BioSci D145 Lecture #5

- Bruce Blumberg (blumberg@uci.edu)
  - 4103 Nat Sci 2 - office hours Tu, Th 3:30-5:00 (or by appointment)
  - phone 824-8573

- TA – Angela Kuo (akuo4@uci.edu)
  - 4311 Nat Sci 2- office hours W 10-12
  - Phone 824-6873

- Updated lectures (with answers) will be posted after lecture
  - http://blumberg-lab.bio.uci.edu/biod145-w20120

- Don't forget to discuss term paper topics with me in office hours or by email

- I will be out of town from 2/1-2/8 but will be checking email.

©copyright Bruce Blumberg 2020. All rights reserved

Potential volunteering opportunity

My name is Rachel Howard and I am reaching out to you from Together We Rise a National Non-Profit Organization working to help children in foster care. We are in the process of launching a new program called STEM Saturdays to help local children in foster care to discover the wonderful world of STEM. This FREE workshop offers hands-on activities and experiments that will open the door to curiosity and imagination to motivate children in foster care to explore STEM-based education. Take a look at our recent STEM Saturday event hosted at our headquarters in Brea, CA! CLICK HERE

Our team would love to partner with you and your students to brainstorm various experiments that we could cover for upcoming STEM Saturdays. Our goal is to demonstrate how much fun STEM can be! This volunteer opportunity would be a great way for your students to use their skills to positively impact their community. Would you consider sharing this opportunity with your students by encouraging them to sign up using the following link? STEM Saturday Brainstorm Volunteer Form

Rachel Howard, Executive Assistant

Office: (714) 784-6760, Office Hours: Mon: 9am - 5:30pm PST; T-F: 6am- 2:30pm PST

Follow us on Facebook Together We Rise

 ©copyright Bruce Blumberg 2020. All rights reserved

Microarray vs. RNAseq

- Microarray
  - Assumes you know all transcripts
  - Any sequence you did not know was expressed will not be there.
    - except whole genome tiling arrays
  - Detection limit issues
    - Signal-noise ratio
  - Well validated , expression analysis can be quantitative
  - Expense of procuring and processing especially for high depth analysis

- RNAseq
  - No assumption re transcripts identity. Illumina requires genome assembly (or de novo)
  - Can discover novel sequences or new splice forms not yet characterized
  - Detection limits are not a problem – can detect small #
  - Expression analysis can be quantitative
  - Cost much lower than microarray especially for high depth analysis. Many replicates possible per analysis

 ©copyright Bruce Blumberg 2020. All rights reserved

Single cell methods for RNA-seq scRNA-seq

- Bulk RNA seq can detect rare transcripts, but not cell types
- Isolating single cells and performing some technique on them is very powerful and hot technique today
  - RNA-seq
  - ATAC-seq
  - Hi-C seq
- Several types of technologies currently employed – dominant types emerging
  - Isolate single cells by flow cytometry
  - Isolate single cells using microfluidics

Table 1 Brief overview of scRNA-seq approaches

Protocol example	*C1 (SMARTer)*	*Smart-seq2*	*MATQ-seq*	*MARS-seq*	*CEL-seq*	*Drop-seq*	*InDrop*	*Chromium*	*SEQ-well*	*SPLIT-seq*
Transcript data	Full length	Full length	Full length	3′-end counting	3′-end counting	3′-end counting	3′-end counting	3′-end counting	3′-end counting	3′-end counting
Platform	Microfluidics	Plate-based	Plate-based	Plate-based	Plate-based	Droplet	Droplet	Droplet	Nanowell array	Plate-based
Throughput (number of cells)	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^3–10^4	10^3–10^4	10^3–10^4	10^3–10^4	10^3–10^5
Typical read depth (per cell)	10^6	10^6	10^6	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4
Reaction volume	Nanoliter	Microliter	Microliter	Microliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Nanoliter	Microliter

 ©copyright Bruce Blumberg 2020. All rights reserved

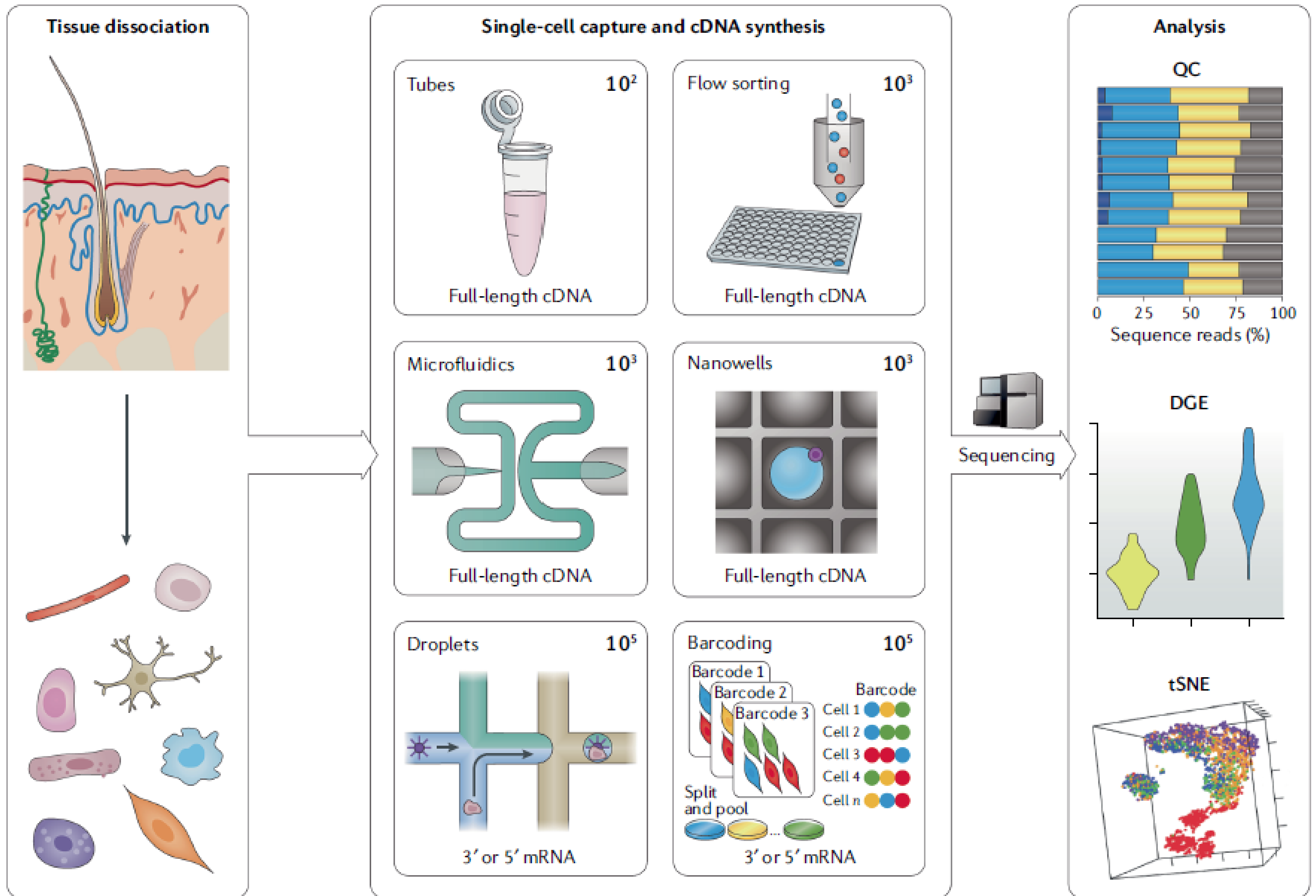
Single cell methods for RNA-seq scRNA-seq

