

## **EVALUATION OF mRNA TRANSCRIPT LEVELS USING ddPCR FOR EARLY POOL AND SINGLE-CELL CLONE ASSESSMENT IN CELL LINE DEVELOPMENT**

### **FIELD OF DISCLOSURE**

[0001] The present application relates generally to methods for selecting single-cell clones and pools of cells which produce high amounts of a protein of interest. More specifically, the application relates to the use of digital PCR techniques, such as droplet digital PCR (ddPCR), for measuring mRNA transcript levels to select high expressing single-cell clones and pools of cells.

### **BACKGROUND**

[0002] Due to their broad applications, biologics are used worldwide in a variety of applications, such as therapeutics and diagnostics. Mammalian cell lines are the predominant expression systems for these biologics, with Chinese hamster ovary (CHO) cells being the predominate cellular factory. See Lalonde et al., 2017, J Biotechnol 251:128-140. Particularly with the advent of biosimilars, speed-to-market and cost-efficiency are now more important than ever before.

[0003] The costs of manufacturing biologics are high due to their complexity of production utilizing a multistep process involving the selection of optimal cell lines, culturing production cells in large quantities, and purification of the desired biologic from the cell harvest. While these costs are decreasing due to improvements in all facets of production, their costs can still be prohibitive in their widespread adoption as front-line therapies.

[0004] In order to make biological therapeutics more accessible to patients, decreasing the cost of goods for the manufacturing process is an attractive proposition. One area of significant interest is in the selection of optimal cell lines. Cell line development (CLD) plays a critical role in generating stable high expressing cell lines for biologics production. In a platform CLD workflow, single-cell clones and stable cell pools from transfected CHO cells are generated and then evaluated for protein production following a production process and then submitted for titer and product quality analysis, which can take several weeks. With an increase in novel and complex molecules formed by difficult to express domains or by assembly of several polypeptide chains, such as antibody chains, investigation of several variables to increase productivity, such as scaffold design, protein of interest nucleotide sequence, vector elements and medium or process parameters is required leading to high numbers of pools or single-cell clones to be generated and screened.

[0005] One approach to identify high expressing cell clones has been to automate titer measurement. See European Patent Application No. EP1901068A1. Another approach to identify optimal cell clones has been to apply Next-generation sequencing (NGS) to genomic DNA to identify genomic positions of transgene integration. See Stadermann et al., 2021, Biotechnology Bioengineering 119:868-880.

[0006] There still exists a need for methods for selecting single-cell clones and pools producing high amounts of proteins of interest.

**SUMMARY**

[0007] The present disclosure relates to a method of facilitating selection of a single cell clone or pools of cells for manufacturing an antigen-binding protein with two to four different antibody chains comprising: a) passaging a single cell clone or pool of cells expressing the antibody chains so that the cells have at least 85% viability and have a doubling time less than 40 hours for at least two consecutive passages; b) extracting mRNA from the cells; c) reverse transcribing the mRNA to generate cDNA; d) performing ddPCR to quantify the transcript levels of the antibody chains and e) eliminating single cell clones or pool of cells having at least the lowest 30<sup>th</sup> percentile of the sum of the transcript levels of all the antibody chains.

[0008] In certain embodiments, at least one of the antibody chains is a heavy chain or heavy chain-ScFv and/or at least one of the antibody chains is a light chain.

[0009] In certain embodiments, the single cell clone or pool of cells expresses an antigen-binding protein with two different antibody chains. In an aspect of this embodiment, the single cell clone or pool of cells expresses an antigen-binding protein with one heavy chain (HC) or heavy chain-scFv (HC-scFv) and one light chain (LC). In a sub-aspect, the method further comprises eliminating cell clones or pools of cells expressing an antigen-binding protein with one heavy chain (or HC-scFv) and one light chain where the ratio of the transcript levels for HC:LC (or HC-scFv:LC) or LC-HC (or LC-HC-scFv) is less than 0.2 or greater than 1.5. This eliminates cell clones or pools of cells having a ratio at either extreme (i.e., too low or too high).

[0010] In certain embodiments, the single cell clone or pool of cells expresses an antigen binding protein with three or four different antibody chains. In an aspect of this embodiment, the cells express an antigen-binding protein with two heavy chains (for example, HC1 and HC2 or HC1 and HC2-scFv) and one light chain (LC). In a sub-aspect, the method further comprises eliminating cell clones or pools of cells where the ratio of the transcript levels for HC1:HC2 or HC2:HC1 is less than 0.2 or greater than 1.5 or the ratio of HC1:HC2-scFv or HC2-scFv:HC1 is less than 0.2 or greater than 1.5. In an aspect of this embodiment, the cells express an antigen-binding protein with two heavy chains (HC1 and HC2) and two light chains (LC1 and LC2). In a sub-aspect, the method further comprises eliminating cell clones or pools of cells where the ratio of the transcript levels for HC1:HC2 or HC2:HC1 is less than 0.2 or greater than 1.5.

[0011] In any of the above embodiments, eliminating single cell clones or pools of cells eliminates those having at least the lowest 50<sup>th</sup> percentile of the sum of the transcript levels of all the chains.

[0012] In certain embodiments, the method further comprises f) subjecting the remaining single cell clones or pools of cells to a fed batch culture for at least an additional 8 days while maintaining the cell viability above 80%.

[0013] In certain embodiments, the ddPCR is one-step ddPCR or two-step ddPCR.

[0014] In certain embodiments, the cells are selected from the group consisting of CHO, HEK293, NS0 or Sp2/0 cells. In one aspect of this embodiment, the cells are CHO cells. In a sub-aspect of this embodiment, the CHO cells are DHFR- (dihydrofolate reductase minus) or GSKO (glutamine synthetase knock out).

[0015] In certain embodiments, the cells have been generated by using single cell printing.

[0016] The present disclosure also relates to a method of facilitating selection of a single cell clone or pools of cells expressing an antigen-binding protein with one or more antibody heavy chains and one or more antibody light chains for stable growth comprising: a) passaging the cells until the cells have a viability greater than 80% and have a doubling time less than 35 to 40 hours for at least two passages, b) extracting mRNA from the cells; c) reverse transcribing the mRNA to generate cDNA; d) performing ddPCR to quantify the transcript levels of the one or more heavy chains and the one or more light chains; and e) eliminating cell clones or pool of cells having at least the lowest 30<sup>th</sup> percentile of the sum of the transcript levels of all antibody chains. In one embodiment, the doubling time has less than 50% deviation between the at least two passages

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

[0017] Figs. 1A-C: Transcript levels vs. Titer (Fed Batch) in various antibody modalities. Fed batch transcript expression sums are represented by grey bars and normalized bulk titer is represented by black bars for (A) a 4-chain heteroIgG, (B) a 3-chain asymmetric antibody with HC-only binder (AmAb-HCAb or [Fab\*VH]-heteroFc), and (C) a 3-chain asymmetric C1mAb. Normalized titer values were obtained by dividing each bulk titer value by the sample set's lowest titer value. Top pool performers selected based on titer and product quality attributes are marked by black stars.

[0018] Figs. 2A-C: Transcript vs. Titer (Norm Passage) in various antibody modalities. Norm-passage transcript expression sum is represented by light-grey bars, fed batch expression sum is represented by grey bars, and the normalized bulk fed batch titer is represented by black bars for (A) a 4- chain heteroIgG, (B) AmAb-HCAb, and (C) a 3-chain asymmetric C1mAb. Normalized titer values were obtained by dividing each bulk titer value by the sample set's lowest titer value. Top pool performers (marked by black stars), selected based on titer and product quality attributes, rank at the top two-thirds most range of transcript expression.

[0019] Figs. 3A-B: Spearman matrix correlations between 3-chain AmAb-HCAb molecule transcript/chain ratio vs. product quality (where HC2 is the shorter heavy chain of the 3-chain AmAb-HCAb). Spearman analysis was performed to identify correlations between transcript expression and fed batch titer and product quality attributes from (A) fed batch and (B) norm-pass culture time points.

[0020] Fig. 4: Transcripts evaluated for bispecific molecules with different molecular structures (B2HmAb, C2HmAb, B2mAb, and C2mAb) with similar binding properties from norm-passage timepoint. Sum of heavy chain (HC) and light chain (LC) transcripts is represented by light-grey bars, normalized bulk fed batch titer is represented by black bars. Selected top pools marked by black star and

backup pools marked by light-grey star. Normalized titer values were obtained by dividing each bulk titer value by the sample set's lowest titer value. Each bar represents the average of four pools with identical vector configurations, each with three replicate expression values; error bars represent the standard deviation.

[0021] Figs. 5A-D: Clone Transcript Expression Analysis for 3-chain AmAb-HCAb using Quick RNA-Extract Method from norm-passage timepoint. (A) Clone expression sum is represented by grey bars, normalized bulk titer is represented by black bars), with the top pool performer (marked by black star). Error bars represent the standard deviation of three replicates. (B) Individual chain expression titer results; clone with the lowest titer marked by a light-grey star. (C) Spearman correlation matrix providing HC1/HC2 ratio (where HC2 is the shorter heavy chain of the 3-chain AmAb-HCAb). (D) HC1:HC2 ratio versus MP and LMW. Top clone marked by a black square. Normalized titer values were obtained by dividing each bulk titer value by the sample set's lowest titer value.

#### DETAILED DESCRIPTION

[0022] The present invention is based, in part, on the discovery that droplet digital PCR (ddPCR) quantitation of mRNA levels in a particular cell correlates with titers of secreted proteins. To streamline this highly resource-intensive process of clonal selection, a strategy for single-cell clone and/or early pool assessment was employed using one-step or two-step RT-ddPCR. One-step RT-ddPCR is a sensitive and high-throughput assay, performed in a single reaction vessel, which can determine mRNA transcript levels of the genes of interest (e.g., each encoding a heavy chain and light chain) within a short time frame. A two-step ddPCR separates the RT step and the PCR step in two different reaction vessels. To evaluate whether analysis of transcript levels of cultures derived from single-cell clones and/or stable pools at early timepoints could serve as a predictor for titer and product quality of the production process, samples from single-cell clones or normal passage pools in a fed batch process were analyzed for mRNA levels and compared using protein titers and final product quality results (e.g. size-exclusion chromatography (SEC) % high molecular weight (HMW) and/or %low molecular weight (LMW) species, non-reduced capillary electrophoresis (nr-CE) % pre peaks and/or post peaks). This assay can serve as a powerful tool to predict the top performing single-cell clones and/or pools and eliminate the poor performing single-cell clones and/or pools before reaching the fed batch production stage. This assay has the potential to save time, cost, and resources since it would further enable next-generation high-throughput cell line development (CLD) workflows before reaching the fed batch production stage. This assay has the potential to save time, cost, and resources since it would further enable next-generation high-throughput CLD workflows.

[0023] In the production of biologics, the invention finds particular utility in assessing transcript levels of a light chain and a heavy chain (from a product comprising both a heavy chain and light chain), or combinations of 3 or 4 heavy or light chains, in different vector configurations to find the best vector configurations.

[0024] Digital PCR is a nucleic acid amplification and detection method that is based on the dilution of template DNA into independent non-interacting partitions. See Sykes, et al., 1992, BioTechniques 13: 444-449. Following Poisson statistics with high dilutions of DNA template, each reaction is independently interrogated for the presence of a nucleic acid at single molecule sensitivity. In recent years several companies have produced commercially accessible ways to automate and expand the range of partitioning. This included droplet digital PCR (ddPCR) systems (e.g. Bio-Rad QX200) that disperse template DNA randomly into emulsion droplets of equal volume. See Hindson et al., 2011, Anal. Chem. 83:8604-8610.

[0025] Recently, digital PCR has seen wider use as an analytical tool for research and clinical applications. For example, digital PCR can be used as a robust tool for analyzing copy number variations seen in the amplifications or deletions of specific genes, detecting mutations, and quantifying specific nucleic acids species.

[0026] While the terminology used in this application is standard within the art, definitions of certain terms are provided herein to assure clarity and definiteness in the meaning of the claims. Units, prefixes, and symbols may be denoted in their SI (International System of Units) accepted form. Numeric ranges recited herein are inclusive of the numbers defining the range and include and are supportive of each integer within the defined range. The methods and techniques described herein are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2001) and Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992), and Harlow and Lane, Antibodies: A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990).

[0027] Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art.

