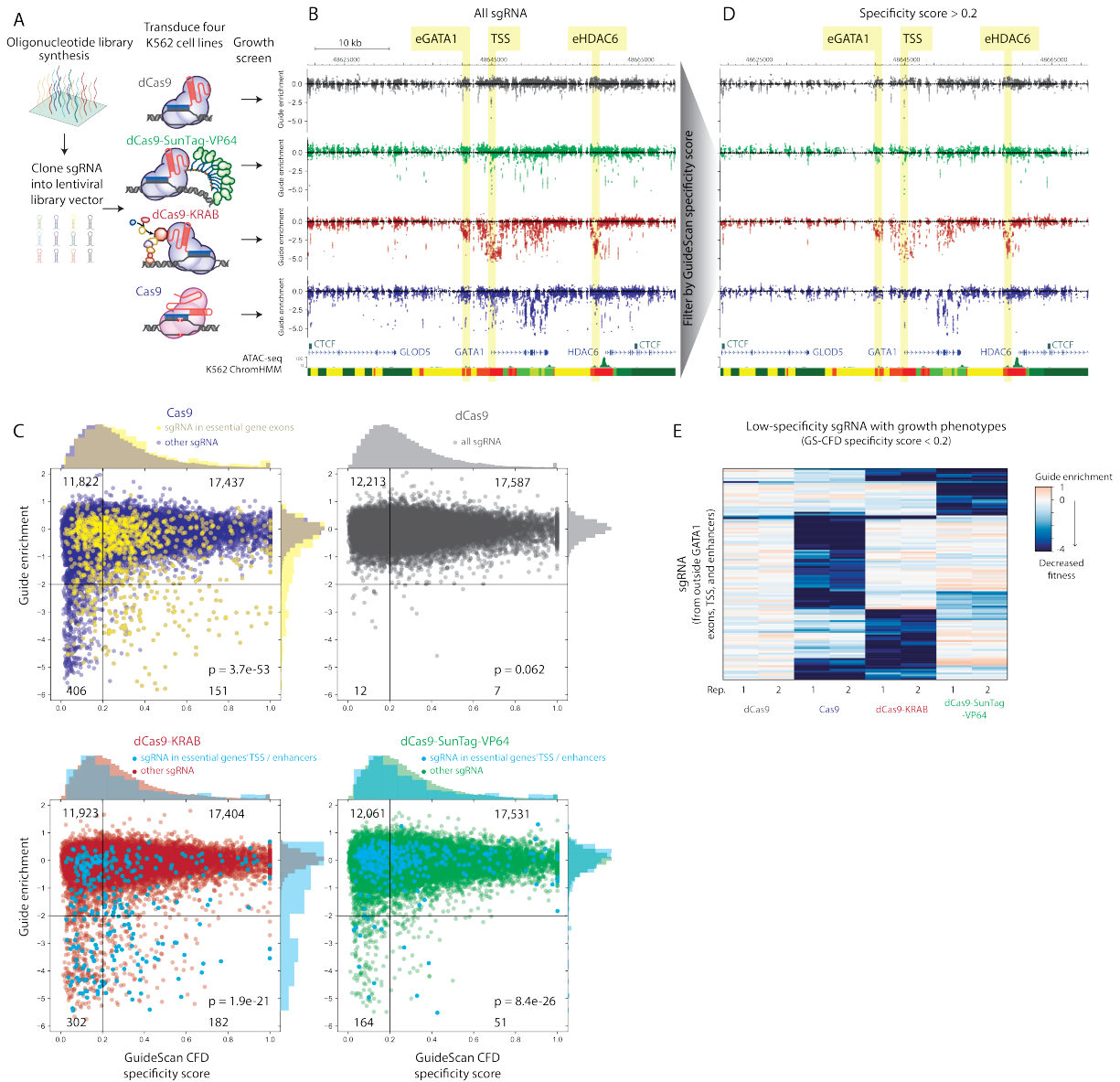


# Summary of progress

During the last year we have continued our efforts towards characterizing the functional candidate cis-regulatory elements (cCREs) in the human genome at a large scale using a variety of high-throughput CRISPR screen strategies. Our research is currently directed in the following directions.