Excellent reviews are

- Hwang et al. Experimental & Molecular Medicine (2018) 50:96
  DOI 10.1038/s12276-018-0071-8
- Ziegenhain et al. Molecular Cell (2017) 65, 631–643
  DOI 10.1016/molcel.2017.01.023
- Haque et al, BMC Genome Medicine (2017) 9:75
  DOI 10.1186/s13073-017-0467-4

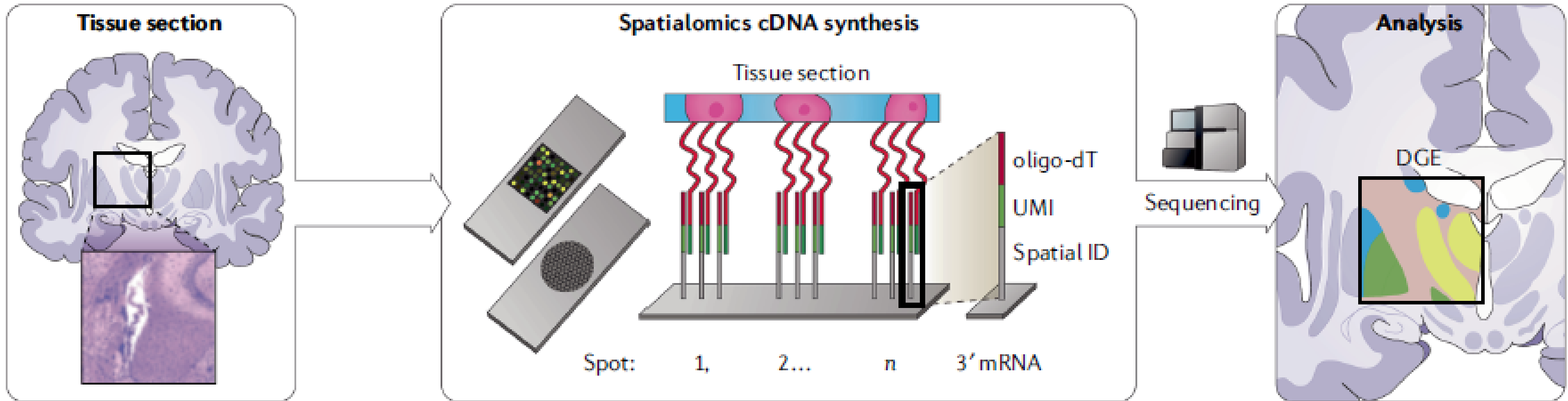
 ©copyright Bruce Blumberg 2020. All rights reserved

Single cell methods – from tissues



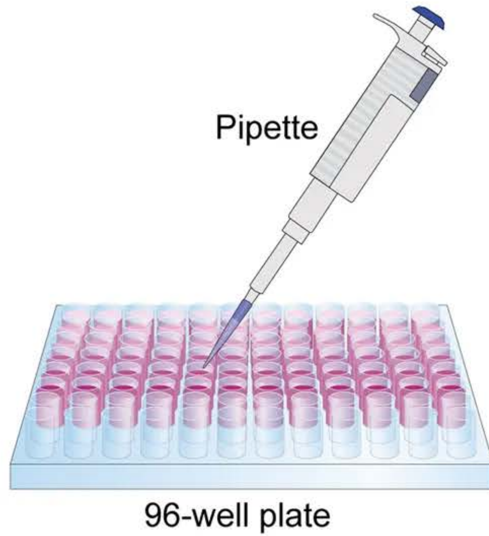
 ©copyright Bruce Blumberg 2020. All rights reserved

