In order of Importance, list of 10 best papers of the applicant highlighting the important discoveries/contributions described in them briefly (Max. 1 MB) *

1. Combination of engineered FnCas9 and extended gRNAs for PAM-flexible, robust and nucleobase specific editing and diagnostics

Sundaram Acharya, Asgar Hussain Ansari, Seiichi Hirano, Sajal Sarkar, Riya Rauthan, Manoj Kumar, Rhythm Phutela, Sneha Gulati, C. Afzal1, Deepanjan Paul, Abdul Rahman, Sudipta Mahato, Savitri Maddileti, Vinay Kumar Pulimamidi, Subhadra Jalali, Hiroshi Nishimasu, Indumathi Mariappan, Osamu Nureki, Souvik Maiti, **Debojyoti Chakraborty**

Nature Biotechnology [under review, 2023; preprint at Research Square [https://doi.org/10.21203/rs.3.rs-3104171/v1]

The clinical success of CRISPR therapies is dependent on the safety and efficacy of Cas proteins. The Cas9 from *Francisella novicida* (FnCas9) has negligible affinity for mismatched substrates enabling it to discriminate off-targets in DNA with very high precision even at the level of binding. However, its cellular targeting efficiency is low, limiting its use in therapeutic applications. In this study, we rationally engineer the protein to develop enhanced FnCas9 (enFnCas9) variants and expand its cellular editing activity to genomic loci previously inaccessible. Notably, some of the variants release the protospacer adjacent motif (PAM) constraint from NGG to NGR/NRG increasing their accessibility across human genomic sites by ~ 3.5-fold. The enFnCas9 proteins harbor single mismatch specificity both *in vitro* and *in cellulo* leading to broadened target range of FnCas9-based CRISPR diagnostics for detection of point mutations and pathogenic DNA signatures. Importantly, they provide superior outcomes in terms of editing efficiency, knock-in rates and off-target specificity over other engineered high-fidelity versions of SpCas9 (SpCas9-HF1 and eSpCas9). Remarkably, enFnCas9 variants can be combined with extended length gRNAs for robust base editing at sites which are inaccessible to PAM-constrained canonical base editors. Finally, we show the complete correction of a disease-specific Retinitis Pigmentosa mutation in patient derived iPSCs using enFnCas9 Adenine Base Editor highlighting its broad application in therapeutics and diagnostics.

2. TOBF1 modulates mouse embryonic stem cell fate through regulating alternative splicing of pluripotency genes

Meghali Aich, Asgar Hussain Ansari, Li Ding, Vytautas Iesmantavicius, Deepanjan Paul, Chunaram Choudhary, Souvik Maiti, Frank Buchholz, **Debojyoti Chakraborty**

Cell Reports [2023, In press]

In this manuscript we show the complete characterization of TOBF1 a novel protein that regulates mouse embryonic stem cell pluripotency. Embryonic stem (ES) cells retain the ability to undergo lineage-specific differentiation that can eventually give rise to different cell types that constitute an organism. Although stem cell specific biological networks of transcription factors and epigenetic modifiers are well established, how the ES cell specific transcriptional and alternative splicing (AS) machinery regulate their expression has not been sufficiently explored in the past. In this study, we show that the lncRNA associated protein TOBF1 regulates the alternative splicing of transcripts necessary for maintaining stem cell identity in mouse ES cells. Among the

genes affected, of particular significance is SRSF1 whose alternative splicing leads to global changes in splicing and expression of a large number of downstream genes involved in the maintenance of ES cell pluripotency. Overlaying information derived from three sources-TOBF1 chromatin occupancy, the distribution of its pluripotency-associated OCT-SOX binding motifs, and transcripts undergoing differential expression and alternative splicing upon its knockout we also unmask local nuclear territories where these distinct events converge. Collectively these contribute to the maintenance of mouse ES cell identity.

3. CriSNPr, a single interface for the curated and de novo design of gRNAs for CRISPR diagnostics using diverse Cas systems.

Asgar Hussain Ansari, Manoj Kumar, Sajal Sarkar, Souvik Maiti and **Debojyoti Chakraborty**

eLife (2023)

