MPRA Methods

0. Primer table

	PRIMER	Use	Legend	Sequence (5'->3')
1	Pool Medium Lib FWD	Oligo nucleotide library amplification	Gibson 5' - Spacer 5' - Amplification arm pool medium GC 5'	TAGAGCATGCACCGGTGATAGCAAAGTGAACACATC GCTAAGCGAAAGCTAAGAGGACCGGATCAACT
2	Pool Medium Lib REV	Oligo nucleotide library amplification	Gibson 3' - Spacer 3' - Amplification arm pool medium GC 3'	CTGCCGGCCGAATTCGTCGATCCTGGCCTAGTTGGA ATTCCCTGCAGG
3	Pool High GC Lib FWD	Oligo nucleotide library amplification	Gibson 5' - Spacer 5' - Amplification arm pool High GC 5'	TAGAGCATGCACCGGTGATAGCAAAGTGAACACATC GCTAAGCGAAAGCTAAGACTGGCCGCTTGACG
4	Pool High GC Lib REV	Oligo nucleotide library amplification	Gibson 3' - Spacer 3' - Amplification arm pool High GC 3'	CTGCCGGCCGAATTCGTCGATCCTGGCCTAGTTGCA GGAGCCGCAGTG
5	Pool High AT Lib FWD	Oligo nucleotide library amplification	Gibson 5' - Spacer 5' - Amplification arm pool high AT 5'	TAGAGCATGCACCGGTGATAGCAAAGTGAACACATC GCTAAGCGAAAGCTAAGCTGCGCCTGATGCAG
6	Pool High AT Lib REV	Oligo nucleotide library amplification	Gibson 3' - Spacer 3' - Amplification arm pool high AT 3'	CTGCCGGCCGAATTCGTCGATCCTGGCCTAGTTGCG ATAGCGAGCACC
7	Gibson FWD	Gibson cloning		TCGACGAATTCGGCCGGCAG
8	Gibson REV	Gibson cloning		TATCACCGGTGCATGCTCTA
9	PolyA FWD	PolyA site cloning		GGATCCCATTGCGTGAACCGATCGACGAATTCGGCC GGCCGCTTCG
10	PolyA REV	PolyA site cloning		GGTACCACCTCCCCCTGAACCTGAAAC
11	SEQ.INI.FWD 1	Library Preparation PE150 & PE25	Illumina P5 - index 1- Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACACGTAGA CCAGCAAAGTGAACACATCGCTAAGCGAAAGC
12	SEQ.INI.FWD 2	Library Preparation PE150 & PE25	Illumina P5 - index 2- Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACAGCGGCG ACCGCAAAGTGAACACATCGCTAAGCGAAAGC
13	SEQ.INI.FWD 3	Library Preparation PE150 & PE25	Illumina P5 - index 3- Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACACTACTTG CTGCAAAGTGAACACATCGCTAAGCGAAAGC
14	SEQ.INI.FWD 4	Library Preparation PE150 & PE25	Illumina P5 - index 4- Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACTGAACTTC TTGCAAAGTGAACACATCGCTAAGCGAAAGC
15	SEQ.INI.FWD 5	Library Preparation PE150 & PE25	Illumina P5 - index 5- Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACGCCAGCG CCGGCAAAGTGAACACATCGCTAAGCGAAAGC
16	SEQ.INI.FWD 6	Library Preparation PE150 & PE25	Illumina P5 - index 6- Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACCTGAACC GCAGCAAAGTGAACACATCGCTAAGCGAAAGC
17	SEQ.INI.FWD 7	Library Preparation PE150 & PE25	Illumina P5 - index 7- Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACGGTTGGC CAAGCAAAGTGAACACATCGCTAAGCGAAAGC
18	SEQ.INI.FWD 8	Library Preparation PE150 & PE25	Illumina P5 - index 8 - Priming seq to Spacer 5'	AATGATACGGCGACCACCGAGATCTACACGCTTAAC CAAGCAAAGTGAACACATCGCTAAGCGAAAGC
19	SEQ.INI.REV.NO. UMI	Library Preparation PE150	Illumina P7 - Gibson 3' - Spacer 3'	CAAGCAGAAGACGGCATACGAGATCTGCCGGCCGA ATTCGTCGATCCTGGCCTAGTTG
20	P5	Library Preparation PE150 & PE25	Illumina P5	AATGATACGGCGACCACCGAGATCTACAC
21	P7	Library Preparation PE150 & PE25	Illumina P7	CAAGCAGAAGACGGCATACGAGAT
22	SEQ.INI.R1	Sequencing PE150 & PE25	Priming seq to Spacer 5'	GCAAAGTGAACACATCGCTAAGCGAAAGCTAAG
23	SEQ.INI.R2	Sequencing PE150	Gibson 3' - Priming seq to Spacer 3'	CTGCCGGCCGAATTCGTCGATCCTGGCCTAGTTG
24	SEQ.MPRA.REV	Library Preparation	Illumina P7 - 10 mer UMI -	CAAGCAGAAGACGGCATACGAGATNNNNNNNNNC

