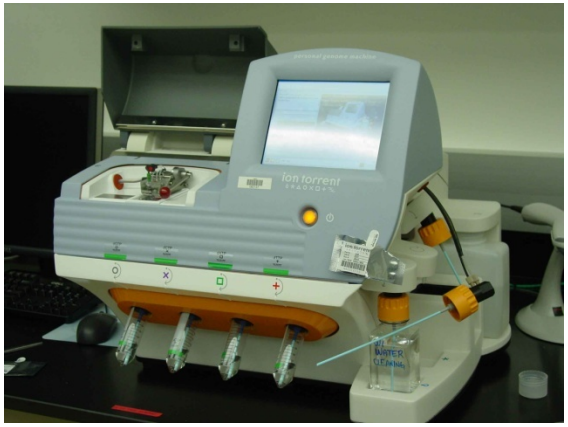
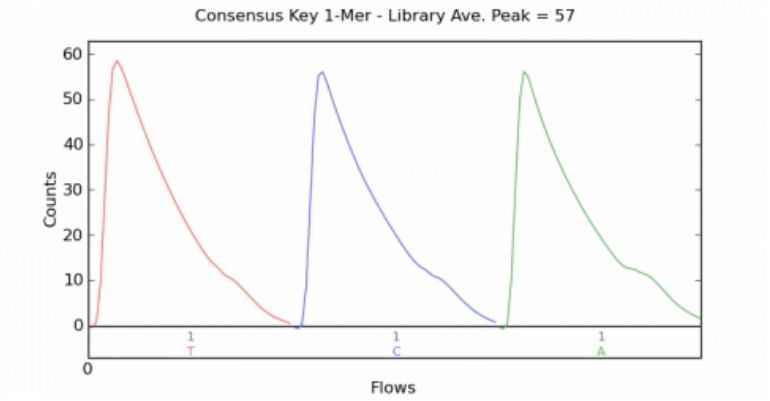
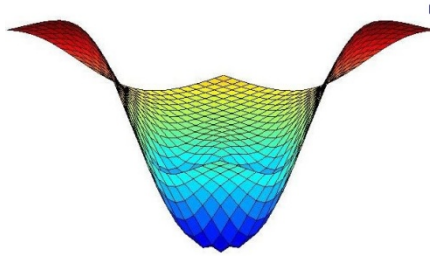
BINF 6010 ITSC 8010
Spring 2010
Biotechnology & Genomics Lab
Experimental Design- Technical

