

UK10K Genome Browser Functionality.....	2
1. Outline	2
2. Requirements.....	2
3. UK10K Genome Browser content.....	2
3.1. Genetic data.....	2
3.2. Phenotypic traits.....	2
3.3. Statistical tests	4
4. User interface overview	5
4.1. Top bar buttons	5
4.2. Managing Displayed Tracks.....	5
4.2.1. UK10K Track Hub	6
4.2.2. UCSC Track Hub.....	6
4.2.3. ENCODE Track Hub.....	6
4.2.4. NIH Roadmap Track Hub.....	6
4.2.5. BLUEPRINT Track Hub	6
4.2.6. Custom Tracks (Binary, DAS).....	6
4.3. Navigation Options	6
4.3.1. Zoom	7
4.3.2. Edit Settings	7
4.3.3. Track Settings.....	7
4.4. Visualisation options.....	7
4.4.1. Merge Tracks	7
4.4.2. Image Export.....	7
4.4.3. Feature Info Box.....	8
4.4.4. SNP Consequence Symbols.....	8
4.4.5. LD Calculation	8
4.5. Keys to navigation commands	10
4.6. Data access	10
4.7. References	10
UK10K Genome Browser introductory step-by-step tutorial.....	10

UK10K Genome Browser Functionality

1. Outline

The UK10K Genome Browser is based on the Biodalliance platform (Down et al., Bioinformatics 2011), designed to facilitate the retrieval of genotype-phenotype association results from the UK10K Cohorts Project (Walter et al, Nature 2015), and their visualisation in the context of different annotation features. In the following sections we describe the content and functionality of the UK10K Genome Browser. This includes navigation around the tracks, selecting tracks to display or adding new data. We also describe how to display LD information and producing high quality images from the current track view.

2. Requirements

The UK10K Genome Browser can be accessed using the latest versions of most major web browsers (Chrome, Firefox, Internet Explorer and Safari).

3. UK10K Genome Browser content

The UK10K Genome Browser was designed to allow exploration and retrieval of phenotype-genotype association results in the UK10K Cohorts dataset. Currently the browser supports the GRCh37/hg19 reference sequence.

3.1. Genetic data

The *cohort arm* of the UK10K study assessed the contribution of genetic variation genome wide to a range of quantitative traits in 3,781 healthy individuals from intensively studied cohorts of European ancestry, namely the Avon Longitudinal Study of Parents and Children (ALSPAC) and TwinsUK . Whole genome sequence (WGS) at 7x read depth was employed as a method maximising total variation detected whilst allowing access to noncoding variation where most GWAS signals lie. The final call set contains over 42M single nucleotide variants (SNVs) and ~3.5M INDELS (**Table 1**).

Table 1. Table of numbers of variants by frequency

Study name and design	N	Sequencing strategy and mean read depth	SNVs/ INDELS	SNVs/INDELS by allele frequency
Cohorts. Unselected samples from two population-based cohorts.	3,781	WGS, 7x Ts/Tv=2.15	42,001,210/ 3,490,825	<1%: 34,247,969/2,296,962 1-5%: 2,298,220/412,168 >5%: 5,869,317/1,496,955

3.2. Phenotypic traits

The study considered a total of 64 different phenotypes, including traits of primary clinical relevance in 11 major phenotypic groups (obesity, diabetes, cardiovascular and blood biochemistry, blood pressure, dynamic measurements, birth, heart, lung, liver and renal function; **Table 2**). Of these, 31 phenotypes are available in both studies, 18 are unique to TwinsUK and 15 are unique to ALSPAC.

Table 2. List of phenotypic traits in the UK10K Cohorts dataset, with number of participants with available whole genome sequence and phenotypic data