		PE25	Priming to cloned polyA site	GAAGCGGCCGGCCGAATTCGTCGA
25	SEQ.MPRA.UMI	Sequencing PE25	Priming to cloned polyA site (Reverse)	TGAACCGATCGACGAATTCGGCCGGCCGCTTCG
26	SEQ.MPRA.R2	Sequencing PE25	Priming to cloned polyA site (Forward)	GCCGGCCGAATTCGTCGATCGGTTCACGCAATG
27	NEW_mRNA_FWD	check for plasmid contamination in the cDNA		ACGTTAAGGGATTTTGGGCC
28	NEW_mRNA_REV	check for plasmid contamination in the cDNA		GGGCATGGCACTCTTGAAAA

1. Backbone vector and polyA site subcloning

The basis for the backbone vector for the assay is the hSTARR-seq_ORI vector (Addgene Plasmid #99296) following the recommendations from (Muerdter et al. 2018). This plasmid does not have a core promoter and uses the ORI instead and does not come with the GFP reporter cloned. It has a 133 bp synthetic intron upstream the Multi Cloning site making it possible to distinguish the pre-mature mRNA (unspliced) from the spliced form of the reporter gene cloned downstream by agarose gel. The GFP was obtained by excising sgGFP from the pSTARR-seq_human vector (Addgene Plasmid #71509). 1 ugr of the pSTARR-seq_human vector was digested overnight at 37℃ with 20 U of Afl II (unless stated otherwise all the restriction enzymes used in this work are the HF version of NEB) and 20 U of AgeI and the product (739 bp) was run in a 2% agarose gel and purified using QIAquick PCR Purification Kit (Qiagen, 28106). In the same way, 1 ugr of hSTARR-seq_ORI vector was digested overnight and the 4kb linearized vector was gel purified (0.8% gel). The ligation was performed overnight at 16℃ using T4 DNA ligase (NEB M0202S) with 50 ng of linearized vector and an insert to vector ratio of 3:1. One-Shot ccdB Survival 2 T1R Competent Cells (Invitrogen A10460) were transformed with 5ul of ligation and grown to perform a mini-prep (Qiagen 27104).

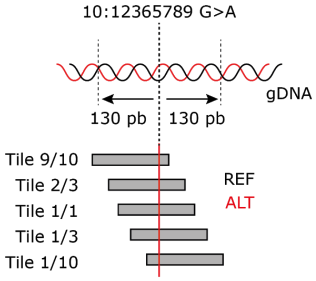
Following an idea proposed in (Muerdter et al. 2018), we excised the polyA site from the pSTARR-seq_human GFP vector we had just obtained to be cloned back at a later stage of our library construction separating the barcode and the candidate regulatory sequence. The excision was carried out by NaeI digestion (60' 10 Units of enzyme) followed by gel purification of the 4.2 kB linearized plasmid (pSTARR-seq_human GFP poly(A) minus and the 368 bp fragment containing the SV40 polyA signal. 50 ng of the linearized plasmid were subsequently recircularized using Blunt-end ligation (NEB M0367S) for 15' at RT. One-Shot ccdB Survival 2 T1R Competent Cells were transformed with 5ul of this ligation and grown overnight to perform a mini-prep.

The SV40 poly(A) signal fragment was gel purified (2% gel) and PCR amplified with the PolyA oligos (primers 9 and 10, Oligos table) and BioTaq (Bioline BIO-21071) that is compatible with TA cloning. PCR conditions included 5 cycles at 58℃ followed by 30 cycles at 63℃ annealing temperatures. The amplified fragment (242 bp) was then cloned in standard conditions in the pGEM T easy system (Promega A1360) at 3:1 insert:vector ratio and 5 ul of the ligation were used to transform JM109 cells which were grown overnight to perform a mini-prep.