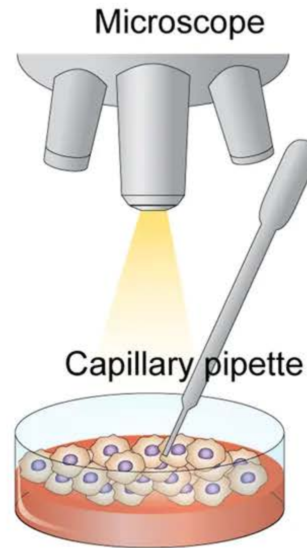
Single cell methods – from tissues

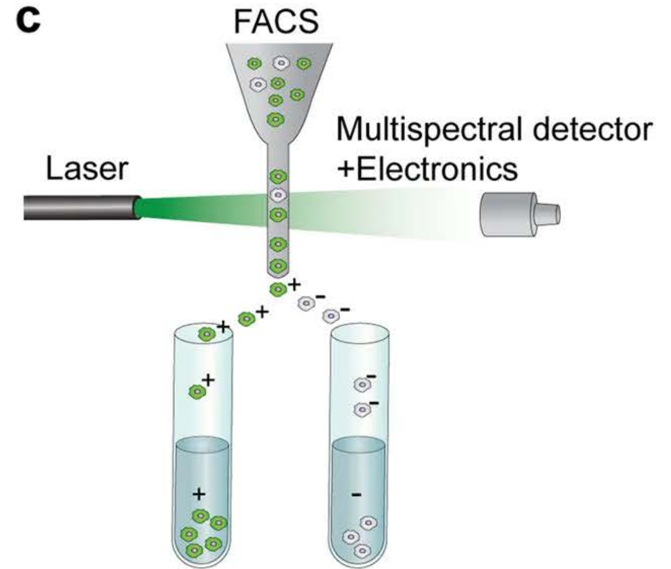


 ©copyright Bruce Blumberg 2020. All rights reserved

Single cell methods – how to get the single cells

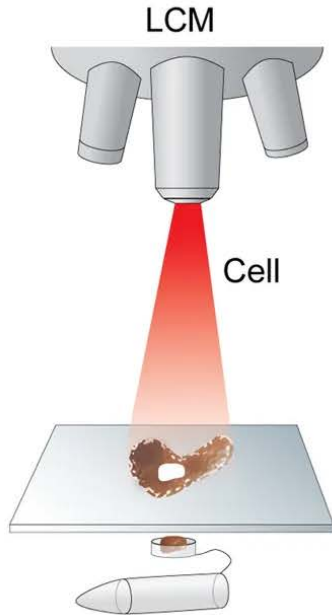


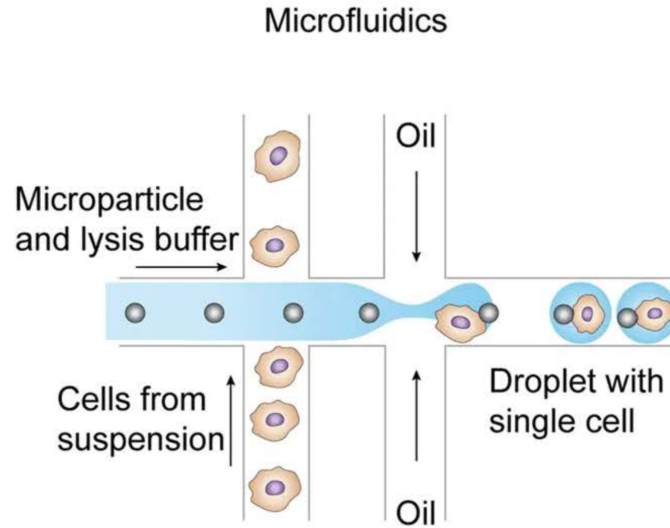


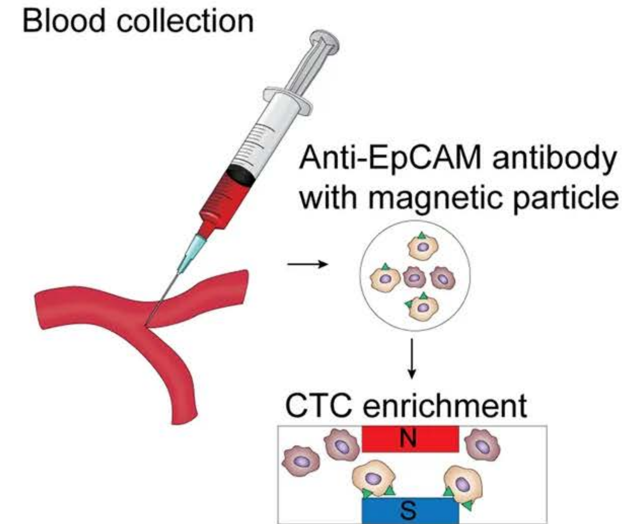


 ©copyright Bruce Blumberg 2020. All rights reserved

Single cell methods – how to get the single cells



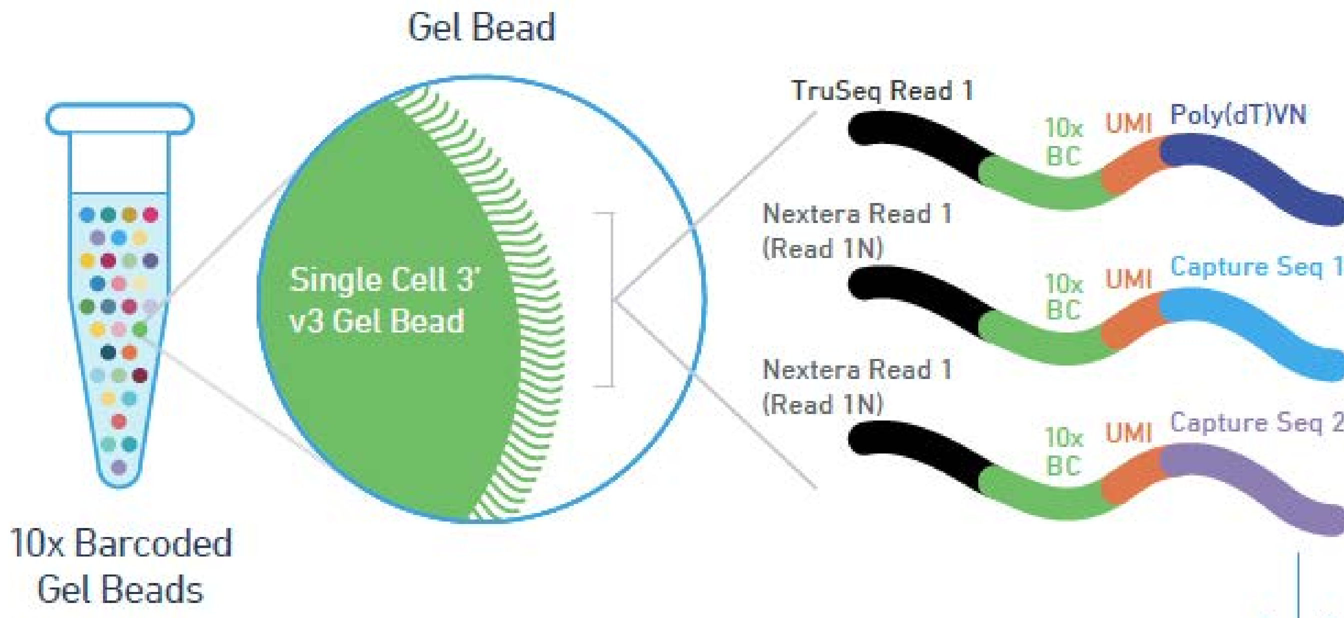




 ©copyright Bruce Blumberg 2020. All rights reserved

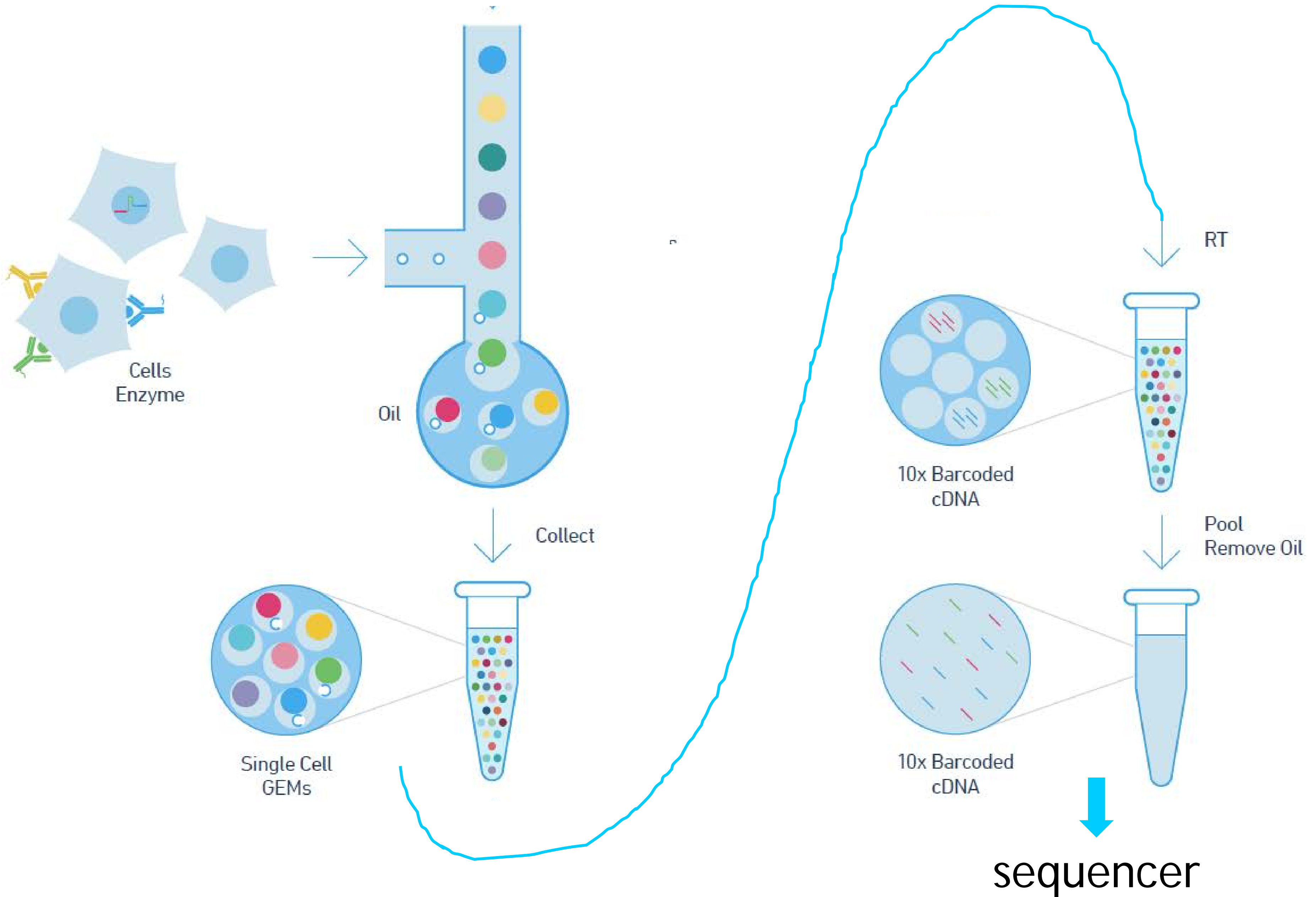
Single cell methods for RNA-seq scRNA-seq - 10x Genomics

- https://10xgenomics.wistia.com/medias/f75ht43w1q
- Separate cells and unique beads into microdroplet of oil/water
  - Collect all of these and run RT reaction
  - Pool 10x barcoded cDNA and remove oil and reagents
  - Use to construct Illumina library
  - Deep sequencing
  - Analysis of resulting sequences - divide into cell types
  - https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_BR025_Chromium-Brochure_Letter_Digital.pdf



 ©copyright Bruce Blumberg 2020. All rights reserved

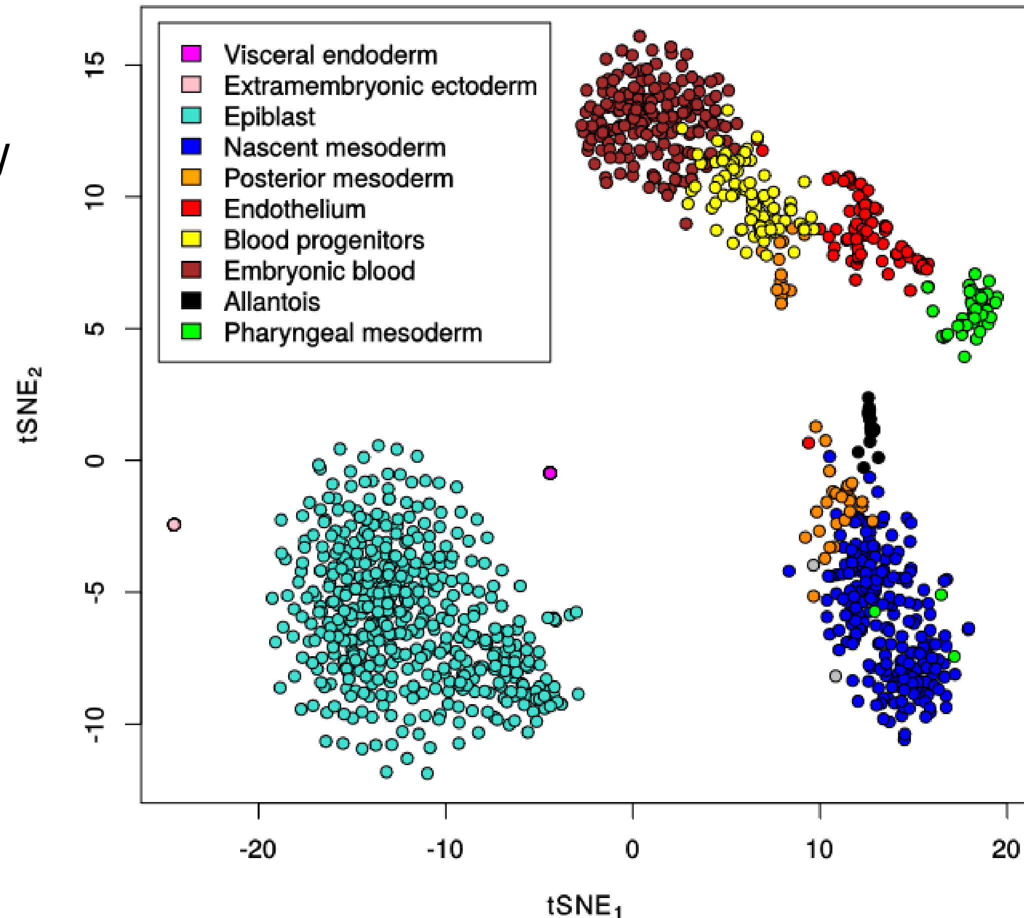
Single cell methods for RNA-seq scRNA-seq - 10x Genomics



 ©copyright Bruce Blumberg 2020. All rights reserved

Single cell methods for RNA-seq scRNA-seq - 10x Genomics

- Uses of scRNA seq
  - Determine transcriptome of individual cells - heterogeneity analysis
  - Identify rare cells in population
  - Determine relationship of cell types to each other (requires time series)
  - Mono-allelic gene expression
- Limitations
  - Dropout - failure to detect specific transcripts - depends on read depth per cell
  - Temporal fluctuation in sequences
  - Changes in cell transcriptome during processing (protease digestion, collection, time till sequencing
- Technology good to isolate single cells or nuclei for many methods
  - basically, any analysis that uses barcoded samples can be performed
  - https://www.10xgenomics.com/resources/product-literature/



 ©copyright Bruce Blumberg 2020. All rights reserved

Methods of profiling gene expression (small number of genes)

- How to evaluate gene expression? Validate RNA-seq or microarray results?
  - Old, low-throughput - prepare RNA sample and perform
    - Northern blot – immobilize RNA on filter, probe
      - Quantitative WHY?