http://webpages.uncc.edu/~jweller2





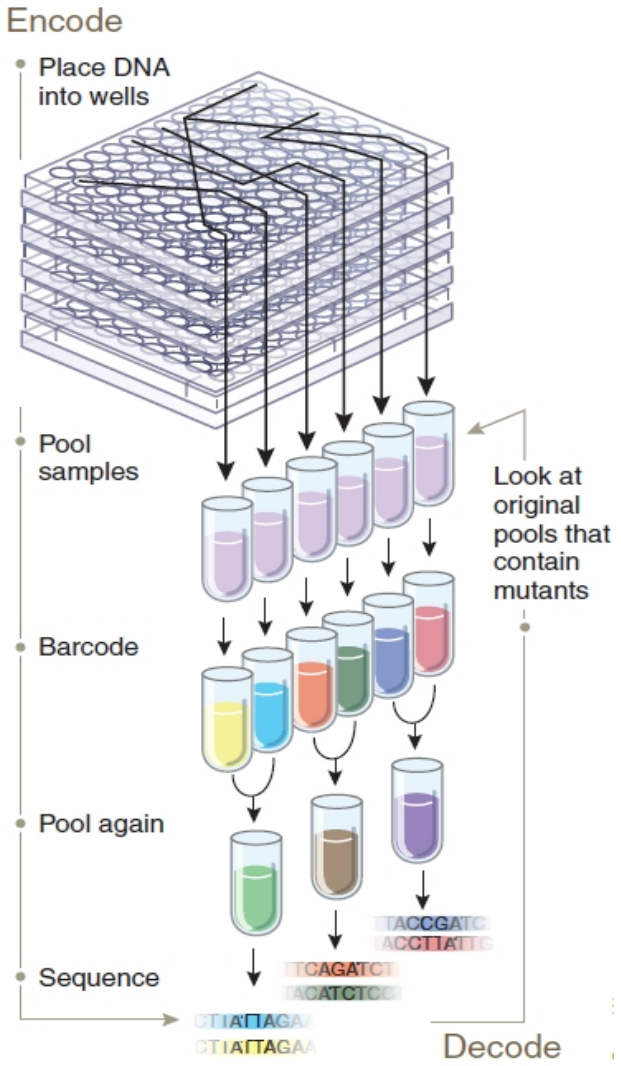


Topics

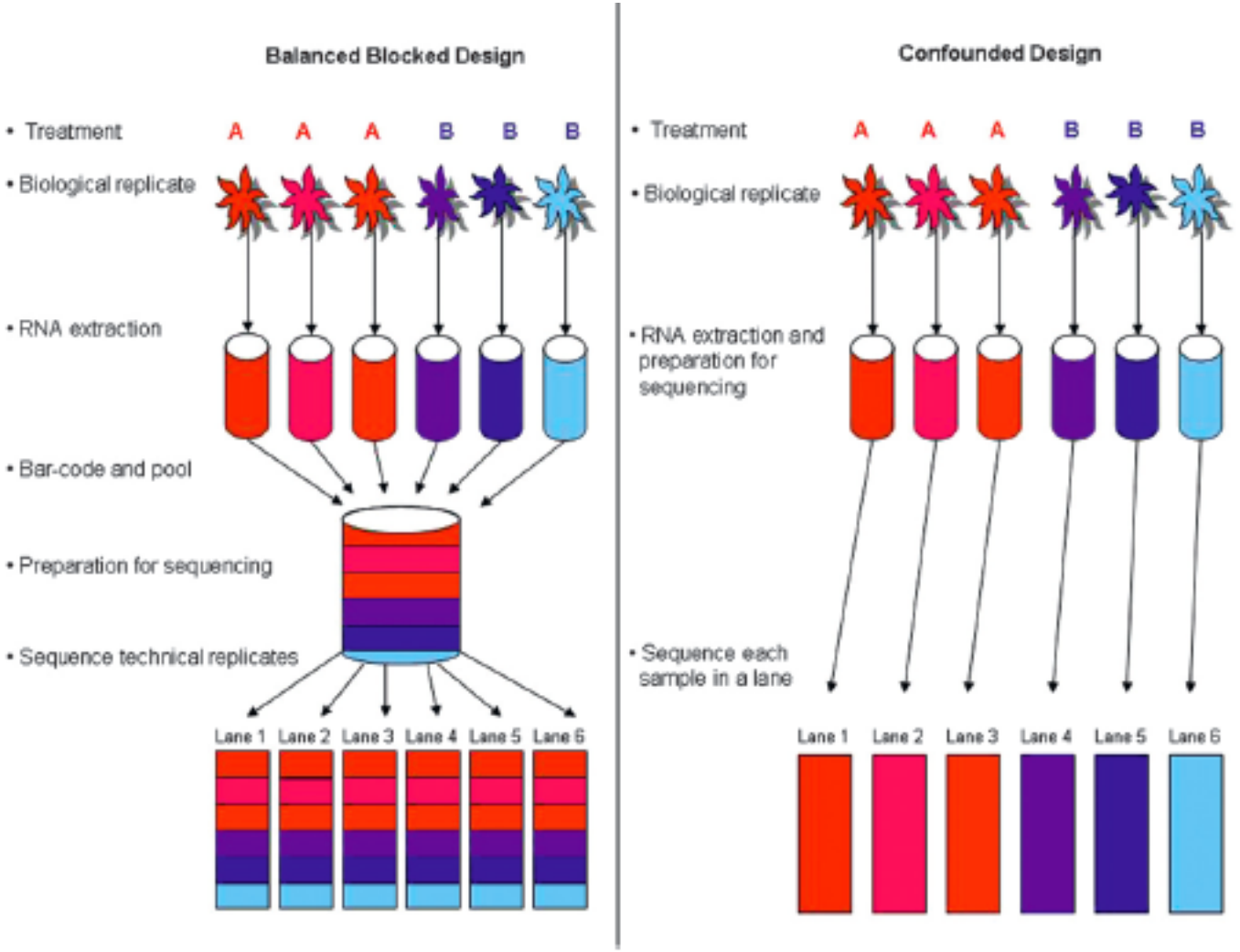
- Experimental Design - sources of technical variation
  - Library construction
  - Platform performance
- Principles of Cell Disruption
  - CTAB extraction
  - Centrifugation

Experimental Design should isolate the dependent and independent variables from all other factors.