2. Oligonucleotide library amplification

We designed a library of 20,340 200-mer oligonucleotides that were synthesized by Twist Bioscience. The library covered 138 SNPs, each one assayed in five partially overlapping tiles, every tile having an alternative and reference allele version. Each reference or alternative allele tile was tagged by 15 unique 11 bp barcodes. The library included 7 MPRA-ASE positive controls from (Ulirsch et al. 2016) and three sequences devoid of CRISPRa activity in K562 as negative controls (Fulco et al. 2016). The structure of the 200-mers is shown below. It includes two 15 bp amplification arms at each end, an 11 bp barcode, the restriction enzyme sites for BamHI and KpnI and 148 bp of candidate regulatory sequence to be assayed. We used three different sets of amplification arms to subdivide the 200-mers according to their GC content in three bins: Medium GC (10,305 oligos, 38.5-54.6 % GC), Low GC (4,770 20.9-38.4 %GC) and High GC content (5,265 oligos, 54.7 - 79.4 %GC). PCRs were done with Kapa HiFi HS Ready Mix (Kapa Biosystems 07983034001), 20 ng input template and 50 ul final volume with the exception of the High GC bin in which 20 ul of KAPA2G GC Buffer (Merck KK5516) was added to a final volume of 100 ul. The first set of 10 cycles at 58℃`annealing temperature for 30'' was followed by 8 cycles in which the annealing temperature was raised to 63℃ for 10 reactions (Medium GC pool) and 5 reactions (Low and High GC bins). After pooling the reactions corresponding to each bin, we performed a digestion with ExoI to eliminate free primers (1 ul of Exo I per 10 ul of PCR), applied Agencourt AMPure XP system (Beckman Coulter BCAG0006) to purify the amplicons in each bin, 1.8 volumes of beads per volume of

PCR, and eluted in 40 ul of Elution buffer (Qiagen) diluted 1/10 in nuclease-free water. The number of amplicons in each bin was quantified using Qubit (Qubit dsDNA BR Assay Kit, Life Technologies Q32850)



7 ASE ctrls[1]

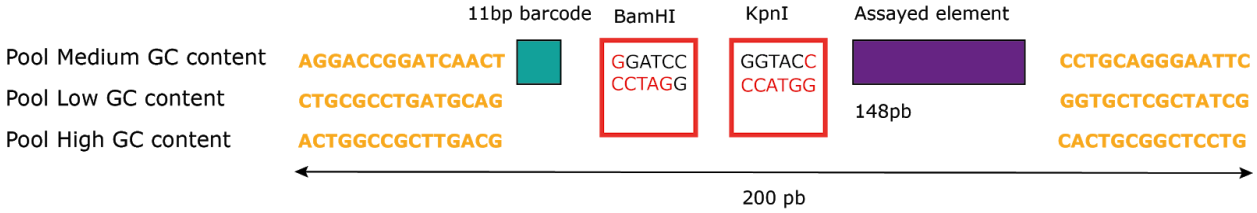
11 undetermined controls[1]

3 Negative ctrls[2]

11 colocalized variants in the ITGA4 locus

106 non coding variants with high effect sizes and PP >= 0.9 in UKBB Full Blood counts

- 73 intronic
- 12 intergenic
- 2 TF binding site
- 4 regulatory region
- 3 upstream
- 2 downstream