[0028] All documents, or portions of documents, cited in this application, including but not limited to patents, patent applications, articles, books, and treatises, are hereby expressly incorporated by reference. What is described in an embodiment of the invention can be combined with other embodiments of the invention.

[0029] The present disclosure provides methods of selecting single-cell clones and/or cell pools expressing high levels of a "protein of interest". A "protein of interest" includes naturally occurring proteins, recombinant proteins, and engineered proteins (e.g., proteins that do not occur in nature and which have been designed and/or created by humans). A protein of interest can, but need not be, a protein that is known or suspected to be therapeutically relevant.

[0030] As used herein, the term “heterologous” used in connection with a nucleic acid means having a nucleic acid not naturally occurring within a host cell. This can include mutated sequences, e.g., sequences differing from the naturally occurring sequence. This can include sequences from other species. This can also include having a sequence at a different position in the genome than those naturally-occurring in the host cell. This generally does not include natural mutations that may occur in a host cell. A cell already containing a heterologous nucleic acid encoding a protein of interest, for example, by stable integration of an expression cassette, would be considered to contain a heterologous nucleic acid sequence. For clarity, a CHO cell or a derivative thereof (e.g., a DHFR- or GS knockout) having a nucleic acid encoding an antigen binding protein would be considered to have a heterologous nucleic acid.

[0031] As used herein, “cell culture” or “culture” or “Norm Passage” refers to the growth and propagation of cells outside of a multicellular organism or tissue. Suitable culture conditions for mammalian cells are known in the art. See e.g. *Animal cell culture: A Practical Approach*, D. Rickwood, ed., Oxford University Press, New York (1992). Mammalian cells may be cultured in suspension or while attached to a solid substrate. Fluidized bed bioreactors, hollow fiber bioreactors, roller bottles, shake flasks, or stirred tank bioreactors, with or without microcarriers, can be used.

[0032] The phrase “cell culture medium” (also called “culture medium,” “cell culture media,” “tissue culture media,”) refers to any nutrient solution used for growing cells, e.g., animal or mammalian cells, and which generally provides at least one or more components from the following: an energy source (usually in the form of a carbohydrate such as glucose); one or more of all essential amino acids, and generally the twenty basic amino acids, plus cysteine; vitamins and/or other organic compounds typically required at low concentrations; lipids or free fatty acids; and trace elements, e.g., inorganic compounds or naturally occurring elements that are typically required at very low concentrations, usually in the micromolar range.

[0033] A “fed-batch culture” refers to a form of suspension culture and means a method of culturing cells in which additional components are provided to the culture at a time or times subsequent to the beginning of the culture process. The provided components typically comprise nutritional supplements for the cells which have been depleted during the culturing process. Additionally, or alternatively, the additional components may include supplementary components (e.g., a cell-cycle inhibitory compound). A fed-batch culture is typically stopped at some point and the cells and/or components in the medium are harvested and optionally purified.

[0034] “Cell density” refers to the number of cells in a given volume of culture medium. “Viable cell density” refers to the number of live cells in a given volume of culture medium, as determined by standard viability assays (such as trypan blue dye exclusion method).

[0035] “Cell viability” means the ability of cells in culture to survive under a given set of culture conditions or experimental variations. The term also refers to that portion of cells which are alive at a particular time in relation to the total number of cells, living and dead, in the culture at that time.

[0036] “Titer” means the total amount of a polypeptide or protein of interest (which may be a naturally occurring or recombinant protein of interest) produced by a cell culture in a given amount of medium volume. Titer can be expressed in units of milligrams or micrograms of polypeptide or protein per milliliter (or other measure of volume) of medium. “Cumulative titer” is the titer produced by the cells during the course of the culture, and can be determined, for example, by measuring daily titers and using those values to calculate the cumulative titer.

[0037] As used herein, the term “host cell” is understood to include a cell that has been genetically engineered to express a polypeptide of interest. Genetically engineering a cell involves transfecting, transforming or transducing the cell with a nucleic acid encoding a recombinant polynucleotide molecule (a “gene of interest”), and/or otherwise altering (e.g., by homologous recombination and gene activation or fusion of a recombinant cell with a non-recombinant cell) so as to cause the host cell to express a desired recombinant polypeptide. Methods and vectors for genetically engineering cells and/or cell lines to express a polypeptide of interest are well known to those of skill in the art; for example, various techniques are illustrated in *Current Protocols in Molecular Biology*, Ausubel et al., eds. (Wiley & Sons, New York, 1988, and quarterly updates); Sambrook et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Laboratory Press, 1989); Kaufman, R.J., *Large Scale Mammalian Cell Culture*, 1990, pp. 15-69. The term includes the progeny of the parent cell, whether or not the progeny is identical in morphology or in genetic makeup to the original parent cell, so long as the gene of interest is present. A cell culture can comprise one or more host cells.

[0038] As used herein, the phrase “transcript expression sum” refers to the sum of transcript expression of the individual antibody chains.

[0039] Workflow

[0040] In a typical workflow, a cell sample is obtained from, for example, a frozen sample from either a single-cell clone or a stable pool and either from Norm Passage or from a fed-batch process. In certain embodiments, the cells are generated by single cell printing, for example, using commercially available equipment from vendors such as NamoCell (San Jose, CA). The cell sample contains nucleic acids expressing a protein of interest. The cell sample may be cultured in order to obtain sufficient numbers of cells for mRNA extraction. Generally, single-cell clones and pools are cultured until they achieve a viability of 85% (~2-4 weeks) with an average doubling time less than 35-40 hours for at least two or three consecutive passages.

[0041] The cell sample is subjected to mRNA extraction. See, e.g., Cheng et al., 2021, *Anal. Methods* 3:289-298; and Svec, 2013, *Front. Oncol.* 3:1-11. For instance, the sample may be treated to disrupt or lyse cells, for example by treating the samples with one or more detergents and/or denaturing agents (e.g., guanidinium agents). Nucleic acids may also be extracted from samples, for example, after detergent treatment and/or denaturing. Total nucleic acid extraction may be performed using known techniques, for example by non-specific binding to a solid phase (e.g., silica). See, e.g., U.S. Pat. Nos. 5,234,809,

6,849,431; 6,838,243; 6,815,541; and 6,720,166. Commercially available kits are available from companies such as Qiagen, Inc. (Germantown, MD). The Qiagen kit uses silica-based extraction based on size >200bp to select for mRNA. First, samples are treated with DNase to get rid of DNA and RNases are inactivated. mRNA is then enriched using the RNeasy silica column.

[0042] After mRNA extraction, the mRNA is reverse transcribed into cDNA using commonly known methods involving a reverse transcriptase. Alternatively, a cell sample can be subjected to a quick lysate kit for cDNA transcription without the need for RNA isolation. An example of such a kit is SingleShot from Bio-Rad (Hercules, CA).

[0043] Finally, the cDNA expression (and mRNA expression) is quantified using digital PCR.

[0044] Digital PCR

[0045] The methods described herein employ digital PCR methods. For digital PCR, a sample containing nucleic acids is separated into a large number of partitions before performing PCR. Partitioning can be achieved in a variety of ways known in the art, for example, by use of micro well plates, capillaries, emulsions, arrays of miniaturized chambers or nucleic acid binding surfaces. Separation of the sample may involve distributing any suitable portion including up to the entire sample among the partitions. Each partition includes a fluid volume that is isolated from the fluid volumes of other partitions. The partitions may be isolated from one another by a fluid phase, such as a continuous phase of an emulsion, by a solid phase, such as at least one wall of a container, or a combination thereof. In certain embodiments, the partitions may comprise droplets disposed in a continuous phase, such that the droplets and the continuous phase collectively form an emulsion.

[0046] The partitions may be formed by any suitable procedure, in any suitable manner, and with any suitable properties. For example, the partitions may be formed with a fluid dispenser, such as a pipette, with a droplet generator, by agitation of the sample (e.g., shaking, stirring, sonication, etc.), and the like. Accordingly, the partitions may be formed serially, in parallel, or in batch. The partitions may have any suitable volume or volumes. The partitions may be of substantially uniform volume or may have different volumes. Exemplary partitions having substantially the same volume are monodisperse droplets. Exemplary volumes for the partitions include an average volume of less than about 100, 10 or 1  $\mu$ L, less than about 100, 10, or 1 nL, or less than about 100, 10, or 1 pL, among others.

[0047] After separation of the sample, PCR is carried out in the partitions. The partitions, when formed, may be competent for performance of one or more reactions in the partitions. Alternatively, one or more reagents may be added to the partitions after they are formed to render them competent for reaction. The reagents may be added by any suitable mechanism, such as a fluid dispenser, fusion of droplets, or the like.

[0048] After PCR amplification, nucleic acids are quantified by counting the partitions that contain PCR amplicons for the target and/or reference polynucleotides. Partitioning of the sample allows

quantification of the number of different molecules by assuming that the population of molecules follows a Poisson distribution. For a description of digital PCR methods, see, e.g., Hindson et al., 2011, *Anal. Chem.* 83:8604-8610; Pohl and Shih, 2004, *Expert Rev. Mol. Diagn.* 4:41-47; Pekin et al., 2011, *Lab Chip* 11: 2156-2166; Pinheiro et al., 2012, *Anal. Chem.* 84:1003-1011; Day et al., 2013, *Methods* 59:101-107; herein incorporated by reference in their entireties.

[0049] Nucleic acid amplification of cDNA from mRNA transcripts involves digital PCR. Digital PCR may include any method, process, and/or protocol, using instruments and/or kits associated with performing such, that can discretely amplify and quantitate a nucleic acid(s) within individual partitions of a sample. The individual partitions for a digital PCR may be generated by a microfluidic process, such as by using a microfluidic device, and/or by a droplet generating process. Generation of individual partitions by a microfluidic process, such as by using a microfluidic device, and/or a droplet generating process to provide a plurality of partitions in the form of droplets and performing nucleic acid amplification thereon has been more particularly described in the art as "droplet digital PCR." The droplets generated for droplet digital PCR may be provided in, for example, a water-in-oil emulsion. In some embodiments, the methods, processes, and/or protocols, and instruments and/or kits for performing nucleic acid amplification on partitions in the form of droplets generated using a microfluidic device/process and/or a droplet generating process, are commercially available, for example, but not limited to, those provided by Bio-Rad, 10X Genomics, Qiagen, and/or ThermoFisher. In an exemplary embodiment, nucleic acid amplification includes droplet digital PCR (ddPCR™) using Bio-Rad's QX100™ or QX200™ Droplet Digital PCR systems, and analysis of nucleic acid amplification products produced by the same but is not limited thereto.

[0050] Reagents necessary for nucleic acid amplification, e.g., oligonucleotide primers, oligonucleotide probes, NTPs/dNTPs, polymerases, etc., may be provided to each of the plurality of partitions, and nucleic acid amplification may be performed on each of the plurality of partitions. In some embodiments, reagents for nucleic acid amplification may be bundled within a partition, e.g., an aqueous phase of an emulsion, e.g., a droplet. Partitions containing nucleic acid amplification reagents/components may be merged/fused with partitions containing a sample/nucleic acid extracted from a sample, and nucleic acid amplification performed on the merged/fused partitions.

[0051] A microfluidic device can be used to merge nucleic acid amplification reagent-containing partitions with sample/nucleic acid extract-containing partitions such that every sample/nucleic acid-containing partition includes nucleic acid amplification components/reagents. The number of partitions provided is not particularly limited. In some embodiments, the number of different sample/polynucleotide-containing partitions, e.g., in a sequencing reaction, can be about, more than about, less than about, or at least about 1000, 5000, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 200,000, 300,000, 400,000, 500,000, 600,000, 700,000, 800,000, 900,000, 1,000,000, 2,000,000, 3,000,000, 4,000,000, 5,000,000, 6,000,000, 7,000,000, 8,000,000, 9,000,000, or 10,000,000 partitions. In some embodiments, about 1000 to about 10,000, about 10,000 to about 100,000,

about 10,000 to about 500,000, about 100,000 to about 500,000, about 100,000 to about 1,000,000, about 500,000 to about 1,000,000, about 1,000,000 to about 5,000,000, or about 1,000,000 to about 10,000,000 partitions may be generated in the methods as set forth herein.

