Molecular Biology 101

Overview

- Principles of Molecular Biology
- Applying the Concepts of Molecular Biology
- Workflow, Contamination, Safety, and Regulation
- Considerations for TB Molecular Testing

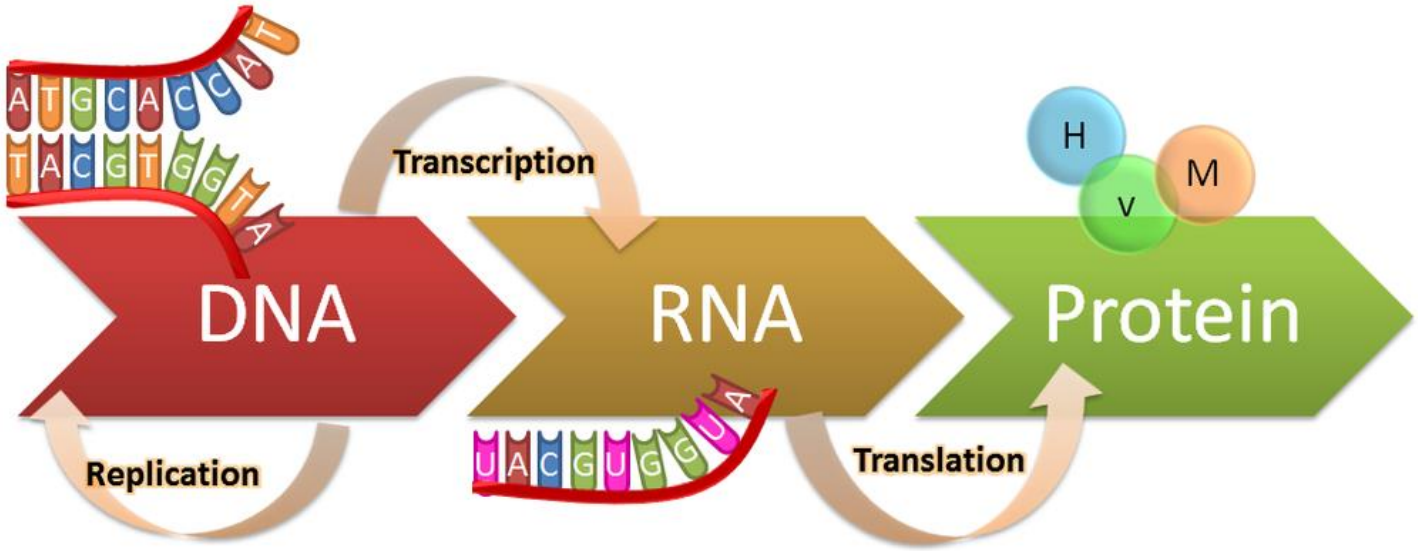
Molecular Biology 101

PRINCIPLES OF MOLECULAR BIOLOGY

Defining Molecular Biology

- The study of the formation, structure, and function of macromolecules essential to life, such as nucleic acids and proteins, and their role in cell replication and the transmission of genetic information

