may provide information that pertains to the function of just a single variant or edit, and may not provide information that pertains to the functions of any other variants. Single-cell RNA sequencing (scRNA-seq) is commercially available and may be used to obtain and sequence the transcriptome from a cell.

SUMMARY

[0005] Analyzing expression of protein-coding variants in cells is provided herein.

[0006] Some examples herein provide a method of analyzing expression of a protein-coding region of DNA in a cell. The method may include replacing a protein-coding region of the DNA in the cell with a donor vector including a variant of the protein-coding region and a first barcode identifying that variant. The cell may generate mRNA including an expression of the variant and an expression of the first barcode. The method may include coupling, to the mRNA, a second barcode corresponding to the cell. The method may include reverse transcribing the mRNA, having the second barcode coupled thereto, into cDNA. The method may include sequencing the cDNA. The method may include sequencing the donor vector or cDNA using amplicon sequencing. The method may include correlating the donor vector sequence and the cDNA sequence to identify the variant and the cell's expression of the variant. In one aspect, the present disclosure refers to a method of analyzing expression of a protein-coding region of DNA in a cell, the method comprising:

replacing a protein-coding region of the DNA in the cell with a donor vector comprising a variant of the protein-coding region and a first barcode identifying that variant,

wherein the cell generates mRNA comprising an expression of the variant and an expression of the first barcode;

coupling, to the mRNA, a second barcode corresponding to the cell;

reverse transcribing the mRNA, having the second barcode coupled thereto, into cDNA;

sequencing the cDNA;

sequencing the donor vector or cDNA using amplicon sequencing; and

correlating the donor vector sequence and the cDNA sequence to identify the variant and the cell's expression of the variant.

[0007] In some examples, the donor vector includes a promoter region. In some examples, the barcode is located between the promoter region and the variant. In some examples, the donor vector includes right and left homology arms, the variant and the first barcode being between the right and left homology arms. In some examples, the promoter region includes a reverse promotor region. In some examples, the reverse promoter region is disposed between the first barcode and the variant. In some examples, the expression of the variant of the protein-coding region is in the forward direction, and wherein the expression of the first barcode is in the reverse direction.

[0008] Additionally, or alternatively, in some examples, the method further includes using a first polymerase chain reaction (PCR) process to generate a first amplicon of the donor sequence that includes the variant, the first barcode, and the right homology arm and substantially excludes the left homology arm. The method may include using a second PCR process to generate a second amplicon of the first amplicon that includes the variant and the first barcode and substantially excludes the right and left homology arms. Additionally, or

protein-coding region and a first barcode identifying that variant. The cells may receive different variants than one another. The method may include obtaining mRNA from the cells. The mRNA from each cell may include an expression of the variant of the protein-coding region in that cell and an expression of the first barcode. The method may include coupling, to the mRNA from each cell, a second barcode corresponding to that cell. The method may include reverse transcribing the mRNA, having the second barcode coupled thereto, into cDNA. The method may include sequencing the cDNA. The method may include sequencing the donor vector or cDNA using amplicon sequencing. The method may include correlating the donor vector sequence and the cDNA sequence to identify the variant in each of the cells and that cell's expression of that variant. In one aspect, the present disclosure refers to a method of analyzing expression of a protein-coding region of DNA in a collection of cells, the method comprising:

replacing the initial protein coding-region of the DNA in each of the cells with a donor vector comprising a variant of the protein-coding region and a first barcode identifying that variant, wherein the cells receive different variants than one another;

obtaining mRNA from the cells, the mRNA from each cell comprising an expression of the variant of the protein-coding region in that cell and an expression of the first barcode;

coupling, to the mRNA from each cell, a second barcode corresponding to that cell;

reverse transcribing the mRNA, having the second barcode coupled thereto, into cDNA;

sequencing the cDNA;

sequencing the donor vector; and

correlating the donor vector sequence and the cDNA sequence to identify the variant in each of the cells and that cell's expression of that variant.