CRISPR-based diagnostics (CRISPRDx) have improved clinical decision-making, especially during the COVID-19 pandemic, by detecting nucleic acids and identifying variants. This has been accelerated by the discovery of new and engineered CRISPR effectors, which have expanded the portfolio of diagnostic applications to include a broad range of pathogenic and non-pathogenic conditions. However, each diagnostic CRISPR pipeline necessitates customized detection schemes based on the fundamental principles of the Cas protein used, its guide RNA (gRNA) design parameters, and the assay readout. This is especially relevant for variant detection, a low-cost alternative to sequencing-based approaches for which no in silico pipeline for the ready-to-use design of CRISPRDx currently exists. In this manuscript, we fill this lacuna using a unified web server, CriSNPr (CRISPR-based SNP recognition), which provides the user with the opportunity to de novo design gRNAs based on six CRISPRDx proteins of choice (*Fn*/en*Fn*Cas9, *Lw*Cas13a, *Lb*Cas12a, *Aa*Cas12b, and Cas14a) and query for ready-to-use oligonucleotide sequences for validation on relevant samples. Furthermore, we provide a database of curated pre-designed gRNAs as well as target/off-target for all human and SARS-CoV-2 variants reported thus far. CriSNPr has been validated on multiple Cas proteins, demonstrating its broad and immediate applicability across multiple detection platforms. CriSNPr can be found at http://crisnpr.igib.res.in/.

4. FnCas9 based CRISPR diagnostic for rapid and accurate detection of major SARS- CoV2 variants on a paper strip.

Manoj Kumar, Sneha Gulati, Asgar Hussain Ansari, Rhythm Phutela, Sundaram Acharya, Mohd. Azhar, Jayaram Murthy, Poorti Kathpalia, Akshay Kanakan, Ranjeet Maurya, Janani Srinivasa Vasudevan, Aparna Murali, Rajesh Pandey, Souvik Maiti, **Debojyoti Chakraborty**

eLife (2021)

The COVID-19 pandemic originating in the Wuhan province of China in late 2019 has impacted global health, causing increased mortality among elderly patients and individuals with comorbid conditions. During the passage of the virus through affected populations, it has undergone mutations, some of which have recently been linked with increased viral load and prognostic complexities. Several of these variants are point mutations that are difficult to diagnose using the gold standard quantitative real-

time PCR (qRT-PCR) method and necessitates widespread sequencing which is expensive, has long turn-around times, and requires high viral load for calling mutations accurately. In this study, we repurposed the high specificity of *Francisella novicida* Cas9 (FnCas9) to identify mismatches in the target for developing a lateral flow assay that can be successfully adapted for the simultaneous detection of SARS-CoV-2 infection as well as for detecting point mutations in the sequence of the virus obtained from patient samples. We report the detection of the S gene mutation N501Y (present across multiple variant lineages of SARS-CoV-2) within an hour using lateral flow paper strip chemistry. The results were corroborated using deep sequencing on multiple wild-type (n = 37) and mutant (n = 22) virus infected patient samples with a sensitivity of 87% and specificity of 97%. The design principle can be rapidly adapted for other mutations (as shown also for E484K and T716I) highlighting the advantages of quick optimization and roll-out of CRISPR diagnostics (CRISPRDx) for disease surveillance even beyond COVID-19.

5. Rapid identification and tracking of SARS-CoV-2 variants of concern

Debojyoti Chakraborty*, Anurag Agrawal and Souvik Maiti.

*Corresponding author

The Lancet (2021)

During the advanced stages of the COVID pandemic, we have seen emergence of clinically important mutations that alter infectivity, severity, or immune susceptibility of SARS-CoV-2. Prominent examples include Asn501Tyr, His69_Val70del**,** and Glu484Lys mutations in the spike protein that have emerged independently in many global strains, such as those from the UK, South Africa, and Brazil, possibly driving resurgence of the pandemic when it appeared to be coming under control.[2] Some of these variants are likely to be resistant to vaccines and capable of reinfections. Future public health policy and pandemic response will need knowledge of the presence of such variants in the local population and their rapid identification on introduction into communities. This is not possible without local sequencing capacity, which is scarce in many vulnerable parts of the globe, where lockdown regulations are not strictly enforced and movement is unrestricted.[3] Even in those low-income and middle-income countries where such capacity is present and high alert is in place, the delay between positive diagnosis and sequencing results leads to an opportunity for a new variant to become established. Hard quarantine, involving strict confinement and isolation for all people with a positive test for SARS-CoV-2 who are at risk of carrying clinically important new variants, until cleared by sequencing, is a public health measure that is difficult to implement. This difficulty arises because long turnaround times that are associated with sequencing might lead to extra pressure on health-care authorities for institutional quarantine or follow-up after release in the event of variant detection in an individual. Diagnostic platforms that are based on sequencing and are suitable for use at the point of care, such as pore-based technologies, are anticipated to contribute substantially to this process in the near future, being capable of diagnosis, variant calling, genealogy, and novel mutant detection. Until then, we propose an alternative approach for low-resolution, yet accurate, early detection of specific variants of concern through clustered interspaced short palindromic repeats (CRISPR) diagnostics, which rely on the specific DNA interrogation properties of

