlation tests, we restricted the comparison to the windows overlapping with genes (and to control windows lacking genes). As an additional control, we performed the correlation tests similarly with nonpolymorphic

SNPs ("_notpoly.*" sheets). Findings were corroborated using a robust scheme, by performing nonparametric Spearman tests using equidistant bins of size 5, 10 or 20 for the distribution of counts of DE probes (always normalized with the number of designed probes per window) and of polymorphic (or non polymorphic) SNPs ("_poly.bin" and "_notpoly.bin" sheets).

1.3.D. eQTL mapping for *ZFP57, HLA-DQA2* and *HLA-DQB2*:

Standard QC measures and filtering on the volunteer data were done as described in (Fairfax et al. 2010). The quantitative trait association was conducted using PLINK with 4 maximum per-SNP missing genotypes (GENO 0.1) and MAF 0.03. For each SNP, PLINK generates a phenotypic mean for the three genotypic states and compares these means using the Wald test statistic to generate a P-value. The Wald test is useful especially in this instance, since it does not require that the data fit a normal distribution.

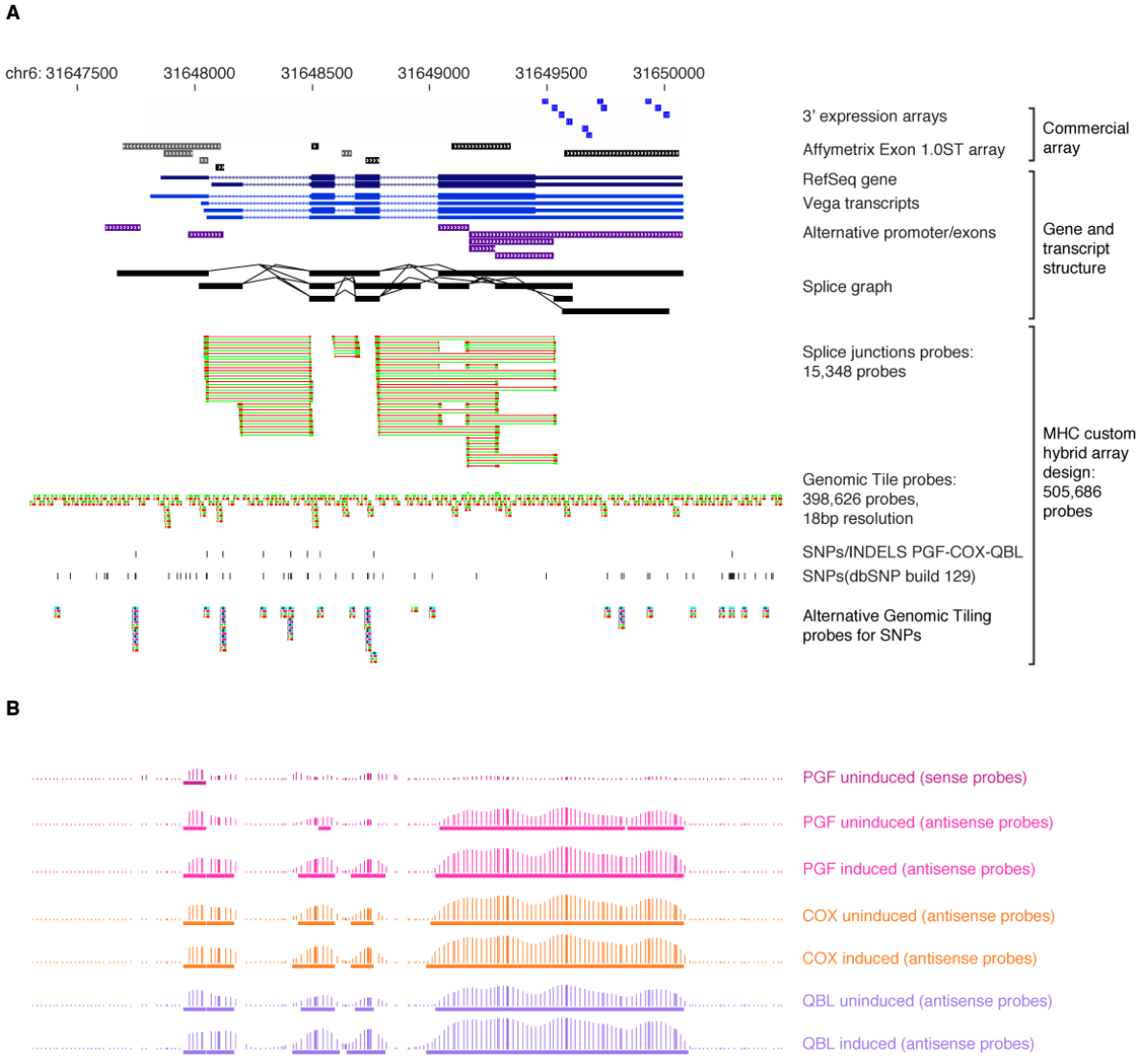
1.3.E. MHC splicing extent analysis in PGF sample:

Exon intensities normalized to the corresponding gene intensity (NIs) were computed as described earlier (see Splicing Computation). A negative NI value indicates the exon may be spliced out whereas a positive NI value indicates the exon may be included. We thus computed the percentage of exons with an absolute NI value different from zero exceeding various thresholds. This was done for both the 226 MHC genes annotated in Vega on the PGF sequence and the 28,454 non-MHC annotated Vega genes in the CDF files.

To test whether the observed difference between the two lists of genes was statistically significant, we used a permutation method to exclude the possibility of influence from any differences in gene structure. We analyzed the 131 MHC genes and the 15,659 non-MHC genes having at least 4 exons, a number that corresponds to the median number of exons per gene in the human genome. First, for the 131 MHC genes, we generated 1,000 samples of 4 exons per gene, drawn randomly with replacement. Using the NI of the selected MHC exons (n=131*4*1,000=524,000), we obtained the median and mean NI for MHC exons. For each sample, we also cumulatively enumerated the "spliced" exons, i.e. exons with an absolute NI above a given threshold (varying from 1 to 4). We then generated bootstrap distributions of the median and mean NI values and of the cumulative counts of spliced exons in non-MHC genes. To this end, we produced 10,000

sets of 131 genes randomly drawn from the 15,659 non-MHC genes having at least four exons. For each set, we generated 1,000 samples of 4 exons per gene and computed the median and mean NI and the total count of spliced exons for each threshold. We used these distributions to assess whether the NI values and the number of spliced exons were significantly different in MHC and non-MHC genes. We similarly compared non-MHC immune genes (733 genes with 4 exons) with non-MHC non-immune ones (sampling from 14,926 genes with 4 exons), and MHC genes with non-MHC immune genes. Immune-related genes were identified based on Gene Ontology classifications (GO:0002376). The comparison for the NI medians is shown graphically in Fig 4.C-D and for the number of spliced exons for different threshold in Supplemental Table 6. Repeating the whole analysis with Ensembl annotations yielded similar results.

2. SUPPLEMENTAL FIGURES AND TABLES



Supplemental Figure 1. Illustration of array design strategy and generated tracks for the *LTA* gene.