3. ***Gibson cloning to obtain Library poly*** (A) ***minus***

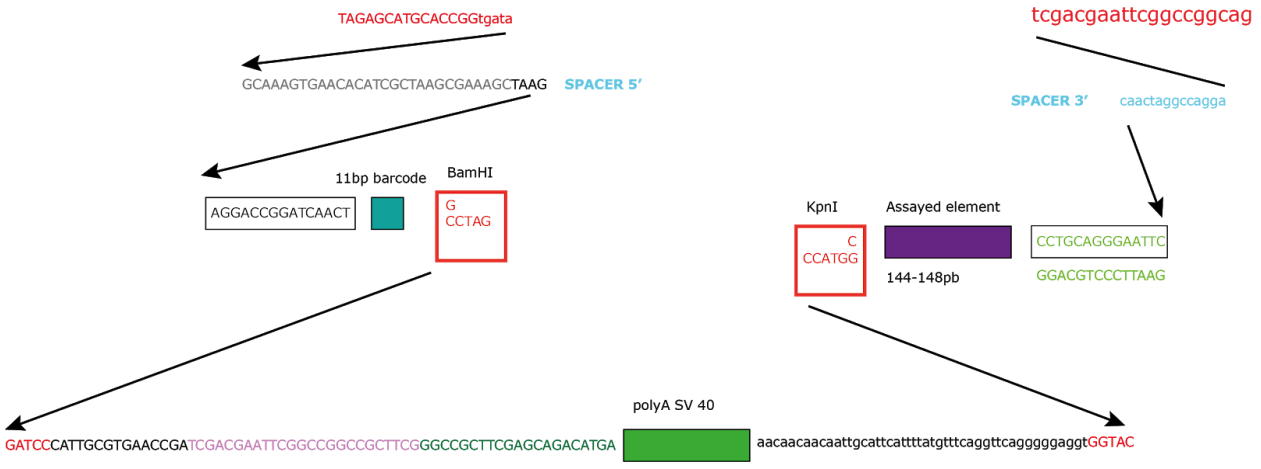
We linearized 1 ugr of the pSTARR-seq_human GFP polyA minus vector using PCR amplification (Primers 7 and 8 of the Oligos table) ramping up from 64℃ to 72℃ annealing temperature across 30 cycles of amplification. Following instructions from the Gibson cloning handbook (NEB E5510S), we digested the PCR result with DpnI and BamHI, 4 ul and 1 ul per 100 ul of PCR respectively, to degrade the circular template and purified the linearized vector (2,8 kB) using Agencourt AMPure XP system (1.5 volumes of beads per volume of PCR). We eluted the linearized vector in 40 ul of Elution buffer diluted 1/10 in nuclease-free water and quantified using Qubit.

The ligation in Gibson mix was done with 100 ng of the linearized vector and a molar insert:vector ratio of 2 for each of the bins (4 reactions for the Medium GC bin and 2 for the other two bins). The reaction was set up on ice followed by 40' at 50℃ and 20' at 40℃. The ligations corresponding to each bin were pooled separately, purified with Agencourt AMPure XP system (1.5 vol beads per vol of Gibson ligation) and eluted in 20 ul of Elution buffer diluted 1/10 in nuclease-free water. 10 ul of each ligation were used to electroporate electrocompetent E. coli (NEB C3020K) at 2000 V 25 uF 200 Ohm. SOC media was added immediately to cells after electroporation and cells were allowed to recuperate for 1 hour at 37℃ and gentle agitation. We performed serial dilutions for each of the bins to ascertain the yield in colony-forming units (CFU) of the Gibson cloning. We seeded 1 million CFU for the Medium GC bin and 0.5 million CFU for each of the other two bins in 245mm Square BioAssay dishes (Corning,431272) with Agar + Ampicillin and allowed them to grow overnight. We kept a ratio of 100 CFU per oligo element in each of the pools throughout the process. The individual colonies in each bin were lysed and the plasmids corresponding to each bin were purified using the maxi prep kit (Qiagen 1216). This intermediate step in the library construction was sequenced (see below) to check for inconsistencies in the barcode - candidate regulatory sequence association.

4. ***PolyA cloning into Library poly*** (A) ***minus to obtain the Final Library***

We digested 4 ugr per reaction of every bin with 50 Units of KpnI overnight at 37℃. Ten reactions were done for the Medium GC content bin and 5 for each of the other two bins. At least 3.6 ugr of each of the resulting linearized vectors (3.082 kB) after gel purification was subjected to a successive digestion with BamHI (50 Units) in the presence of Shrimp Phosphatase (NEB M0371S) for 2h at 37℃. After enzyme inactivation by heating at 65℃ for 5' we purified each bin reaction mix with Agencourt AMPure XP system (1.5 vol beads per vol of reaction), eluted in 40 ul of Elution Buffer and quantified using Qubit. In parallel, we digested overnight 20ugr of the pGEMT-polyA