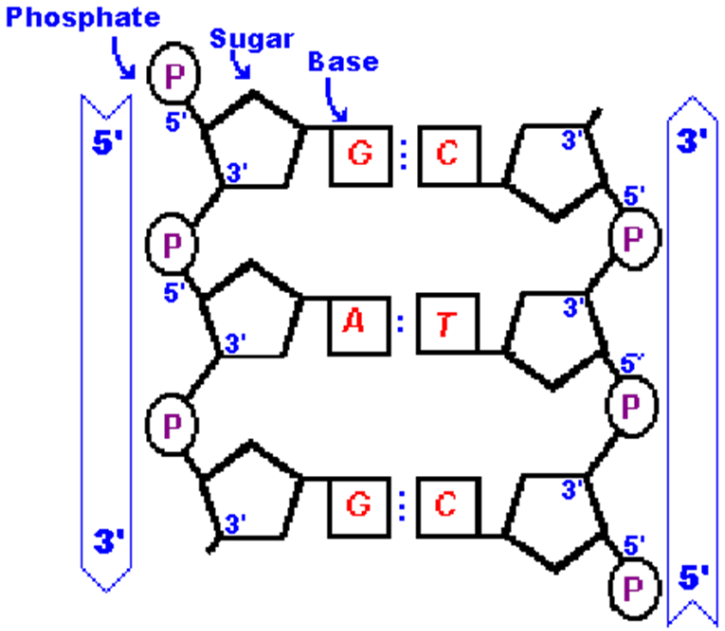
Central Dogma of Molecular Biology



http://lit.genius.com/Biology-genius-the-central-dogma-annotated

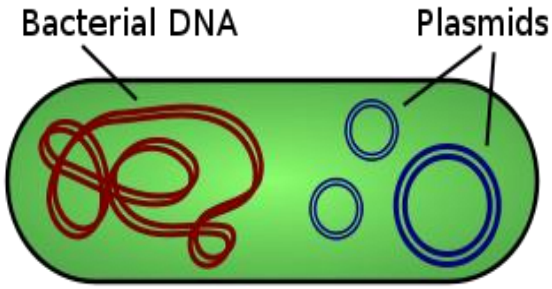
Deoxyribonucleic Acid (DNA)

- Polynucleotide chains consisting of 4 bases
  - **A**denine, **T**hymine, **G**uanine, **C**ytosine
  - There are approximately 4 million base pairs (bps) in *M. tuberculosis* genome (~4,000 genes)
- Double-stranded
  - Chains are held together by complementary base pairing
    - A-T
    - C-G
- Complimentary and anti-parallel
  - Strands of nucleic acids have their **sugar-phosphate backbone** arranged in **opposite directions**, creating a stable **double helix**
- Replication driven by DNA Polymerase, in the 5’ to 3’ direction



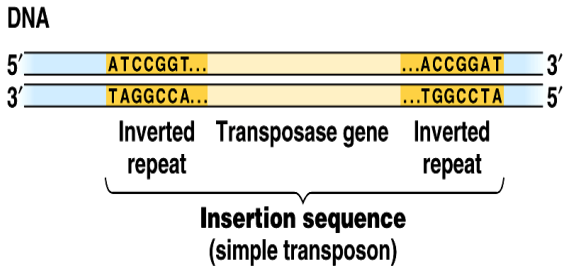
Different Types of Bacterial DNA

- **Chromosomal DNA**
  - Double-stranded chromosome
  - Varies in length between genera and species
  - Contains genes required for bacterial survival



- **Plasmids**
  - Circular double stranded DNA (dsDNA)
  - Replicate independent of the bacterial chromosome
  - Smaller than chromosomal DNA