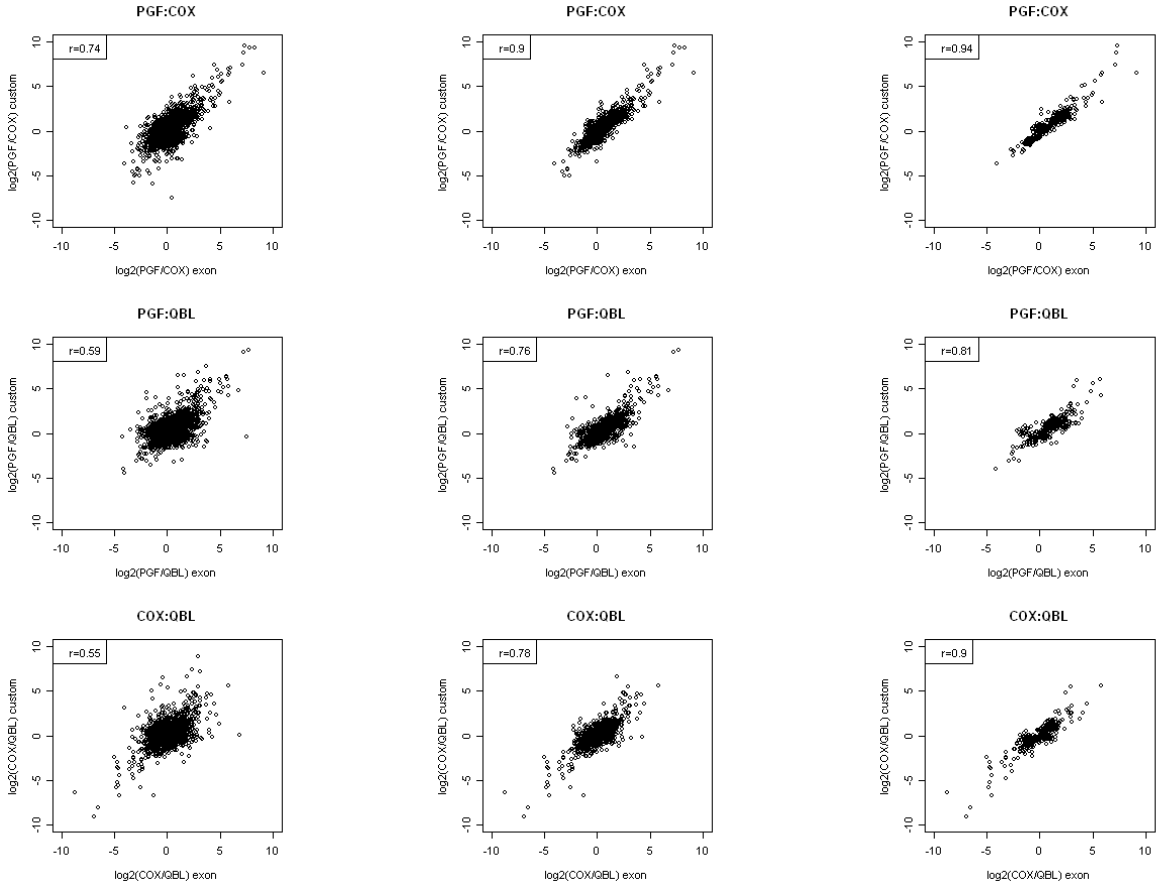
(A) To characterize the expression and splicing phenotypes of the MHC, as depicted here for the *LTA* gene, available commercial arrays with probes at the 3' end of transcripts were inappropriate. Recently, the Affymetrix GeneChip Human Exon 1.0 ST Array with probes covering the full length of transcripts was released, but its design does not account for the full complexity of gene expression, including alternative promoters or exons. Furthermore, commercial expression arrays are designed to the human reference assembly sequence and do not account for genetic diversity. Therefore, a custom array on the Affymetrix platform was conceived, with an original hybrid combination of two main probe sets. A set of 15,348 probes present in 4 copies aims at monitoring any possible known or predicted splice

events. On average 6 probes were designed at the centre of each junction (range 2-42), corresponding to 1,043 junctions of 78 genes. For each single probe, its reverse complement was also designed (red versus green color). In addition, a tiling set of 398,626 overlapping probes covers both strands of the genomic MHC region with a final resolution of 18 bases, enabling accurate transcript profiling and discovery. Strand specificity can be determined as reverse-complement tiling probes were also designed. The design also takes into account genetic polymorphism as alternate probes were specifically designed for any SNPs or haplotypic segmental duplications in the region. The array also includes 26,484 probes for relevant non-MHC genes involved in alternative splicing or immune response. Finally, 19,184 control probes were included for signal processing (for calibration assessment, background correction) and for other specific applications of the array. The complete array comprises 505,686 probes. Criteria, including uniqueness against the genome and transcriptome, and structural conformation were carefully considered during the design process (cf. Material and Methods).

(B) Example of custom tracks for shared probes between the three haplotypes as they appear in the UCSC browser. For each sample, the smoothed intensity signal is displayed with the corresponding TARs at a FDR of 1% shown below. This matches the known exon structure of the *LTA* gene. As the *LTA* gene is transcribed on the forward strand, the signal is expected to be on antisense probes as observed for example in the PGF sample. For the three cell lines, antisense probes tracks are shown for either unstimulated or stimulated conditions, revealing a more pronounced induction of *LTA* gene expression for PGF and QBL than in COX whose unstimulated expression was already higher.

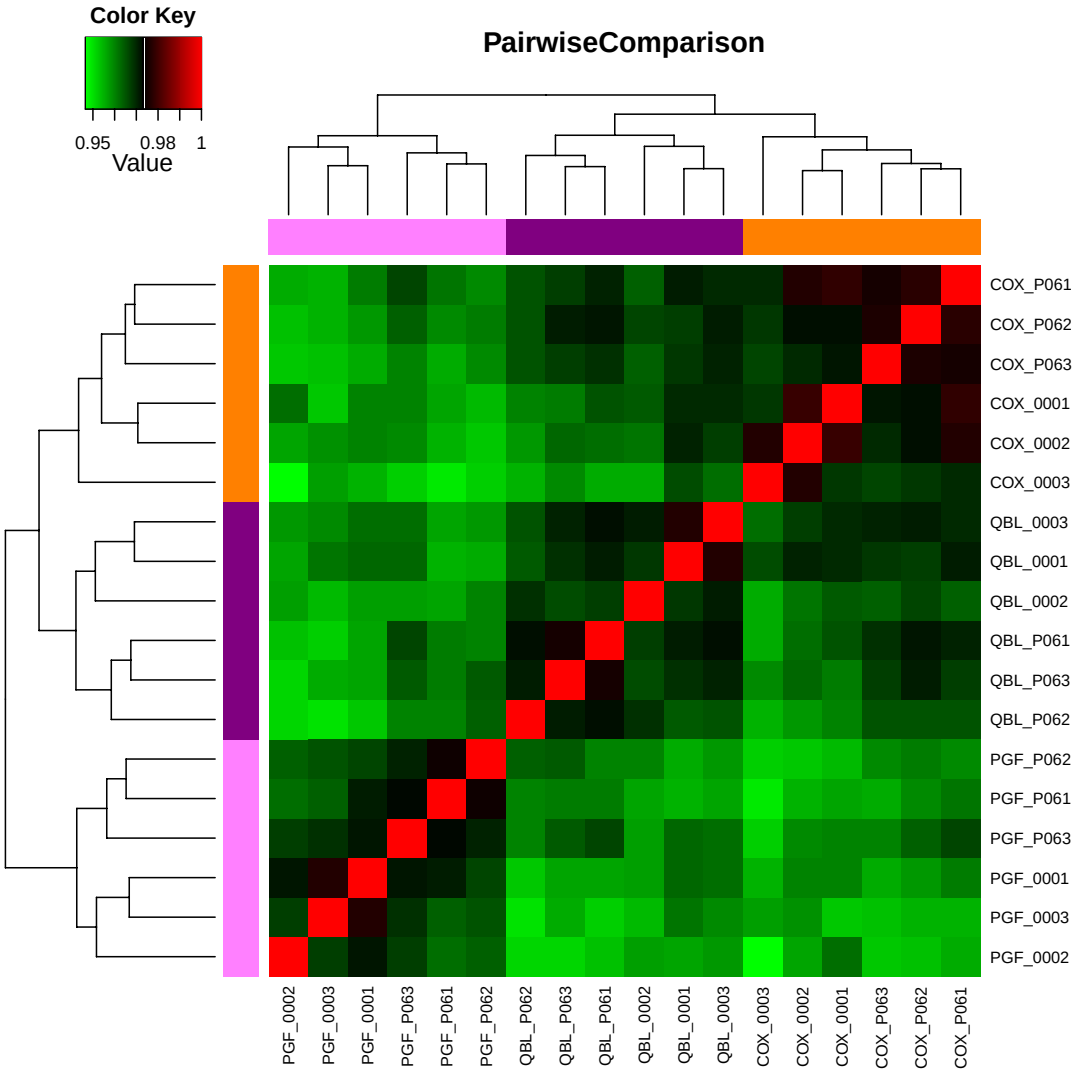
Supplemental Table 1. Correlation of intensity data for 10,572 probes shared between the MHC and the Affymetrix Exon 1.0 ST arrays.