vector to release the SV40 polyA signal (230 bp) with 500 Units of KpnI and BamHI and gel purified it. For each bin we used 100 ngr of the input vector and a 2:1 insert:vector ratio in the T4 ligase 16℃ overnight reactions. For the medium GC bin, we performed 8 ligation reactions and 4 for each of the other two bins. After ligation, we pooled together the mixes corresponding to each bin and used Agencourt AMPure XP system to purify the ligation product (1.5 vols of beads per vol of reaction) and eluted each one in 20 ul of Elution buffer diluted 1/10 in nuclease-free water. We took 10 ul of each ligation bin and electroporated electrocompetent E. coli in the same conditions specified for the Gibson cloning. We performed 2 electroporations for the Medium GC bin and one for each of the other two bins. After evaluating the efficiency of the transformation we seeded 8 245mm Square BioAssay dishes (Corning,431272), with Agar + Ampicillin for the medium GC bin and 4 for each of the other two bins. Each plate was seeded with 0.5 million CFU's. The resulting individualized colonies were lysed and the final library plasmids were purified using the Endo-free Plasmid Mega prep kit (Qiagen 12381). The final quantities for each of the library components were: 3.5 mgr of plasmid for the medium GC bin, 2 mgr for the High GC bin and 1 mgr for the low GC bin. The structure of the final library that was used for the nucleofection of the cells is shown below:



5. ***PE150 sequencing of Library poly (A) minus***

To address the correct matching of the barcode and regulatory candidate sequence we sequenced the poly (A) minus library. For the library preparation, we amplified 25 ngr of each of the poly (A) minus library bins (Medium GC, High GC and Low GC content) with oligos 11 and 19 from the Oligo table for 15 cycles with annealing and extension done at 72℃ in a combined step for 1'. The libraries were quantified afterwards with KAPA Illumina SYBR Universal Lib Q. Kit. (Roche. 07960140001), adjusted at 4nM and pooled together for a final volume of 40 ul. The sample was subsequently sequenced in a MiSeq (MiSeq Reagent Kit v2 300 cycle Illumina BCSO0012) with the first 15 cycles of read 1 set to dark cycling and using custom primers 22 and 23 from the oligo table.

6. ***Cell culture***

K562 (ATCC® CCL-243™) cells were culture as indicated by the distributor, 1x RPMI 1640 media with L-glutamine (Gibco Medium.: 52400025), supplemented with 10% FBS (Gibco, A31604-02) and 1x penicillin/streptomycin (Gibco, 15070-063). Cells were maintained up to a confluence of $1x10^6$ cells/ml and then reseeded at $1x10^5$ cells/ml. To prepare for the transfection cells were seeded at $1.5x10^5$ /ml into 2x T175 flasks (Corning, CLS431085-50EA) in 60 ml of fresh medium (9 million cells each flask), and cultured for 48hrs changing half of their media 24hrs after seeding.

7. ***Electroporation***.

The electroporation was performed using the Neon Transfection System (ThermoFisher, Cat# MPK5000). After 48hrs cells were collected, spin down (300 G 5') washed once with PBS (Gibco, Cat # 14190-094), and counted on a haemocytometer (NanoEntek, Cat # DHC-N01).We electroporated 5 million cells with 25 ugr of a pool of the three plasmid bins per reaction (12.5 ugr of the Medium GC bin and 6.25 ugr of each of the other two bins). We included the following controls: untransfected cells, mock transfected cells, cells transfected with the hSTARR-seq_ORI vector GFP vector, and cells transfected with the pmaxGFP green vector (Lonza Cat #: VDC-1040). The efficiency of the transfection was evaluated at 24h and 48h post nucleofection in the pmaxGFP plate with a Countess II

FL Automated Cell Counter (Cat # AMQAF1000). For every replica, the efficiency of transfection was above 90%. The nucleofection was performed following the manufacturer instructions using buffer R and 1.45 v, 10 ms pulse width and 3 pulses. After the electroporation, the cells were recovered and cultured in 8ml of DMEM without antibiotics for 2 days.