[0052] Methods for generating droplets are described, for example, in U.S. Patent Application Publication No. 2011/0053798, hereby incorporated herein by reference in its entirety. In some embodiments, for example, an inner droplet (or partition) can be fused with an outer droplet (or partition) by heating/cooling to change temperature, applying pressure, altering composition (e.g., via a chemical additive), applying acoustic energy (e.g., via sonication), exposure to light (e.g., to stimulate a photochemical reaction), applying an electric field, or any combination thereof. In some cases, the inner droplet may fuse to the outer droplet spontaneously. The treatment may be continuous or may vary temporally (e.g., pulsatile, shock, and/or repetitive treatment). The treatment may provide a gradual or rapid change in an emulsion parameter, to effect steady state or transient initiation of droplet fusion. The stability of the partitions, and their responsiveness to a treatment to induce droplet fusion, may be determined during their formation by selection of an appropriate surfactant type, surfactant concentration, critical micelle concentration, ionic strength, etc., for one or more phases of the inner/outer partition.

[0053] Fusion may occur spontaneously, such that no treatment, other than a sufficient time delay (or no delay), is necessary before processing fused droplets. Alternatively, the inner/outer droplet may be treated to controllably induce fusion of droplets to form assay mixtures.

[0054] The emulsion resulting from the fusion may be processed. Processing may include subjecting the emulsion to any condition or set of conditions under which at least one reaction of interest can occur (and/or is stopped), and for any suitable time period. Accordingly, processing may include maintaining the temperature of the emulsion near a predefined set point, varying the temperature of the emulsion between two or more predefined set points (such as thermally cycling the emulsion), exposing the emulsion to light, changing a pressure exerted on the emulsion, adding at least one chemical substance to the emulsion, applying an electric field to the emulsion, or any combination thereof, among others.

[0055] Signals may be detected from the emulsion after and/or during processing. The signals may be detected optically, electrically, chemically, or a combination thereof, among others. The detected signals may include test signals that correspond to at least one reaction of interest performed in the emulsion. Alternatively, or in addition, the detected signals may include code signals that correspond to codes present in the emulsion. Test signals and code signals generally are distinguishable and may be detected using the same or distinct detectors. For example, the test signals and code signals each may be detected as fluorescence signals, which may be distinguishable based on excitation wavelength (or spectrum), emission wavelength (or spectrum), and/or distinct positions in a fused droplet (e.g., code signals may be detectable as more localized than test signals with respect to fused droplets), among others. As another example, the test signals and code signals may be detected as distinct optical characteristics, such as test

signals detected as fluorescence and code signals detected as optical reflectance. As a further example, the test signals may be detected optically and the code signals electrically, or vice versa.

[0056] Partitions can be formed by any mode of separating that can be used for digital PCR. A partition can be a microfluidic channel, a well on a nano- or microfluidic device or on a microtiter plate, or a reaction chamber in a microfluidic device. A partition can be an area on an array surface. A partition can be an aqueous phase of an emulsion (e.g., a droplet).

[0057] The droplets described herein can include emulsion compositions (or mixtures of two or more immiscible fluids) described in U.S. Patent No. 7,622,280, and droplets generated by devices described in International Patent Application Publication No. WO 2010/036352. The term emulsion, as used herein, can refer to a mixture of immiscible liquids (such as oil and water). Oil-phase and/or water-in-oil emulsions can allow for the compartmentalization of reaction mixtures within aqueous droplets. In some embodiments, the emulsions can include aqueous droplets within a continuous oil phase. In other embodiments, the emulsions provided herein are oil-in-water emulsions, wherein the droplets are oil droplets within a continuous aqueous phase. The droplets provided herein can be used to prevent mixing between compartments, and each compartment can protect its contents from evaporation and coalescing with the contents of other compartments. One or more enzymatic reactions can occur in a droplet.

[0058] Methods of generating partitions/droplets, additives, preparation of primers, probes, polymerases for PCR, etc. are described, for example, in U.S. Patent No. 9,347,059.

[0059] Splitting a sample into small reaction volumes as described herein can enable the use of reduced amounts of reagents, thereby lowering the material cost of the analysis. Reducing sample complexity by partitioning can also improve the dynamic range of detection, since higher-abundance molecules can be separated from low-abundance molecules in different compartments, thereby allowing lower-abundance molecules greater proportional access to reaction reagents, which in turn can enhance the detection of lower-abundance molecules. A computer can be used to store and process the data. A computer-executable logic can be employed to perform such functions as grouping and/or analyzing the data. A computer can be useful for displaying, storing, retrieving, or calculating diagnostic results from the molecular profiling; displaying, storing, retrieving, or calculating raw data; or displaying, storing, retrieving, or calculating any sample or patient information useful in the methods described herein.

[0060] In some embodiments, a reference sequence that is present at two copies per diploid genome can be used, e.g., a housekeeping gene (e.g., a gene that is required for the maintenance of basic cellular function). Dividing the concentration or amount of the target by the concentration or amount of the reference can yield an estimate of the number of target copies per genome.

[0061] A housekeeping gene that can be used as reference in the methods described herein can include a gene that encodes a transcription factor, a transcription repressor, an RNA splicing gene, a translation factor, tRNA synthetase, RNA binding protein, ribosomal protein, RNA polymerase, protein processing protein, heat shock protein, histone, cell cycle regulator, apoptosis regulator, oncogene, DNA

repair/replication gene, carbohydrate metabolism regulator, citric acid cycle regulator, lipid metabolism regulator, amino acid metabolism regulator, nucleotide synthesis regulator, NADH dehydrogenase, cytochrome C oxidase, ATPase, mitochondrial protein, lysosomal protein, proteosomal protein, ribonuclease, oxidase/reductase, cytoskeletal protein, cell adhesion protein, channel or transporter, receptor, kinase, growth factor, tissue necrosis factor, etc. Specific examples of housekeeping genes that can be used in the methods described include, e.g., human hydroxymethylbilane synthase (HMBS), or BRAF genes.

[0062] Methods of detecting nucleic acids are well-known in the art and may include specific hybridization of the probe to nucleic acid sequences and, for example, detecting fluorescence emission from a fluorescently tagged or labeled oligonucleotide probe that hybridizes to a nucleic acid and/or a nucleic acid amplification product and gives off/releases fluorescence emission as a result of hybridization with the nucleic acid and/or the nucleic acid amplification product.

[0063] Any measurement tool known in the art may be used to analyze nucleic acids present in a sample and/or subject as described above, e.g., a spectrophotometer for absorption or calorimetric measurements, a fluorometer or flow cytometer for fluorescence measurements, a scintillation or gamma counter for radioactive measurements, and an automated cell counter, automated plate counter, or manual plate counter for cell number measurements. As a further example, a microwell reader can be used for fluorescence, absorbance or calorimetric measurements. In some embodiments, measurements to detect nucleic acids/nucleic acid amplification products may include analysis by digital PCR. In some embodiments, measurements to detect nucleic acids/nucleic acid amplification products may include analysis by droplet digital PCR using, for example, but not limited to, ddPCR™ using Bio-Rad's QX100™ or QX200™ Droplet Digital PCR systems. Measurement and analysis of amplification products may include analysis of 1D and/or 2D plots of fluorescence amplitude exhibited in a droplet digital PCR amplification.

[0064] In some embodiments, measurements to identify nucleic acids/nucleic acid amplification products may include analysis by digital PCR. In some embodiments, measurements to identify nucleic acids/nucleic acid amplification products may include analysis by droplet digital PCR using, for example, but not limited to, ddPCR™ using Bio-Rad's QX100™ or QX200™ Droplet Digital PCR systems. Measurement and analysis of amplification products may include analysis of 1D and/or 2D plots of fluorescence amplitude exhibited in a droplet digital PCR amplification. In some embodiments, amplification, detection, analysis and/or identification of nucleic acids/nucleic acid amplification products according to the concepts described herein may be performed as described by Maggi et al., 2020. J. Microbiol. Methods. 176, 106022.

[0065] In some embodiments, capturing and concentrating of nucleic acids and/or nucleic acid sequences using highly porous hydrogel particles can be accomplished using, for example, NANOTRAP® technology (Ceres Nanosciences, Inc. (Manassas, VA)), including NANOTRAP®

particles and/or NANOTRAP® magnetic particles, such as are described in, for example, U.S. Patent Application Publication Nos. 2009/0087346, 2009/0148961, 2012/0164749, and 2014/0045274. Nanoparticles, for example, NANOTRAP® particles and NANOTRAP® magnetic particles, functionalized with affinity baits, affinity groups, and/or affinity ligands that have very high affinities for the particular nucleic acids and/or nucleic acid sequences that are biomarkers for a particular fastidious microorganism may be used, for example, to capture the particular nucleic acids and/or nucleic acid sequences on the nanoparticles under conditions that the particular nucleic acids and/or nucleic acid sequences bind to/have high affinities for the nanoparticles, followed by separating/isolating the nanoparticles including the captured nucleic acids and/or nucleic acid sequences bound to the nanoparticles from the sample, and eluting the captured nucleic acids and/or nucleic acid sequences from the nanoparticles to provide an enrichment of/an enriched sample including the particular nucleic acids and/or nucleic acid sequences. The enriched sample may be suitable further analysis, such as detection and/or analysis of nucleic acids/nucleic acid sequences by the amplification, detection and identification methods described herein.

[0066] The methods described herein can utilize at least two sets of primers for separately amplifying a region of interest in a target polynucleotide sequence and a reference polynucleotide sequence used for comparison. Primers can be designed with a 5' tail sequence to allow discrimination of amplicons for the target polynucleotide and the reference polynucleotide. A first set of primers having tails of a certain length can be used for amplification of the region of interest in the target polynucleotide sequence, and a second set of primers having tails of a different length from that of the first set of primers can be used for amplification of the reference polynucleotide sequence, such that amplifying the nucleic acids with the two sets of primers results in amplicons of the target polynucleotide and the reference polynucleotide having detectably different lengths.

[0067] A fluorescent DNA dye can be used for detection of the amplicons resulting from the PCR reaction. Any DNA dye can be used that binds nonspecifically to DNA (i.e. binds DNA of any sequence), which permits discrimination of the amplicons by length and quantitative measurement of the nucleic acids. Exemplary DNA dyes that can be used include EvaGreen dye (EG) dye, SYBR green, SYBR green II, SYBR gold, Oxazole yellow (YO), YOYO, Thiazole orange (TO), PicoGreen (PG), and SYTO dyes.

[0068] Primers can be designed with a region that is complementary to a portion of the template nucleic acid to be amplified to allow initiation of the polymerization reaction. Noncomplementary nucleotides can be added to the 5' ends of primers to create the 5' tails. Typically, the lengths of primers range between 7-100 nucleotides in length, such as 15-60, 20-40, and so on, more typically in the range of between 20-40 nucleotides long, and any length between the stated ranges.

[0069] It is desirable that the amplification efficiencies of the target and reference sequences be similar or approximately equal, in order to allow comparison for quantitative analyses. For this reason, primers and amplification conditions should be selected to obtain this result. For example, the lengths of the

primers can be extended or shortened at the 5' end or the 3' end to produce primers with desired melting temperatures. Although the complimentary regions of primers designed for amplification of the target and reference sequences may be the same length, their non-complimentary tail sequences will generally have different lengths to allow discrimination of amplicons for the target and reference polynucleotides. Primers with shorter tails can be designed with a higher GC content, whereas primers with longer tails can be designed with a higher AT content to approximately match melting temperatures for the primer-template complexes for target and reference polynucleotides, preferably within 3°C of each other. The amplification efficiency of any pair of primers can be readily determined using routine techniques (see e.g., Furtado et al., "Application of real-time quantitative PCR in the analysis of gene expression." DNA amplification: Current Technologies and Applications. Wymondham, Norfolk, UK: Horizon Bioscience p. 131-145 (2004)).

[0070] Primers can readily be synthesized by standard techniques, e.g., solid phase synthesis via phosphoramidite chemistry, as disclosed in U.S. Pat. Nos. 4,458,066 and 4,415,732, incorporated herein by reference; Beaucage et al., *Tetrahedron* (1992) 48:2223-2311; and Applied Biosystems User Bulletin No. 13 (1 Apr. 1987). Other chemical synthesis methods include, for example, the phosphotriester method described by Narang et al., *Meth. Enzymol.* (1979) 68:90 and the phosphodiester method disclosed by Brown et al., *Meth. Enzymol.* (1979) 68:109. Poly(A) or poly(C), or other non-complementary nucleotide extensions may be incorporated into oligonucleotides using these same methods. Hexaethylene oxide extensions may be coupled to the oligonucleotides by methods known in the art. Cload et al., *J. Am. Chem. Soc.* (1991) 113:6324-6326; U.S. Pat. No. 4,914,210 to Levenson et al.; Durand et al., *Nucleic Acids Res.* (1990) 18:6353-6359; and Horn et al., *Tet. Lett.* (1986) 27:4705-4708.