Platform	Comparison	Number of correlations	Median correlation coefficient	Minimum correlation coefficient	Maximum correlation coefficient
MHC	Within haplotypes	9	0.98	0.97	0.98
Exon	Within haplotypes	9	0.92	0.90	0.94
MHC	Between haplotypes	27	0.95	0.93	0.96
Exon	Between haplotypes	27	0.89	0.84	0.91
Exon-MHC	Within haplotypes	27	0.88	0.83	0.91
Exon-MHC	Between haplotypes	27	0.86	0.80	0.89
Exon-MHC	Same Sample	9	0.88	0.83	0.91



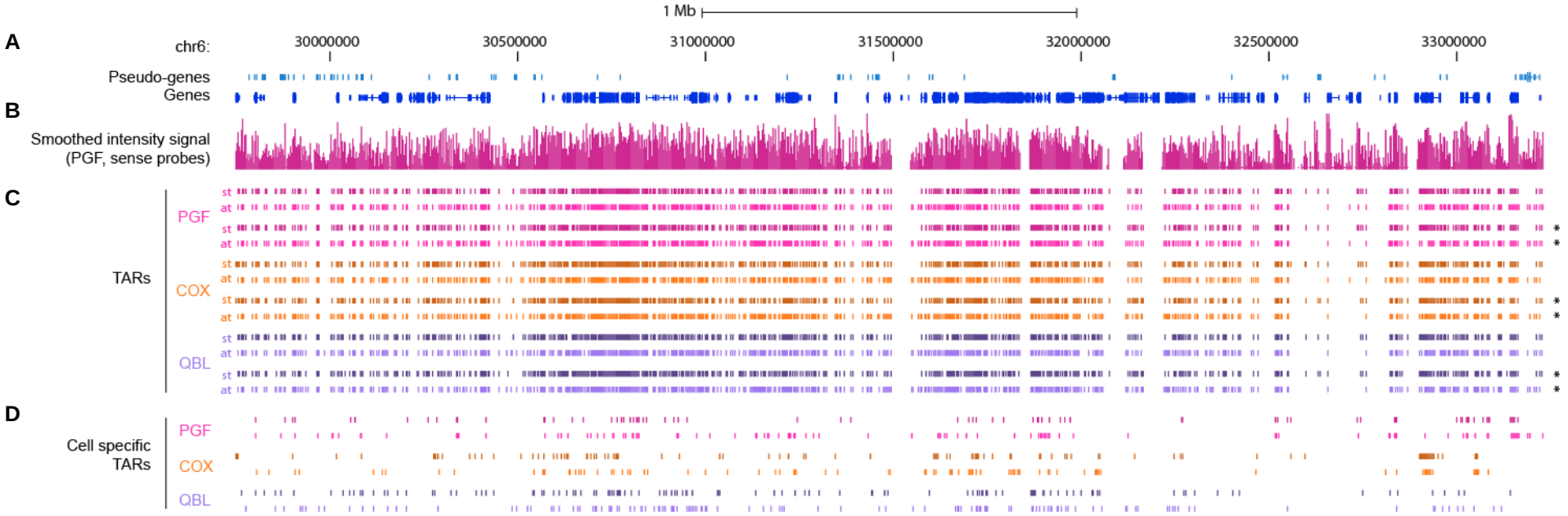
Supplemental Figure 2. Correlation of the haplotypic differences between the MHC array and the Affymetrix Exon 1.0 ST Array.

The fold changes between PGF-COX (top), PGF-QBL (middle) and COX-QBL (bottom) was computed using Exon 1.0 ST array data (x axis) and plotted against the same fold changes computed using MHC array data (y axis). Left column includes data for the 2,129 most varying probes (standard deviation greater 0.5 in both Exon 1.0 ST and MHC array datasets); middle column for 924 probes additionally having mean intensity >6 on Exon array; right column for 324 probes additionally having mean intensity >8 on the Exon array. The strong positive Pearson correlation indicates that similar results are obtained on both platforms, and increases when low intensity probes are removed.



Supplemental Figure 3. Heatmap of between-array pairwise Pearson correlation coefficients.

The color key for the correlation coefficient is given on the top left-hand corner. The dendograms illustrate the relationship between samples. The bars below the dendograms are colored in orange for COX samples, in purple for QBL samples and in pink for PGF samples. All biological replicates from each cell line are found clustered together. The names of the samples are displayed on each row and column. The suffixes stand for the culture condition (000=unstimulated; P06=stimulated), followed by one digit for the biological replicate number.



Supplemental Figure 4. Transcriptionally active regions across the MHC on the "shared path" relative to the human reference sequence.

(A) Vega genes (dark blue) and pseudo-genes (light blue). **(B)** Smoothed intensity signals from probe hybridization with single-strand ccDNA of PGF unstimulated cells on the forward strand. **(C)** Transcriptionally active regions (TAR) in unstimulated or stimulated (asterisk on the right-hand side) PGF (pink), COX (orange), and QBL (purple) in antisense (dark hue) or sense (light hue) orientation. TARs are defined by the presence of at least 3 probes above background per 51-base window whose median intensity is above thresholds corresponding to a false discovery rate of 1%. **(D)** Cell-specific TARs found only in one cell line (color code as in c). Number of cell-specific TARs for PGF, COX and QBL unstimulated cells are respectively: 419, 519 and 289 in antisense orientation, and 661, 544 and 270 in sense orientation.

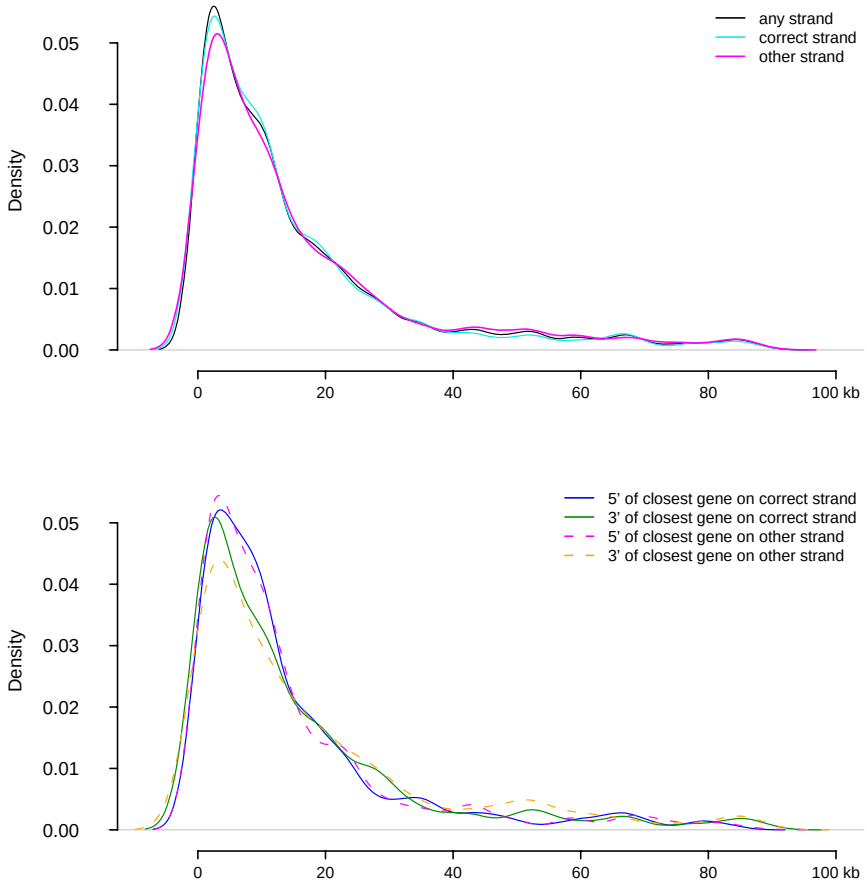
Supplemental Table 2. List of transcribed blocks on each haplotypic path

This supplemental element made of an excel workbook of 7 spreadsheets is provided to the manuscript as an independent zip file.