- Biological variation asks whether there a link between a biological factor and the attribute of interest?
  - What is the effect size?
  - What factors might impact the outcome?
  - Is the event rare in a cell, in a population?
- Technical variation reflects transformation steps; controls preclude falsely asserting that an observation arises from biological properties of the sample.
  - How complete is each conversion?
  - What losses are typical of each method or platform?
  - Are there ways to pool samples and steps minimize variation?

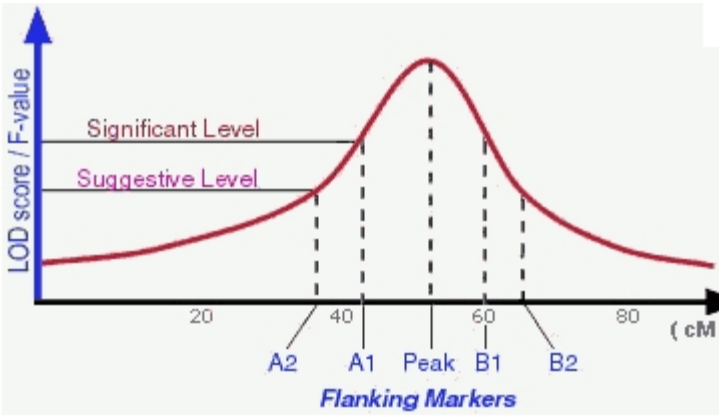


- Simulation can be used to figure out designs for multi-factorial experiments
  - Model the environment in a test bed that allows control and controlled variation of all important factors including changes in the environment caused by the process.
  - Randomization, replication and local control strategies are applied

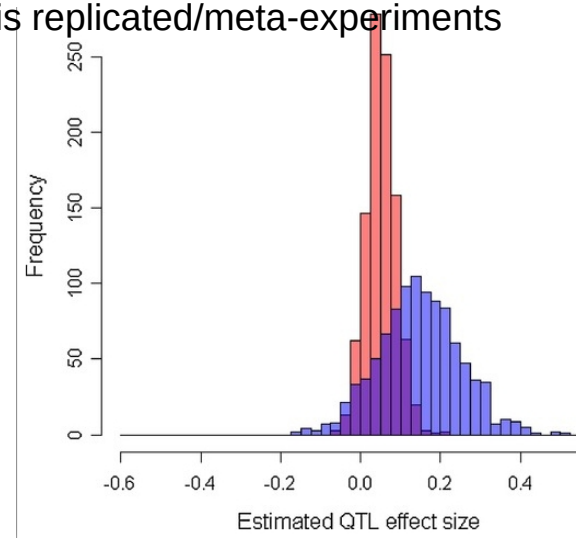


The effect size quantifies the difference between two groups if you are looking at QTLs or SNPs there is often a small individual effect.

- Please see the white paper by Dr. Coe for a full explanation : http://www.leeds.ac.uk/educol/documents/00002182.htm
- For a given factor, how well does it work in a range of contexts?
  - Note that statistical significance conflates effect size and sample size.
  - Variation is used is quantifying the effect.
    - Effect size = [$Mean_E$ – MeanC]/SD
    - SD is the standard deviation in the population, or some pooled equivalent in the samples
    - Effect sizes can be combined when an experiment is replicated/meta-experiments