enzymes, such as FnCas9, Cas12, or Cas13, to identify variants of concern through fluorescence or paper strip-based diagnosis (appendix).[4] Such tests are rapid, inexpensive, and especially suited for low-income countries. Even where sequencing is being done, CRISPR diagnostics can help to isolate variants in the first instance, which can then be sequenced to validate and map coexisting mutations (appendix). We have used this approach to identify the Asn501Tyr variant of concern, starting from RNA.[5] The variant detection strategy (ie, rapid variant assay) can be readily combined with a CRISPR diagnostics platform that is already approved as an equivalent diagnostic method to quantitative real-time PCR in India, providing diagnosis and identification of one variant of concern in less than 90 min from sample to result, at a test cost of less than US$15.

6. Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis

Azhar M, Phutela R, Kumar M, Ansari AH, Rauthan R, Gulati S, Sharma N, Sinha D, Sharma S, Singh S, Acharya S, Sarkar S, Paul D, Kathpalia P, Aich M, Sehgal P, Ranjan G, Bhoyar RC; Indian CoV2 Genomics & Genetic Epidemiology (IndiCovGEN) Consortium; Singhal K, Lad H, Patra PK, Makharia G, Chandak GR, Pesala B, **Chakraborty D***, Maiti S*.

*Co-Corresponding author

Biosensors and Bioelectronics (2021)

Rapid detection of DNA/RNA pathogenic sequences or variants through point-of-care diagnostics is valuable for accelerated clinical prognosis, as witnessed during the recent COVID-19 outbreak. Traditional methods relying on qPCR or sequencing are tough to implement with limited resources, necessitating the development of accurate and robust alternative strategies. Here, we report FnCas9 Editor Linked Uniform Detection Assay (FELUDA) that utilizes a direct Cas9 based enzymatic readout for detecting nucleobase and nucleotide sequences without trans-cleavage of reporter molecules. We also demonstrate that FELUDA is 100% accurate in detecting single nucleotide variants (SNVs), including heterozygous carriers, and present a simple web-tool JATAYU to aid end-users. FELUDA is semi-quantitative, can adapt to multiple signal detection platforms, and deploy for versatile applications such as molecular diagnosis during infectious disease outbreaks like COVID-19. Employing a lateral flow readout, FELUDA shows 100% sensitivity and 97% specificity across all ranges of viral loads in clinical samples within 1hr. In combination with RT-RPA and a smartphone application True Outcome Predicted via Strip Evaluation (TOPSE), we present a prototype for FELUDA for CoV-2 detection closer to home.

7. Voices of biotech research

Annabi N, Baker M, Boettiger A, **Chakraborty D**, Chen Y, Corbett KS, Correia B, Dahlman J, de Oliveira T, Ertuerk A, Yanik MF, Henaff E, Huch M, Iliev ID, Jacobs T, Junca H, Keung A, Kolodkin-Gal I, Krishnaswamy S, Lancaster M, Macosko E, Martínez-Núñez MA, Miura K, Molloy J, Cruz AO, Platt RJ, Posey AD Jr, Shao H, Simunovic M, Slavov N, Takebe T, Vandenberghe LH, Varshney RK, Wang J.

Nature Biotechnology (2021)

In this article, *Nature Biotechnology* asks a selection of faculty about the most exciting frontier in their field and the most needed technologies for advancing knowledge and applications. What will be the most important areas of research in biotech over the coming years? Which technologies will be most important to advance knowledge and applications in these areas? *Nature Biotechnology* reached out to a set of faculty doing outstanding work in research areas representative of the journal's remit and asked them to contribute their vision of where their fields are going.

Debojyoti Chakraborty: Detecting and correcting diseases requires precise molecular tools. The promise shown by ongoing gene editing trials for hemoglobinopathies has truly put CRISPR on track for therapeutic interventions. With the development of novel editors, cleavage-free genome engineering and robust delivery options, the coming years would see active clinical evaluation of in vivo genome editing — a challenging frontier. At the same time, the evolution of more sensitive and inexpensive CRISPR diagnostics platforms suited to a wide range of diseases would bring the benefits of early detection of disorders. This may be invaluable in developing countries.

8. Terminal Uridylyl Transferase Mediated Site-Directed Access to Clickable Chromatin Employing CRISPR-dCas9

George JT, Azhar M, Aich M, Sinha D, Ambi UB, Maiti S, Chakraborty D*, Srivatsan SG.