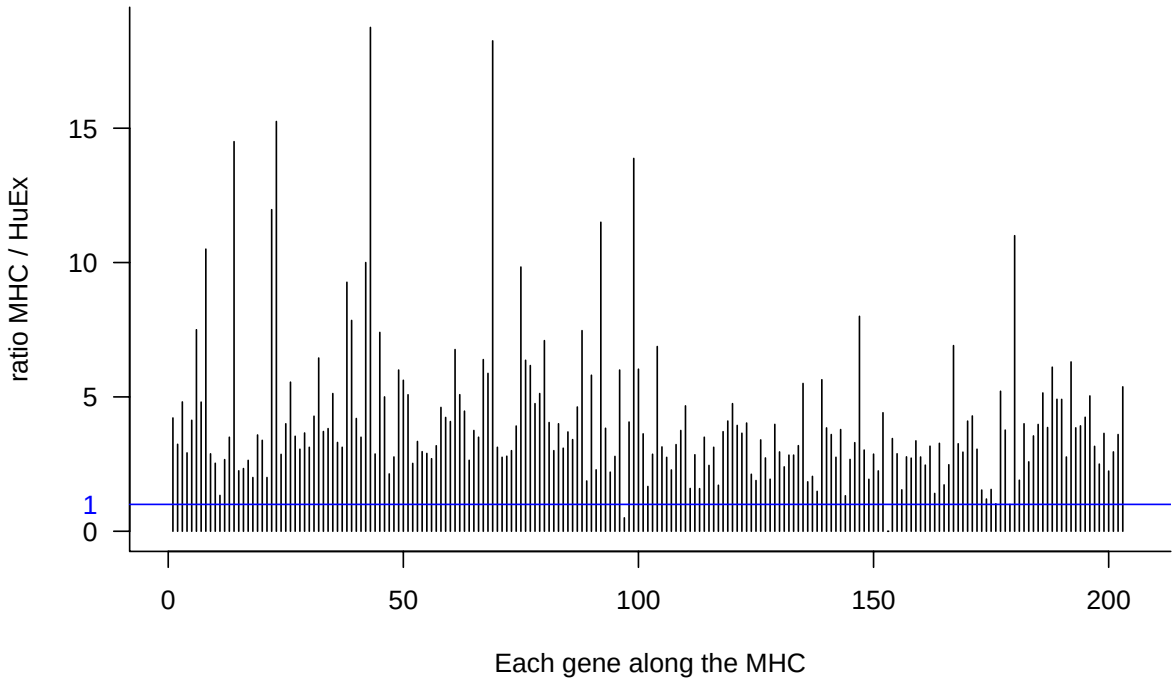
Supplemental Table 3. Statistics of gene annotations including at least one TAR at a FDR of 5%.

	PGF		COX		QBL	
	transcribed	untranscribed	transcribed	untranscribed	transcribed	untranscribed
Total genes and pseudogenes	197	29	192	36	187	27
Total genes	150	9	146	11	141	9
KNOWN_processed_transcript	11	0	11	0	10	0
KNOWN_protein_coding	122	4	118	6	117	3
NOVEL_processed_transcript	7	2	7	3	6	0
NOVEL_protein_coding	1	0	1	0	1	2
PUTATIVE_processed_transcript	9	3	9	2	7	4
Total pseudogenes	47	20	46	18	46	18
processed_pseudogene	17	8	17	8	17	8
transcribed_processed_pseudogene	0	0	1	0	1	0
transcribed_unprocessed_pseudogene	5	1	3	0	3	0
unprocessed_pseudogene	24	11	24	10	24	10
KNOWN_polymorphic_pseudogene	1	0	1	0	1	0

Note:
The selection for TARs using a FDR of 5% rather than 1% was adopted here as we were not aiming to discover new transcripts but to define expression of known and predicted transcripts, which allowed us to be less stringent on threshold inclusion.



Supplemental Figure 5. Distance of intergenic TARs from closest gene.



Supplemental Figure 6. Increased number of probes per gene in the MHC array compared with the Human Exon 1.0 ST array.

For each of the 204 genes of the MHC (x-axis, from telomere to centromere) present in the CDF libraries from the Microarray Lab (University of Michigan), we computed the ratio of the number of probes in our MHC array versus that in the Exon 1.0 ST array. The horizontal blue line indicates an equal number of probes.

Supplemental Table 4. Top 30 MHC genes showing significant differential expression between haplotypes after Benjamini-Hochberg adjustment using the Affymetrix Exon 1.0 ST array.

Pseudogenes are indicated by an asterisk.

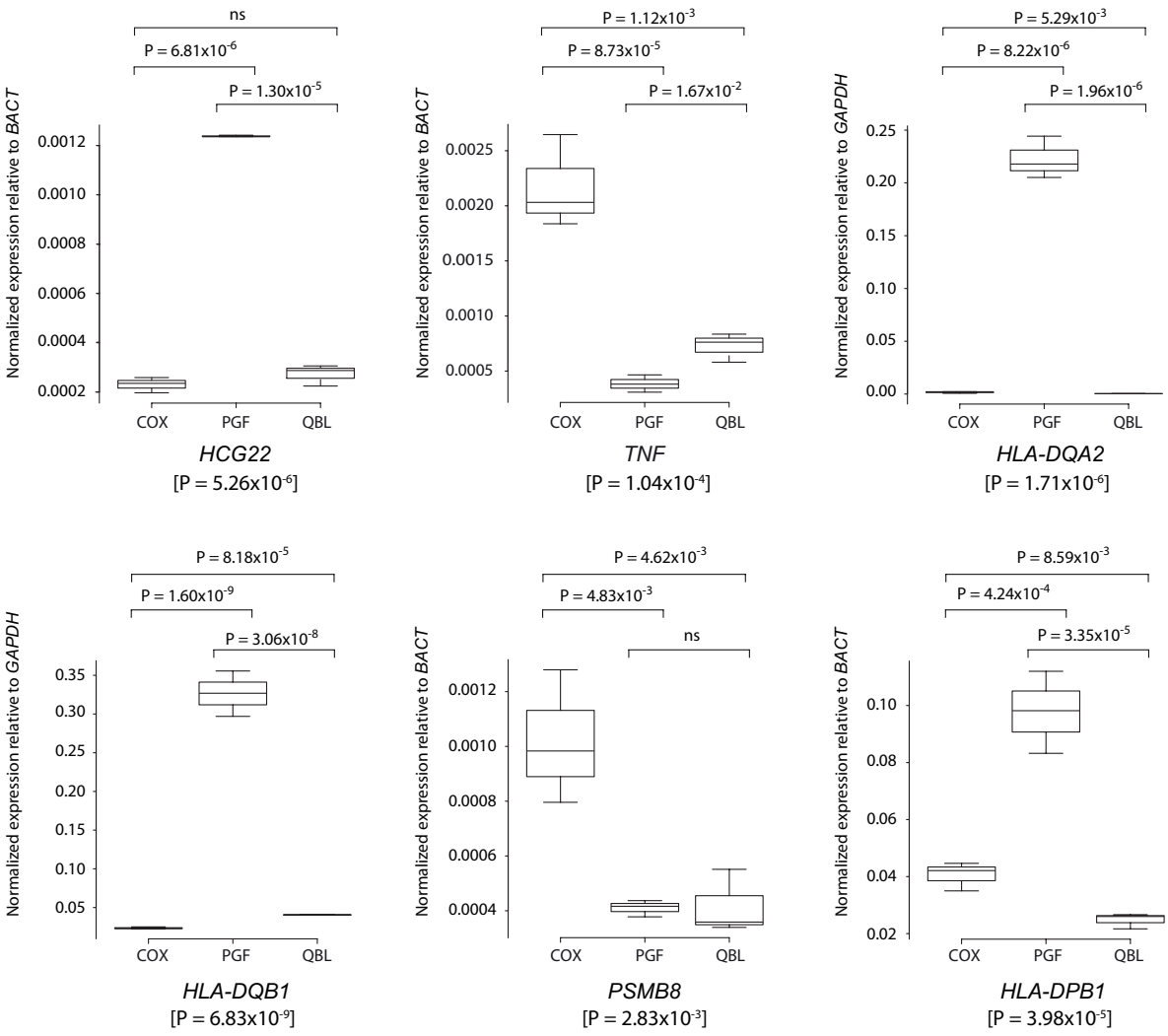
rank at genome scale [a]	rank with MHC array [b]	Gene Name [c]	Class [c]	Log2 (Fold Change) [c]			Adj.P.Val [c]
				COX vs PGF	QBL vs PGF	QBL vs COX	
13	NA	*HLA-DRB5*	II	-3.62	-3.19	0.43	$3.55x10^{-8}$
23	5	*HLA-U* *	I	-3.53	-0.91	2.61	$7.81x10^{-8}$
104	7	*HLA-DPB1*	II	-1.56	-0.88	0.68	$1.30x10^{-6}$
105	8	*RPL32P1* *	II	-1.64	-1.05	0.59	$1.30x10^{-6}$
144	4	*HLA-DQB2*	II	-1.77	-1.87	-0.10	$3.91x10^{-6}$
235	13	*HCG22*	I	-1.59	-1.59	0.00	$1.53x10^{-5}$
263	NA	*HLA-DOB*	II	-2.04	-1.16	0.88	$1.82x10^{-5}$
286	18	*HLA-DOA*	II	-1.26	-1.13	0.13	$1.82x10^{-5}$
331	54	*HLA-DRA*	II	-1.00	-1.15	0.15	$2.45x10^{-5}$
387	10	*HLA-A*	I	-0.32	-0.82	-0.50	$2.45x10^{-5}$
354	2	*HLA-DPB2* *	II	-1.00	-0.94	0.06	$2.57x10^{-5}$
372	40	*HLA-DRB1*	II	-1.81	-0.94	0.06	$3.59x10^{-5}$
471	25	*HLA-DMA*	II	-0.62	-0.83	-0.22	$4.90x10^{-5}$
589	80	*HLA-DMB*	II	-0.94	-0.37	0.57	$1.15x10^{-4}$
561	1	*ZFP57*	I	1.75	0.17	-1.58	$1.15x10^{-4}$
639	62	*IER3*	I	-0.75	-0.98	-0.23	$1.15x10^{-4}$
730	11	*HLA-L* *	I	-0.87	-1.14	-0.27	$1.16x10^{-4}$
923	29	*PSMB9*	II	0.47	-0.18	-0.65	$3.80x10^{-4}$
1094	28	*HLA-C*	I	0.02	-0.78	-0.79	$7.45x10^{-4}$
1185	55	*HLA-DPA1*	II	-0.57	-0.52	0.05	$7.77x10^{-4}$
1090	6	*TNF*	III	1.47	0.41	-1.06	$7.77x10^{-4}$
1149	3	*HLA-DQA2*	II	-1.06	-0.92	0.14	$8.68x10^{-4}$
1294	17	*HLA-F*	I	-0.08	-0.70	-0.62	$1.06x10^{-4}$
1346	NA	*FLOT1*	I	-0.47	-0.64	-0.17	$1.06x10^{-4}$
1254	15	*LTA*	III	0.92	0.05	-0.87	$1.06x10^{-4}$
1332	24	*CLIC1*	III	0.56	-0.10	-0.66	$1.06x10^{-4}$
1277	39	*AIF1*	III	-1.21	0.56	0.65	$1.10x10^{-4}$
1677	129 ns	*GPSM3*	III	0.02	-0.42	-0.44	$1.17x10^{-4}$
1526	42	*HLA-DQB1*	II	-0.98	-1.18	-0.20	$1.18x10^{-4}$
1617	19	*TAP1*	II	0.53	-0.38	-0.90	$1.19x10^{-4}$