Category	Name	TwinsUK	ALSPAC	Total
Obesity/anthropometry	Body mass index (BMI)	1,747	1,791	3,538
	Height (HT)	1,747	1,794	3,541
	Weight (WT)	1,747	1,812	3,559
	Hip circumference (HIP)*	1,266	1,808	3,074
	Waist circumference (WST)*	1,265	1,807	3,072
	Waist hip ratio (WHR)*	1,265	1,806	3,071
	Total fat mass (TFM)	1,716	1,683	3,399
	Total lean mass (TLM)	1,716	1,683	3,399
	Trunk fat mass (TRFM)	1,514	1,683	3,197
	Forearm length (FAL)	-	1,760	1,760
	Head circumference (HCRF)	-	1,762	1,762
	Leg length (LL)	-	1,764	1,764
	Sitting height (SHT)	-	1,764	1,764
	Upperarm length (UAL)	-	1,762	1,762
	Diabetes Biochemistry	Adiponectin (ADIPO)	864	1,461
Leptin (LEPTN)		958	1,459	2,417
Glucose (GLU)*		1,701	1,224	2,925
HOMA-B (HOMA-B)*		1,669	1,219	2,888
HOMA-IR (HOMA-IR)*		1,577	1,219	2,796
Heart function	Insulin (INS)*	1,676	1,220	2,896
	Heart rate (HRT)	1,385	1,590	2,975
CVD hypertension	Diastolic blood pressure (DBP)	1,536	1,773	3,309
	Systolic blood pressure (SBP)	1,536	1,773	3,309
CVD Biochemistry	High density lipoprotein cholesterol (HDL)	1,713	1,497	3,210
	Low density lipoprotein cholesterol (LDL)	1,696	1,495	3,191
	Total cholesterol (TC)	1,711	1,495	3,206
	Triglycerides (TG)	1,705	1,497	3,202
	Very low density lipoprotein (VLDL)	1,700	1,497	3,197
	Apolipoprotein A1 (ApoA1)	1,449	1,465	2,914
	Apolipoprotein B (ApoB)	1,443	1,468	2,911
	Homocysteine (HCY)	1,279	-	1,372
Blood Biochemistry	HsCRP (CRP)	879	1,167	2,046
	Hemoglobin (HGB)	1,553	1,524	3,077
	Mean corpuscular hemoglobin (MCH)	1,549	-	1,549
	Mean corpuscular hemoglobin concentration (MCHC)	942	-	942
	Mean corpuscular volume (MCV)	1,548	-	1,548
	Packed cell volume (PCV)	1,555	-	1,555
	Platelet counts (PLT)	1,553	-	1,553
	Red blood cell counts (RBC)	1,561	-	1,561
Liver Function	White blood cell counts (WBC)	1,551	-	1,551
	Interleukin 6 (IL6)	-	1,480	1,480
	Albumin (ALB)	1,713	-	1,713
	Alkaline phosphatase (ALP)	1,702	-	1,702
Renal Function	Bilirubin (BIL)	1,702	-	1,702
	Gamma glutamyl transpeptidase (GGT)	1,699	-	1,699
	Bicarbonate (BIC)	1,714	-	1,714
	Creatinine (CRT)	1,707	-	1,707
	Phosphate (PHPT)	1,392	-	1,392
	Sodium (SOD)	1,683	-	1,683
	Urea (UR)	1,697	-	1,697
Lung Function	Uric acid (UA)	1,305	-	1,305
	FEV/FVC ratio (FEV1-FVC)	1,676	1,604	3,280
	Forced Vital Capacity (FVC)	1,679	1,606	3,285
Birth	Forced Expiratory Volume (FEV1)	1,681	1,606	3,287
	Birth weight (BWT)	-	1,691	1,691
	Birth length (BL)	-	1,137	1,137
	Gestational age (GA)	-	1,712	1,712
	Ponderal index (PI)	-	1,122	1,122
Dynamic	Placental weight (PLWT)	-	703	703
	Grip strength (GRP)	1,514	1,682	3,196
	Ever broken bone**	-	1,756	1,756
	Eye preference**	-	1,671	1,671
	Handedness tasks**	-	1,700	1,700
	Handedness drawing**	-	1,676	1,676

*Browser also contains BMI-adjusted data for these traits

** These binary traits are not reported in the UK10K Genome Browser

3.3. Statistical tests

Different testing strategies were employed for analysis (**Table 3**), which generated overall 1,460 million individual association statistics.

Single-variant tests. We fitted linear models on standardised traits, residualised for relevant covariates, to test associations of allele dosages with 13,074,236 SNVs and 1,122,542 biallelic INDELS (MAF \geq 0.1%) and 18,739 large deletions in whole-genome sequenced samples.

Rare variant tests. For each of the four scenarios below, we applied two separate statistical models with different properties to rare variants (MAF $<$ 1%): sequence kernel association tests (SKAT) and burden tests implemented in SKAT and SKAT-O^{9,10}. SKAT is a variance-component multiple regression test which retains power in settings where neutral variants or variants with opposite direction of effects could result in loss of power. SKAT-O represents the best linear combination of SKAT and burden tests. Analyses were either exome-wide, or genome-wide, as detailed below.

For *exome-wide* rare variants tests, we focused on variants in coding exons and untranslated regions (UTRs), with three different variant aggregation strategies.

1. *Loss-of-function* tests: including loss-of-function (LoF) variants, defined as predicted to cause essential splice site donor or acceptor changes, stop codon gains and frameshift mutations
2. *Functional* tests: including missense and LoF variants.
3. *Naïve* tests: all variants in exons, UTRs and essential splice sites, where all variants are given equal weight of being causal.

For *genome-wide* analyses, we partitioned the genome into 3kb half-overlapping tiling windows with an average of 37 variants per window. In total, for each trait, we generated SKAT and SKAT-O p-values and their corresponding MetaSKAT p-values for more than 1.8 million windows across the genome.

For 31 overlapping traits ('core' traits) with phenotype data in both ALSPAC and TwinsUK, associations were run independently in the two cohorts and meta-analysed using inverse variance models (single-variant tests), or directly meta-analysed using MetaSKAT for rare variant tests. For these 31 traits, in the browser we report the association p-values (-log₁₀ scale) for the meta-analysis of the two studies. For traits with data available in only ALSPAC or TwinsUK, we report association statistics for that study.