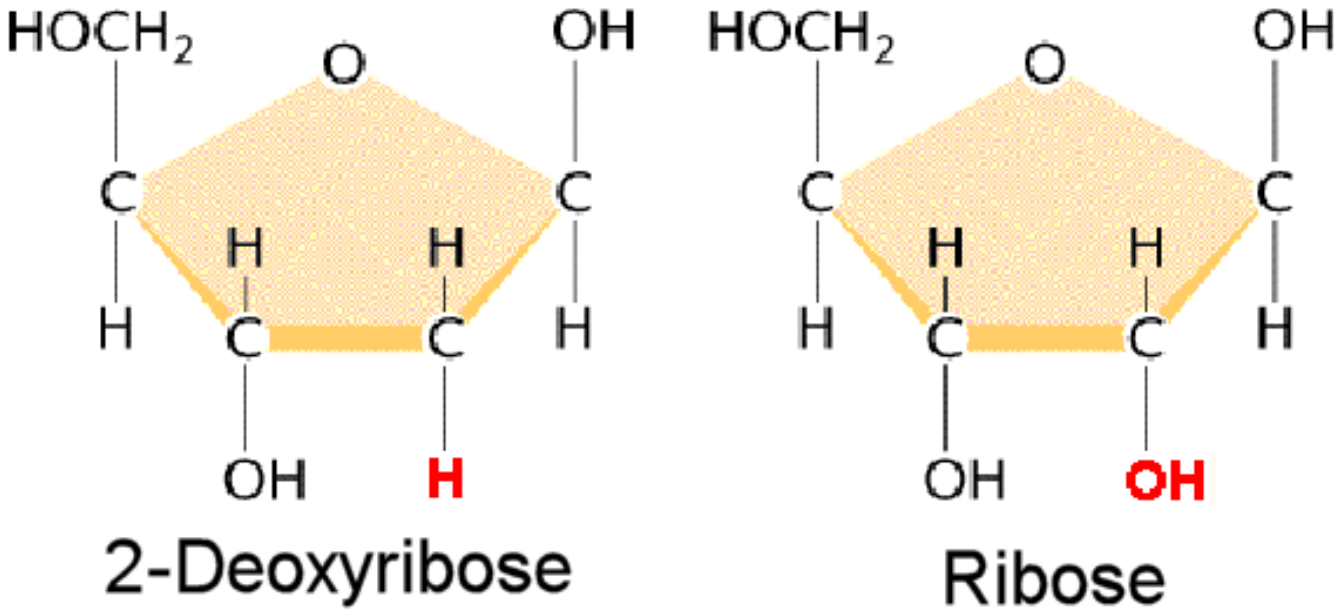
- **Insertion sequences**
  - Segments of DNA in a bacterial chromosome flanked by inverted repeats that can move from spot to spot within the chromosome
  - Interrupt coding sequence/block gene expression depending on where inserted within the genome



Ribonucleic Acid (RNA)

- Polynucleotide formed by 4 bases
  - **A**denine, **C**ytosine, **G**uanine, **U**racil
- Single-stranded
- Transcribed from DNA by RNA polymerase
- Can form hydrogen bonds with DNA, other RNAs, and itself
- Less stable than DNA
  - hydroxyl group on the 2’ carbon of the ribose

Deoxyribose vs Ribose sugars



2-deoxy-ribose *in* ***DNA*** *is replaced by* ***Ribose*** *in* ***RNA****.*

The extra ***-O-*** *in the ribose backbone prevents formation of stable double-helices.*

http://www.mun.ca/biology/scarr/Deoxyribose_vs_Ribose.gif

Important Definitions

Gene

A distinct sequence of nucleotides along a segment of DNA that provide the coded instructions for synthesizing a protein or RNA molecule

Locus

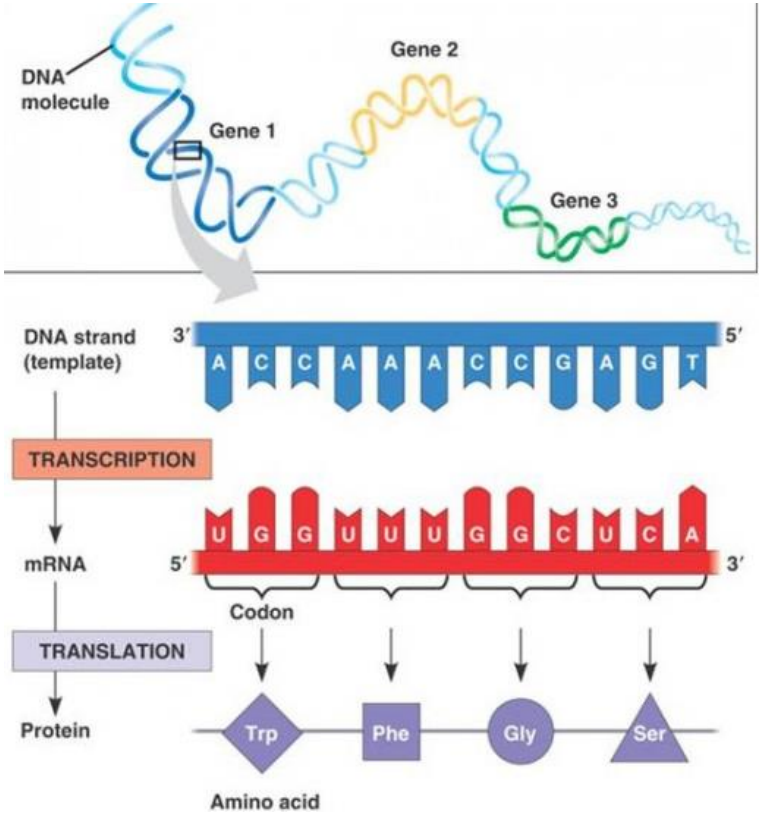
A region of interest in a gene

Codon

A combination of three consecutive nucleotides within a gene that specifies a particular amino acid