Notes:

[a] rank obtained after comparing differential expression between unstimulated cell lines at the genome scale

[b] For comparison, this is the rank obtained after comparing differential expression between unstimulated samples hybridized on the MHC array. Twelve genes that had been found with a significant differential expression using the MHC array are not among the top 30 genes using the Exon array but were found as follows: *HLA-B* (43^{rd} with Adj.P = 0.015); *XXbac-BPG254F23.6* (61^{st}, with Adj.P = 0.038); *XXbac-BPG254F23.5* (No metaprobesets available); *NCR3* (not significant (n.s)); *LTB* (n.s); *LST1* (n.s); *DAQB-335A13.8* (n.s); *TCF19* (39^{th}, Adj.P = 0.009); *BRD2* (48^{th}, Adj.P = 0.017) ; *NRM* (64^{th}, Adj.P = 0.044); *HCG27* (n.s). NA: not applicable, concerns 3 genes that are not present on the three annotated haplotype sequences, and thus that were not considered when running the comparison between haplotypes with the MHC array.

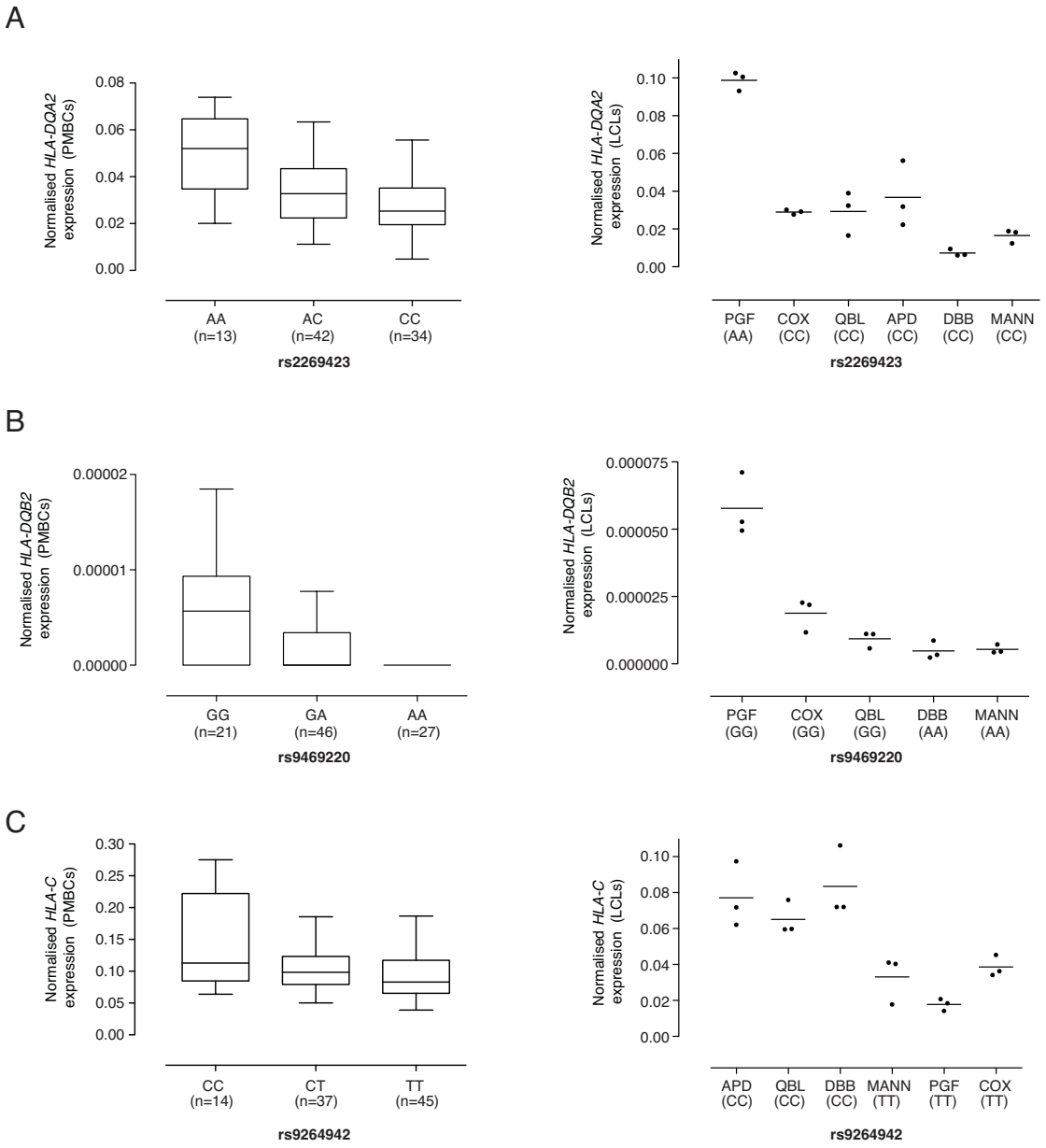
[c] results obtained comparing differential expression between unstimulated cell lines on the extracted MHC genes only.



Supplemental Figure 7. Validation of differential expression between COX, PGF and QBL samples. Three other genes were also tested, *NCR3*, *CLIC1* and *TCF19* but although there was a similar pattern of haplotypic expression as in the array, this was not significant.

Supplemental Table 5. Correlation between distribution of differentially expressed probes and polymorphic SNPs

This supplemental element made of an excel workbook of 9 spreadsheets is provided to the manuscript as an independent zip file.



Supplemental Figure 8. Validation data for differentially expressed genes.

(A) Expression of *HLA-DQA2* (relative to *ACTB*) determined by quantitative real time RT-PCR in peripheral blood mononuclear cells of 89 healthy volunteers plotted by genotype for rs2269423 (Kruskal-Wallis test on genotypes P=1.5x10-3) and for MHC-homozygous lymphoblastoid cell lines. **(B)** Expression of *HLA-DQB2* (relative to *ACTB*) determined by quantitative real time RT-PCR in peripheral blood mononuclear cells of 94 healthy volunteers plotted by genotype for rs9469220 in 94 healthy volunteers (Kruskal-Wallis test on genotypes $P=7x10^{-4}$) and for MHC-homozygous lymphoblastoid cell lines. **(C)** Expression of *HLA-C* (relative to *GAPDH*) determined by quantitative real time RT-PCR in peripheral blood mononuclear cells of 96 healthy volunteers plotted by genotype for rs9264942 (Kruskal-Wallis test on genotypes P=0.034) and for MHC-homozygous lymphoblastoid cell lines. Among the lymphoblastoid cell lines, genotypes that were unknown for rs2269423 (DBB cell line) and for rs9264942 (APD) were determined by direct sequencing.