Table 3. Summary of statistical tests implemented, and convention for naming UK10K Cohorts results tracks

Track name	Test	Variant selection strategy	N windows (N variants) tested
Single-variant association tests			
single_variant	Linear regression	Single-variant tests for MAF $>$ 0.1%	14,196,778 variants (median = 13,933,511, range = 13,450,148-15,336,389)
Exome-wide association tests			
EW_LoF_skat	MetaSKAT	Functional exome-wide rare variant SKAT tests for MAF $<$ 1% (loss-of-function)	3,208 (9,113)
EW_LoF_skat-o	MetaSKAT-O	Functional exome-wide rare variant SKAT-O tests for MAF $<$ 1% (loss-of-function)	3,208 (9,113)
EW_functional_skat	MetaSKAT	Functional exome-wide rare variant SKAT tests for MAF $<$ 1% (loss-of-function and missense)	14,909 (256,733)
EW_functional_skat-o	MetaSKAT-O	Functional exome-wide rare variant SKAT-O tests for MAF $<$ 1% (loss-of-function and missense)	14,909 (256,733)
EW_naive_skat	MetaSKAT	Naïve exome-wide rare variant SKAT tests for MAF $<$ 1%	50,717 (1,783,548)
EW_naive_skat-o	MetaSKAT-O	Naïve exome-wide rare variant SKAT-O tests for MAF $<$ 1%	50,717 (1,783,548)
Genome-wide association tests			
GW_skat	MetaSKAT	Genome-wide rare variant SKAT tests for MAF $<$ 1%	1,845,982 (35,858,684)
GW_skat-o	MetaSKAT-O	Genome-wide rare variant SKAT-O tests for MAF $<$ 1%	1,845,982 (35,858,684)

4. User interface overview

The user interface (**Figure 1**) can generally be divided into three regions, the top bar, the main display area and a sidebar area (which collapses while not in use).

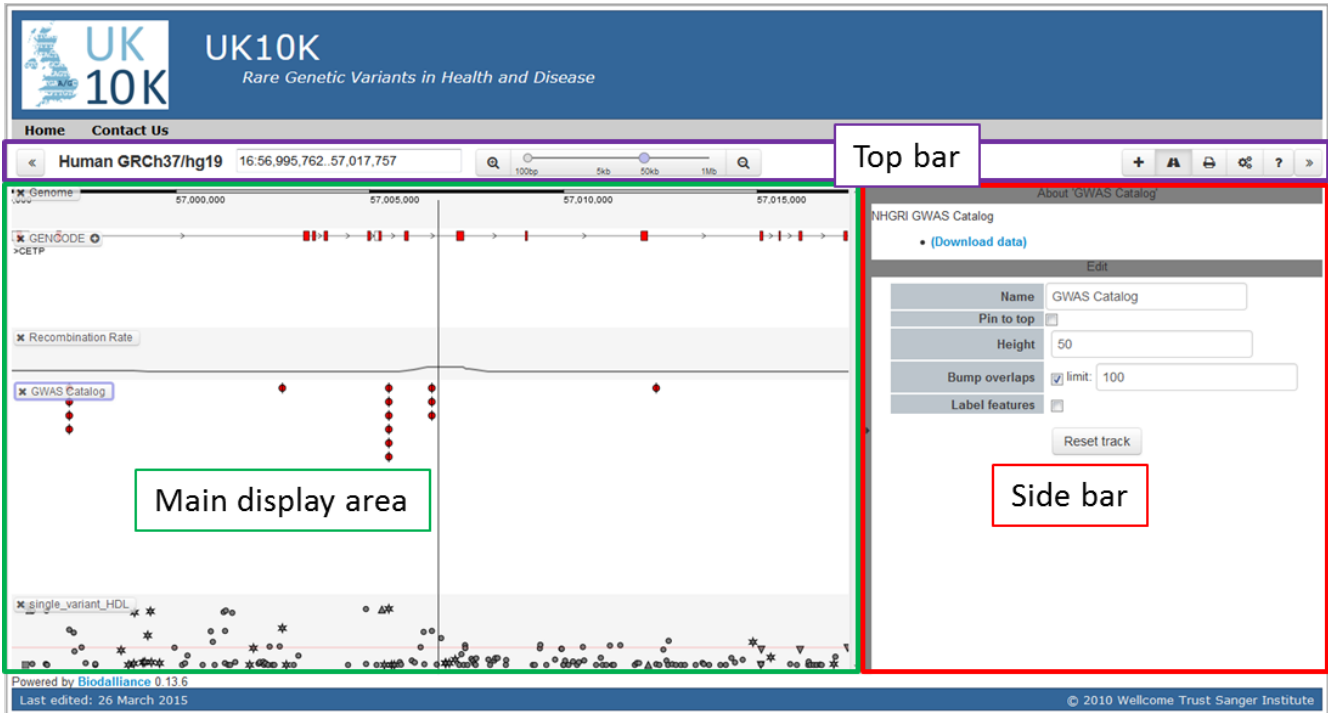


Figure 1. The user interface overview, showing how the browser is divided into three main sections

The top bar fulfils two functionalities. First, it provides functionality for navigating around the track view. Second, on the right side it holds buttons for accessing a number of additional sidebars. The different sidebars are activated by clicking on the respective buttons and are used for showing additional information, editing settings or managing the displayed data. The main display area is used for visualizing the genome track data.