Weller UNCC

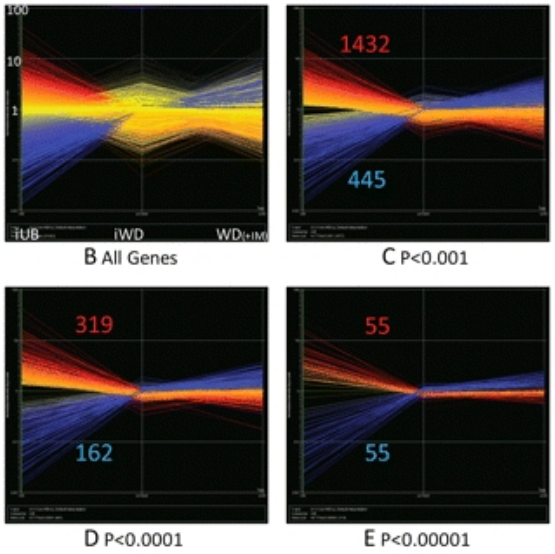


There are six steps in designing a genomics experiment

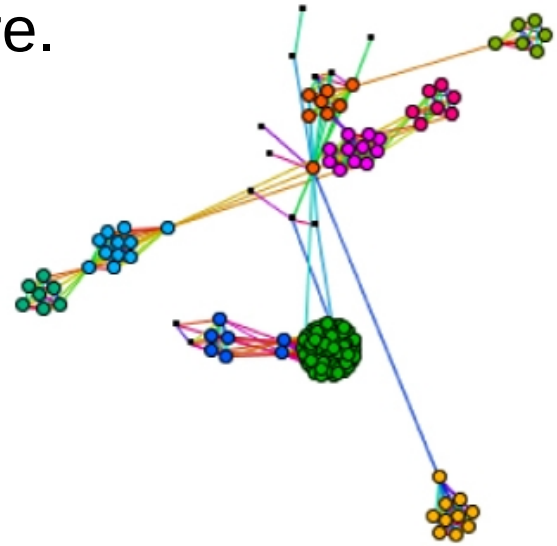
- Decide what material will be used and how you will measure outcomes (platform)
  - This requires that the assay design be performed or confirmed
  - Ensure sensitivity and specificity are sufficient for the effect size
  - IRB certification and approvals, SOPs, etc.
- Acquire samples of the correct types
  - Ensure sufficient sampling, no bias, and replicates
  - Is blinding needed? How is collection of sample details assured?
- Process down to analytes
  - Ensure technical variance does not swamp the biological signal – replication by stage/process
- Perform the assays
  - Allocate samples to different runs
- Perform the analyses

There are three common classes of genomics experiments.

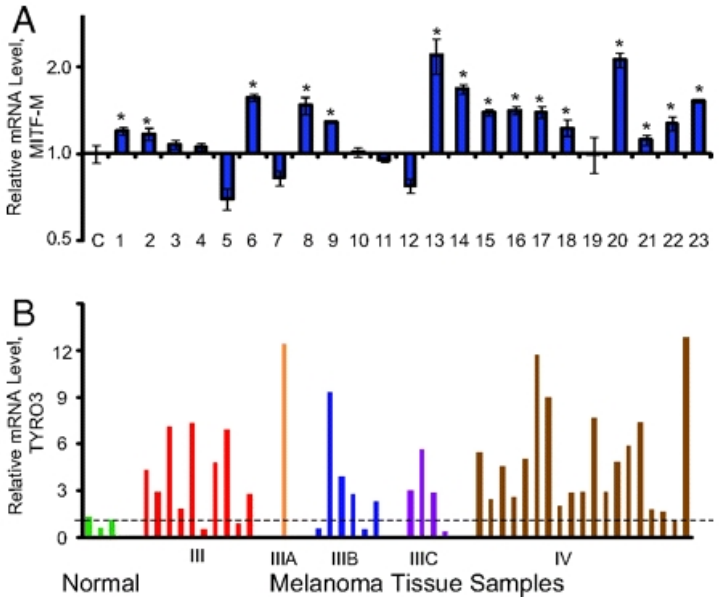
- Comparative: compare levels of x and y in samples from either assigned treatment or observed phenotypic states.
- Associative: identify concentrations of x that correlate with some quantitative feature (time to flowering).
  - Often samples are classified a priori
- Exploratory: search for patterns and clusters that correlate with some quantitative feature.