Supplemental Table 6. Permutation results to evaluate the extent of alternative splicing between the MHC genes and non MHC genes.

For the mean and median SI values, we counted the number of sets with a higher value than in the tested set. For the different NI thresholds we counted how many of the 10,000 compared sets had more exons above the given threshold than in the tested set.

		MHC versus non-MHC			Non-MHC Immune versus Non-MHC Non-Immune			MHC versus non-MHC immune		
		Value for MHC set	Number of sets with a value > MHC set	*P*-value	Value for Immune set	Number of sets with a value > Immune set	*P*-value	Value for MHC set	Number of sets with a value > MHC set	*P*-value
Vega-based probesets	Mean NI	1.13	1	0.0001	1.00	4753	0.4753	1.13	0	<0.0001
	Median NI	0.88	3	0.0003	0.76	3148	0.3148	0.88	0	<0.0001
	NI >1	226619	40	0.0040	1131879	4567	0.4576	226736	28	0.0028
	NI>1.5	135367	78	0.0078	646512	5551	0.5551	135789	40	0.0040
	NI>2	76672	306	0.0306	360474	6098	0.6098	76958	188	0.0188
	NI>3	32409	8	0.0008	111275	5789	0.5789	32647	0	<0.0001
	NI>4	12188	43	0.0043	35556	3886	0.3886	12263	3	0.0003
Ensembl-based probesets	Mean NI	1.08	0	<0.0001	0.95	2877	0.29	1.08	0	<0.0001
	Median NI	0.79	228	0.0228	0.72	2711	0.27	0.79	164	0.0164
	NI >1	159168	86	0.0086	603940	1970	0.20	159316	137	0.0137
	NI>2	54684	52	0.0052	185148	1502	0.15	54735	68	0.0068
	NI>3	22443	0	<0.0001	45523	8888	0.89	22509	0	<0.0001
	NI>4	10421	0	<0.0001	14986	5794	0.58	10583	0	<0.0001

Supplemental Table 7. Variation of splicing between haplotypes.

Top 30 exons showing significant differential expression between haplotypes after Benjamini-Hochberg adjustment are listed. For each biological replicate of each cell line, the exon level intensity was normalized against the gene level intensity to generate the normalized intensity (NI). Intensities are expressed using the log2 scale. Gene and exon levels of genes shared by the three haplotypes were computed from the signal intensity of the probes matching uniquely and perfectly to their haplotype sequence. Pseudogenes are indicated by an asterisk.

Gene Name	**Exon** (Transcript.Exon ID)	**Class**	**Gene Intensity**			**Exon Intensity**			**NI**			**Adj.P.Val**
			PGF	**COX**	**QBL**	**PGF**	**COX**	**QBL**	**PGF**	**COX**	**QBL**	
HLA-C	012.2	I	13.15	13.20	14.26	14.35	14.69	9.84	1.20	1.49	-4.42	7.76×10^{-17}
HLA-G	004.4/001.5/006.3/002.6/005.4	I	10.21	9.83	9.81	14.63	10.19	8.46	4.42	0.36	-1.35	5.70×10^{-16}
*HLA-T**	001.2	I	8.43	7.85	7.95	9.49	15.03	11.29	1.06	7.18	3.34	2.77×10^{-14}
*HLA-DPB2**	001.4	II	11.02	7.83	8.00	12.49	14.55	14.88	1.47	6.72	6.88	2.26×10^{-13}
HLA-DQA2	001.4	I	9.81	7.36	8.19	13.97	15.27	15.54	4.16	7.91	7.35	3.14×10^{-13}
HLA-DPA1	001.5	II	11.03	10.42	10.53	15.98	10.54	15.55	4.95	0.12	5.02	4.74×10^{-13}
CYP21A2	008.1/001.1/009.1	III	8.31	8.10	8.21	12.60	6.77	9.76	4.29	-1.33	1.55	7.04×10^{-13}
*HLA-P**	001.3	I	7.35	7.26	7.36	14.21	10.16	10.15	6.86	2.90	2.79	4.41×10^{-12}
HLA-G	001.4/004.3/005.3/002.5	I	10.21	9.83	9.81	10.74	13.68	8.88	0.53	3.85	-0.93	4.41×10^{-12}
LSM2	004.1	III	11.45	11.47	11.45	11.11	6.71	11.23	-0.34	-4.76	-0.22	5.34×10^{-12}
HLA-B	004.4	I	13.50	13.44	12.31	10.57	10.09	12.88	-2.93	-3.35	0.57	5.93×10^{-12}
*HLA-DPB2**	001.3	II	11.02	7.83	8.00	14.43	13.76	14.03	3.41	5.93	6.03	1.06×10^{-11}
HLA-C	001.1	I	13.15	13.20	14.26	12.66	12.86	9.67	-0.49	-0.34	-4.59	1.25×10^{-11}
HLA-DQB2	004.4	II	12.12	9.38	9.54	14.65	7.77	10.26	2.53	-1.61	0.72	1.87×10^{-11}
EHMT2	003.19	III	9.86	10.00	9.96	11.33	11.57	9.12	1.47	1.57	-0.84	8.64×10^{-11}
*DDX39BP1**	001.2	I	6.88	7.38	7.27	10.71	14.68	14.68	3.83	7.30	7.41	1.34×10^{-10}
HLA-A	001.2	I	12.75	11.24	10.89	6.82	7.63	8.81	-5.93	-3.61	-2.08	1.91×10^{-10}
HLA-C	005.1	I	13.15	13.20	14.26	12.96	12.94	10.88	-0.19	-0.26	-3.38	2.16×10^{-10}
HLA-DQA2	001.3	II	9.81	7.36	8.19	15.24	14.84	15.20	5.43	7.48	7.01	7.88×10^{-10}
HLA-C	007.2	I	13.15	13.20	14.26	14.46	15.36	14.19	1.31	2.16	-0.07	8.83×10^{-10}
HLA-C	010.2	I	13.15	13.20	14.26	9.96	11.68	9.96	-3.19	-1.52	-4.30	1.40×10^{-09}
HLA-A	005.2/006.2/007.2	I	12.75	11.24	10.89	14.50	11.35	10.59	1.75	0.11	-0.30	1.56×10^{-09}
EHMT2	003.15	III	9.86	10.00	9.96	9.73	10.25	8.31	-0.13	0.25	-1.65	1.82×10^{-09}
EHMT2	008.23	III	9.86	10.00	9.96	12.11	12.25	10.96	2.25	2.25	1.00	2.02×10^{-09}
EHMT2	009.14	III	9.86	10.00	9.96	9.98	10.31	8.31	0.12	0.31	-1.65	2.09×10^{-09}
HLA-B	008.1	I	13.50	13.44	12.31	11.51	11.65	11.96	-1.99	-1.79	-0.35	2.18×10^{-09}
HLA-B	007.1	I	13.50	13.44	12.31	11.29	11.40	11.52	-2.21	-2.04	-0.79	5.02×10^{-09}
*HLA-DPA3**	001.1	II	8.11	7.72	8.18	9.20	11.28	9.54	1.09	3.56	1.36	7.60×10^{-09}
ZFP57	001.2	I	7.59	10.36	7.60	7.39	8.99	7.52	-0.20	-1.37	-0.08	8.88×10^{-09}
HLA-G	004.1/005.1/003.1/006.1	I	10.21	9.83	9.81	12.04	13.48	14.30	1.83	3.65	4.49	1.13×10^{-08}

Supplemental Table 8. Variation of junction intensities between haplotypes.

Genes showing a differential NI and having at least one junction showing differential expression between haplotypes are listed. The other significant junctions of these genes are also displayed, several being mutually exclusive (same acceptor or donor site, but not both). For each biological replicate of each cell line, the junction level intensity was normalized against the gene level intensity to generate the junction normalized gene intensity (JNI). Gene and junction levels of genes shared by the three haplotypes were computed from the signal intensity of the probes matching uniquely and perfectly to their haplotype sequence. Names of 5' donor exons and 3' acceptor exons are listed with reference to their transcript name (Transcript.ExonID). P values after Benjamini Hochberg adjustment are displayed. Overall, 31 genes had a junction showing differential expression between haplotypes, 27 of them also showed a differential exon NI. These are: *LST1, PPT2, AIF1, LTA, DDX39B, CLIC1, SLC44A4, XXbac-BPG296P20.15, EHMT2, C6orf25, DDAH2, CSNK2B, C6orf48, BAG6, PRRT1, DAQB-331I12.5, NOTCH4, ABHD16A, APOM, STK19, LSM2, TNF, PRRC2A, SKIV2L, EGFL8, NCR3* and *NFKBIL1* (ordered by the minimal Adj.P. Value).