Amino Acids

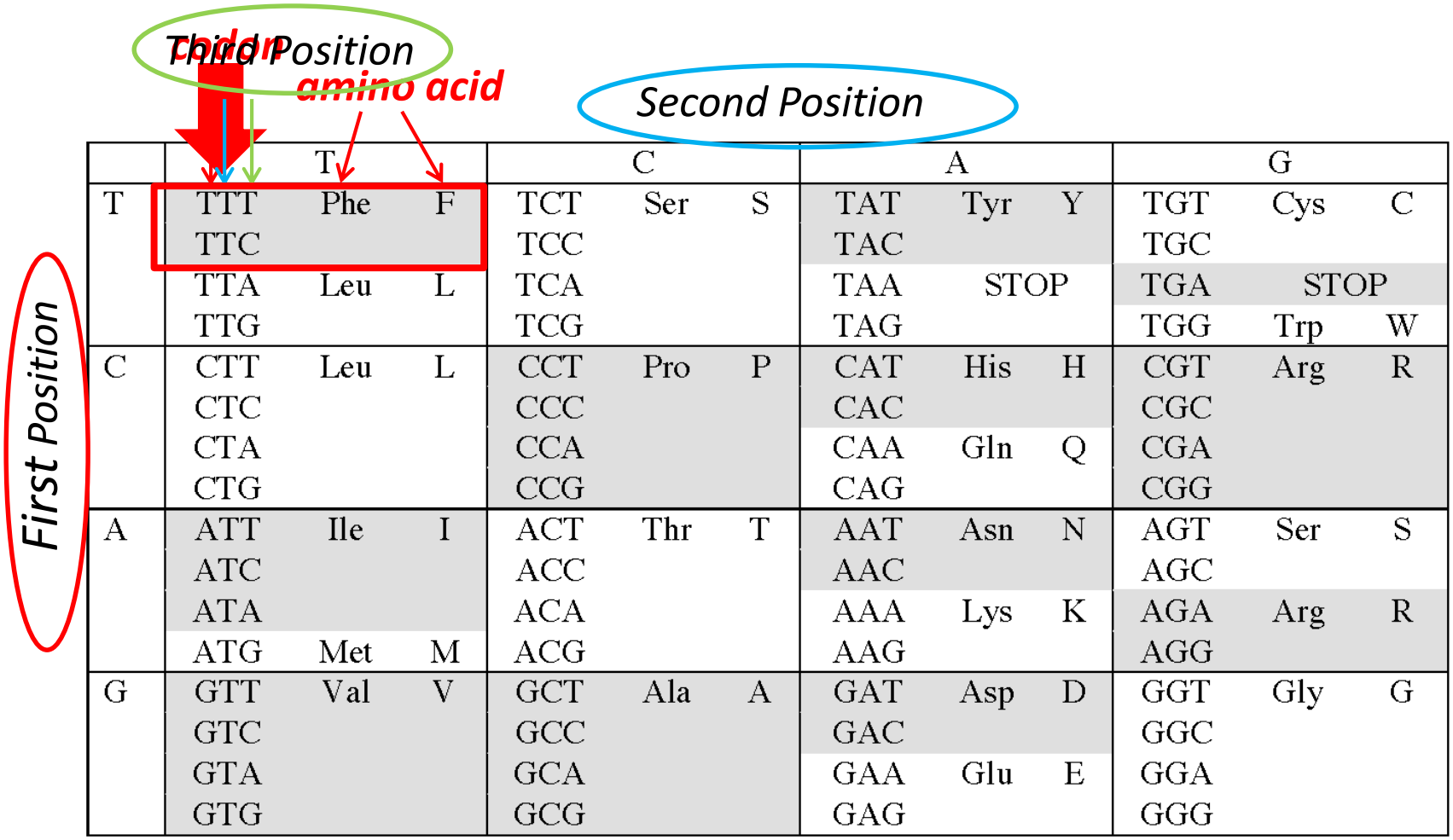
- Amino acids are the building blocks of proteins
- They are encoded by DNA
- Each group of three nucleotides within the open reading frame* makes up a codon
- Each codon encodes a specific amino acid based on the universal genetic code