8. ***Parallel extraction of genomic DNA and mRNA from transfected cells***

48 h post-electroporation, cells were collected, spun down, and treated with DNase I (NEB M0303L) at 37℃ for 15' using 10 U of enzyme per ug of plasmid transfected in a final volume of 2 ml of DPBS. After this, the cells were spun down at 300 G for 5' and washed twice with DPBS. Cells were lysed in 600 ul of Buffer RLT Plus (Qiagen) with added β-mercaptoethanol and homogenized using QIAshredder columns (Qiagen 79654). DNA and total RNA were extracted using AllPrep DNA/RNA Kit (Qiagen, Cat# 80204) according to the manufacturer's instructions. In the RNA preparation, a step of on-column DNase I treatment was performed for all samples (Rnase_Free Dnase Set, Qiagen 79254). We isolated mRNA from total RNA using the Oligotex mRNA Kit (Qiagen 70042) followed by a final treatment with Turbo Dnase (Invitrogen AM1907). DNA and mRNA quantifications were done using Qubit Assay Kits (ThermoFisher, Q32854 and Q32852 respectively). Supernatant aliquots pre and post extracellular treatment with DNAse I were assayed by doing a qPCR using primers 27 and 28 from the oligo table using StepOne Real-Time PCR System and Kapa SYBR Fast (SigmaAldrich KK4601) at 40 cycles. The results indicated that the treatment worked in eliminating any residual non incorporated extracellular DNA plasmid.

9. ***Retrotranscription***

We retro-transcribed 1-1,5 ug of mRNA per replica following the protocol for SuperScript IV (ThermoFisher 18090050) using a reporter specific RT primer (primer 24 in the oligo table) at 2uM carrying the 10-mer UMIs. As a means of controlling the overall possibility of episomal DNA contamination in the cDNA samples, ⅛ of every sample was amplified with primers 27 and 28. Because these primers flank the synthetic intron present in the GFP we were able to distinguish the pre-spliced and the mature forms of GFP and therefore ascertain possible contaminations (change of 133 bp in band size). None of the cDNA libraries showed appreciable contamination of episomal DNA.

10. ***cDNA Library Preparation***

We split retrotranscription samples for PCR amplification so as the RT template would represent 10% of the final volume of the PCR (50 ul). We performed the first round of amplification in which we introduced the sample index primer, i5-i8 (primers 15-18 from the oligo table) for the cDNA samples of four replicas (we did two batches of four replicas). As a reverse primer, we used P7 (primer 21 from the oligo table). The PCR was carried out using Kapa HiFi HS Ready Mix, and 65℃ annealing temperature for a total of 3 cycles. The amplification from each replica was then pooled, purified using Agencourt AMPure XP system (1.5 volumes of beads per volume of reaction) and eluted in 129 of Elution Buffer. We used these eluates to assess the minimum amount of cycles for a second round of PCR by q-PCR with P5 and P7 (20 and 21, oligo table) with StepOne Real-Time PCR System and Kapa SYBR Fast. We determined that between 11-13 cycles were enough to keep the second round PCR from plateauing. We split each of the replicas cDNA into 8 reactions and performed the second round PCR for the cycles determined with the P5 and P7 primers and Kapa HiFi HS Ready Mix at 64℃ annealing temperature. We then pooled the reactions from each replica, purified using Agencourt AMPure XP system (1.2 volumes of beads per volume of reaction) and eluted in 60 ul of Elution Buffer.

11. ***Episomal DNA Library Preparation***

The gDNA fraction from the All Prep Qiagen kit was used as a source for the episomal plasmid nucleofected in every replica. For every replica, we used 12 ugr of gDNA that was split into 24 PCR reactions following (Klein et al. 2020) after measuring quantity using Qubit. In this first round of PCR, we introduced the sample index and the UMIs. We used primers i1-i4 (11-14 from oligo table) as forward primers and primer 24 carrying the 10-mer UMIs as the reverse primer. The 8 replicas were prepared in two batches.The first round of PCR run for 3 cycles at 65℃ annealing temperature. The reactions corresponding to each replica were pooled, purified using Agencourt AMPure XP system (1.8 volumes of beads per volume of reaction) and eluted in 320 ul of Elution Buffer. As in the cDNA library preparation we assessed by qPCR the number of cycles to keep the second round PCR from plateauing. We determined 10-11 cycles for the second round PCR. We split each of the replicas cDNA into 29 reactions and performed the second round PCR for the cycles determined with the P5 and P7 primers and Kapa HiFi HS Ready Mix at 72℃ annealing temperature. Finally, we pooled the reactions for each replica, purified 200 ul of the mix using Agencourt AMPure XP system (1.2 volumes of beads per volume of reaction) and eluted in 30 ul of Elution Buffer.