B All Genes C P<0.001 D P<0.0001 E P<0.00001



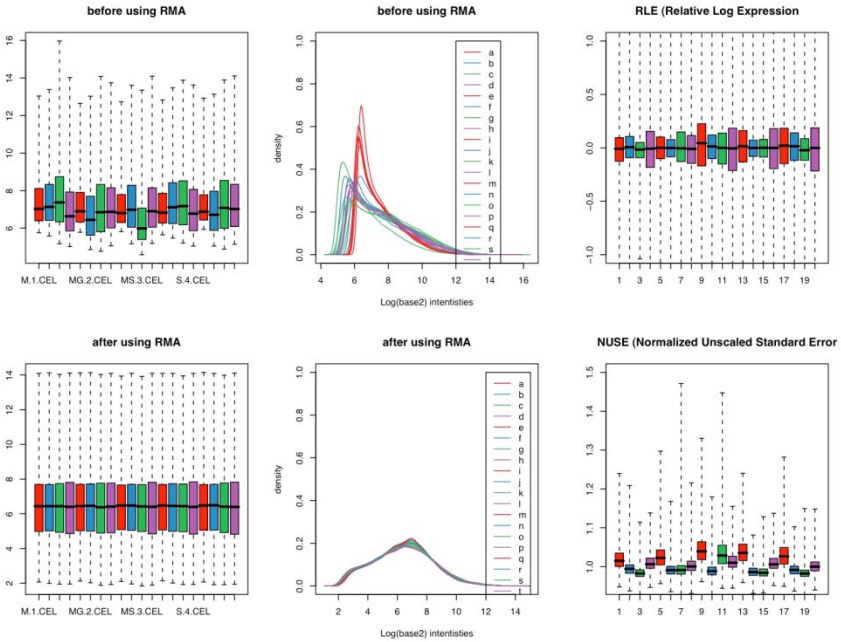
Genomics experiments are often screens rather than classical experiments.



- Reduce the field of variables to a manageable and high-likelihood set
- Design combines aspects of classical experimental design with epidemiological design
  - There is a sequential aspect
  - For example:
    - var{cell}<var{multiple cells}<var {tissues}<var{organs}
    - var {cell lines} < var{inbredorganisms} < var{outbredorganisms} < var{diseased outbred org}

The majority of genomics platforms do not have calibration controls or standards.

- Performance of the instrument and reagents cannot be assured, or compared between different labs.
  - This places extra burden on the experimental design.
  - Blocking: making sure you control for factors like instrument and technician.



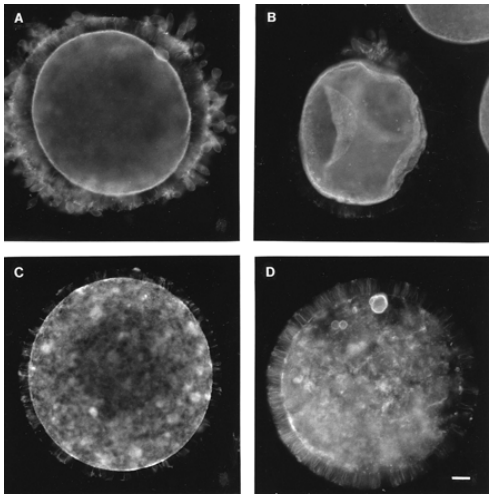
3 technicians, at least 2 scanners, 4 time periods, 5 conditions (each of 3 pooled individuals)

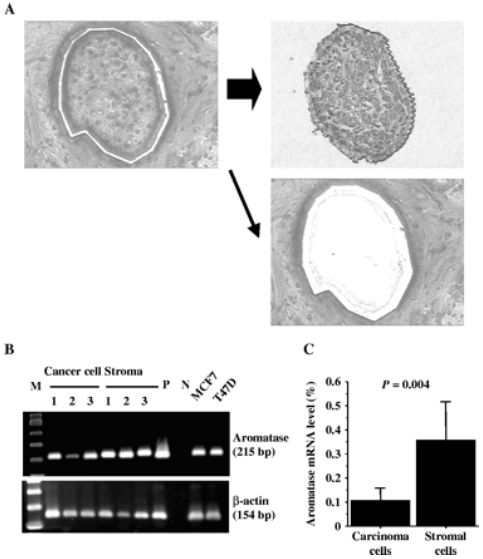
A sequencing *library* is the set of all target molecules that you intend to sequence, modified to allow the sequencing to occur.