http://bio1152.nicerweb.com/Locked/media/ch17/central_dogma.html

*regions of sequences that code for amino acids

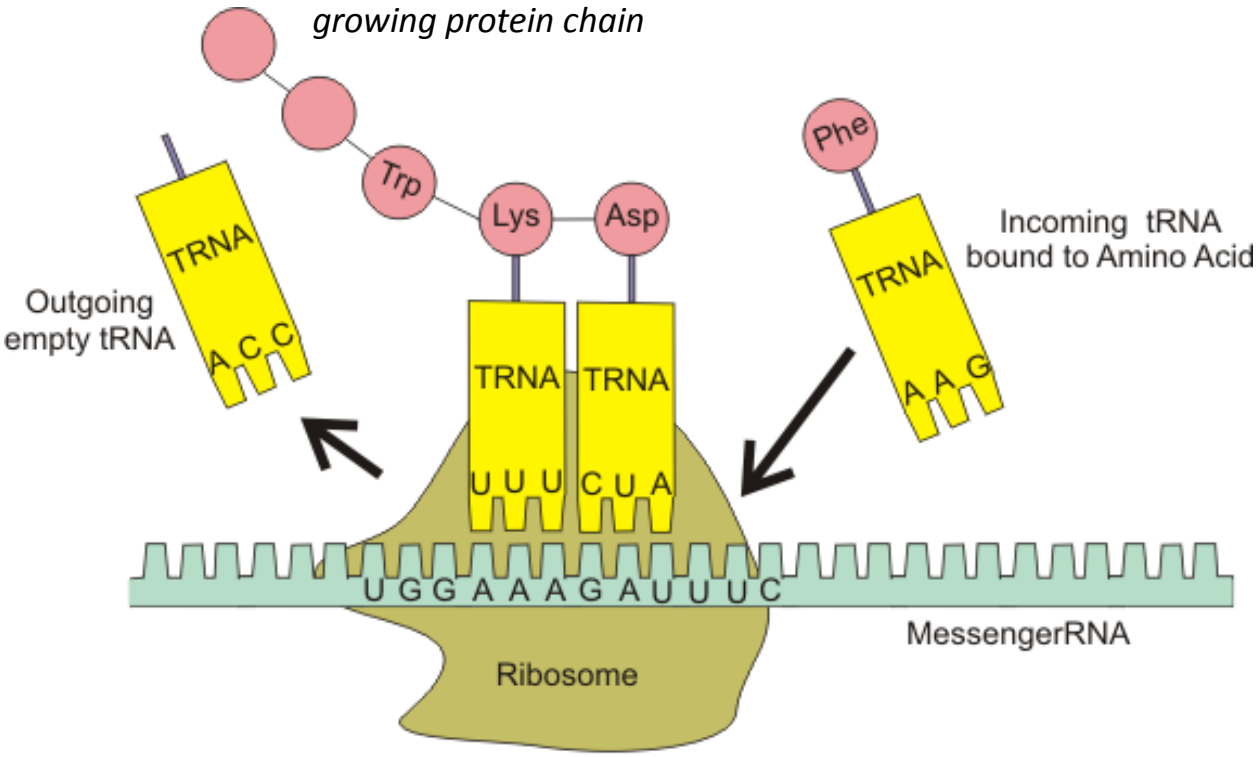
Universal Genetic Code



First Position	T			C			A			G		
T	TTT	Phe	F	TCT	Ser	S	TAT	Tyr	Y	TGT	Cys	C
	TTC			TCC			TAC			TGC		
	TTA	Leu	L	TCA			TAA	STOP		TGA	STOP	
	TTG			TCG			TAG			TGG	Trp	W
C	CTT	Leu	L	CCT	Pro	P	CAT	His	H	CGT	Arg	R
	CTC			CCC			CAC			CGC		
	CTA			CCA			CAA	Gln	Q	CGA		
	CTG			CCG			CAG			CGG		
A	ATT	Ile	I	ACT	Thr	T	AAT	Asn	N	AGT	Ser	S
	ATC			ACC			AAC			AGC		
	ATA			ACA			AAA	Lys	K	AGA	Arg	R
	ATG	Met	M	ACG			AAG			AGG		
G	GTT	Val	V	GCT	Ala	A	GAT	Asp	D	GGT	Gly	G
	GTC			GCC			GAC			GGC		
	GTA			GCA			GAA	Glu	E	GGA		
	GTG			GCG			GAG			GGG		

Proteins

Codons in the mRNA are read by translation machinery (e.g., ribosomes) to incorporate the corresponding amino acid in the formation of a nascent protein



http://www.genomebc.ca/files/5012/7439/7100/6.6.28.Translation.png

Mutations

- A mutation is a change in the DNA or RNA sequence resulting in variation from previous generations that may be transmitted to subsequent generations
- Range in size from a single base to a large segment of DNA
- Possible causes of mutations
  - Mutagens (UV-light)
  - Selective pressure (antibiotics)
  - Spontaneous, due to replication errors

Types of Mutations

- **Point mutation**—a single base in the sequence is changed. Also called a **S**ingle **N**ucleotide **P**olymorphism (SNP)

- **Deletion**—a single base or set of bases is deleted from the sequence

- **Insertion**—a new base or set of bases is inserted into the sequence

Influence of Mutations

- **Silent/synonymous mutation—a** nucleotide change that *does not* affect the resulting amino acid sequence
- **Nonsense mutation**—a mutation that results in a stop codon.