[0071] In addition, one or more PCR additives or enhancing agents may be included to improve the yield of the amplification reaction, for example, by reducing secondary structure in a nucleic acid or mispriming events. Such additives or enhancing agents include, but are not limited to, dimethyl sulfoxide (DMSO), N,N,N-trimethylglycine (betaine), formamide, glycerol, nonionic detergents (e.g., Triton X-100, Tween 20, and Nonidet P-40 (NP-40)), 7-deaza-2'-deoxyguanosine, bovine serum albumin, T4 gene 32 protein, polyethylene glycol, 1,2-propanediol, and tetramethylammonium chloride.

[0072] As explained above, primers designed with 5'-tails may be used in polymerase chain reaction (PCR)-based techniques to distinguish amplicons resulting from different polynucleotide sequences based on length. PCR can be used for amplifying a desired target nucleic acid sequence contained in a nucleic acid molecule or mixture of molecules. In PCR, a pair of primers is employed in excess to hybridize to the complementary strands of the target nucleic acid. The primers are each extended by a polymerase using the target nucleic acid as a template. The extension products become target sequences themselves after dissociation from the original target strand. New primers are then hybridized and extended by a polymerase, and the cycle is repeated to geometrically increase the number of target sequence molecules. The PCR method for amplifying target nucleic acid sequences in a sample is well known in the art and has been described in, e.g., Innis et al. (eds.) *PCR Protocols* (Academic Press, N Y 1990); Taylor (1991)

Polymerase chain reaction: basic principles and automation, in PCR: A Practical Approach, McPherson et al. (eds.) IRL Press, Oxford; Saiki et al. (1986) Nature 324:163; as well as in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,889,818, all incorporated herein by reference in their entireties.

[0073] In particular, PCR uses relatively short oligonucleotide primers which flank the target nucleotide sequence to be amplified, oriented such that their 3' ends face each other, each primer extending toward the other. The polynucleotide sample is extracted and denatured, preferably by heat, and hybridized with the first and second primers that are present in molar excess. Polymerization is catalyzed in the presence of the four deoxyribonucleotide triphosphates (dNTPs—dATP, dGTP, dCTP and dTTP) using a primer- and template-dependent polynucleotide polymerizing agent, such as any enzyme capable of producing primer extension products, for example, *E. coli* DNA polymerase I, Klenow fragment of DNA polymerase I, T4 DNA polymerase, thermostable DNA polymerases isolated from *Thermus aquaticus* (Taq), available from a variety of sources (for example, Perkin Elmer), *Thermus thermophilus* (United States Biochemicals), *Bacillus stearothermophilus* (Bio-Rad), or *Thermococcus litoralis* ("Vent" polymerase, New England Biolabs). This results in two "long products" which contain the respective primers at their 5' ends covalently linked to the newly synthesized complements of the original strands. The reaction mixture is then returned to polymerizing conditions, e.g., by lowering the temperature, inactivating a denaturing agent, or adding more polymerase, and a second cycle is initiated. The second cycle provides the two original strands, the two long products from the first cycle, two new long products replicated from the original strands, and two "short products" replicated from the long products. The short products have the sequence of the target sequence with a primer at each end. On each additional cycle, an additional two long products are produced, and a number of short products equal to the number of long and short products remaining at the end of the previous cycle. Thus, the number of short products containing the target sequence grows exponentially with each cycle. Preferably, PCR is carried out with a commercially available thermal cycler, e.g., Perkin Elmer. The amplification products can be detected in solution or using solid supports.

[0074] Applications

[0075] The methods described herein can be used to quantitate mRNA transcript levels of heterologous sequences encoding a protein of interest. Gene copy number and mRNA transcript levels have been shown to be correlated with antibody production. See Jiang et al., 2006, Biotechnology Progress 22:313-318. In certain embodiments, the protein of interest will be an antigen binding protein with a heavy chain and a light chain. In such embodiments, the mRNA transcript levels can be measured for both the heavy chain and light chain. In such embodiments, it may be desirable to obtain a ratio of 1:1 for the heavy chain and light chain in order to achieve optimal formation of antibody. In other embodiments, the protein of interest will be an antigen binding protein with two heavy chains (HC1 and HC2) and one light chain (LC). In yet other embodiments, the protein of interest will be an antigen binding protein with two heavy chains (HC1 and HC2) and two light chains (LC1 and LC2). For three- and four-chain molecules, any combination of ratios of transcript levels can be used to assess potential candidates, e.g., HC1:HC2.

[0076] Due to possible issues with individual chain(s), such as antibody chains, that may affect product quality (e.g., a product-related impurity resulting from improperly folded domains that may lead to protein aggregation (HMW impurities), incomplete molecules (half molecules), or improperly assembled chains (incorrect pairing of heavy chains, heavy chain to light chain (LMW impurities)), efficiency of transcription to translation, secretion efficiency, etc. and the possibility of missing an optimal ratio for a specific chain pair, a preferred workflow selects and propagates those cells producing product that meet process and product quality requirements. In one workflow, cells expressing the lowest levels of transcripts are eliminated from further processing. For example, single-cell clones or pools of cells with the lowest 80<sup>th</sup>, 70<sup>th</sup>, 60<sup>th</sup>, 50<sup>th</sup>, 40<sup>th</sup>, or 30<sup>th</sup> percentile of transcript levels can be eliminated from the workflow.

[0077] Since having widely disparate ratios of chains, such as antibody chains, is indicative of a low likelihood of single-cell clones or pool of cells being a high expresser of the desired protein of interest and/or a high likelihood of producing high levels of product-related impurities, single-cell clones or pools of cells where the ratio of the transcript levels for any two antibody chains is less than 0.2 or greater than 1.5, can be eliminated. Antibody chain ratios for 3 or 4 chain would be the same for any two chains. Chain ratios of expressed polypeptides, such as antibody chains, can be measured using techniques known in the art. After a percentage of the poorer expressing single-cell clones or pools of cells are eliminated, the remaining single-cell clones or pools of cells are cultured by performing fed-batch for at least an additional 8, 9, or 10 days to assess the best performing ~~cell~~ single-cell clones or pools of cells. Preferably the viability of the cells is above 80%.

[0078] As discussed in greater detail below, an expression vector can be optimized to achieve the desired ratio of heavy chain to light chain through the use of different promoter combinations, signal sequences, enhancer elements, and the like.

[0079] Due to the design of the expression vectors, it is expected that the protein of interest will be expressed at high transcript levels. Transcript expression is normalized to transcript expression of a reference housekeeping gene so that apparent differences in RNA levels in different samples are not distorted by differences in sample amounts. Expression of each chain can be summed in the same sample and compared across different samples to determine relative expression.

[0080] The methods described herein can be adapted for multiplex PCR, for example, to detect and/or quantify multiple target polynucleotides simultaneously. Thus, a plurality of primer sets, comprising forward and reverse primers, can be used in each reaction mixture, each set of primers directed to different target polynucleotide sequences and comprising 5' tails of detectably different lengths to allow the amplicons produced from the different target polynucleotides to be distinguished based on the intensity of their fluorescent signals after binding a nonspecific fluorescent DNA dye. In some cases, multiplexing is performed such that all species of various target polynucleotides of interest can be simultaneously detected and/or quantified from a single sample. In certain embodiments, a plurality of primers sets is

used to simultaneously detect and/quantify different target alleles at the same locus or at different target loci. In other embodiments, a plurality of primers sets is used to simultaneously quantify transcript level expression of different target polynucleotide sequences (e.g., LC, HC).

[0081] A specific embodiment is a method wherein at least one step is performed in multi-well plates as well as a method wherein at least step b) is performed in multi-well plates. The multi-well plates can be 96-well plates or 384-well plates, preferably 384-well plates.

[0082] A further embodiment is a method wherein the single-cell clones or pools of cells are monitored over a period of time that is sufficient to obtain batch-like titer curves, preferably over a period of 5-15 days with samples taken every 2-3 days.

[0083] A further embodiment is a method wherein steps a) to e) are performed in 96-well plates and steps g) to j) are performed in 384-well plates.

[0084] For ddPCR, sample tracking can be ensured by barcoded plates and barcode reader using methods well known in the art.

#### **Generation of Mammalian Host Cells Expressing a Protein of Interest**

[0085] Expression of a protein of interest in a cell can be achieved by well-known methods, either transiently or by stable expression (Davis *et al.*, Basic Methods in Molecular Biology, 2<sup>nd</sup> ed., Appleton & Lange, Norwalk, Conn., 1994; Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001).

[0086] Methods for stable integration are well known in the art. Briefly, stable integration is commonly achieved by transiently introducing a heterologous polynucleotide or a vector containing the heterologous polynucleotide into the host cell, which facilitates the stable integration of said heterologous polynucleotide into the cell genome. Typically, the heterologous polynucleotide is flanked by homology arms, i.e., sequences homologous to the region upstream and downstream to the integration site. Before their introduction into the mammalian host cell, circular vectors may be linearized to facilitate integration into the cell genome. Methods for the introduction of vectors into cells are well known in the art and include transfection with biological methods, such as viral delivery, with chemical methods, such as using cationic polymers, calcium phosphate, cationic lipids or cationic amino acids; with physical methods, such as electroporation or microinjection; or with mixed approaches, such as protoplast fusion.

[0087] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection

(e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die), among other methods.

[0088] A specific method of stable integration uses recombinase mediated cassette exchange (RMCE; Bode and Baer, 2001, *Curr Opin Biotechnol.* 12:473-80, and Bode et al., 2000, *Biol. Chem.* 381:801- 813) for site-specific integration in the genome (also termed "targeted integration"). Site-specific recombinases such as Flp and Cre mediate recombination between two copies of their target sequence termed FRT and loxP, respectively. The use of two incompatible target sequences, for example FRT in combination with F3 (Schlake and Bode, 1994, *Biochemistry*, 33:12746-51) as well as inverted recognition target sites (Feng et al., 1999, *J. Mol. Biol.* 292:779-85) allows the insertion of DNA segments into a predefined chromosomal locus carrying target sequences in a similar configuration. See also European Patent No. EP1781796B1 and European Patent Application Publication No. EP2789691A1.

[0089] Insertion of RMCE into a specific site in the genome can be mediated by nucleases (e.g., zinc finger protein (ZFP), transcription activator-like effector nuclease (TALEN), clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9)) that can be engineered to create single- and double-stranded breaks (SSBs/DSBs) in the genome. There are two major and distinct pathways to repair DSBs -- homologous recombination and non-homologous end-joining (NHEJ). Homologous recombination requires the presence of a homologous sequence as a template (e.g., "donor" containing RMCE) to guide the cellular repair process and the results of the repair are error-free and predictable. In the absence of a template (or "donor") sequence for homologous recombination, the cell typically attempts to repair the DSB via the unpredictable and error-prone process of non-homologous end-joining (NHEJ).

[0090] A vector may be any molecule or entity (e.g., nucleic acid, plasmid, bacteriophage, transposon, cosmid, chromosome, virus, virus capsid, virion, naked DNA, complexed DNA and the like) suitable for use to transfer and/or transport protein encoding information into a host cell and/or to a specific location and/or compartment within a host cell. Vectors can include viral and non-viral vectors, non-episomal mammalian vectors. Vectors are often referred to as expression vectors, for example, recombinant expression vectors and cloning vectors. The vector may be introduced into a host cell to allow replication of the vector itself and thereby amplify the copies of the polynucleotide contained therein. The cloning vectors may contain sequence components generally including, without limitation, an origin of replication, promoter sequences, transcription initiation sequences, enhancer sequences, and selectable markers. These elements may be selected as appropriate by a person of ordinary skill in the art.

[0091] Vectors are useful for transformation of a host cell and contain nucleic acid sequences that direct and/or control (in conjunction with the host cell) expression of one or more heterologous coding regions operatively linked thereto. An expression construct may include, but is not limited to, sequences that affect or control transcription, translation, and, if introns are present, affect RNA splicing of a coding region operably linked thereto. "Operably linked" means that the components to which the term is applied

are in a relationship that allows them to carry out their inherent functions. For example, a control sequence, e.g., a promoter, in a vector that is “operably linked” to a protein coding sequence are arranged such that normal activity of the control sequence leads to transcription of the protein coding sequence resulting in recombinant expression of the encoded protein.