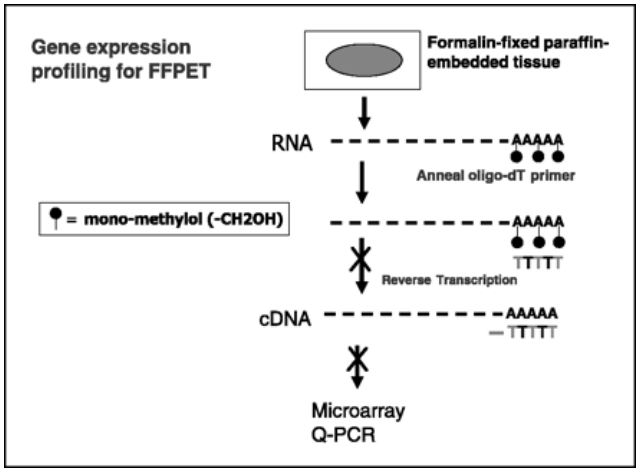
- Purified starting material: gDNA or cDNA (RNA that is converted to DNA).
- Produce fragment lengths compatible with the platform (25-nt to ~300nt).
- Modify the ends so that the adaptors specific to the platform can be added
- Ligate the adaptors to the fragments
  - Pooling strategies = bar codes
  - Clean-up or enrichment strategies
- Produce sufficient input mass.

The steps in library preparation will lead to some representational distortion in the sample.

- How is the sample stored?
- What subset of organs/tissues/cells will be used?
- What molecule will be profiled – how complete is its extraction?





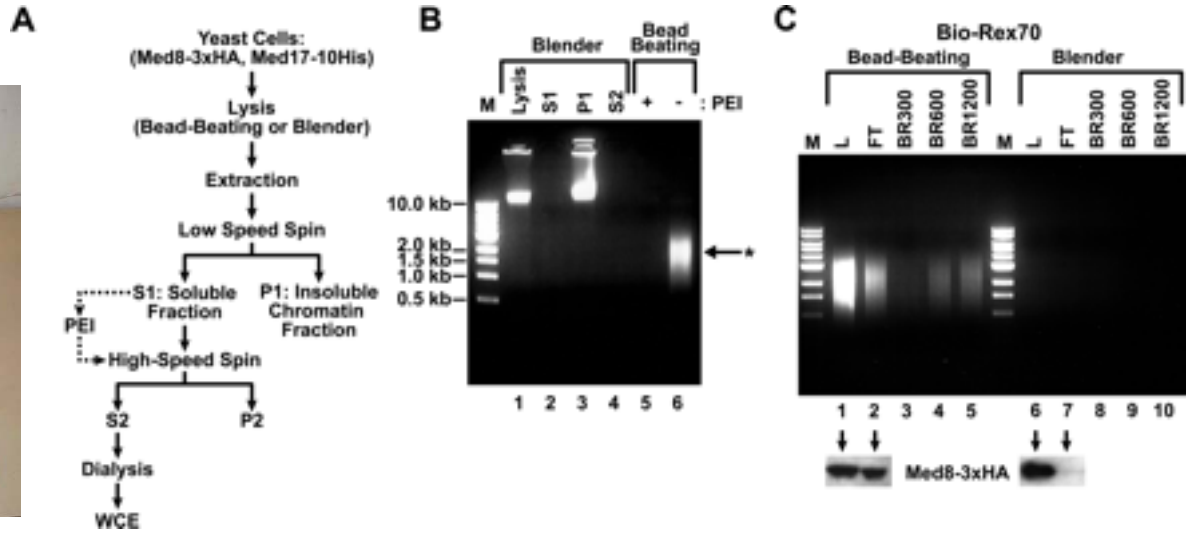


Sample sub-selection is the first step in the processing pipeline

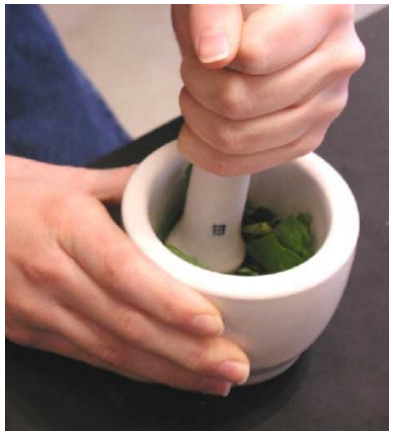
- Selection and capture of target material
  - Tissue- based uses large morphological features
  - Cell-based requires tissue preparation and then identification
    - Storage methods: cryogenic, embedding, inhibiting solutions (isotonic)
    - Identification: Microscopy and LCM, Antibody and FACS or column capture

Tissue Disruption is the second step in the processing pipeline.