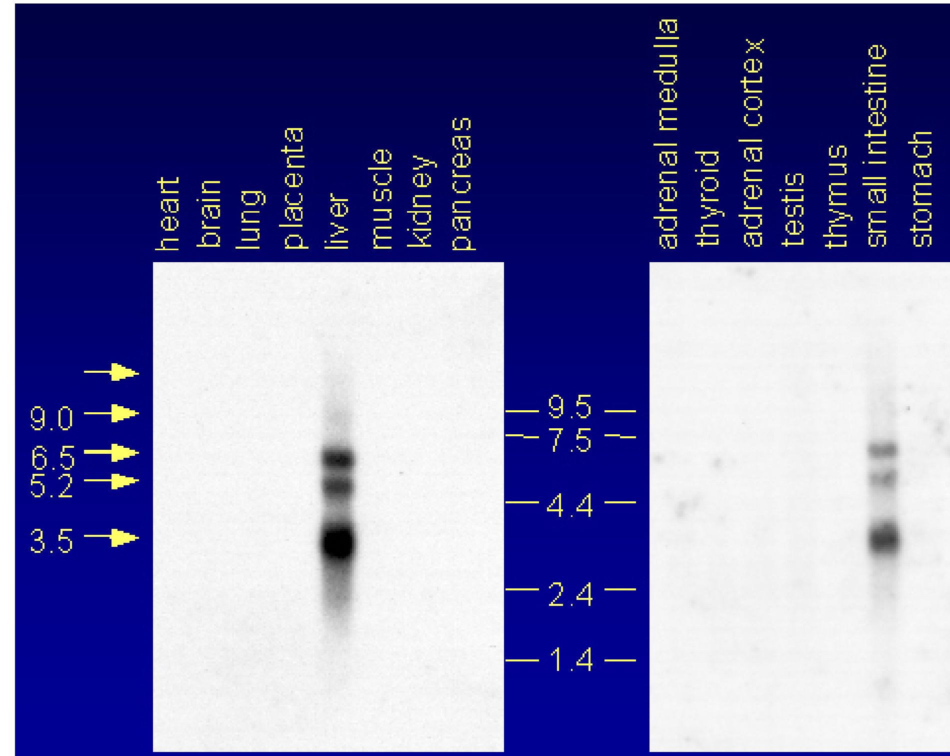
Probe is in excess

    - Nuclease protection
      - quantitative
    - In situ hybridization
      - Not quantitative – enzymatic reaction
  - Newer, high throughput methods
    - RT-PCR
      - Can be quantitative
    - Quantitative real time RT-PCR
  - Or prepare protein samples and evaluate proteins
    - Western blot - detect protein of interest with specific antibody.
    - ELISA – enzyme linked immunosorbent assay quantitative
    - RIA – radioimmunoassay - quantitative

 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - size and splicing

- Quantitation of mRNA levels
  - possible methods
    - Northern analysis
    - nuclease protection
    - RT-PCR
  - measure steady state mRNA levels (production/degradation)

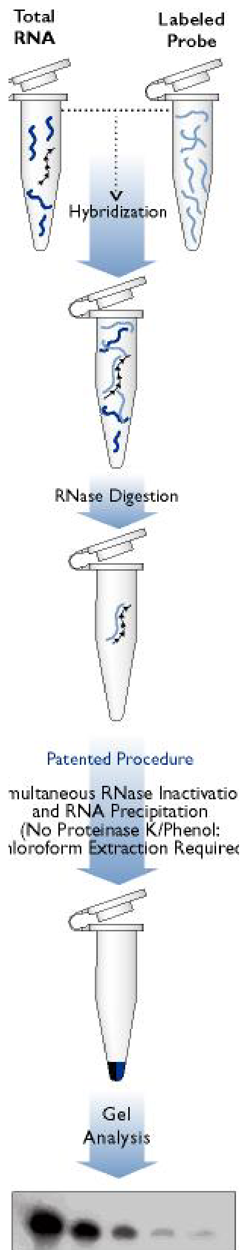


- mRNA size determination –
  - Northern blot only way
  - good RNA size markers = accurate sizing
  - which to use, poly A^+ or total RNA?
    - A^+ much more sensitive (50-100x)
      - what about mRNAs with no or short tails?
    - total RNA much simpler
      - gel limitations – 20 µg/lane is practical limit
  - what is a key factor in sizing mRNAs?

Appropriate size standards larger and smaller than target

 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)

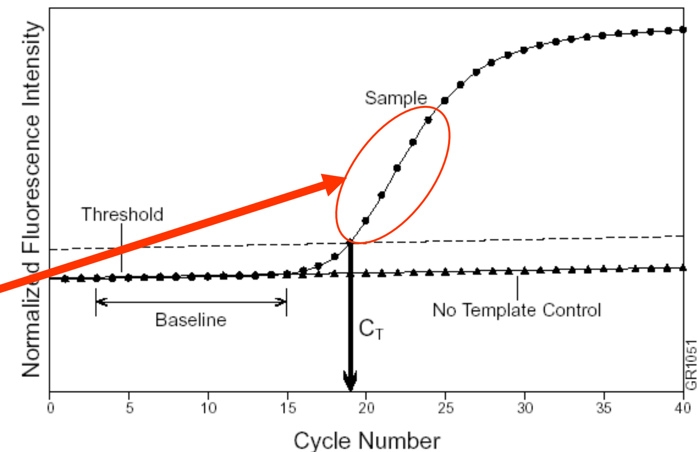


- Nuclease protection assays
  - approach
    - hybridize a single-stranded (SS) probe (DNA or RNA) to RNA sample
      - probe must be larger than protected region
    - digest remaining single stranded regions
    - electrophorese on denaturing polyacrylamide gel
  - advantages
    - less sensitive to slightly degraded mRNA
    - absolutely quantitative
    - can tolerate large amounts of RNA (100+ µg)
      - allows detection of rare transcripts
      - but gives high background
    - multiple simultaneous detection
  - disadvantages
    - more tedious than Northern
    - no blot to reuse
    - multiple simultaneous detection hard to optimize

 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)