Gene	Donor site	Acceptor site	5' exon^3' exon	JNI			Log2 Fold Change			Adj.P.Val
				PGF	COX	QBL	COX vs PGF	QBL vs PGF	QBL vs COX	
LST1	31663074	31663699	017.1^2, 011.2^3, 020.2^3, 004.2^3, 006.1^2	-1.91	-1.85	-0.50	0.06	1.41	1.34	$1.83x10^{-8}$
	31663074	31663396	013.2^3, 022.1^2, 001.2^3, 002.2^3, 018.2^3, 016.2^3, 005.2^3, 019.1^2, 023.1^2	-0.67	-1.36	-2.19	-0.69	-1.52	-0.83	$1.83x10^{-6}$
	31662872	31662955	016.1^2, 014.1^2, 020.1^2, 005.1^2	0.30	-0.26	0.72	-0.56	0.42	0.98	$5.16x10^{-6}$
	31663722	31664252	022.3^4, 006.2^3	0.72	0.82	0.06	0.10	-0.66	-0.76	$1.41x10^{-5}$
	31663074	31664273	014.2^3, 003.2^3, 015.1^2, 012.2^3	-0.01	0.10	0.76	0.11	0.77	0.66	$1.98x10^{-5}$
	31662862	31662955	018.1^2	-0.39	-0.60	0.18	-0.21	0.56	0.77	$1.32x10^{-4}$
	31663722	31664318	004.3^4, 019.3^4	2.74	3.15	2.41	0.41	-0.33	-0.74	$6.04x10^{-4}$
	31663074	31664318	008.2^3	-0.11	-0.13	0.66	-0.02	0.77	0.79	$6.11x10^{-4}$
	31663489	31664252	002.3^4	-2.17	-1.96	-2.53	0.21	-0.36	-0.57	$1.46x10^{-3}$
	31663489	31663699	022.2^3, 019.2^ 3, 016.3^4, 001.3^4	-0.43	-1.06	-1.29	-0.63	-0.86	-0.23	0.002

	31663489	31664273	005.3^4	-0.80	-0.55	-1.13	0.25	-0.33	-0.58	0.012
	31662071	31662955	011.1^2, 004.1^2, 012.1^2, 001.1^2, 010.1^2, 009.1^2, 008.1^2	-1.53	-1.37	-1.88	0.16	-0.35	-0.51	0.013
	31663722	31664273	011.3^4, 16.4^5, 021.2^3, 017.2^3, 001.4^5, 020.3^4	1.26	0.52	0.41	-0.74	-0.85	-0.11	0.015
	31663489	31663571	018.3^4	-1.36	-1.57	-1.75	-0.21	-0.39	-0.18	0.016
	31663489	31664318	023.2^3	0.16	-0.14	-0.40	-0.30	-0.56	-0.26	0.019
	31663078	31663306	010.2^3	-2.21	-2.13	-2.61	0.08	-0.40	-0.48	0.025
PPT2	32231538	32231625	001.4^5, 005.4^5, 010.4^5, 009.3^4, 003.4^5, 002.5^6, 007.5^6, 008.3^4, 006.4^5, 004.4^5, 011.3^4	1.06	1.32	-0.10	0.26	-1.16	-1.42	1.25×10^{-6}
	32233679	32238322	014.2^3, 004.7^8, 001.7^8, 003.7^8, 005.7^8, 006.7^8, 002.8^9, 016.2^3	1.17	1.34	0.45	0.17	-0.72	-0.89	5.16×10^{-6}
	32238377	32238561	004.8^9, 003.8^9, 001.8^9, 002.9^10	-0.35	0.10	-0.73	0.45	-0.38	-0.83	9.28×10^{-5}
	32231733	32233391	005.5^6, 001.5^6, 011.4^5, 010.5^6, 009.4^5, 003.5^6, 002.6^7, 008.4^5, 006.5^6, 004.5^6	0.53	0.69	0.07	0.16	-0.46	-0.62	0.0016
	32229391	32229753	002.1^2	-2.05	-2.14	-1.64	-0.09	0.41	0.50	0.0020
	32233475	32233594	014.1^2, 005.6^7, 001.6^7, 003.6^7, 002.7^8, 009.5^6, 006.6^7, 015.1^2, 016.1^2, 008.5^6, 004.6^7	-0.97	-0.70	-1.31	0.27	-0.34	-0.61	0.0025
	32229391	32230341	008.1^2, 001.1^2	-1.44	-1.79	-1.04	-0.35	0.40	0.75	0.0029
	32229391	32229950	007.1^2	-2.04	-2.10	-1.65	-0.06	0.39	0.45	0.029

	32230001	32230341	004.1^2	2.43	2.46	2.95	0.03	0.52	0.49	0.0047
	32247312	32247477	006.20^21	1.90	0.26	0.68	-1.64	-1.22	0.42	0.0061
	32245760	32246169	006.19^20	-1.91	-1.99	-1.55	-0.08	0.36	0.44	0.066
	32245256	32245662	006.18^19	-1.82	-2.13	-1.76	-0.31	0.06	0.37	0.021
AIF1	31692262	31692571	001.5^6, 004.5^6, 005.2^3, 006.3^4, 002.2^3	1.93	0.51	1.80	-1.42	-0.13	1.29	$2.89x10^{-6}$
	31691901	31692099	004.4^5, 001.4^5, 002.1^2	2.64	1.35	2.31	-1.29	-0.33	0.96	$6.37x10^{-4}$
	31691492	31692099	005.1^2	-2.51	-1.80	-2.27	0.71	0.24	-0.47	0.0021
	31691049	31691276	004.1^2, 003.1^2	-2.96	-2.33	-2.82	0.63	0.14	-0.49	0.0028
LTA	31648058	31648489	002.1^2, 003.1^2	-0.82	-2.22	-1.54	-1.40	-0.72	-0.68	$1.14x10^{-5}$
DDX39B	31617389	31616290	016.1^2	-2.97	-3.10	-1.84	-0.13	1.13	1.26	$1.45x10^{-5}$
	31616908	31616420	012.2^3	-3.62	-4.43	-3.75	-0.81	-0.13	0.68	$2.44x10^{-5}$
	31617151	31616420	018.1^2	-3.54	-4.16	-3.45	-0.62	0.09	0.71	$5.28x10^{-4}$
	31614902	31614695	013.3^4, 010.3^4	-1.00	-1.72	-1.72	-0.72	0.00	0.00	$7.78x10^{-4}$
	31617705	31616420	010.1^2, 003.3^4, 019.1^2, 001.1^2	0.85	-0.08	0.37	-0.93	-0.48	0.45	0.0012
	31606534	31606206	004.9^10, 001.10^11, 022.5^6, 002.10^11, 006.9^10, 003.12^13, 007.3^4	3.84	3.24	3.79	-0.60	-0.05	0.55	0.0013
	31617588	31616420	014.1^2	-3.16	-3.96	-3.45	-0.80	-0.28	0.52	0.0020
	31606808	31606705	006.8^9	-3.44	-4.20	-3.53	-0.76	-0.09	0.67	0.0022
	31617151	31616290	017.1^2	-3.52	-4.06	-3.40	-0.54	0.12	0.66	0.0025
	31621864	31617904	003.2^3	-1.07	-1.64	-1.08	-0.57	-0.01	0.56	0.011
	31610241	31608667	020.1^2	-2.87	-3.27	-2.80	-0.40	0.07	0.47	0.025
CLIC1	31812106	31815019	003.2^003.3	-4.41	-5.68	-5.11	-1.27	-0.70	0.57	$2.39x10^{-5}$
	31812106	31812551	002.1^002.2	-4.25	-5.73	-5.18	-1.48	-0.93	0.55	$3.32x10^{-5}$
	31809717	31809909	004.3^004.2, 003.4^003.5, 002.3^002.4, 001.2^001.3	3.23	2.44	2.78	-0.78	-0.45	0.33	$1.91x10^{-4}$
	31808154	31809321	004.5^004.6, 001.4^001.5, 003.6^003.7, 002.5^002.6	1.85	1.34	1.44	-0.51	-0.41	0.11	$1.72x10^{-3}$