- Cell disruption releases components
  - Break up cells so material is not trapped
    - Mechanical: homogenize, sonicate, vortex, French press, bead beating, grinding/macerating.
      - Fresh versus frozen tissue.
      - Inclusion of detergents and inhibitors
    - Enzymatic: target cell wall/membrane connecting molecules, use
  - Minimize degradation with bio/chemical reagents and/or physical techniques



Selective extraction of the target molecule is the third step in the processing pipeline.



- Selective solubility: molecules partition between solvents depending on their relative solubility. Additives can help modify this.
- Selective insolubility separates molecules that remain in solution from those that do not.
- CTAB: a cationic surfactant that forms complexes with negatively charged molecules
  - Proteins and carbohydrates remain soluble when the salt is below 0.5M and nucleic acid is insoluble, but becomes soluble above ~1.4M salt.
- Isopropanol/Ethanol: at higher concentration (>35% for isopropanol, >75% for ethanol).
- Phenol-$CHCl_3$: denatures proteins – they then become insoluble – they don't dissolve in either the buffer OR the phenol.
- NaCl: nucleic acids will not make a tight pellet if the phosphate backbone is not neutralized. Proteins are denatured in high salt.
- Beta-mercaptoethanol – helps unfold proteins, also helps prevent oxidation of small molecules like terpenes and other phenolics.
- SDS: a detergent that dissolves cell membranes, denatures proteins

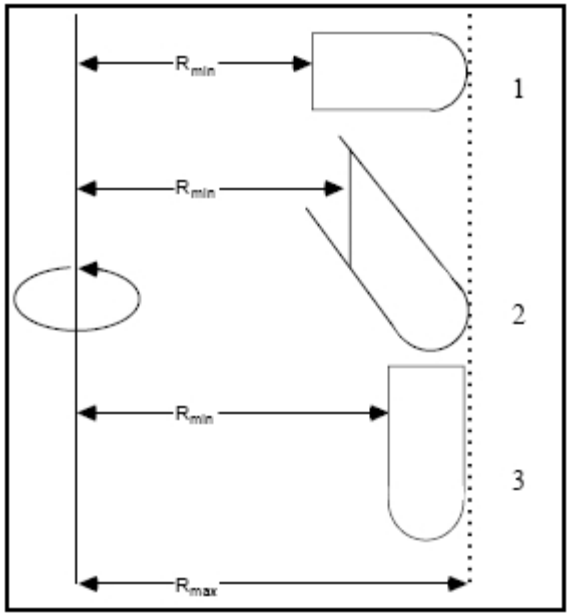
Solvent phases are usually separated by centrifugation.