12. ***PE25 sequencing of Final Library***

We sequenced the 8 replicas in two batches, each one comprising the 4 cDNA and 4 gDNA samples of four replicas. Before sequencing every batch all the libraries were quantified using KAPA Illumina SYBR Universal Lib Q. Kit.and adjusted to 4nM and pooled afterwards in 40 ul final volume. We used a HiSeq 2500 RR for sequencing each batch with the Hiseq Rapid PE Cluster Kit V2 (Illumina BCSO1412) and the HiSeq Rapid SBS Kit v2 200 cycles (Illumina BCSO1432). The recipe included the first 15 cycles of read 1 set to dark cycling and used custom primers 22, 25 and 26 from the oligo table.

13. *Bioinformatics pipeline*

The 6 nucleotides corresponding to BamHI restriction site in the read 2 were eliminated using Trimmomatic (Bolger, Lohse, and Usadel 2014) before merging read 1 and read 2 in one unique amplicon with flash2 (Magoč and Salzberg 2011) with minimum overlap set to 10 bp. Every group of reads sharing the same ID and corresponding to the merged amplicon plus the UMI identifier were quantified against a reference of all the barcodes included in the design of the library. Reads sharing the same UMI and barcode were collapsed to obtain the raw counts for every replica cDNA or episomal DNA sample. A series of in-house R scripts were used to put together the episomal and cDNA samples of all the replicas and fill in with 0's the missing data for dropout barcodes to obtain a regular matrix. This matrix was then filtered, only barcodes having at least 5 counts in all the replicas in the cDNA and episomal DNA samples were kept and transformed into the input format accepted by MPRAnalyze (Ashuach et al. 2019). Samples were then normalized using upper Quartile normalization. We calculated mean FoldChanges of cDNA over episomal DNA as well as a mean measure of Allele Specific Expression (ASE) introduced in (Vockley et al. 2015) described as RNA REF allele counts/DNA REF allele counts)/(RNA total counts/DNA total counts for every tile. As a quality control step .MPRAnalyze was used to calculate the statistically significant tiles that exhibited enhancer properties and ASE. As a quality control measure we excluded tiles with less than 0.1 normalized counts and tiles with an imbalance between reference and alternative allele higher than 3 in the nucleofected plasmids. For the enhancer analysis, the three negative controls were used to set the baseline of the analysis. For the ASE we set the reducedDesign parameter to ~1.

Bibliography

Ashuach, Tal, David S. Fischer, Anat Kreimer, Nadav Ahituv, Fabian J. Theis, and Nir Yosef. 2019. "MPRAnalyze: Statistical Framework for Massively Parallel Reporter Assays." *Genome Biology* 20 (1): 183.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.

Fulco, Charles P., Mathias Munschauer, Rockwell Anyoha, Glen Munson, Sharon R. Grossman, Elizabeth M. Perez, Michael Kane, Brian Cleary, Eric S. Lander, and Jesse M. Engreitz. 2016. "Systematic Mapping of Functional Enhancer–promoter Connections with CRISPR Interference." *Science* 354 (6313): 769–73.

Klein, Jason C., Vikram Agarwal, Fumitaka Inoue, Aidan Keith, Beth Martin, Martin Kircher, Nadav Ahituv, and Jay Shendure. 2020. "A Systematic Evaluation of the Design and Context Dependencies of Massively Parallel Reporter Assays." *Nature Methods*, October. https://doi.org/10.1038/s41592-020-0965-y.

Magoč, Tanja, and Steven L. Salzberg. 2011. "FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies." *Bioinformatics* 27 (21): 2957–63.

Muerdter, Felix, Łukasz M. Boryń, Ashley R. Woodfin, Christoph Neumayr, Martina Rath, Muhammad A. Zabidi, Michaela Pagani, et al. 2018. "Resolving Systematic Errors in Widely Used Enhancer Activity Assays in Human Cells." *Nature Methods* 15 (2): 141–49.

Ulirsch, Jacob C., Satish K. Nandakumar, Li Wang, Felix C. Giani, Xiaolan Zhang, Peter Rogov, Alexandre Melnikov, et al. 2016. "Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits." *Cell* 165 (6): 1530–45.

Vockley, Christopher M., Cong Guo, William H. Majoros, Michael Nodzenski, Denise M. Scholtens, M. Geoffrey Hayes, William L. Lowe, and Timothy E. Reddy. 2015. "Massively Parallel Quantification of the Regulatory Effects of Noncoding Genetic Variation in a Human Cohort." *Genome Research*. https://doi.org/10.1101/gr.190090.115.