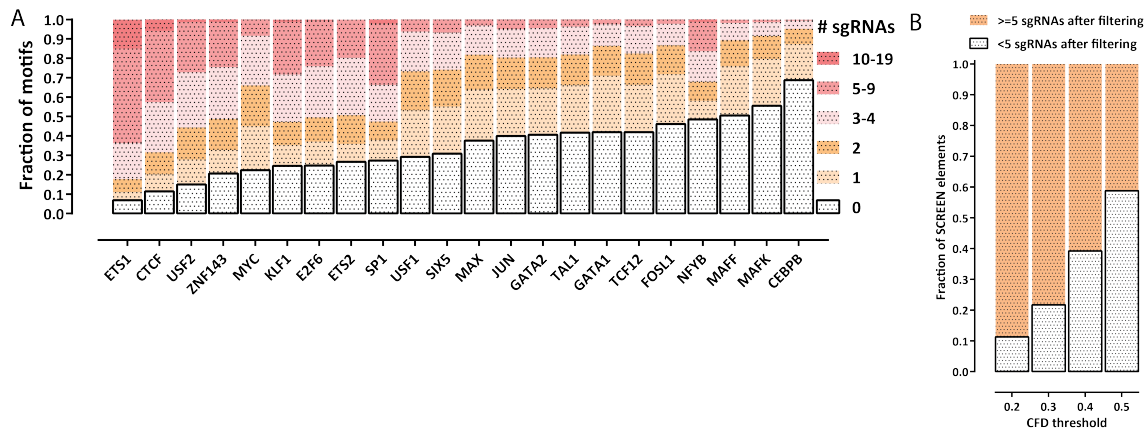
First, in the course of a systematic CRISPR perturbation comparison effort involving tiling screens in K562 cells, and of some of our other screens, we have developed a much better understanding of off-target CRISPR effects in the context of non-coding screens and epigenetic perturbations, with significant implications for future screen designs and interpretation (Summarized in Figure 1). We find that, contrary to prior expectations, off-target effects are not limited to CRISPRk (where cellular toxicity is induced by DNA damage at multiple sites when a guides has many off-targets), but are also observed when catalytically inactive Cas9 is used to generate epigenetic perturbations (such as CRISPRi and CRISPRa). Importantly,



**Figure 1: Low-specificity sgRNA are enriched in tiling screens for essential non-coding elements when using Cas9, CRISPRi, and CRISPRa.** a) Four parallel screens were conducted tiling the loci of *GATA1*, *MYB*, and *ZMYND8*; b) Zoom-in view of screen data around the essential gene *GATA1*; numerous guides with strong fitness effects but with little concordance with nearby guides or genome and chromatin state annotations are observed. c) Enrichment of fitness effects among low-specificity sgRNA; d) Filtering out low-specificity guides reduces noise; e) Clustering of low-specificity sgRNA shows that a unique subset of such sgRNAs exhibits fitness effects with each type of perturbation.

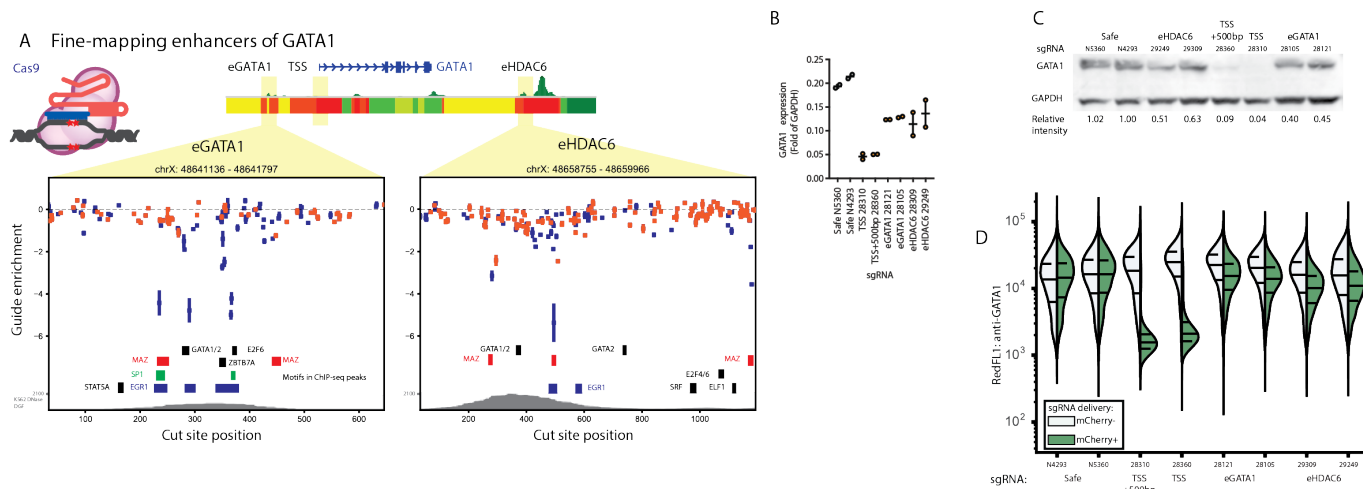
off-target effects appear to be specific to each perturbation type.