4.1. Top bar buttons

The top bar contains a numerous buttons that allow the user to navigate the browser, alter the display and add data tracks (**Figure 2**).

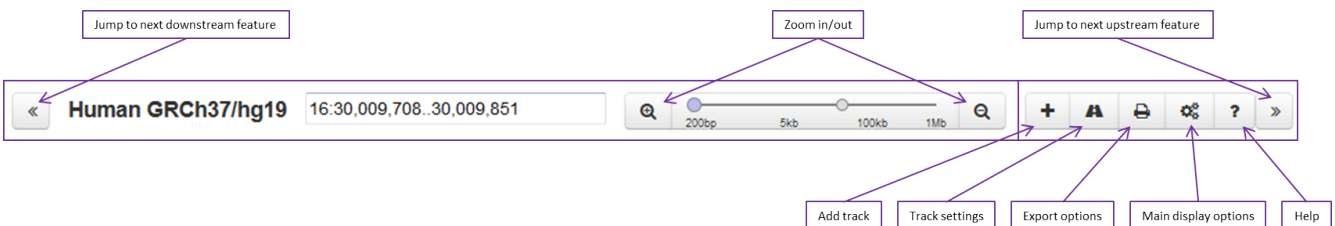


Figure 2. The top bar showing the functions of the various buttons and the genome build currently used.

4.2. Managing Displayed Tracks

The 'Add track' sidebar holds various options to show/hide, add/remove tracks to the track hub. In general, available data tracks are organized in tab categories. A chosen track from a tab category is displayed or hidden according to the checkbox in front of the track name.

4.2.1. UK10K Track Hub

There are nine track hubs representing each of the nine categories of association tests summarised in **Table 3**. Each hub contains association results for that particular test and for all 60 (out of 64) traits presented (meta-analysis or individual cohort depending on trait, see **Table 2**).

4.2.2. UCSC Track Hub

The UK10K Genome Browser has support for connecting to UCSC browser track hubs. Select the +tab in the 'add track' sidebar, specify the URL of the "hub.txt" file and click the add-button.

4.2.3. ENCODE Track Hub

Tracks derived from the ENCODE project data can be added using this tab. Categories include DNase-sequencing, FAIRE-sequencing, histone modification, RNA-sequencing, transcription factors and genome segmentations (from http://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration_data_jan2011/hub.txt).

4.2.4. NIH Roadmap Track Hub

Tracks derived from the NIH Roadmap Project can be added using this tab. The complete epigenome atlas hosted by Washington University in St. Louis is available (from <http://vizhub.wustl.edu/VizHub/RoadmapReleaseAll.txt>).

4.2.5. BLUEPRINT Track Hub

Tracks derived from the BLUEPRINT project data can be added using this tab (data from ftp://ftp.ebi.ac.uk/pub/databases/blueprint/releases/current_release/homo_sapiens/hub/hub.txt).

4.2.6. Custom Tracks (Binary, DAS)

Genomic data in a range of supported formats (including bigWig, bigBed, plain bed and wig, BAM and VCF) can be added to the browser via the "Binary"-tab. Data may be stored by the user on another webserver or on the local hard-drive. Data created using the DAS protocol (<http://biodas.org/>) are also supported and can be added using the 'DAS' tab.

4.3. Navigation Options

The user is provided with various methods for navigating through the association results data tracks. See **Figure 4**. LD (r^2) scoring colour scheme when multiple (two to five allowed) SNVs are selected as references ('make ref' option for SNV group 1, and 'add ref' buttons for groups 2-5). Solid colours represent a score of one, with the SNV becoming gradually more transparent as the LD score approaches 0.2. SNVs with low linkage disequilibrium ($r^2 < 0.2$) are displayed in grey.

Keys to navigation commands for a full list of navigation shortcuts.

1. *Direct Navigation.* The top bar of the browser view holds a text field for direct navigation. Initially, the text field shows the displayed region of the genome. To navigate to a different region, the user can type a genome region (e.g. "16:29,993,244..30,023,244") or genome position (e.g. "16:29,993,244") and navigate there by pressing [Enter]. It is also possible to search for rsIDs (e.g. type "rs10") or genes (e.g. type "CETP") in the same way.
2. *Scroll Navigation.* To scroll navigate through the track data, the user can either use the keyboard arrow keys, horizontal track pad scroll or drag the view with a mouse click. For convenience, in navigation directions can be inverted in the browser settings (main display options).
3. *Threshold-leap Navigation.* The UK10K Genome Browser allows to search a quantitative track for features scoring above a given threshold value. Search in upwards or downwards direction is initiated by clicking the arrow buttons in the top bar (or press Ctrl+LeftArrow / Ctrl+RightArrow). The threshold value can be adjusted via the track settings sidebar.