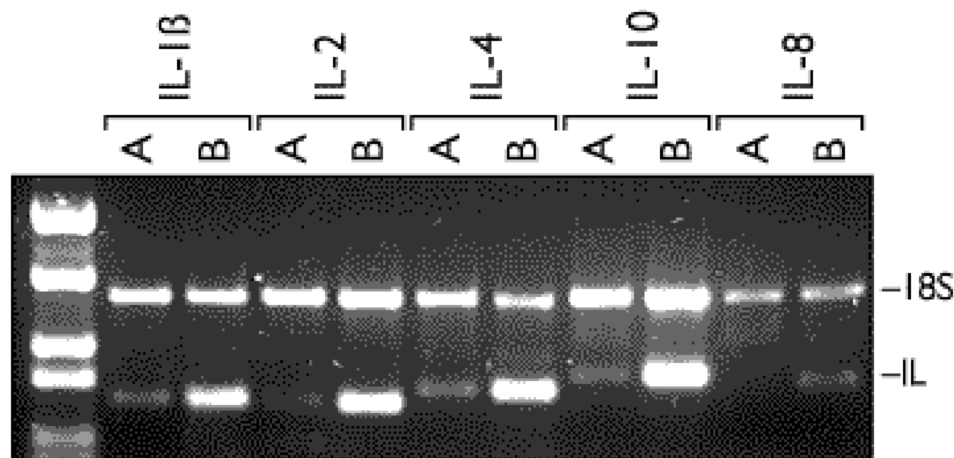
- RT-PCR - reverse transcriptase mediated PCR
  - approach
    - reverse transcribe mRNA -> cDNA
    - amplify with specific primers
    - quantitate
  - flavors
    - relative quantitation – compare to invariant gene
    - absolute quantitation
      - by comparison to synthetic reference
      - competitive PCR
      - various fluorescent dye mediated methods
  - advantages
    - very fast and simple
    - works with tiny amounts of material
  - limitations
    - RT efficiency differs by mRNAs
    - Must be in linear amplification range
    - Errors increase exponentially with amplification



 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)

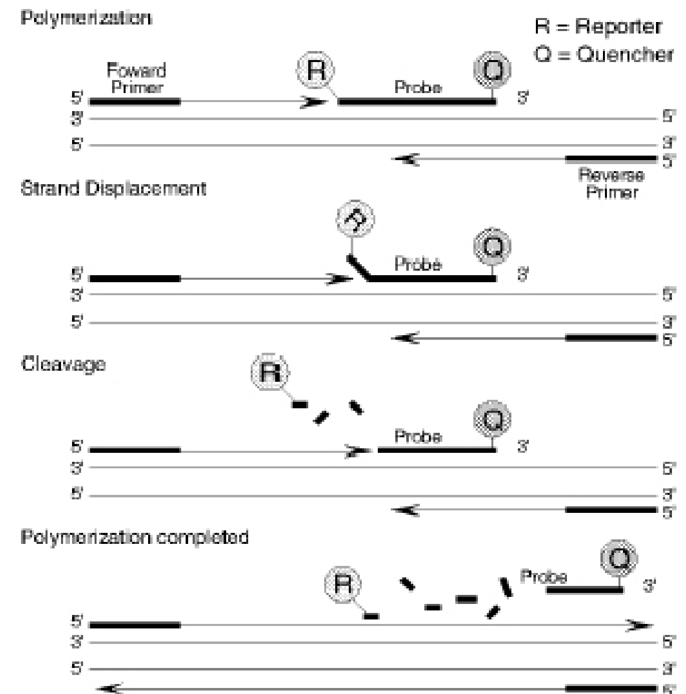
- RT-PCR reverse transcriptase mediated PCR
  - relative concentration determination
    - perform multiplex reaction using two primer sets
      - 1 for reference, 1 experimental
    - advantages
      - no fancy equipment required
    - disadvantages
      - careful attention to linear region for both primer sets
      - often must add one set during reaction
        - » companies claim to have products that eliminate this need
        - » more than 2 primer sets are not reliable



 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)

- RT-PCR (contd)
  - absolute concentration determination real time PCR
    - Taqman, molecular beacons
      - Fluorescent methods that allow direct quantitation of PCR product
    - approach
      - special oligonucleotide that has a fluor and a quenching group on it.
        - » When whole, no fluorescence
      - perform PCR reaction, if primer anneals, Taq polymerase removes the reporter group which can now fluoresce



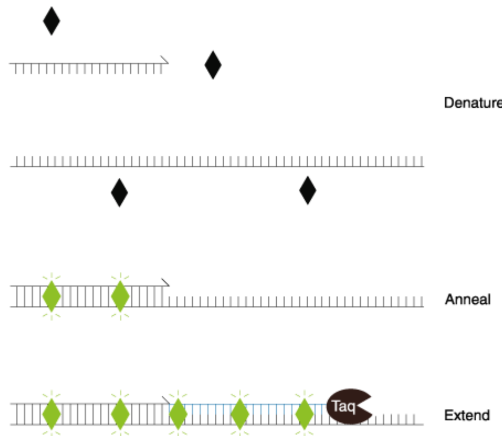
 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)

- RT-PCR (contd)
  - absolute concentration determination - Taqman, etc
    - Fluorescence detected continuously in real time
    - advantages
      - can be detected in real time with proper instrument
      - no difficulties with linearity
      - multiplexing of probes possible (limited by available dyes)
      - very good for clinical diagnostics
    - disadvantages
      - requires instrument
        - » varies from expensive to extremely expensive
        - » Not of equal quality
      - need to make custom oligos - can be expensive
      - must know something about relative abundance of mRNAs before setting up reactions
      - careful optimization required for best results
        - » primer concentrations
        - » target concentrations

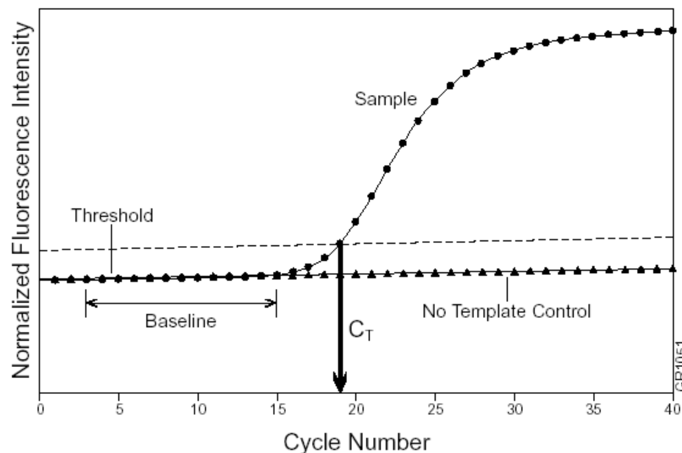
 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)



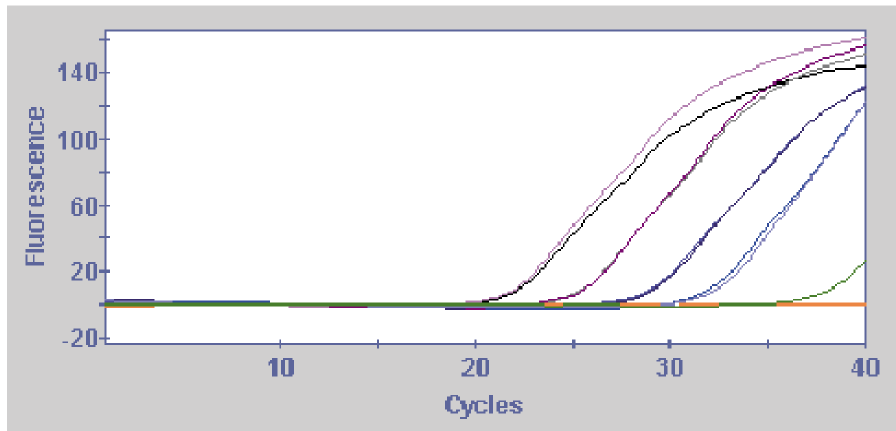
- RT-PCR (contd)
  - absolute concentration determination – Sybr Green
    - Alternative real time RT-PCR utilizes a single dye
    - approach
      - Extend a single template
      - Detect ds DNA with a specific dye

Real Time Detection The threshold cycle or C_T value is the cycle at which a statistically significant increase in ΔR_n is first detected. Threshold is defined as the average standard deviation of R_n for the early cycles, multiplied by an adjustable factor. On the graph shown below, the threshold cycle occurs when the Sequence Detection Application begins to detect the increase in signal associated with an exponential growth of PCR product.