	31806759	31807972	004.6^004.7, 002.6^002.7, 001.5^001.6, 003.7^003.8	1.80	1.10	1.48	-0.70	-0.32	0.38	$8.88x10^{-3}$
SLC44A4	31947317	31950215	004.7^004.8, 001.7^001.8	1.81	2.32	2.84	0.51	1.03	0.52	$2.44x10^{-5}$
	31945393	31946567	006.1^006.2	0.49	1.04	0.70	0.55	0.21	-0.34	0.030
	31950602	31950684	004.5^004.6, 001.5^001.6, 002.5^002.6, 003.5^003.6	1.48	1.95	1.30	0.47	-0.18	-0.65	0.04
XXbac-BPG296P20.15	31618152	31618464	002.1^002.2	0.46	0.65	-0.57	0.18	-1.03	-1.22	$3.72x10^{-5}$
EHMT2	31968318	31968449	001.6^001.7, 003.4^003.5, 010.2^010.3, 008.5^008.6, 009.5^009.6, 007.6^007.7	0.49	0.34	1.09	-0.15	0.60	0.75	$3.72x10^{-5}$
	31956020	31956428	007.26^007.27, 001.27^001.28, 005.8^005.9, 008.26^008.27, 009.26^009.26, 004.9^004.10, 003.25^003.26, 006.8^006.9	1.09	1.00	0.79	-0.10	-0.30	-0.21	0.047
C6orf25	31799742	31800501	004.2^004.3	1.26	1.54	-1.42	0.28	-2.69	-2.96	$9.28x10^{-5}$
			003.2^003.3	-4.60	-4.26	-4.73	0.35	-0.12	-0.47	0.046
DDAH2	31804280	31804401	006.2^006.3, 010.2^010.3, 001.3^001.4, 008.3^008.4, 002.2^002.3, 004.3^004.4, 007.3^007.4	-0.91	-1.69	-2.07	-0.79	-1.16	-0.37	$3.93x10^{-4}$
C6orf10	32407809	32411191	002.14^002.15, 001.15^001.16	0.60	0.03	0.00	-0.57	-0.60	-0.03	$4.32x10^{-4}$
CSNK2B	31744928	31745074	001.5^001.6, 007.5^007.6, 008.5^008.6, 004.5^004.6, 002.5^002.6, 009.5^009.6	2.13	1.71	2.29	-0.43	0.15	0.58	0.0012
	31742243	31742576	009.1^009.2	-3.00	-3.20	-2.59	-0.19	0.42	0.61	0.0043
	31744410	31744852	001.4^001.5, 007.4^007.5, 008.4^008.5, 004.4^004.5, 002.4^002.5, 009.4^009.5	3.46	3.17	3.54	-0.28	0.09	0.37	0.047
C6orf48	31913191	31915298	007.4^007.5, 001.3^001.4, 004.3^004.4, 008.2^008.3, 006.4^006.5, 009.1^009.2, 005.3^005.4, 002.2^002.3, 003.4^003.5	3.63	2.72	3.03	-0.91	-0.60	0.31	0.0015
	31910956	31911166	006.1^006.2, 003.1^003.2, 007.1^007.2, 001.1^001.2	0.77	-0.07	0.27	-0.84	-0.50	0.34	0.0030

	31911207	31912991	008.1^008.2, 004.2^004.5, 005.2^005.3, 001.2^001.3	1.21	0.34	0.69	-0.87	-0.52	0.35	0.0127
	31911207	31912054	006.2^006.3	-2.50	-3.04	-2.67	-0.54	-0.17	0.37	0.049
BAG6	31719950	31720062	004.12^004.13	0.46	0.43	-0.49	-0.03	-0.95	-0.92	0.0015
	31714982	31715255	031.1^031.2, 029.3^029.4, 002.24^002.25, 003.24^003.25, 001.24^001.25	1.86	1.46	1.31	-0.39	-0.55	-0.15	0.0022
	31714982	31715954	028.5^028.6, 030.3^0.30^4	0.52	0.89	1.10	0.37	0.58	0.21	0.0077
	31716062	31716400	0.30.2^030.3	-3.38	-3.32	-2.88	0.06	0.50	0.44	0.0147
PRRT1	32227930	32229491	005.2^005.3	2.56	2.01	1.95	-0.55	-0.62	-0.06	0.0030
	32229591	32229992	008.1^008.2, 012.1^012.2, 005.1^005.2	3.76	3.18	3.42	-0.59	-0.34	0.25	0.0041
	32225153	32225291	006.2^006.3, 001.5^005.2, 002.3^002.4, 003.2^003.3, 007.1^007.2	-3.39	-2.00	-3.12	1.39	0.27	-1.12	0.0145
DAQB-331I12.5	31959567	31959638	001.1^001.2	-0.01	-1.07	-0.14	-1.06	-0.13	0.93	0.0077
NOTCH4	32296039	32296159	001.6^001.7, 002.6^002.7	-1.27	-0.36	-1.08	0.91	0.19	-0.73	0.0088
	32272176	32272679	005.1^005.2, 003.8^003.9, 001.28^001.29	0.21	-0.14	-0.43	-0.35	-0.64	-0.29	0.0159
	32277255	32277830	001.21^001.22.	1.79	2.32	1.87	0.54	0.08	-0.46	0.039
	32278354	32279524	001.20^001.21	-1.86	-1.67	-1.16	0.19	0.70	0.51	0.046
ABHD16A	31767436	31767559	005.6^005.7, 004.6^004.7, 001.8^001.9, 013.2^013.3, 002.7^002.8, 006.1^006.2, 003.8^003.9	0.28	0.70	0.28	0.42	0.00	-0.42	0.014
	31767674	31768782	005.5^005.6, 004.5^004.5, 013.1^013.2, 002.6^002.7, 001.7^001.8, 003.7^003.8	0.56	1.23	0.83	0.67	0.27	-0.40	0.049
APOM	31728307	31732227	003.1^003.2, 002.1^002.2	-1.13	-1.57	-1.60	-0.44	-0.47	-0.03	0.015
STK19	32048675	32054658	010.1^010.2	-1.38	-1.62	-1.23	-0.25	0.14	0.39	0.017
	32048267	32048376	004.2^004.3, 007.1^007.2, 001.2^001.3, 005.1^005.2, 002.2^002.3	5.20	4.70	5.31	-0.49	0.11	0.60	0.031
	32054754	32055169	003.2^003.3, 010.2^010^3, 009.2^009.3, 001.4^001.5,	-1.83	-1.98	-1.55	-0.15	0.28	0.43	0.034

			008.2^008.3, 005.4^005.5							
LSM2	31881898	31882510	002.1^002.2, 001.2^001.3, 003.1^003.2	-0.79	-1.06	-0.60	-0.27	0.20	0.47	0.024
TNF	31652570	31652871	001.3^001.4	-0.65	-0.23	-0.30	0.42	0.35	-0.06	0.032
PRRC2A	31711505	31711722	002.24^002.25, 001.24^001.25	1.80	1.39	1.50	-0.41	-0.30	0.11	0.038
SKIV2L	32044294	32044378	010.1^010.2, 002.21^002.22	0.49	-0.01	0.36	-0.50	-0.13	0.37	0.046
EGFL8	32242375	32242455	001.3^001.4, 002.3^002.4, 003.1^003.2, 005.2^005.3, 004.3^004.4	-0.95	-1.05	-1.46	-0.51	-0.51	-0.41	0.047
NCR3	31665389	31665537	006.2^006.3, 001.2^001.3, 007.2^007.3, 002.2^002.3, 003.2^003.3, 004.3^004.4	1.38	1.83	1.74	0.45	0.36	-0.09	0.048
NFKBIL1	31622729	31623918	003.1^003.2	-1.35	-0.98	-0.96	0.37	0.39	0.02	0.049

3. SUPPLEMENTAL REFERENCES

Aho AV, Hopcroft JE, Ullman JD. 1983. *Data Structures and Algorithms*. Pearson Education

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H et al. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* **33**(20): e175.

Fairfax BP, Vannberg FO, Radhakrishnan J, Hakonarson H, Keating BJ, Hill AV, Knight JC. 2010. An integrated expression phenotype mapping approach defines common variants in LEP, ALOX15 and CAPNS1 associated with induction of IL-6. *Hum Mol Genet* **19**(4): 720-730.

Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**: S96-104.

Mei R, Hubbell E, Bekiranov S, Mittmann M, Christians FC, Shen MM, Lu G, Fang J, Liu WM, Ryder T et al. 2003. Probe selection for high-density oligonucleotide arrays. *Proc Natl Acad Sci U S A* **100**(20): 11237-11242.

Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A et al. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* **3**(7): 511-518.

Smyth GK. 2004. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology* **3**(1): art3.