4.3.1. Zoom

It is possible to increase or decrease the size of the displayed genomic region. To zoom in or out the user can either use the zoom slider in the top bar, or press the + / - buttons. In addition, a base resolution zoom can be toggled by pressing the [Space] bar. To zoom back out again, press the [Space] bar a second time.

4.3.2. Edit Settings

The UK10K Genome Browser allows for minor configuration to respond to user inputs via the configuration sidebar. Also, the track editing sidebar allows for adjusting track settings individually for each track. The 'main display options' sidebar holds options to invert scrolling directions. The user can also choose the horizontal position of the vertical score guideline or turn it off completely. The Reset button can be used to revert the settings to default.

4.3.3. Track Settings

Each displayed track in the track view can be configured to some extent by using the 'Track settings' sidebar. The sidebar holds options to change the track name, color, height or style (histogram, line plot, ribbon, scatter). Furthermore, the displayed horizontal value range can be adjusted and a threshold value for threshold navigation can be specified. Changes made to the tracks are applied instantly; there is no need to reset the track before altering any settings.

There also is an option to 'pin to top' that will move the selected track to the uppermost position in the main display. Using the 'reset track' button will cause the selected tracks' settings to revert to the defaults, but it will not however revert to its original position in the main display, as this may not exist anymore if other tracks have been moved.

4.4. Visualisation options

4.4.1. Merge Tracks

For visualisation purposes, multiple tracks can be merged/overlaid into a new single track. To do so, first select the tracks to be merged (hold Shift to select multiple tracks), then press "ctrl + m". A new track will appear (below the existing selected tracks) showing an overlay of the selected tracks.

4.4.2. Image Export

The UK10K Genome Browser allows the user to export a track view as a scalable vector graphic (SVG) file. To create and export the SVG file showing the current track view, select the 'Export options' button [printer icon],

which toggles the image export sidebar and click the export button. A download link to the created SVG file will appear as well a preview link.

4.4.3. Feature Info Box

When clicking on a track feature a feature info box with additional information about the clicked feature pops up. For UK10K single point trait tracks this box holds fields like rsID, effect allele, effect allele frequency, VQSLOD score and indicators if the SNP is discovered in HapMap or 1000Genomes.

4.4.4. SNP Consequence Symbols

SNP consequences are indicated by the symbol displayed for that SNP. By default, SNPs are displayed as a circle, whereas squares indicate an untranslated region SNP, upwards pointing triangles indicate a splice region SNP, downwards pointing triangles indicate a missense region SNP and stars indicate a regulatory region SNP.

4.4.5. LD Calculation

The UK10K Genome Browser supports display of linkage disequilibrium (LD) statistics r^2 , which is calculated on the fly from UK10K Cohorts whole-genome sequencing data. LD can be referenced to one, or multiple, reference SNVs of choice. To display LD scores with respect to one chosen reference SNV, click on the SNV symbol to open the SNV info box, then click the “Make ref.”-button. Once the LD scores have been calculated, SNV are coloured as illustrated in **Figure 3**.

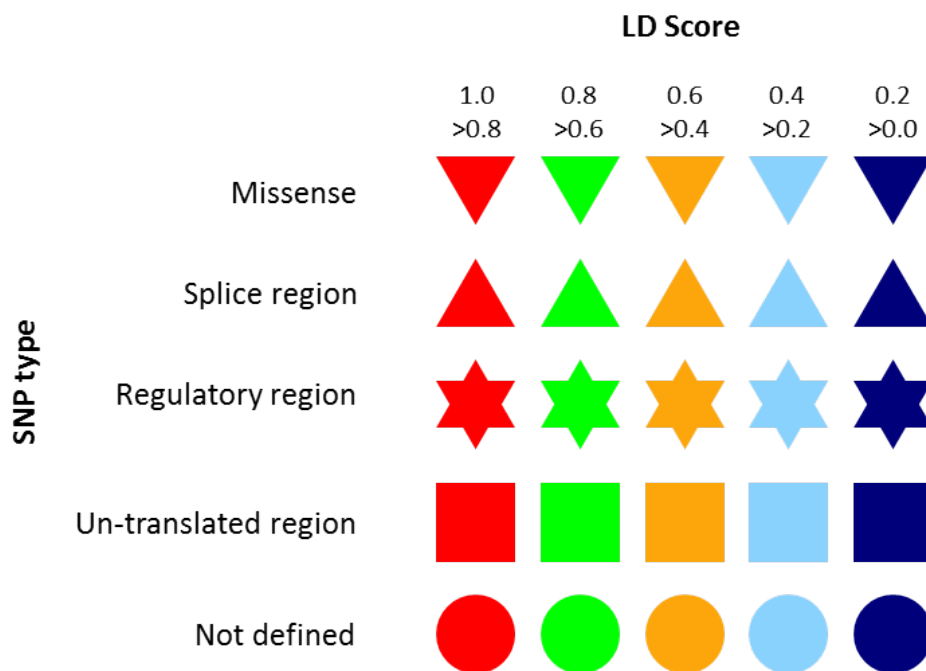


Figure 3. LD scoring colour scheme for the case where a single SNV is made the reference (‘Make ref.’ option). The selected SNV will be solid red, and all others SNVs are coloured by LD score groups relative to that SNV; red ($r^2 > 0.8$), yellow ($r^2 = 0.6-0.8$), green ($r^2 = 0.4-0.6$), light blue ($r^2 = 0.2-0.4$) and dark blue (< 0.2).