- **Missense/non-synonymous mutation—a** nucleotide change that results in a change in the amino acid sequence
- **Frame Shift**—a mutation that affects the **reading frame** of the gene (insertion or deletion not divisible by 3)

	No Mutation	Point Mutation		
		Silent	Nonsense	Missense
DNA level	TTC	TT**T**	**A**TC	T**G**C
Protein level	Lys	**Lys**	**STOP**	**Thr**

Mutations Illustrated

Normal Gene: AS THE MAN SAW THE DOG HIT THE CAN END ITI

Point Mutation: AS THE MAN SAW THE BIG HIT THE CAN END ITI

Frameshift Mutation: AS THE MAN SAW THE BIG GUY HIT THE CAN END ITI

Adapted From: http://www.brooklyn.cuny.edu/bc/ahp/BioInfo/MUT/Mut.Types.html

Mutations Illustrated (2)

Normal Gene:

AS THE MAN SAW THE DOG HIT THE CAN END ITI S

Point Mutation:

AS THE MAN SAW THE DOT HIT THE CAN END ITI S

Deletion:

*AS THE MAN SAW THE*HIT THE CAN END ITI S*

Insertion:

*AS THE MAN SAW THE **FAT** DOG HIT THE CAN END ITI S*

Frame Shift:

*AS THE MAN SAW THE ***OGH ITT HEC ANE NDI TIS***

Adapted From: http://www.brooklyn.cuny.edu/bc/ahp/BioInfo/MUT/Mut.Types.html

Naming Mutations

- Mutations may be named at the DNA or protein level in reference to the gene of interest.
- Substitutions, insertions and deletions in DNA can be indicated by naming the base position in combination with the specific change that has occurred
  - Substitution: 76A>C
  - Insertion: 76_77insT
  - Deletion: 76_78del
- If the mutation occurs before the start of the open reading frame, a negative number is used to indicate the position relative to the start codon
  - C(-15)T
- Changes in the codon sequence are designated by indicating the original amino acid followed by codon number and the resulting amino acid based on changes in nucleotide sequence
  - His526Asp or H526D for a missense mutation
  - Phe514Phe or F514F for a silent mutation