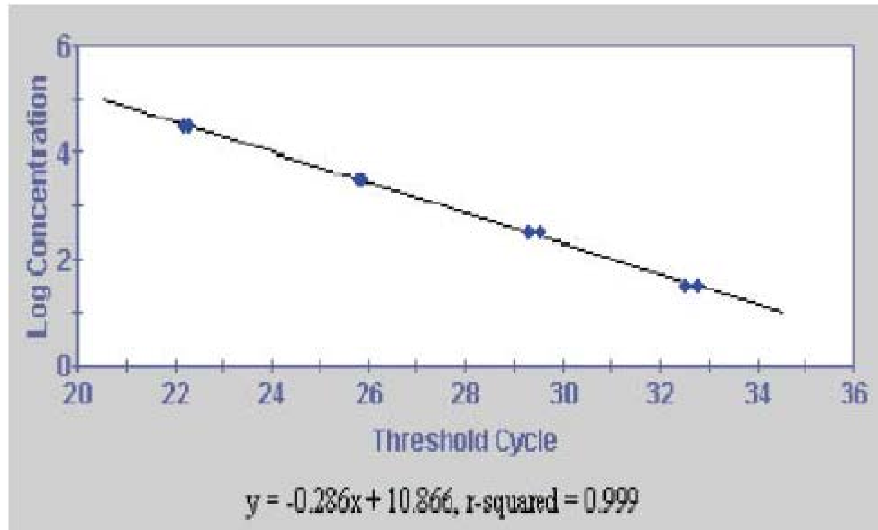


 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)



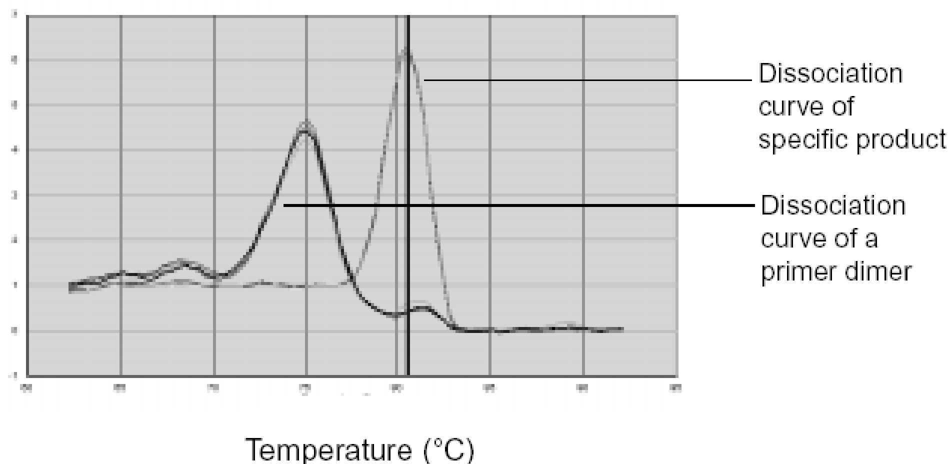
- RT-PCR (contd)
  - absolute concentration determination – Sybr green
    - Plot lift off time
    - Generate standard curve



 ©copyright Bruce Blumberg 2020. All rights reserved

Analysis of mRNA - quantitation (contd)

- RT-PCR Sybr Green (contd)
  - Advantages
    - No special primers needed
    - Single dye, simple
    - Fast, robust and quantitative
    - Good for routine use
  - Disadvantages
    - Need instrument
    - Single dye, can't multiplex
    - Problems with multiple fragments
      - » Melting curves required
    - Absolute quantitation requires std curve



 ©copyright Bruce Blumberg 2020. All rights reserved

1. (7 points) There exists a group of Antarctic fish (the Notothenioids), sometimes called "crocodile icefish" inhabiting the Southern Ocean that have some quite unusual, and even bizarre properties. Among these are the ability to live in water that is -1.8°C, below the freezing point of fresh water, or vertebrate blood, partly because their blood contains "antifreeze proteins". Notothenioids have also lost their swim bladders (buoyancy organ) and are the only known vertebrates to lack red blood cells, hemoglobin or myoglobin (oxygen binding protein in muscle). Notothenioid fishes evolved over the past 20 million years or so and are useful for studying how a diverse group can evolve from a common ancestor in a relatively brief time period, so we will be studying them during this quarter's D145 laboratory. Dr. Jacob Daane, a new faculty member in the Dept of Ecology and Evolutionary Biology recently identified a new type of icefish on a research trip to the Antartcic. Jacob has generously agreed to let our D145 class aid in the characterization of this species of fish. Let's get started.

a) (3 points) The Family Channichthyidae is composed of 11 genera. Dr. Daane believes (on morphological evidence) that his new fish is a member of the genus *Chaenocephalus* and has named it *Chaenocephalus albennetti* after the former Dean and chair of EcoEvo, Al Bennett. **Describe how you would test molecularly whether *C. albennetti* was more closely related to other members of this genus than it is to other 10 genera in the family Channichthyidae.**

The best approach would probably be to do a whole genome sequencing of members of the 11 genera as well as *C. albennetti*. You will want to use at least NextGen sequencing to create a draft genome, then compare the sequences with each other along the lines of the Lindblad-Toh paper and determine where *C. albennetti* falls on the on the tree you generate.

 ©copyright Bruce Blumberg 2020. All rights reserved

b) (4 points) While you are at it, why not test whether *C. albennetti* is even a member of the Family Channichthyidae as opposed to one of the other 4 families of Notothenioids? Even better, why not do a proper phylogeny of *C. albennetti?* After all, anatomists who made the original taxonomic classification may have mistaken small differences in morphology and misclassified the specimens. **Describe how you would 1) determine which of the 5 families of Notothenioids *C. albennetti* is most closely related to, 2) assess whether *C. albennetti* is even a member of the Suborder Notothenioidei, vs. being a member of one of other 18 other Suborders in the order Perciformes.**

Same approach as 1a except that you will want to include members of the 5 families of Notothenioids for the first section and members of each suborder of the order Perciformes in the second. There are probably already genome sequences availabe for some of these, so I'd check what is available first before doing more sequencing. In both cases, determine relatedness of *C. albennetti* with each of the other specimens, build a tree and see where *C. albennetti* falls on the tree.