Multiple reference SNVs (maximum of 5) can be created by making the first SNV reference as explained previously, and then by selecting and adding extra SNVs via the “Add ref.” button-. LD scores for multiple reference SNVs are illustrated in **Figure 4**.

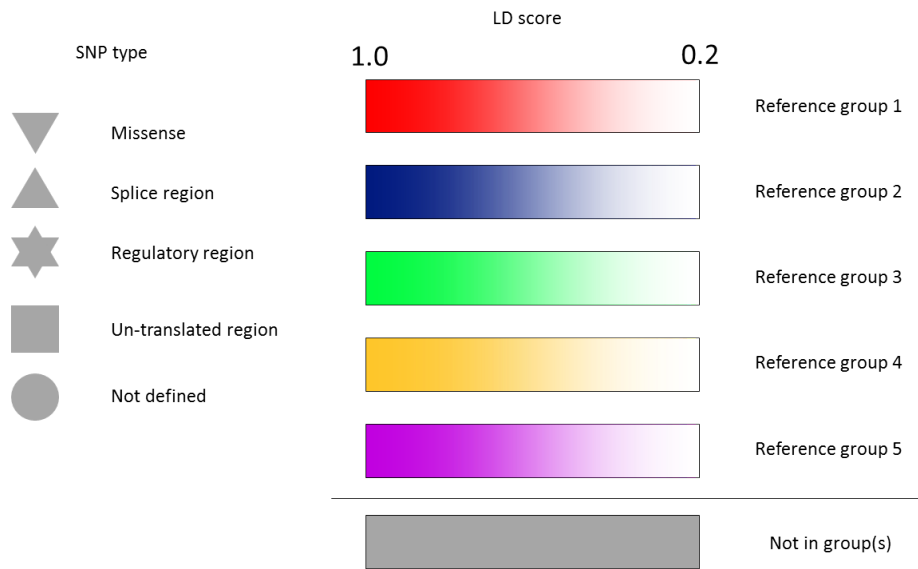


Figure 4. LD (r^2) scoring colour scheme when multiple (two to five allowed) SNVs are selected as references ('make ref' option for SNV group 1, and 'add ref' buttons for groups 2-5). Solid colours represent a score of one, with the SNV becoming gradually more transparent as the LD score approaches 0.2. SNVs with low linkage disequilibrium ($r^2 < 0.2$) are displayed in grey.

4.5. Keys to navigation commands

The key navigation commands are summarised in **Table 4**. Please note that the mouse must be within the browser window to navigate; click on track label if the position has been lost.

Table 4. Navigation shortcuts for UK10K Genome Browser

Global navigation	
← →	Move left right (decrease increase genomic coordinates)
ctrl ← ctrl → (doesn't work for all data sources)	Next feature left right (above threshold) in selected track
Space bar	Toggle zoom to base pair coordinates and back
Track Navigation (Highlighted track label is selected track)	
↑ / w ↓ / s	Select track above below
shift ↑ shift ↓	Expand contract selected track
t shift t	For tracks with ⊕ toggle expansion of selected all tracks (genes ↔ transcripts)
ctrl+m	Merge selected tracks
shift+left mouse click	Select multiple tracks
Information	
I	Toggle description of selected track
H	Toggle help panel
U	Toggle ruler location indicator

4.6. Data access

The flat files BigBed files for the traits can be downloaded from the 'Track settings' by selecting the desired track and clicking on 'Download data'. Raw data (genotypes and phenotypes) used to generate summary statistics can be accessed by application to the UK10K Project data access committee http://www.uk10k.org/data_access.html.

4.7. References

For use of the UK10K Cohorts Genome Browser, please cite: 'An interactive genome browser of association results from the UK10K cohorts project'. Geihs et al. *Bioinformatics* (2015)

For use of the UK10K sequence data, please cite: 'The UK10K Project: Rare variants in health or disease'. Walter et al. *Nature* (2015)

UK10K Genome Browser introductory step-by-step tutorial

This short guide will allow you to re-create an approximation of **Figure 1** from the UK10K Genome Browser manuscript (Geihs et al, Bioinformatics 2015) as a way of demonstrating some of its functionality. Although this walkthrough can be used in isolation, it is strongly recommended that you read the user documentation in the previous section.