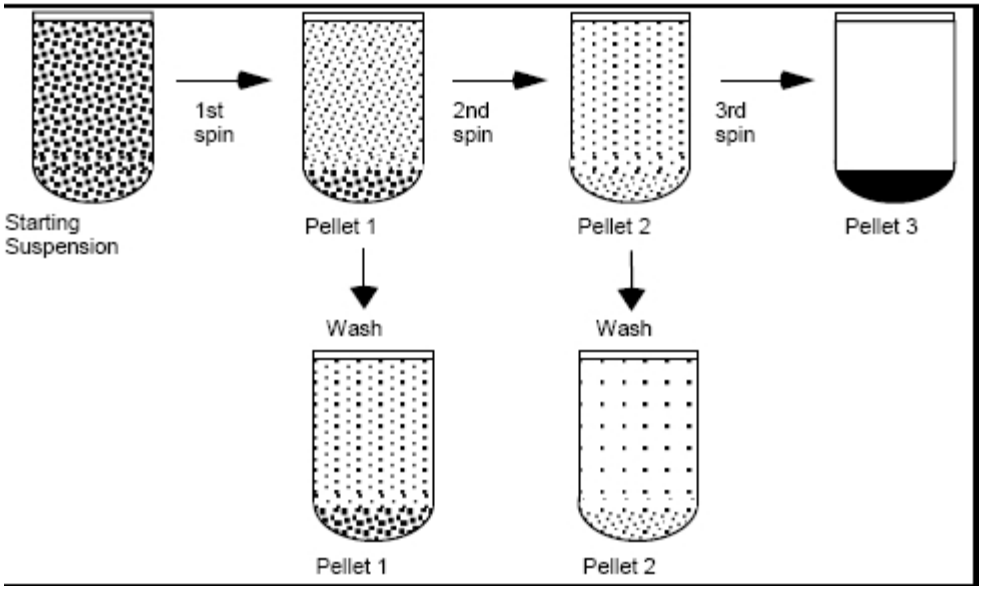
- The solvents are immiscible
- The solvents have different densities.
- Centrifugation: use centrifugal force to increase effective gravitational force – more-dense components migrate away from the axis and less-dense components towards the axis.
  - Pelleting: an insoluble material is raidly collected at the bottom of the tube
  - Density gradient: dissolved material migrates to the point where its density is equal to the solvent, where it bands.
  - The rate of centrifugation: angular velocity (rpm) or acceleration expressed as g.
    - Converting between RPM and g requires knowing the radius of the sample in the rotor.

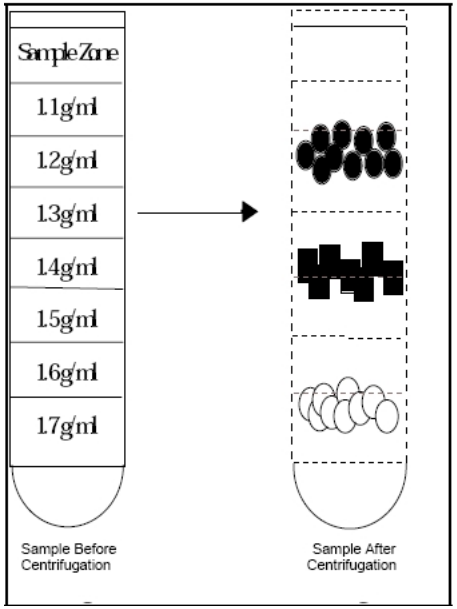
Pellet: hard-packed concentration of particles in a tube or rotor after centrifugation.
Supernatant: The clarified liquid above the pellet.
Adapter: A device used to fit smaller tubes or centrifugal devices in the rotor cavities.
RPM: Revolutions Per Minute (Speed).
R_{max}: Maximum radius from the axis of rotation in centimeters.
R_{min}: Minimum radius from the axis of rotation in centimeters.
RCF: Relative centrifugal Force. RCF = 11.17 x Rmax $(RPM/1000)^2$
K-factor: Pelleting efficiency of a rotor. Smaller the K-factor, better the pelleting efficiency.

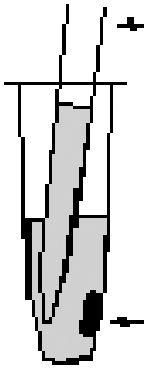
$$K = \frac{2.53 \times 10^{11} \, \mathrm{Ln}(R_{max}/R_{min})}{(RPM)^2}$$

	Centrifuge Classes		
	Lowspeed	**High-speed**	**Ultra/micro-ultra**
Maximum Speed (rpm x10^3)	10	28	100/150
Maximum RCF (x10^3)	7	100	800/900
Pelleting applications			
Bacteria	Yes	Yes	(Yes)
Animal and plant cells	Yes	Yes	(Yes)
Nuclei	Yes	Yes	(Yes)
Precipitates	Some	Most	(Yes)
Membrane fractions	Some	Some	Yes
Ribosomes/Polysomes	-	-	Yes
Macromolecules	-	-	Yes
Viruses	-	Most	Yes









Pellet is likely to one side (Point hinge away from axis)
Kepp pipette tip away from the pellet
Pipette solution up slowly and away from liquid surface

Set speed (note rotor and tube limits)
Balance tubes: 2 tubes with the same mass are 180° from each other
Set speed
Set temperature
Set rate