2. (8 points) Good job, your group proved molecularly that *C. albennetti* is most closely related to other members of the Genus *Chaenocephalus* and has probably been correctly named. Let's start digging deeply into the biology of this fish. As mentioned in question 1, the family Channichthyidae is characterized by a total loss of red blood cells, hemoglobin and myoglobin.

2a) (4 points) Describe how you could determine whether genes for α-globin, β-globin or myoglobin are present in the *C. albennetti* genome?

Since you sequenced the genome of *C. albennetti* in answering 1a, all you will need to do is check the sequences and determine whether any of these genes is present. If you didn't sequence the genome of *C. albennetti* in the answer to question 1, then you will want to sequence it here and look for those genes.

2b) (4 points) Great, you found that there are indeed some globin genes present in the *C. albennetti* genome but that they are non-functional. **Please describe how you could determine whether the protein coding sequence of C. *albennetti* β-globin has mis-sense or nonsense mutations compared with β-globin genes from the related, red blooded Notothenioid species *Notothenia coriiceps and Notothenia angustat.***

This is an application for amplicon sequencing Β-globin mRNAs from *C. albennetti*, *N. coriiceps* and *N. angustat*. Determine the sequences and compare to identify mutations. You could determine this from the whole genome sequence, but the mRNA sequencing is better because you are only guessing at the locations of exons in the genome sequencing without comparing them with an mRNA sequence.

 ©copyright Bruce Blumberg 2020. All rights reserved

3. (4 points) Another intriguing feature of the Channichthyidae is that while they were derived from a benthic (bottom dwelling) ancestor, many members of the family have occupied pelagic niches in the local ecosystem. This means that they can be found at all depths in the water column, not just at the bottom. Since these fish have all lost their swim bladders (which would allow them to maintain their position in the water column without expending energy), they must have some other adaptation that allows this. Dr. Daane found that these fish have bones that are porous and filled with fat as well as deposits of fat within their muscle. This fat makes the fish more buoyant and compensates for the loss of a swim bladder. It turns out that mesenchymal stem cells (MSCs) are the common precursor to both fat and bone cells. **Assuming that you can isolate MSCs from the fatty bones of *C. albennetti,* describe how you would identify, to a high degree of certainty, all of the transcripts that are expressed in these cells. Be sure to note any key materials you need to accomplish this.**

To a high degree of certainty excludes microarray analysis since you will undoubtedly miss some transcripts. You will want to perform extensive RNA-seq analysis of the transcripts expressed in cultured MSCs from *C. albennetti*. You will need the MSCs and the usual materials to perform RNA-seq analysis. Of course, this will depend on the type of sequencing you choose, but the most likely answer would be for RNA-seq using Illumina sequencing.

 ©copyright Bruce Blumberg 2020. All rights reserved

4. (8 points) Excellent, now you know what genes are expressed in MSCs from *C. albennetti*. The next important question is why these MSCs make fat cells inside the bone, where they ordinarily would not. Intriguingly, osteoporosis has been called the "obesity of bone" in humans with osteoporosis because the bones of such patients are often brittle and have fat where bone should be. Maybe *C. albennetti* could serve as a model for the development of osteoporosis in humans?

a) (4 points) First you need some idea of why the *C. albennetti* MSCs produce fat cells instead of bone. Describe how you would identify differences in the MSC transcriptomes between *C. albennetti* and the related species *Notothenia coriiceps*, which does not have fatty bones. Be sure to state any assumptions you need to make.

To answer this question you are first going to want to determine the transcriptomes between the MSCs derived from bones of *C. albennetti* and *N. coriiceps.* The assumption you need to make is that these cells can be isolated from the respective organisms. Perform RNA-seq and compare the transcriptomes, looking for genes that are differentially expressed in *C. albennetti* vs. *N. coriiceps.*

3b) (4 points) Hmm. You did not find any dramatic changes in the transcriptomes between MSCs from the 2 test species noted in 4a. **How would you test the hypothesis that MSCs in *C. albenneti* contain a subpopulation of cells that are predisposed to form fat, rather than bone, compared with MSCs from *N. coriiceps?***

In order to determine whether the isolated MSCs from each group have a one or more subpopulations, you will want to perform single cell RNA-seq of MSCs from each organism. The results of this analysis will tell you whether the cells are homogeneous, or whether they contain one or more subpopulations. If subpopulations are identified, you will want to look carefully at the genes expressed and see if these are characteristic of cells that are predisposed to form fat (preadipocytes).

 ©copyright Bruce Blumberg 2020. All rights reserved

5. (8 points) Even more intriguingly, fat cells in the muscle of C. albennettii resemble a type of mammalian fat called beige fat. Beige fat is derived from white adipose tissue (WAT) but differs from WAT in that beige fat is thermogenic, i.e., it can uncouple ATP production from energy production, leading to the generation of heat. But this is very strange because all fishes, including *C. albennettii* are ectothermic; i.e., their body temperature is regulated by the temperature of their environment, rather than internally controlled as in endothermic animals such as mammals. Could it be that part of the reason that icefish can tolerate such cold temperatures is because they can generate internal heat?

a) (4 points) Ok, fine, the fat cells in the muscle of *C. albennetti* look like beige adipocytes - they are smaller and have many fat globules rather than 1 and have more mitochondria than WAT. But are they, in fact beige adipocytes? The genes Fgf21, P2rx5, Pat2, Car5, Ucp1 and CIDEA are all expressed in mammalian beige fat. **Please describe how you can test whether beige fat cells from *C. albennetti* express these genes and at what levels?**

You need to test the expression levels of six different genes from fat cells in the muscles of *C. albennetti*. First you need to isolate these cells from the muscle. Then design primers for QPCR analysis of these genes and test whether they are expressed in the putative beige fat cells and at what levels compared with a housekeeping gene. Or you could perform a direct quantitation compared with a standard. If you wanted to work hard, you could make RNA from these cells, run gels, perform Northern blot analysis and determine the levels of mRNA expression as well as the size of the mRNAs. I'd do qPCR, myself.

 ©copyright Bruce Blumberg 2020. All rights reserved

b) (4 points) Indeed, you found that the apparent beige fat cells in the muscle of *C. albennetti* express beige marker genes. Nice. You decide to test whether these beige adipocytes can generate heat in culture. **Assuming that you can culture the cells at -1.8°C, and that you can measure the heat produced by each cell, how could you test whether expression of *Fgf21*, *Ucp1* or *CIDEA* was associated with heat generation in the cultured cells?**

Culture the cells at -1.8°C. Measure the heat produced by the cells and assess whether the expression of the 3 genes in question is associated with heat production. You might want to culture the cells at somewhat warmer temperatures to test whether the expression of the genes is higher at lower temperatures, as might be expected for beige adipocytes.

 ©copyright Bruce Blumberg 2020. All rights reserved