[0092] Vectors may be selected to be functional in the particular host cell employed (i.e., the vector is compatible with the host cell machinery, permitting amplification and/or expression of the gene to occur). In some embodiments, vectors are used that employ protein-fragment complementation assays using protein reporters, such as dihydrofolate reductase (see, for example, U.S. Pat. No. 6,270,964). Suitable expression vectors are known in the art and are also commercially available.

[0093] Typically, vectors used in host cells will contain sequences for plasmid maintenance and for cloning and expression of exogenous nucleotide sequences. Such sequences will typically include one or more of the following nucleotide sequences: a promoter, one or more enhancer sequences, an origin of replication, transcriptional and translational control sequences, a transcriptional termination sequence, a complete intron sequence containing a donor and acceptor splice site, various pre- or pro-sequences to improve glycosylation or yield, a native or heterologous signal sequence (leader sequence or signal peptide) for polypeptide secretion, a ribosome binding site, a polyadenylation sequence, internal ribosome entry site (IRES) sequences, an expression augmenting sequence element (EASE), tripartite leader (TPA) and VA gene RNAs from Adenovirus 2, a polylinker region for inserting the polynucleotide encoding the polypeptide to be expressed, and a selectable marker element. Vectors may be constructed from a starting vector such as a commercially available vector, additional elements may be individually obtained and ligated into the vector. Methods used for obtaining each of the components are well known to one skilled in the art.

[0094] Vector components may be homologous (i.e., from the same species and/or strain as the host cell), heterologous (e.g., from a species other than the host cell species or strain), hybrid (i.e., a combination of flanking sequences from more than one source), synthetic or native. The sequences of components useful in the vectors may be obtained by methods well known in the art, such as those previously identified by mapping and/or by restriction endonucleases. In addition, they can be obtained by polymerase chain reaction (PCR) and/or by screening a genomic library with suitable probes.

[0095] A ribosome-binding site is usually necessary for translation initiation of mRNA and is characterized by a Shine-Dalgarno sequence (prokaryotes) or a Kozak sequence (eukaryotes). The element is typically located 3' to the promoter and 5' to the coding sequence of the polypeptide to be expressed.

[0096] An origin of replication aids in the amplification of the vector in a host cell. They may be included as part of commercially available prokaryotic vectors and may also be chemically synthesized based on a known sequence and ligated into the vector. Various viral origins (e.g., SV40, polyoma, adenovirus, vesicular stomatitis virus (VSV), or papillomaviruses such as HPV or BPV) are useful for cloning vectors in mammalian cells.

[0097] Transcriptional and translational control sequences for mammalian host cell expression vectors can be excised from viral genomes. Commonly used promoter and enhancer sequences are derived from polyoma virus, adenovirus 2, simian virus 40 (SV40), and human cytomegalovirus (CMV). For example, the human CMV promoter/enhancer of immediate early gene 1 may be used. See e.g. Patterson *et al.*, 1994, *Applied Microbiol. Biotechnol.* 40:691-98. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites can be used to provide other genetic elements for expression of a structural gene sequence in a mammalian host cell. Viral early and late promoters are particularly useful because both are easily obtained from a viral genome as a fragment, which can also contain a viral origin of replication (Fiers *et al.*, 1978, *Nature* 273:113; Kaufman, 1990, *Meth. in Enzymol.* 185:487-511). Smaller or larger SV40 fragments can also be used, provided the approximately 250 bp sequence extending from the Hind III site toward the BglI site located in the SV40 viral origin of replication site is included.

[0098] A transcription termination sequence is typically located 3' to the end of a polypeptide coding region and serves to terminate transcription. Usually, a transcription termination sequence in prokaryotic cells is a G-C rich fragment followed by a poly-T sequence. While the sequence is easily cloned from a library or even purchased commercially as part of a vector, it can also be readily synthesized using methods for nucleic acid synthesis known to those of skill in the art.

[0099] A selectable marker gene encoding a protein necessary for the survival and growth of a host cell grown in a selective culture medium. Typical selection marker genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, tetracycline, or kanamycin for prokaryotic host cells; (b) complement auxotrophic deficiencies of the cell; or (c) supply critical nutrients not available from complex or defined media. Specific selectable markers are the kanamycin resistance gene, the ampicillin resistance gene, and the tetracycline resistance gene. Advantageously, a neomycin resistance gene may also be used for selection in both prokaryotic and eukaryotic host cells.

[0100] Other selectable marker genes may be used to amplify the gene that will be expressed. Amplification is the process wherein genes that are required for production of a protein critical for growth or cell survival are reiterated in tandem within the chromosomes of successive generations of recombinant cells. Examples of suitable selectable markers for mammalian cells include glutamine synthase (GS), dihydrofolate reductase (DHFR), and promoterless thymidine kinase genes. Mammalian cell transformants are placed under selection pressure wherein only the transformants are uniquely adapted to survive by virtue of the selectable marker gene present in the vector. Selection pressure is imposed by culturing the transformed cells under conditions in which the concentration of selection agent in the medium is successively increased, thereby leading to the amplification of both the selectable marker gene and the DNA that encodes a protein of interest. As a result, increased quantities of a polypeptide of interest are synthesized from the amplified DNA.

[0101] In some cases, such as where glycosylation is desired in a eukaryotic host cell expression system, one may manipulate the various pre- or pro-sequences to improve glycosylation or yield. For example, one may alter the peptidase cleavage site of a particular signal peptide, or add prosequences, which also may affect glycosylation. The final protein product may have, in the -1 position (relative to the first amino acid of the mature protein), one or more additional amino acids incident to expression, which may not have been totally removed. For example, the final protein product may have one or two amino acid residues found in the peptidase cleavage site, attached to the amino-terminus. Alternatively, use of some enzyme cleavage sites may result in a slightly truncated form of the desired polypeptide if the enzyme cuts at such area within the mature polypeptide.

[0102] Expression and cloning will typically contain a promoter that is recognized by the host organism and operably linked to the molecule encoding a protein of interest. Promoters are untranscribed sequences located upstream (i.e., 5') to the start codon of a structural gene (generally within about 100 to 1000 bp) that control transcription of the structural gene. Promoters are conventionally grouped into one of two classes: inducible promoters and constitutive promoters. Inducible promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as the presence or absence of a nutrient or a change in temperature. Constitutive promoters, on the other hand, uniformly transcribe a gene to which they are operably linked, that is, with little or no control over gene expression. A large number of promoters, recognized by a variety of potential host cells, are well known.

[0103] Suitable promoters for use with mammalian host cells are well known and include, but are not limited to, those obtained from the genomes of viruses such as polyoma virus, fowlpox virus, adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, retroviruses, hepatitis-B virus, and Simian Virus 40 (SV40). Other suitable mammalian promoters include heterologous mammalian promoters, for example, heat-shock promoters and the actin promoter.

[0104] Additional promoters which may be of interest include, but are not limited to: SV40 early promoter (Benoist and Chambon, 1981, *Nature* 290:304-310); CMV promoter (Thomsen *et al.*, 1984, *Proc. Natl. Acad. U.S.A.* 81:659-663); the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto *et al.*, 1980, *Cell* 22:787-797); herpes thymidine kinase promoter (Wagner *et al.*, 1981, *Proc. Natl. Acad. Sci. U.S.A.* 78:1444-1445); glyceraldehyde-3-phosphate dehydrogenase (GAPDH); promoter and regulatory sequences from the metallothionein gene (Prinster *et al.*, 1982, *Nature* 296:39-42); and prokaryotic promoters such as the beta-lactamase promoter (Villa-Komaroff *et al.*, 1978, *Proc. Natl. Acad. Sci. U.S.A.* 75:3727-3731); or the tac promoter (DeBoer *et al.*, 1983, *Proc. Natl. Acad. Sci. U.S.A.* 80:21-25). Also of interest are the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: the elastase I gene control region that is active in pancreatic acinar cells (Swift *et al.*, 1984, *Cell* 38:639-646; Ornitz *et al.*, 1986, *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409; MacDonald, 1987, *Hepatology* 7:425-515); the insulin gene control region that is active in pancreatic beta cells (Hanahan, 1985, *Nature* 315:115-122); the immunoglobulin gene control region that is active in lymphoid cells (Grosschedl *et al.*,

1984, Cell 38:647-658; Adames *et al.*, 1985, Nature 318:533-538; Alexander *et al.*, 1987, Mol. Cell. Biol. 7:1436-1444); the mouse mammary tumor virus control region that is active in testicular, breast, lymphoid and mast cells (Leder *et al.*, 1986, Cell 45:485-495); the albumin gene control region that is active in liver (Pinkert *et al.*, 1987, Genes and Devel. 1:268-276); the alpha-feto-protein gene control region that is active in liver (Krumlauf *et al.*, 1985, Mol. Cell. Biol. 5:1639-1648; Hammer *et al.*, 1987, Science 253:53-58); the alpha 1-antitrypsin gene control region that is active in liver (Kelsey *et al.*, 1987, Genes and Devel. 1:161-171); the beta-globin gene control region that is active in myeloid cells (Mogram *et al.*, 1985, Nature 315:338-340; Kollias *et al.*, 1986, Cell 46:89-94); the myelin basic protein gene control region that is active in oligodendrocyte cells in the brain (Readhead *et al.*, 1987, Cell 48:703-712); the myosin light chain-2 gene control region that is active in skeletal muscle (Sani, 1985, Nature 314:283-286); and the gonadotropic releasing hormone gene control region that is active in the hypothalamus (Mason *et al.*, 1986, Science 234:1372-1378).

[0105] An enhancer sequence may be inserted into the vector to increase transcription by higher eukaryotes. Enhancers are cis-acting elements of DNA, usually about 10-300 bp in length, that act on the promoter to increase transcription. Enhancers are relatively orientation and position independent, having been found at positions both 5' and 3' to the transcription unit. Several enhancer sequences available from mammalian genes are known (e.g., globin, elastase, albumin, alpha-feto-protein and insulin). Typically, however, an enhancer from a virus is used. The SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer, and adenovirus enhancers known in the art are exemplary enhancing elements for the activation of eukaryotic promoters. While an enhancer may be positioned in the vector either 5' or 3' to a coding sequence, it is typically located at a site 5' from the promoter.

[0106] A sequence encoding an appropriate native or heterologous signal sequence (leader sequence or signal peptide) can be incorporated into an expression vector, to promote extracellular secretion of the protein of interest. The choice of signal peptide or leader depends on the type of host cells in which the protein of interest to be produced, and a heterologous signal sequence can replace the native signal sequence. Examples of signal peptides that are functional in mammalian host cells include the following: the signal sequence for interleukin-7 described in U.S. Patent No. 4,965,195; the signal sequence for interleukin-2 receptor described in Cosman *et al.*, 1984, Nature 312:768; the interleukin-4 receptor signal peptide described in EP Patent No. 0367 566; the type I interleukin-1 receptor signal peptide described in U.S. Pat. No. 4,968,607; the type II interleukin-1 receptor signal peptide described in EP Patent No. 0 460 846.

[0107] Additional control sequences shown to improve expression of heterologous genes from mammalian expression vectors include such elements as the expression augmenting sequence element (EASE) derived from CHO cells (Morris *et al.*, in Animal Cell Technology, pp. 529-534 (1997); U.S. Patent Nos. 6,312,951 B1, 6,027,915, and 6,309,841 B1) and the tripartite leader (TPL) and VA gene RNAs from Adenovirus 2 (Gingeras *et al.*, 1982, J. Biol. Chem. 257:13475-13491). The internal ribosome entry site (IRES) sequences of viral origin allow bicistronic mRNAs to be translated efficiently (Oh and

Sarnow, 1993, *Current Opinion in Genetics and Development* 3:295-300; Ramesh *et al.*, 1996, *Nucleic Acids Research* 24:2697-2700).

[0108] Following construction, one or more vectors may be inserted into a suitable cell for amplification and/or polypeptide expression. The transformation of an expression vector into a selected cell may be accomplished by well-known methods including transfection, infection, calcium phosphate co-precipitation, electroporation, nucleofection, microinjection, DEAE-dextran mediated transfection, cationic lipids mediated delivery, liposome mediated transfection, microprojectile bombardment, receptor-mediated gene delivery, delivery mediated by polylysine, histone, chitosan, and peptides. The method selected will in part be a function of the type of host cell to be used. These methods and other suitable methods are well known to the skilled artisan and are set forth in manuals and other technical publications, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2001).