We have systematically compared many guide design models and determined which are most useful for generating high quality, interpretable non-coding CRISPR screen data. We have also examined potential off-target effects issues in the context of future screen designs (Figure 2). We find that targeting individual transcription factor (TF) binding sites with CRISPRk is to a significant extent confounded by low-specificity guides for many human TFs. However, at the level of candidate *cis*-regulatory elements (cCREs) most can be targeted with enough high-specificity sgRNAs using CRISPRi/a approaches.



**Figure 2: Low-specificity sgRNA in motif-level and cCRE-level screen designs.** a) High-specificity sgRNA targeting possibilities for individual occupied transcription factor sites of 22 different human TFs in K562 cells b) High-specificity sgRNA targeting possibilities for ENCODE SCREEN cCRE elements.

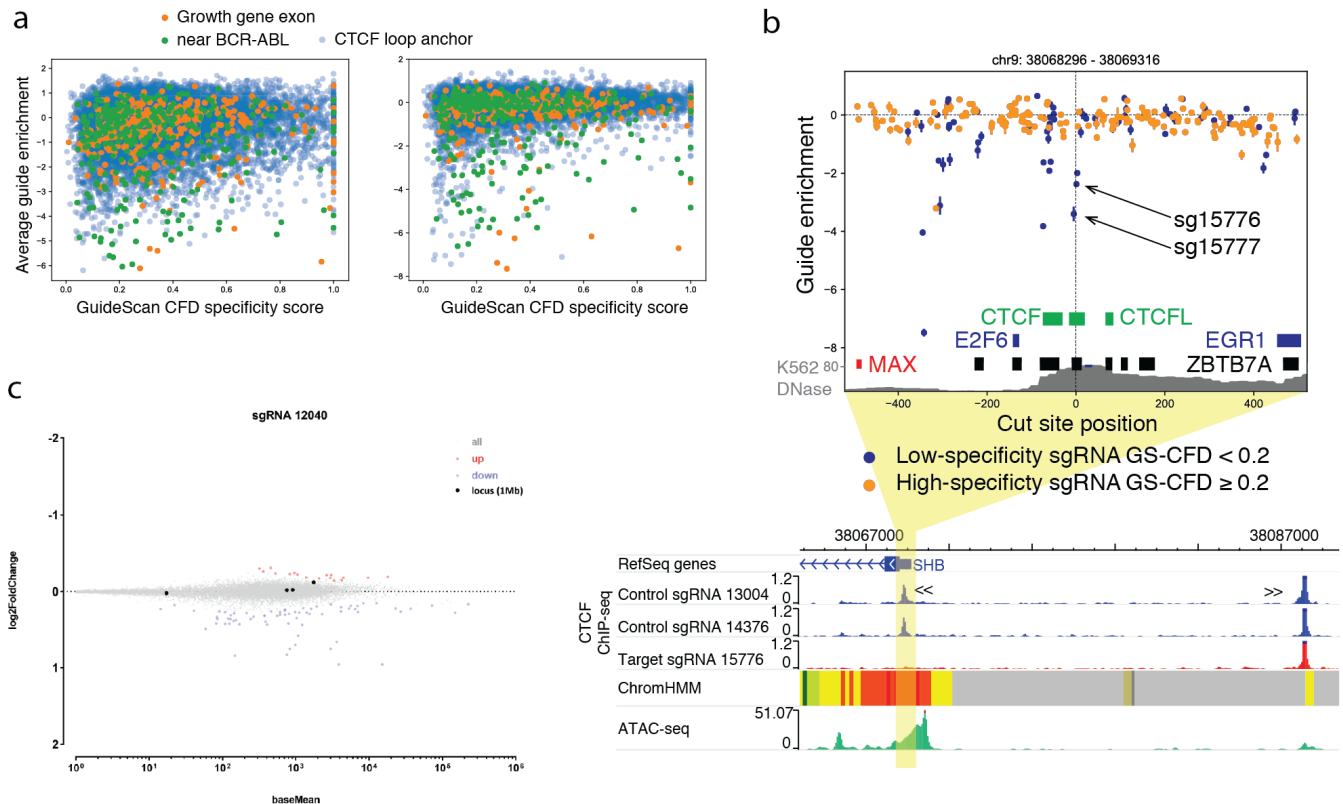
Following up from our tiling screen effort, we have also generated and characterized cell lines carrying out individual sgRNA modifications. An example is shown in Figure 3, where we characterize the effect of sgRNAs targeting the TSS and two of the enhancer elements of the GATA1 gene. These measurements allow us to quantify the relative effects of perturbing enhancer and promoter elements on the expression of their cognate genes.