*Co-Corresponding author

J Am Chem Soc. (2020)

Locus-specific interrogation of target genes employing functional probes such as proteins and small molecules is paramount in decoding the molecular basis of gene function and designing tools to modulate its downstream effects. In this context, CRISPR-based gene editing and targeting technologies have proved tremendously useful, as they can be programmed to target any gene of interest by simply changing the sequence of the single guide RNA (sgRNA). Although these technologies are widely utilized in recruiting genetically encoded functional proteins, display of small molecules using CRISPR system is not well developed due to the lack of adequate techniques. Here, we have devised an innovative technology called sgRNA-Click (sgR-CLK) that harnesses the power of bioorthogonal click chemistry for remodeling guide RNA to display synthetic molecules on target genes. sgR-CLK employs a novel posttranscriptional chemoenzymatic labeling platform wherein a terminal uridylyl transferase (TUTase) was repurposed to generate clickable sgRNA of choice by site-specific tailoring of multiple azide-modified nucleotide analogues at the 3' end. The presence of a minimally invasive azide handle assured that the sgRNAs are indeed functional. Notably, an azide-tailed sgRNA targeting the telomeric repeat served as a Trojan horse on the CRISPR-dCas9 system to guide synthetic tags (biotin) site-specifically on chromatin employing copper-catalyzed or strain-promoted click reactions. Taken together, sgR-CLK presents a significant advancement on the utility of bioorthogonal chemistry, TUTase, and the CRISPR toolbox, which could offer a simplified solution for site-directed display of small molecule probes and diagnostic tools on target genes.

9. Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing

Acharya S, Mishra A, Paul D, Ansari AH, Azhar M, Kumar M, Rauthan R, Sharma N, Aich M, Sinha D, Sharma S, Jain S, Ray A, Jain S, Ramalingam S, Maiti S**, Chakraborty D**.

Proc Natl Acad Sci U S A. (2019)

Genome editing using the CRISPR/Cas9 system has been used to make precise heritable changes in the DNA of organisms. Although the widely used *Streptococcus pyogenes* Cas9 (SpCas9) and its engineered variants have been efficiently harnessed for numerous gene-editing applications across different platforms, concerns remain regarding their putative off-targeting at multiple loci across the genome. Here we report that *Francisella novicida* Cas9 (FnCas9) shows a very high specificity of binding to its intended targets and negligible binding to off-target loci. The specificity is determined by its minimal binding affinity with DNA when mismatches to the target single-guide RNA (sgRNA) are present in the sgRNA:DNA heteroduplex. FnCas9 produces staggered cleavage, higher homology-directed repair rates, and very low nonspecific genome editing compared to SpCas9. We demonstrate FnCas9-mediated correction of the sickle cell mutation in patient-derived induced pluripotent stem cells and propose that it can be used for precise therapeutic genome editing for a wide variety of genetic disorders.

10. lncRNA Panct1 Maintains Mouse Embryonic Stem Cell Identity by Regulating TOBF1 Recruitment to Oct-Sox Sequences in Early G1

Chakraborty D, Paszkowski-Rogacz M, Berger N, Ding L, Mircetic J, Fu J, Iesmantavicius V, Choudhary C, Anastassiadis K, Stewart AF, Buchholz F.

Cell Rep. (2017)

Long noncoding RNAs (lncRNAs) have been implicated in diverse biological processes, including embryonic stem cell (ESC) maintenance. However, their functional mechanisms remain largely undefined. Here, we show that the lncRNA Panct1 regulates the transient recruitment of a putative X-chromosome-encoded protein A830080D01Rik, hereafter referred to as transient octamer binding factor 1 (TOBF1), to genomic sites resembling the canonical Oct-Sox motif. TOBF1 physically interacts with Panct1 and exhibits a cell-cycle-specific punctate localization in ESCs. At the chromatin level, this correlates with its recruitment to promoters of pluripotency genes. Strikingly, mutating an octamer-like motif in Panct1 RNA abrogates the strength of TOBF1 localization and recruitment to its targets. Taken together, our data reveal a tightly controlled spatial and temporal pattern of lncRNA-mediated gene regulation in a cell-cycle-dependent manner and suggest that lncRNAs might function as barcodes for identifying genomic addresses for maintaining cellular states.

11. Combined RNAi and localization for functionally dissecting long noncoding RNAs

Chakraborty D, Kappei D, Theis M, Nitzsche A, Ding L, Paszkowski-Rogacz M, Surendranath V, Berger N, Schulz H, Saar K, Hubner N, Buchholz F

Nat Methods. (2012)

Whereas methods to comprehensively study cellular roles of protein-coding genes are available, techniques to systematically investigate long noncoding RNAs (lncRNAs), which have been implicated in diverse biological pathways, are limited. Here we report combined knockdown and localization analysis of noncoding RNAs (c-KLAN) that merges functional characterization and localization approaches to study lncRNAs. Using this technique we identified transcripts that regulate mouse embryonic stem cell identity.