[0109] The term “transformation” refers to a change in a cell's genetic characteristics, and a cell has been transformed when it has been modified to contain new DNA or RNA. For example, a cell is transformed where it is genetically modified from its native state by introducing new genetic material via transfection, transduction, or other techniques. Following transfection or transduction, the transforming DNA can recombine with that of the cell by physically integrating into a chromosome of the cell or can be maintained transiently as an episomal element without being replicated or can replicate independently as a plasmid. A cell is considered to have been “stably transformed” when the transforming DNA is replicated with the division of the cell.

[0110] The term “transfection” refers to the uptake of foreign or exogenous DNA by a cell. A number of transfection techniques are well known in the art and are disclosed herein. See, *e.g.*, Graham *et al.*, 1973, *Virology* 52:456; Sambrook *et al.*, 2001, *Molecular Cloning: A Laboratory Manual*, supra; Davis *et al.*, 1986, *Basic Methods in Molecular Biology*, Elsevier; Chu *et al.*, 1981, *Gene* 13:197.

[0111] The term “transduction” refers to the process whereby foreign DNA is introduced into a cell via viral vector. See Jones *et al.*, (1998). *Genetics: principles and analysis*. Boston: Jones & Bartlett Publ.

#### **Proteins of Interest**

[0112] A “protein of interest” includes naturally occurring proteins, recombinant proteins, and engineered proteins (*e.g.*, proteins that do not occur in nature and which have been designed and/or created by humans) or polypeptides. Polypeptides and proteins of interest can be of scientific or commercial interest. A protein of interest can, but need not be, a protein that is known or suspected to be therapeutically relevant.

[0113] Proteins of interest include, among other things, secreted proteins, non-secreted proteins, intracellular proteins or membrane-bound proteins. Polypeptides and proteins of interest can be produced by recombinant animal cell lines using cell culture methods and may be referred to as “recombinant

proteins". The expressed protein(s) may be produced intracellularly or secreted into the culture medium from which it can be recovered and/or collected. The term "isolated protein" or "isolated recombinant protein" refers to a polypeptide or protein of interest, that is purified away from proteins or polypeptides or other contaminants that would interfere with its therapeutic, diagnostic, prophylactic, research or other use. Proteins of interest include proteins that exert a therapeutic effect by binding a target, particularly a target among those listed below, including targets derived therefrom, targets related thereto, and modifications thereof.

[0114] Proteins of interest include "antigen-binding proteins". Antigen-binding protein refers to proteins or polypeptides that comprise an antigen-binding region or antigen-binding portion that has affinity for another molecule to which it binds (antigen). Antigen-binding proteins encompass antibodies, peptibodies, antibody fragments, antibody derivatives, antibody analogs, fusion proteins (including single-chain variable fragments (scFvs), double-chain (divalent) scFvs, and IgGscFv (see, e.g., Orcutt et al., 2010, *Protein Eng Des Sel* 23:221-228), hetero-IgG (see, e.g., Liu et al., 2015, *J Biol Chem* 290:7535-7562), muteins, and antibodies made by the XmAb<sup>®</sup> process (Xencor, Inc., Monrovia, CA). Examples of antigen binding proteins include a human antibody, a humanized antibody; a chimeric antibody; a recombinant antibody; a single chain antibody; a diabody; a triabody; a tetrabody; a Fab fragment; a F(ab')<sub>2</sub> fragment; an IgD antibody; an IgE antibody; an IgM antibody; an IgG1 antibody; an IgG2a antibody; an IgG3 antibody; or an IgG4 antibody, and fragments thereof. Also included are bispecific T cell engager molecules (BiTE<sup>®</sup>), bispecific T cell engager molecules having extensions, such as half-life extensions, for example HLE BiTE<sup>®</sup> molecules, HeteroIg BiTE<sup>®</sup> and others, chimeric antigen receptors (CARs, CAR Ts), and T cell receptors (TCRs). As used herein, B1mAb refers to [Fab-scFv\*Fab]-heteroFc, B1HmAb refers to [Fab-VH\*Fab]-heteroFc; B2mAb refers to a Fab-scFv-Fc; B2HmAb refers to Fab-VH-Fc; C1mAb refers to a Fab-heteroFc-[scFv\*]; C2mAb refers to a IgG-scFv; C2HmAb refers to a IgG-VH.

[0115] As used herein, the term "antigen binding protein" is used in its broadest sense and means a protein comprising a portion that binds to an antigen or target and, optionally, a scaffold or framework portion that allows the antigen binding portion to adopt a conformation that promotes binding of the antigen binding protein to the antigen. The antigen binding protein can comprise, for example, an alternative protein scaffold or artificial scaffold with grafted CDRs or CDR derivatives. Such scaffolds include, but are not limited to, antibody-derived scaffolds comprising mutations introduced to, for example, stabilize the three-dimensional structure of the antigen binding protein as well as wholly synthetic scaffolds comprising, for example, a biocompatible polymer. See, e.g., Korndorfer et al., 2003, *Proteins: Structure, Function, and Bioinformatics*, 53(1):121-129; Roque et al., 2004, *Biotechnol. Prog.* 20:639-654. In addition, peptide antibody mimetics ("PAMs") can be used, as well as scaffolds based on antibody mimetics utilizing fibronectin components as a scaffold.

[0116] An antigen binding protein can comprise one or more antibody chains. A typical antibody has four antibody chains from the expression of one heavy chain and one light chain. As used herein, an

“antibody chain” or “chain” refers to antibody light chains, antibody heavy chains, antibody light chain fusion proteins, and antibody heavy chain fusion proteins, scFv-Fc fusion, VHH fusion, and the like. The terms “antibody heavy chain” and “antibody light chain” have the standard meaning in the art and include, for example, the various antibody heavy and light chains described elsewhere herein (e.g. heavy and light chains of IgG1, IgG2, IgG3, and IgG4 mAbs). The terms “antibody heavy chain” and “antibody light chain” include both standard full-length antibody heavy chains and light chains, as well as derivatives thereof that contain at the respective variable region (VL or VH). The terms “antibody heavy chain fusion” and “antibody heavy chain fusion protein” refer to a polypeptide that contains an antibody heavy chain covalently linked to one or more additional proteins or peptides. For example, an “antibody heavy chain fusion protein” can be an antibody heavy chain covalently linked to a cytokine. The linkage may be direct, or via a peptide linker (e.g. a glycine-serine linker). In an antibody heavy chain fusion protein, the antibody heavy chain may be linked to additional protein(s) at the N-terminus or the C-terminus of the heavy chain (or both locations). The terms “antibody light chain fusion protein” and “antibody light chain fusion” have the same meaning as described immediately above for “antibody heavy chain fusion protein”, except for an antibody light chain. As used herein, an “antibody fusion protein” refers to an antibody as provided herein which is covalently linked to one or more additional proteins or polypeptides (e.g. via a heavy chain or light chain of the antibody). Thus, an antibody fusion protein contains at least an antibody heavy chain fusion protein or an antibody light chain fusion protein as one of the polypeptides of the antibody fusion protein. Most commonly, an antibody fusion protein is a molecule that contains two antibody light chains, one antibody heavy chain, and one antibody heavy chain fusion protein, such that the additional protein is linked to one of the heavy chains of the antibody.

[0117] An antigen binding protein can have, for example, the structure of a naturally occurring immunoglobulin. An “immunoglobulin” is a tetrameric molecule. In a naturally occurring immunoglobulin, each tetramer is composed of two identical pairs of polypeptide chains, each pair having one “light” chain (about 25 kDa) and one “heavy” chain (about 50-70 kDa). The amino-terminal portion of each chain includes a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The carboxy-terminal portion of each chain defines a constant region primarily responsible for effector function. Human light chains are classified as kappa and lambda light chains. Heavy chains are classified as mu, delta, gamma, alpha, or epsilon, and define the antibody’s isotype as IgM, IgD, IgG, IgA, and IgE, respectively.

[0118] Naturally occurring immunoglobulin chains exhibit the same general structure of relatively conserved framework regions (FR) joined by three hypervariable regions, also called complementarity determining regions or CDRs. From N-terminus to C-terminus, both light and heavy chains comprise the domains FR1, CDR1, FR2, CDR2, FR3, CDR3 and FR4. The assignment of amino acids to each domain can be done in accordance with the definitions of Kabat et al. in *Sequences of Proteins of Immunological Interest*, 5th Ed., US Dept. of Health and Human Services, PHS, NIH, NIH Publication no. 91-3242, (1991). As desired, the CDRs can also be redefined according to an alternative nomenclature scheme.

such as that of Chothia (see Chothia and Lesk, 1987, J. Mol. Biol. 196:901-917; Chothia et al., 1989, Nature 342:878-883 or Honegger and Pluckthun, 2001, J. Mol. Biol. 309:657-670).

[0119] In the context of the instant disclosure, an antigen binding protein is said to “specifically bind” or “selectively bind” its target antigen when the dissociation constant (KD) is  $\leq 10^{-8}$  M. The antibody specifically binds antigen with “high affinity” when the KD is  $\leq 5 \times 10^{-9}$  M, and with “very high affinity” when the KD is  $\leq 5 \times 10^{-10}$  M.

[0120] The term “antibody” includes reference to both glycosylated and non-glycosylated immunoglobulins of any isotype or subclass or to an antigen-binding region thereof that competes with the intact antibody for specific binding, unless otherwise specified. Additionally, the term “antibody” refers to an intact immunoglobulin or to an antigen binding portion thereof that competes with the intact antibody for specific binding, unless otherwise specified. Antigen binding portions can be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies and can form an element of a protein of interest. Unless otherwise specified, antibodies include human, humanized, chimeric, multi-specific, monoclonal, polyclonal, heteroIgG, bispecific, and oligomers or antigen binding fragments thereof. Antibodies include the IgG1-, IgG2-, IgG3- or IgG4-type. Also included are proteins having an antigen binding fragment or region such as Fab, Fab', F(ab')<sub>2</sub>, Fv, diabodies, Fd, dAb, maxibodies, single chain antibody molecules, single domain VHH, complementarity determining region (CDR) fragments, scFv, diabodies, triabodies, tetrabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to a target polypeptide.

[0121] An antigen binding protein can have one or more binding sites. If there is more than one binding site, the binding sites can be identical to one another or can be different. For example, a naturally occurring human immunoglobulin typically has two identical binding sites, while a “bispecific” or “bifunctional” antibody has two different binding sites.

[0122] A Fab fragment is a monovalent fragment having the VL, VH, CL and CH1 domains; a F(ab')<sub>2</sub> fragment is a bivalent fragment having two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment has the VH and CH1 domains; an Fv fragment has the VL and VH domains of a single arm of an antibody; and a dAb fragment has a VH domain, a VL domain, or an antigen-binding fragment of a VH or VL domain (U.S. Pat. Nos. 6,846,634, 6,696,245, U.S. Patent Application Publication Nos. 2005/0202512, 2004/0202995, 2004/0038291, 2004/0009507, 2003/0039958, Ward et al., 1989, Nature 341:544-546).

[0123] A single-chain antibody (scFv) is an antibody in which a VL region and a VH region are joined via a linker (e.g., a synthetic sequence of amino acid residues) to form a continuous protein chain wherein the linker is long enough to allow the protein chain to fold back on itself and form a monovalent antigen binding site (see, e.g., Bird et al., 1988, Science 242:423-26 and Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-83), U.S. Patent Nos. 7,741,465, and 6,319,494 as well as Eshhar et al., 1997, Cancer

Immunol Immunotherapy 45:131-136. An scFv retains the parent antibody's ability to specifically interact with target antigen.

[0124] Diabodies are bivalent antibodies comprising two polypeptide chains, wherein each polypeptide chain comprises VH and VL domains joined by a linker that is too short to allow for pairing between two domains on the same chain, thus allowing each domain to pair with a complementary domain on another polypeptide chain (see, e.g., Holliger et al., 1993, Proc. Natl. Acad. Sci. USA 90:6444-48; and Poljak et al., 1994, Structure 2:1121-23). If the two polypeptide chains of a diabody are identical, then a diabody resulting from their pairing will have two identical antigen binding sites. Polypeptide chains having different sequences can be used to make a diabody with two different antigen binding sites. Similarly, tribodies and tetrabodies are antibodies comprising three and four polypeptide chains, respectively, and forming three and four antigen binding sites, respectively, which can be the same or different.