**Figure 3: Fine mapping and characterization of GATA1 enhancer elements.** a) Fine mapping of GATA1 enhancers (“eGATA1” and “eHDAC6”); b) Effects of individual sgRNAs targeting the TSS or one of the eGATA1 and eHDAC6 enhancers on GATA1 expression (measured by qPCR); c) Effects of individual sgRNAs targeting the TSS or one of the eGATA1 and eHDAC6 enhancers on GATA1 protein levels d) Effects of individual sgRNAs on the distribution of GATA1 levels in the cell population. Shown is the distribution of GATA1 levels for cells carrying the sgRNA (mCherry positive cells) and cells without it (mCherry negative cells).

Second, we have completed our characterization of CTCF-occupied insulator sites/TAD anchor loops in K562 cells

using a combination of sgRNA, paired-guide (pgRNA)-mediated excision, and “fine mapping” (complete tiling across a region) screens using multiple different perturbations, both genetic (CRISPRk) and epigenetic (CRISPRi, CRISPRa and CRISPRd). As with our tiling screen, we find these screens confounded by low-specificity guides (Figure 4A). Strikingly, filtering out potential artifacts due to low-specificity guides eliminates nearly all “hits” from CTCF screens, meaning that individual CTCF sites have little to no fitness effects in the context of K562 cells. We characterized individual sgRNA cell lines, and found that CTCF perturbations abolish CTCF binding but lead to no changes in gene expression in the vicinity (examples shown in Figure 4B-C), consistent with CTCF perturbations having little effect on gene expression and fitness effects being due to off-target guide activities. These observations are corroborated by further screens we carried out using additional selective pressures such as ricin and hydrogen peroxide (resistance to which can only arise due to changes in gene expression/functionality of a well defined set of genes). The results from these efforts have been drafted and will be posted on bioRxiv and submitted for publication in the immediate future.



**Figure 4: Functional characterization of insulator elements in K562 cells.** a) CTCF site-targeting screens are confounded by low-specificity guides; b) and c) Examples of characterization of K562 cell lines carrying out individual CTCF site perturbations.

Third, we have carried out a fine-mapping screen targeting thousands of promoter elements in K562 in order to understand the *cis*-regulatory sequences controlling gene expression at human promoters. We are in the process of analyzing the resulting data, and we are also preparing for submitting the data from this and our other screens to the ENCODE DCC.

## Ongoing and future efforts

Fourth, we are carrying out paired-guide RNA (pgRNA) screens aimed at understanding the combinatorial action of regulatory elements using CRISPRi perturbations. These screens employ multiple guides targeting each cCREs in all possible combinations (with each other and with “safe” guides), and then either selecting for growth (when studying essential genes in K562 cells, e.g. GATA1 or MYC) or staining at the protein or RNA level and sorting using FACS. We have designed such screens for more than dozen genes in K562 cells. We expect these screens to allow us to quantitatively understand the contributions of individual cCREs to the expression of human genes, individually and in combination with each other.

Fifth, we are carrying out nucleotide-level dissections of regulatory activity by employing base editing CRISPR fusions

("CRISPR-X"). Our strategy uses all possible pairs of sgRNAs targeting different positions within a given cCRE, which introduce a wide diversity of single-base substitutions, staining at the protein/RNA level and cell sorting, and then targeted amplicon sequencing to identify enriched/depleted variants. We are focusing our initial efforts in this direction on GATA1, MYC and the  $\beta$ -globin loci, which are previously well characterized loci, and we plan to expand them to a much larger set of genes in the future. These experiments will enable us to finely map the contribution of individual transcription factor binding sites to the activity of regulatory elements, in way that is not confounded by low-specificity sgRNAs.

Finally, we have designed motif-level libraries for dozens of transcription factors active in K562 in order to assess the regulatory significance of these motifs, in an analogous fashion to our efforts targeting CTCF, and plan to carry out screens using these libraries in the next year. While our motif-level screen for CTCF revealed that CTCF site perturbations do not have fitness effects or effects on gene expression in the immediate genomic neighborhood, our tiling screens do identify numerous highly specific sgRNAs targeting cCREs with strong fitness effects. We therefore expect that further motif screens will provide illuminating insights into the role of individual transcription factors in regulating gene expression.