Molecular Biology 101

APPLYING THE CONCEPTS OF MOLECULAR BIOLOGY

Definitions

- **Clinical specimen—**specimen taken directly from the patient (e.g., sputum, CSF, pleural fluid, etc.)
- **Isolate—**organism isolated from culture of a clinical specimen
- **Direct detection—**detection of RNA or DNA sequences of interest in organisms present in a clinical specimen
- **Nucleic acid amplification—**exponential amplification of a specific sequence of nucleic acid
- **Amplicon—**the product of nucleic acid amplification
- **Laboratory developed test (LDT)—** *in vitro* diagnostic test that is intended for clinical use and is designed, manufactured, validated and used within a single laboratory

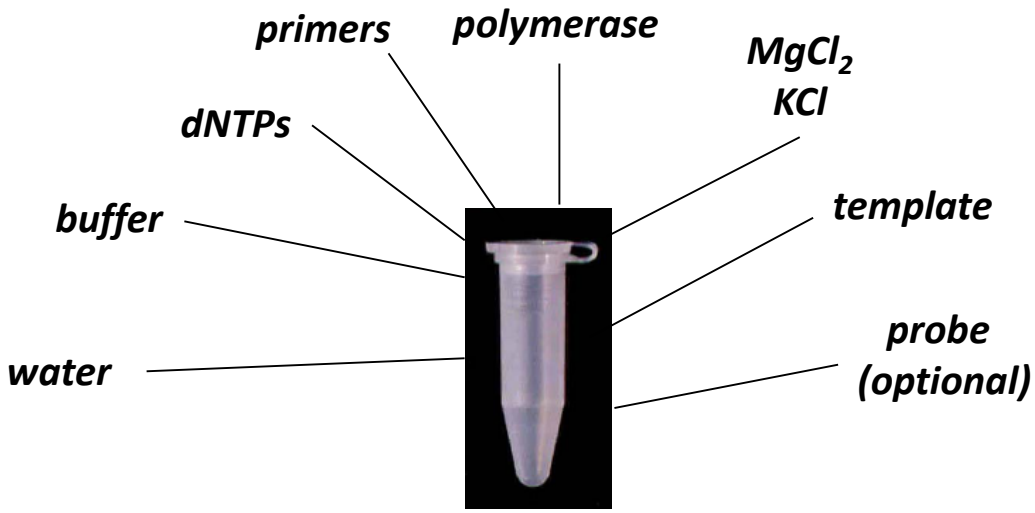
Nucleic Acid Amplification

- Nucleic acid amplification (NAA) increases sensitivity and specificity of a molecular-based assay, especially when only a few organisms may be present in a sample
- 2 types of NAA
  - Polymerase Chain Reaction (PCR) – utilizes thermal cycling to denature DNA, anneal primers, and elongate DNA amplicons; uses DNA polymerase
  - Transcription Mediated Amplification (TMA) – isothermal, produces an RNA amplicon; uses RNA polymerase

Polymerase Chain Reaction (PCR)