[0125] For purposes of clarity, and as described herein, it is noted that an antigen binding protein can, but need not, be of human origin (e.g., a human antibody), and in some cases will comprise a non-human protein, for example a rat or murine protein, and in other cases an antigen binding protein can comprise a hybrid of human and non-human proteins (e.g., a humanized antibody).

[0126] A protein of interest can comprise a human antibody. The term "human antibody" includes all antibodies that have one or more variable and constant regions derived from human immunoglobulin sequences. In one embodiment, all of the variable and constant domains are derived from human immunoglobulin sequences (a fully human antibody). Such antibodies can be prepared in a variety of ways, including through the immunization with an antigen of interest of a mouse that is genetically modified to express antibodies derived from human heavy and/or light chain-encoding genes, such as a mouse derived from a Xenomouse®, UltiMab™, or Velocimmune® system, or a rat derived from UniRat®. Phage-based approaches can also be employed.

[0127] Alternatively, a protein of interest can comprise a humanized antibody. A "humanized antibody" has a sequence that differs from the sequence of an antibody derived from a non-human species by one or more amino acid substitutions, deletions, and/or additions, such that the humanized antibody is less likely to induce an immune response, and/or induces a less severe immune response, as compared to the non-human species antibody, when it is administered to a human subject. In one embodiment, certain amino acids in the framework and constant domains of the heavy and/or light chains of the non-human species antibody are mutated to produce the humanized antibody. In another embodiment, the constant domain(s) from a human antibody are fused to the variable domain(s) of a non-human species. Examples of how to make humanized antibodies can be found in U.S. Pat. Nos. 6,054,297, 5,886,152 and 5,877,293.

[0128] Also included are modified proteins, such as proteins modified chemically by a non-covalent bond, covalent bond, or both a covalent and non-covalent bond. Also included are proteins further comprising one or more post-translational modifications which may be made by cellular modification

systems or modifications introduced ex vivo by enzymatic and/or chemical methods or introduced in other ways.

[0129] Proteins of interest may also include recombinant fusion proteins comprising, for example, a multimerization domain, such as a leucine zipper, a coiled coil, an Fc portion of an immunoglobulin, and the like. Also included are proteins comprising all or part of the amino acid sequences of differentiation antigens (referred to as CD proteins) or their ligands or proteins substantially similar to either of these.

[0130] In some embodiments, proteins of interest may include colony stimulating factors, such as granulocyte colony-stimulating factor (G-CSF). Such G-CSF agents include, but are not limited to, Neupogen® (filgrastim) and Neulasta® (pegfilgrastim). Also included are erythropoiesis stimulating agents (ESA), such as Epogen® (epoetin alfa), Aranesp® (darbepoetin alfa), Dynepo® (epoetin delta), Mircera® (methoxy polyethylene glycol-epoetin beta), Hematide®, MRK-2578, INS-22, Retacrit® (epoetin zeta), Neorecormon® (epoetin beta), Silapo® (epoetin zeta), Binocrit® (epoetin alfa), epoetin alfa Hexal, Abseamed® (epoetin alfa), Ratioepo® (epoetin theta), Eporatio® (epoetin theta), Biopoin® (epoetin theta), epoetin alfa, epoetin beta, epoetin zeta, epoetin theta, and epoetin delta, epoetin omega, epoetin iota, tissue plasminogen activator, GLP-1 receptor agonists, as well as the molecules or variants or analogs thereof and biosimilars of any of the foregoing.

[0131] In some embodiments, proteins of interest may include proteins that bind specifically to one or more CD proteins, HER receptor family proteins, cell adhesion molecules, growth factors, nerve growth factors, fibroblast growth factors, transforming growth factors (TGF), insulin-like growth factors, osteoinductive factors, insulin and insulin-related proteins, coagulation and coagulation-related proteins, colony stimulating factors (CSFs), other blood and serum proteins blood group antigens; receptors, receptor-associated proteins, growth hormones, growth hormone receptors, T-cell receptors; neurotrophic factors, neurotrophins, relaxins, interferons, interleukins, viral antigens, lipoproteins, integrins, rheumatoid factors, immunotoxins, surface membrane proteins, transport proteins, homing receptors, addressins, regulatory proteins, and immunoadhesins.

[0132] In some embodiments proteins of interest bind to one of more of the following, alone or in any combination: CD proteins including but not limited to CD3, CD4, CD5, CD7, CD8, CD19, CD20, CD22, CD25, CD30, CD33, CD34, CD38, CD40, CD70, CD123, CD133, CD138, CD171, and CD174, HER receptor family proteins, including, for instance, HER2, HER3, HER4, and the EGF receptor, EGFRvIII, cell adhesion molecules, for example, LFA-1, Mol, p150,95, VLA-4, ICAM-1, VCAM, and alpha v/beta 3 integrin, growth factors, including but not limited to, for example, vascular endothelial growth factor ("VEGF"); VEGFR2, growth hormone, thyroid stimulating hormone, follicle stimulating hormone, luteinizing hormone, growth hormone releasing factor, parathyroid hormone, mullerian-inhibiting substance, human macrophage inflammatory protein (MIP-1-alpha), erythropoietin (EPO), nerve growth factor, such as NGF-beta, platelet-derived growth factor (PDGF), fibroblast growth factors, including, for instance, aFGF and bFGF, epidermal growth factor (EGF), Cripto, transforming growth factors (TGF),

including, among others, TGF- $\alpha$  and TGF- $\beta$ , including TGF- $\beta$ 1, TGF- $\beta$ 2, TGF- $\beta$ 3, TGF- $\beta$ 4, or TGF- $\beta$ 5, insulin-like growth factors-I and -II (IGF-I and IGF-II), des(1-3)-IGF-I (brain IGF-I), and osteoinductive factors, insulins and insulin-related proteins, including but not limited to insulin, insulin A-chain, insulin B-chain, proinsulin, and insulin-like growth factor binding proteins; (coagulation and coagulation-related proteins, such as, among others, factor VIII, tissue factor, von Willebrand factor, protein C, alpha-1-antitrypsin, plasminogen activators, such as urokinase and tissue plasminogen activator ("t-PA"), bombazine, thrombin, thrombopoietin, and thrombopoietin receptor, colony stimulating factors (CSFs), including the following, among others, M-CSF, GM-CSF, and G-CSF, other blood and serum proteins, including but not limited to albumin, IgE, and blood group antigens, receptors and receptor-associated proteins, including, for example, flk2/flt3 receptor, obesity (OB) receptor, growth hormone receptors, and T-cell receptors; neurotrophic factors, including but not limited to, bone-derived neurotrophic factor (BDNF) and neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6); relaxin A-chain, relaxin B-chain, and prorelaxin, interferons, including for example, interferon-alpha, -beta, and -gamma, interleukins (ILs), e.g., IL-1 to IL-10, IL-12, IL-15, IL-17, IL-23, IL-12/IL-23, IL-2Ra, IL-1-R1, IL-6 receptor, IL-4 receptor and/or IL-13 to the receptor, IL-13RA2, or IL-17 receptor, IL-1RAP; viral antigens, including but not limited to, an AIDS envelope viral antigen, lipoproteins, calcitonin, glucagon, atrial natriuretic factor, lung surfactant, tumor necrosis factor-alpha and -beta, enkephalinase, BCMA, IgKappa, ROR-1, ERBB2, mesothelin, RANTES (regulated on activation normally T-cell expressed and secreted), mouse gonadotropin-associated peptide, DNase, FR-alpha, inhibin, and activin, integrin, protein A or D, rheumatoid factors, immunotoxins, bone morphogenetic protein (BMP), superoxide dismutase, surface membrane proteins, decay accelerating factor (DAF), AIDS envelope, transport proteins, homing receptors, MIC (MIC-a, MIC-B), ULBP 1-6, EPCAM, addressins, regulatory proteins, immunoadhesins, antigen-binding proteins, somatropin, CTGF, CTLA4, eotaxin-1, MUC1, CEA, c-MET, Claudin-18, GPC-3, EPHA2, FPA, LMP1, MG7, NY-ESO-1, PSCA, ganglioside GD2, ganglioside GM2, BAFF, OPGL (RANKL), myostatin, Dickkopf-1 (DKK-1), Ang2, NGF, IGF-1 receptor, hepatocyte growth factor (HGF), TRAIL-R2, c-Kit, B7RP-1, PSMA, NKG2D-1, programmed cell death protein 1 and ligand, PD1 and PDL1, mannose receptor/hCG $\beta$ , hepatitis-C virus, mesothelin dsFv[PE38] conjugate, Legionella pneumophila (lly), IFN gamma, interferon gamma induced protein 10 (IP10), IFNAR, TALL-1, thymic stromal lymphopoietin (TSLP), proprotein convertase subtilisin/Kexin Type 9 (PCSK9), stem cell factors, Flt-3, calcitonin gene-related peptide (CGRP), OX40L,  $\alpha$ 4 $\beta$ 7, platelet specific (platelet glycoprotein IIb/IIIb (PAC-1), transforming growth factor beta (TGF $\beta$ ), Zona pellucida sperm-binding protein 3 (ZP-3), TWEAK, platelet derived growth factor receptor alpha (PDGFR $\alpha$ ), sclerostin, and biologically active fragments or variants of any of the foregoing.

[0133] In another embodiment, proteins of interest include abciximab, adalimumab, adecatumumab, aflibercept, alemtuzumab, alirocumab, anakinra, atacicept, basiliximab, belimumab, bevacizumab, biosozumab, blinatumomab, brentuximab vedotin, brodalumab, cantuzumab mertansine, canakinumab, cetuximab, certolizumab pegol, conatumumab, daclizumab, denosumab, eculizumab, edrecolomab,

efalizumab, epratuzumab, etanercept, evolocumab, galiximab, ganitumab, gemtuzumab, golimumab, ibritumomab tiuxetan, infliximab, ipilimumab, ixekizumab, lerdelimumab, lumiliximab, mapatumumab, motesanib diphosphate, muromonab-CD3, natalizumab, nesiritide, nimotuzumab, nivolumab, ocrelizumab, ofatumumab, omalizumab, oprelvekin, palivizumab, panitumumab, pembrolizumab, pertuzumab, pexelizumab, ranibizumab, rilatumumab, rituximab, romiplostim, romosozumab, sargamostim, tocilizumab, tositumomab, trastuzumab, ustekinumab, vedolizumab, visilizumab, volociximab, zanolimumab, zalutumumab, and biosimilars of any of the foregoing.

[0134] Proteins of interest according to the invention encompass all of the foregoing and further include antibodies comprising 1, 2, 3, 4, 5, or 6 of the complementarity determining regions (CDRs) of any of the aforementioned antibodies. One or more CDRs can be incorporated into a molecule either covalently or noncovalently to make it an antigen binding protein. An antigen binding protein can incorporate the CDR(s) as part of a larger polypeptide chain, can covalently link the CDR(s) to another polypeptide chain, or can incorporate the CDR(s) noncovalently. The CDRs permit the antigen binding protein to specifically bind to a particular antigen of interest. Also included are variants that comprise a region that is 70% or more, especially 80% or more, more especially 90% or more, yet more especially 95% or more, particularly 97% or more, more particularly 98% or more, yet more particularly 99% or more identical in amino acid sequence to a reference amino acid sequence of a protein of interest. Identity in this regard can be determined using a variety of well-known and readily available amino acid sequence analysis software. Preferred software includes those that implement the Smith-Waterman algorithms, considered a satisfactory solution to the problem of searching and aligning sequences. Other algorithms also may be employed, particularly where speed is an important consideration. Commonly employed programs for alignment and homology matching of DNAs, RNAs, and polypeptides that can be used in this regard include FASTA, TFASTA, BLASTN, BLASTP, BLASTX, TBLASTN, PROSRCH, BLAZE, and MPSRCH, the latter being an implementation of the Smith-Waterman algorithm for execution on massively parallel processors made by MasPar.

[0135] An “Fc” region, as the term is used herein, comprises two heavy chain fragments comprising the CH2 and CH3 domains of an antibody. The two heavy chain fragments are held together by two or more disulfide bonds and by hydrophobic interactions of the CH3 domains. Proteins of interest comprising an Fc region, including antigen binding proteins and Fc fusion proteins, form another aspect of the instant disclosure.