[0015] In some examples, the different variants are saturationally mutagenized.

[0016] Some examples herein provide a collection of cells. The DNA of each of the cells in the collection may include a variant of a protein-coding region and a first barcode identifying that variant. The cells may have different variants than one another.

[0017] In some examples, the different variants are saturationally mutagenized.

[0018] Some examples herein provide a collection of polynucleotides from a collection of cells. The polynucleotides may include first and second mRNA molecules from each of the cells. For each cell, the first mRNA molecule includes a first molecule of a barcode corresponding to that cell and an expression of a variant in that cell, and the second mRNA molecule includes the barcode corresponding to that cell and an expression of a first barcode corresponding to the variant. In one aspect, the present disclosure refers to a collection of polynucleotides from a collection of cells, the polynucleotides comprising first and second mRNA molecules from each of the cells, wherein, for each cell: the first mRNA molecule comprises a first molecule of a barcode corresponding to that cell and an expression of a variant in that cell, and the second mRNA molecule comprises the barcode corresponding to that cell and an expression of a first barcode corresponding to the variant.

[0019] In some examples, the different variants are saturationally mutagenized.

[0020] Some examples herein provide a method. The method may include providing a barcoded homology donor vector including a semi-random barcode on termini of a foreign transcript. The donor vector may include homology arms and mutations. The method may include knocking-in the barcoded homology donor vector to the vicinity of an exon to be edited to create a variant on the exon. The method may include cleaving the variant using a CRISPR-associated protein guide RNA ribonucleoprotein (Cas-gRNA RNP). In one aspect, the present disclosure refers to a method, comprising: providing a barcoded homology donor vector comprising a semi-random barcode on termini of a foreign transcript, the donor vector including homology arms and mutations; knocking-in the barcoded homology donor vector to the vicinity of an exon to be edited to create a variant on the exon; and cleaving the variant using a CRISPR-associated protein guide RNA ribonucleoprotein (Cas-gRNA RNP).

[0021] In some examples, the barcode is placed on UTR termini of the donor vector so that it may be expressed and detectable in scRNA-seq.

[0022] In some examples, the donor vector includes a puromycin resistance gene.

[0023] In some examples, providing the barcoded homology donor vector may include using a first polymerase chain reaction (PCR) to specifically amplify the knocked-in region with a genomically edited allele; using a second PCR, using the product of the first PCR as a template, to link the barcode with variants in an amplicon; and performing amplicon sequencing using the product from the second PCR.

[0024] In some examples, the amplicon sequencing covers both the barcode and the variants.

[0025] Some examples herein provide a method. The method may include adding semi-random variant barcodes to UTR regions of a saturationally mutagenized variant library. The method may include coupling cell barcodes to the variant barcodes. The method may include reading the variant barcodes out in scRNA-seq. The method may include linking the variant barcodes to the variants of the library using a separate sequencing operation. In one aspect, the present disclosure refers to a method, comprising: adding semi-random variant barcodes to UTR regions of a saturationally mutagenized variant library; coupling cell barcodes to the variant barcodes; reading the variant barcodes out in scRNA-seq; and linking the variant barcodes to the variants of the library using a separate sequencing operation.

[0026] In some examples, the semi-random variant barcode may be placed downstream of promoters or upstream of terminators of the variant library.

[0027] In some examples, linking the variant barcodes to the variants of the library may include generating tiled polymerase chain reaction (PCR) amplicons by using one set of primers to amplify the barcode on one side, and another set of primers to amplify the variants on the other side, such that each amplicon links a respective segment of the variant to the barcode.

[0028] Some examples herein provide a lentiviral vector including a semi-random barcode.

[0029] Some examples herein provide a composition that includes a plurality of lentiviral vectors, each of the lentiviral vectors including a different semi-random barcode.

[0030] In some examples, the composition further includes a mutagenically saturated variant library in contact with the plurality of lentiviral vectors.

[0031] It is to be understood that any respective features/examples of each of the aspects of the disclosure as described herein may be implemented together in any appropriate