1. Load the browser interface (<http://www.uk10k.org/dalliance.html>)
 - Note that if you have cookies enabled the default view will be replaced with the last view settings used. To reset go to 'Main display options' and click the 'reset browser' button.
2. Re-order the tracks by selecting them one at a time (click on the track name, it will then have a highlighted border) and dragging into position (from top to bottom); GWAS Catalog, single_variant_HDL, GENCODE, Genome.
3. Remove the 'Recombination' track by clicking on the 'x' in the track name displayed in the main display section.
4. Find the *CETP* gene 30kb region by typing 'CETP' into the chromosome co-ordinate search box next to the 'Human GRCh37/hg19' label and pressing enter.
5. You now have the following view. To collapse the GWAS catalog track select the 'GWAS catalogue' track, go to the 'track settings' sidebar and uncheck 'Bump overlaps'. Note that the changes are applied instantly, there is no need to reset the track for them to take place.

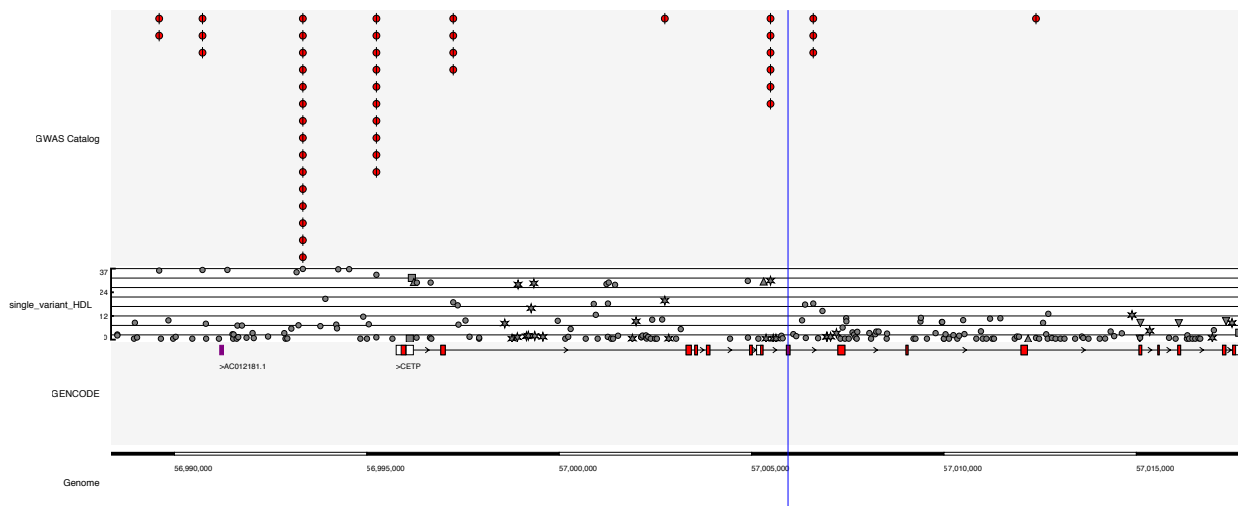


Figure A. Display of 30kb region surrounding the *CETP* gene and displaying (top to bottom). (i) GWAS Catalog track. In its native configuration, it shows all the GWAS catalog hits as a bead on a string, where each bead is a separate study, ordered by publication date. Clicking on a bead will display information about the study and SNV, when collapsed only the oldest study details will be provided. (ii) Single-variant association statistics for HDL cholesterol, where each symbol represents a separate SNV. The y-axis scale is proportional to the $-\log_{10}(p\text{-value})$ for association. (iii) GENCODE gene annotation track. (iv) Genome coordinates (hg19).

6. To move the vertical guideline, go to the 'main display options' sidebar and change the vertical guideline to 'none' in the dropdown menu. The blue line should disappear and the HDL cholesterol y-axis will move to the left edge.
7. To make the significant HDL association SNVs more agreeable to view, select the track and go to the track options then increase the 'max.value' (the p.value) to 45.
8. In order to colour the SNVs as in the manuscript they need to be made into references, and in this case we are adding more than one ref SNV so they are grouped by colour (see the legend description in the user documentation and **Figures B and C**). In the search box enter 'rs72786786' and press enter.
9. That SNV is now centred on the screen, clicking on the SNV in the HDL cholesterol track will bring up the descriptives and options for the SNV, do this and click 'Make ref.' All the SNVs will now have changed colour based on their LD score relative to the ref. SNV (see **Figure B** for a similar example).

10. Next find SNV rs9939224 and this time click 'Add ref.' Now there are two colors groups, red for the first ref and blue for the second.