[0136] A “hemibody” is an immunologically functional immunoglobulin construct comprising a complete heavy chain, a complete light chain and a second heavy chain Fc region paired with the Fc region of the complete heavy chain. A linker can, but need not, be employed to join the heavy chain Fc region and the second heavy chain Fc region. In particular embodiments, a hemibody is a monovalent form of an antigen binding protein disclosed herein. In other embodiments, pairs of charged residues can be employed to associate one Fc region with the second Fc region. A hemibody can be a protein of interest in the context of the instant disclosure.

[0137] As used herein, the terms “a” and “an” mean one or more unless specifically indicated otherwise. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art.

[0138] All documents, or portions of documents, cited in this application, including but not limited to patents, patent applications, articles, books, and treatises, are hereby expressly incorporated by reference. What is described in an embodiment of the invention can be combined with other embodiments of the invention.

[0139] The present invention is not to be limited in scope by the specific embodiments described herein that are intended as single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

## EXAMPLES

### Example 1

[0140] A predictive ddPCR assay for screening pools in cell line development prior to fed batch was developed. Samples from both normal passage (norm-pass) and fed batch (FB) cultures were analyzed for mRNA transcript levels and compared to final product quality and titer results.

#### Materials and Methods

##### [0141] Transfections and cell culture

[0142] Serum-free suspension GSKO CHO cells were transfected with 25 µg of plasmid DNA representing an expression vector for a heavy chain and a light chain of an antibody. For each transfection,  $20 \times 10^6$  cells were electroporated with the settings of 3175 µF, 200 V, and 700 Ω using GenePulser Xcell (BioRad). Transfected cells were placed into 20 mL cell culture media inside a 50 mL conical tube stored within a Kuhner shaker at 36.5°C and 5% CO<sub>2</sub>. Three days post selection, cells were transferred to selection media without glutamine. The pools of clones were cultured until they achieved a viability >85% (~2-4 weeks).

##### [0143] Single-cell cloning

[0144] Pools were single-cell cloned using the Berkeley Lights Instrument (Emeryville, CA), exported into 96-well plates, hitpicked or cherrypicked using Biomek FX<sup>P</sup> (Beckman Coulter) to 24-well plate Gas

Permeable Rapid Expansion (G-rex®, Wilson Wolf, St. Paul, MN) and then scaled up and passaged in 24-well deep-well plates.

[0145] Fed batch production and analysis

[0146] Fed-batch was performed once the single-cell clones achieved >90% viability with stable doubling time to evaluate protein expression and product transcript levels from FB single cell clones and norm-pass pools at early timepoints. HC1/HC2 transcript ratio was analyzed for correlations with PQA (product quality attributes).

[0147] Cells were passaged into FB media and in process monitoring occurred on days 3, 6, 8, and 10. Cultures were fed with nutrients on day 3, 6, and 8. Supernatant was harvested on day 10, and the conditioned media was analyzed for titer and product quality assessment on ATOLL-purified material.

[0148] Cell sample preparation

[0149] CHO pools expressing three different types of molecules were evaluated: 4-chain heteroIgG, 3-chain AmAb, and 3-chain C1mAb.  $1 \times 10^6 - 5 \times 10^6$  cells were collected during normal pool passaging and fed batch production with a criterion of  $\geq 80\%$  cell viability at harvest. Cells were spun down at  $300 \times g$  for 5 min and then a wash step in 1x DPBS (Life Technologies) was performed to remove any residual media. Cell samples were frozen in either 1 mL cryopreservation media (cell culture media plus 10% DMSO) or 350  $\mu$ L of RNeasy kit RLT buffer (QIAGEN) and stored at  $-80^\circ\text{C}$  until ready for cDNA generation.

[0150] cDNA generation

[0151] To extract the mRNA, the RNeasy kit (QIAGEN) was used following manufacturer's protocols. To remove any remaining DNA, a digestion step was performed using FastDigest enzyme (Thermo Scientific) according to manufacturer's protocols. Alternatively, RNA quick extract kit (BioRad SingleShot Cell Lysis) was used to prepare RNA for use in RT-PCR directly from cell culture samples. To generate the cDNA, a reverse transcription step was performed using iScript cDNA synthesis kit (BIORAD) with an incubation time of 20 minutes at  $46^\circ\text{C}$  using a Thermal Cycler (Applied Biosystems). The mRNA and cDNA were quantified and checked for quality by the 260/280 absorbance ratio using a NanoDrop spectrophotometer following mRNA isolation and cDNA synthesis.

[0152] Primer and probe set design

[0153] Primer and probe sets specific to unique sequences in each heavy chain and light chain of each molecule were designed using PrimerExpress or Integrated DNA Technologies (IDT) PrimerQuest Tool and ordered from IDT. The reference gene CHO  $\beta$  actin (ChoBAct) as used to normalize DNA loading across different samples.

[0154] ddPCR assay to quantify mRNA transcript levels

[0155] Two-step RT-ddPCR was used to rapidly quantify mRNA transcript levels (e.g., each heavy chain and light chain of a multispecific antibody). Each reaction consisted of 0.2 ng/ $\mu$ L cDNA, 18  $\mu$ M of each primer and 5  $\mu$ M of the probe for both the target and ChoBAct reference gene (Sigma-Aldrich) combined with Supermix for probes (no dUTP) (BIORAD) and water to reach a final volume of 20  $\mu$ l per well. Droplets were generated using the AutoDG system (BIORAD) and PCR amplification was performed using the following thermal cycling conditions: 95°C for 10 minutes, 40 cycles of denaturation 94°C for 30 seconds, annealing and extension at 60°C for 1 minute (heated lid: 105°C; sample volume: 40  $\mu$ L), finishing with one cycle at 98°C for 10 minutes and holding at 4°C using a Thermal Cycler (Applied Biosystems). Droplets were measured using the Qx200 Droplet reader (Bio-Rad) and quantities were analyzed using QuantaSoft software (Bio-Rad).

[0156] Statistical data analysis

[0157] Spearman correlation matrix in R (available from the R Project for Statistical Computing) was applied to analyze correlations between transcript levels and functional fed batch titers and product quality attributes.

[0158] Results

[0159] Protein expression and product transcript levels from FB pools are shown in Figure 1 and norm-pass pools are shown in Figure 2. Both protein expression and product transcript levels were found to positively correlate with titer results for (A) 4- chain heteroIgG, (B) a 3-chain molecule, and (C) an asymmetric C1mAb.

[0160] Spearman correlation analysis on 3-chain molecule results indicates a strong correlation for the sum of expression with titer and heavy chain (HC) 1 showing the highest positive correlation with titer, where HC1 is the longer of two HCs in the 3-chain molecule. HC1/HC2 ratio is positively proportional to main peak and high molecular weight (HMW) but inversely proportional to the pre-peaks and low molecular weight (LMW) from the size exclusion chromatography (SEC) assay. Consistent trends were observed at both FB and norm-pass timepoints as shown in Figures 3A and 3B.

[0161] Transcripts were evaluated for structurally different bispecific molecules (B2HmAb, C2HmAb, B2mAb, and C2mAb) with similar antigen binding properties. The sum of heavy chain (HC) and light chain (LC) transcripts positively correlates with normalized bulk fed batch titer as shown in Figure 4. Despite structural differences, selected top pool and backup pool fall in the top half of samples evaluated for transcript expression.

[0162] Clone transcript expression analysis using Quick RNA-Extract method shows a positive correlation for clone expression sum (grey bars) compared to normalized bulk titer (black bars), with the top pool performer (marked by black star) for a 3-chain molecule. See Figure 5A. Error bars represent the standard deviation of three replicates. Individual chain expression results in Fig. 5B show that the clone with the lowest titer (marked by a light grey star) had little to no production of HC1, which could explain

the poor clone titer performance, supporting the predictivity of the assay. Spearman correlation matrix in Fig. 5C indicates HC1/HC2 ratio is positively proportional to main peak, but inversely proportional to LMW (SEC) (where HC2 is the shorter of two HC in the 3-chain molecule). Fig. 5D shows a strong correlation between HC1:HC2 ratio versus MP and LMW. Top clone falls at around the 0.6 HC1:HC2 ratio, marked by a black square.

[0163] This assay can be used to predict not only titers from pools and clones but also product quality attributes (PQA). For example for a 3-chain molecule, HC1/HC2 transcript ratio was analyzed for correlations with PQA and found to be positively proportional to %main peak and %high molecular species (HMW) (as measured by size-exclusion chromatography, SEC) and inversely proportional to %pre-peaks (as measured by non-reduced capillary electrophoresis, nr-CE) and %low molecular weight species (LMW, as measured by SEC) at both fed batch and norm-pass culture timepoints. See Figures 3A-B. Overall, this assay could serve as a powerful tool to predict top performing pools before FB assessment with potential applications in clone screening as well.

What is claimed is:

1. A method of facilitating selection of a single cell clone or pools of cells for manufacturing an antigen-binding protein with two to four different antibody chains comprising:
  - a) passaging a single cell clone or pool of cells expressing the antigen-binding protein to at least 85% viability with a doubling time less than 40 hours for at least two consecutive passages,
  - b) extracting mRNA from the cells;
  - c) reverse transcribing the mRNA to generate cDNA;
  - d) performing ddPCR to quantify the transcript levels of the antibody chains and
  - e) eliminating single cell clones or pool of cells within at least the lowest 30<sup>th</sup> percentile of the sum of the transcript levels of all the chains.
2. The method of claim 1, wherein at least one of the antibody chains is a heavy chain or heavy chain-ScFv.
3. The method of claim 1, wherein at least one of the antibody chains is a light chain.
4. The method of claim 1, wherein the single cell clone or pool of cells expresses an antigen-binding protein with two different antibody chains.
5. The method of claim 4, wherein the single cell clone or pool of cells expresses an antigen-binding protein with one heavy chain (HC) or heavy chain-ScFv and one light chain (LC).
6. The method of claim 1, wherein the single cell clone or pool of cells expresses an antigen binding protein with three or four different antibody chains.
7. The method of any of claims 1-6, wherein eliminating single cell clones or pools of cells is for single cell clones or pools of cells within the lowest 50<sup>th</sup> percentile of the sum of the transcript levels of all the chains.
8. The method of claim 1, further comprising f) subjecting the remaining single cell clones or pools of cells to a fed batch culture for at least an additional 8 days while maintaining the cell viability above 80%.
9. The method of claim 5, further comprising eliminating cell clones or pools of cells expressing an antigen-binding protein with one heavy chain and one light chain where the ratio of the transcript levels for HC:LC or LC-HC is less than 0.2 or greater than 1.5.
10. The method of claim 6, wherein the cells express an antigen-binding protein with two heavy chains selected from HC1 and HC2 and a HC1 and HC2-scFv and one light chain (LC).

11. The method of claim 10, further comprising eliminating single cell clones or pools of cells where the ratio of the transcript levels for HC1:HC2 or HC2:HC1 is less than 0.2 or greater than 1.5 or the ratio of HC1:HC2-scFv or HC2-scFv:HC1 is less than 0.2 or greater than 1.5.
12. The method of claim 6, wherein the cells express an antigen-binding protein with two heavy chains (HC1 and HC2) and two light chains (LC1 and LC2).
13. The method of claim 12, further comprising eliminating single cell clones or pools of cells where the ratio of the transcript levels for HC1:HC2 or HC2:HC1 is less than 0.2 or greater than 1.5.
14. The method of claim 1, wherein the ddPCR is one-step ddPCR.
15. The method of claim 1, wherein the ddPCR is two-step ddPCR.
16. The method of claim 1, where the cells are selected from the group consisting of CHO, HEK293, NS0 or Sp2/0 cells.
17. The method of claim 16, wherein the cells are CHO cells.
18. The method of claim 17, wherein the CHO cells are DHFR- (dihydrofolate reductase minus) or GSKO (glutamine synthetase knock out).
19. The method of claim 1, wherein the cells have been generated by using single cell printing.
20. A method of facilitating selection of a single cell clone or pools of cells expressing an antigen-binding protein with one or more antibody heavy chains and one or more antibody light chains for stable growth comprising:
  - a) passaging the cells until viability is greater than 80% with a doubling time less than 35 to 40 hours for at least two passages,
  - b) extracting mRNA from the cells;
  - c) reverse transcribing the mRNA to generate cDNA;
  - d) performing ddPCR to quantify the transcript levels of the one or more heavy chains and the one or more light chains; and
  - e) eliminating single cell clones or pool of cells within at least the lowest 30<sup>th</sup> percentile of the sum of the transcript levels of all chains.
21. The method of claim 20, wherein the doubling time has less than 50% deviation between consecutive passages.