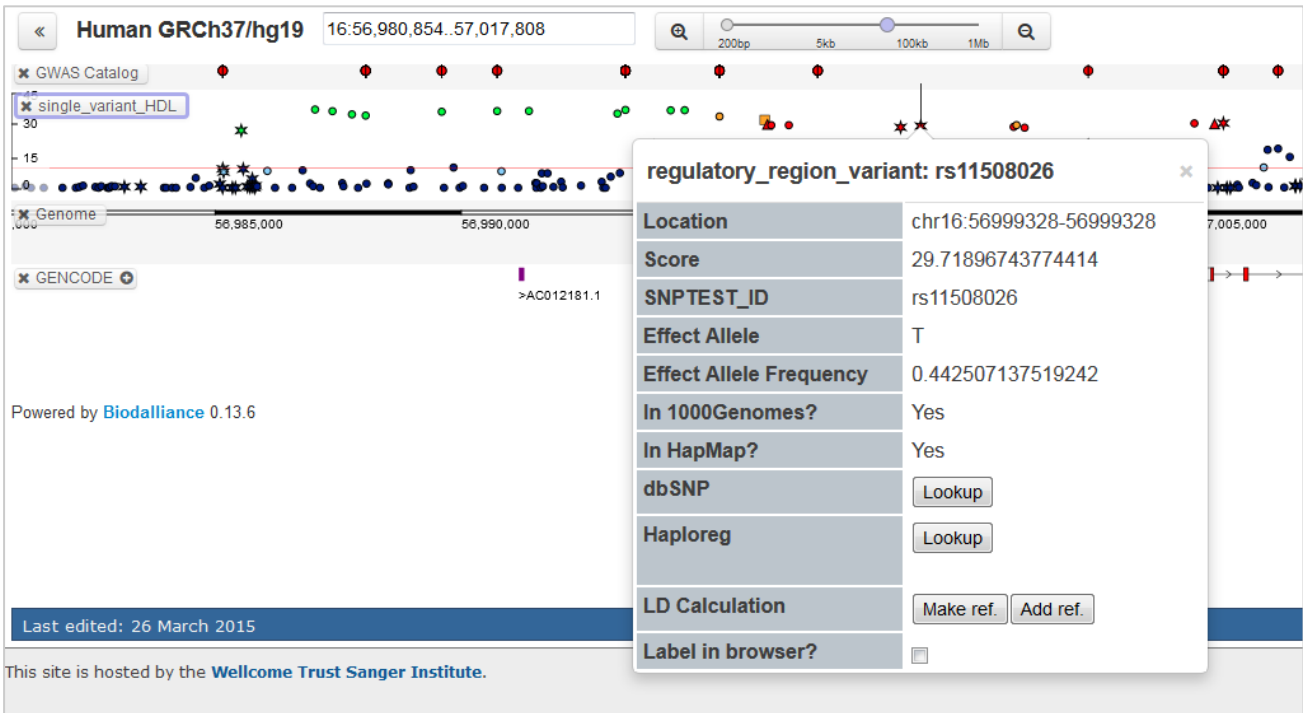



Figure B. An example where one SNV (rs11508026) has been made the reference by clicking on the 'Make ref.' button, all others SNVs are coloured by LD score groups relative to that SNV; red ($r^2=0.8-1$), yellow ($r^2=0.6-0.8$), green ($r^2=0.4-0.6$), light blue ($r^2=0.2-0.4$) and dark blue ($r^2=0-0.2$).

11. Repeat this for SNVs rs 289717 (green) and rs12720889 (yellow) to add the last two groupings. Up to five SNVs groups can be displayed at any time (see **Figure C** for a five SNV groups example)

12. Finally, to remove the pink highlight of the gene area click on the 'eraser' button .

You should now have a view that is largely identical to the figure in the manuscript. If you wish to export the image or other information you can click on the 'export options' button and to the left of the screen a drop down with the various export formats will appear. On the right is the 'export' button, clicking on this brings up the options to preview or download the selected file.

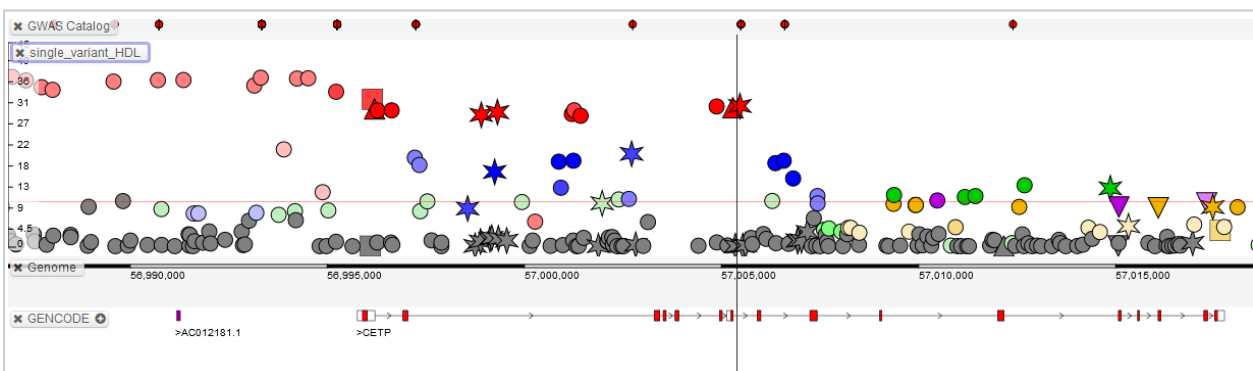


Figure C. In this instance five SNVs have been made into references, the correlated (LD) SNVs are now coloured by group and the intensity of the colour represents the LD score from dark ($r^2=1$) to light ($r^2=0.2$). In this notation, independent SNVs ($r^2<0.2$) from the selected reference SNVs are in grey. Note that the track height has been increased to better illustrate this.