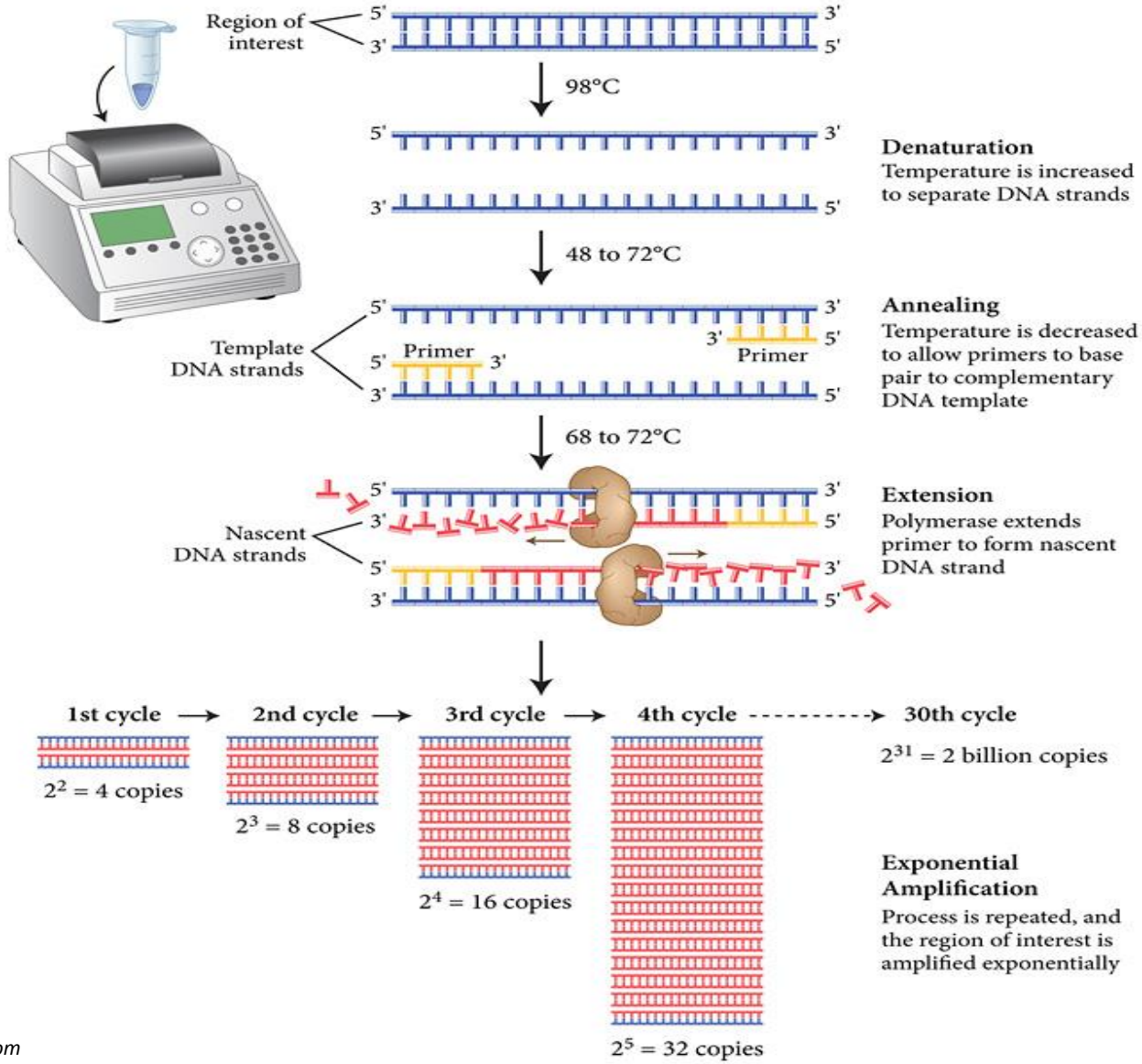
- A technique in which specific regions of DNA are amplified by the enzyme DNA polymerase
- Requires repeated cycles of heating and cooling which can be automatically performed in a thermal cycler
- The quantity of DNA doubles with each amplification cycle
- At completion, the presence, size, or intensity of the PCR products are analyzed by a variety of methods

Components of a PCR Reaction



COMPONENT	PURPOSE
Primers	Directs DNA synthesis to the desired region
dNTPs	Oligonucleotides; building blocks that extend the primers
KCl	Monovalent cation (salt) for optimal hybridization of primers to template
Tris buffer	Maintains optimal pH for enzyme reaction
$MgCl_2$	Divalent cation, required by the enzyme
Polymerase	Enzyme that extends the primers (adds dNTPs)
Template	Sample DNA that is being tested
Probe (optional)	May be added for detection of amplification in real-time PCR

Elements of a PCR Cycle

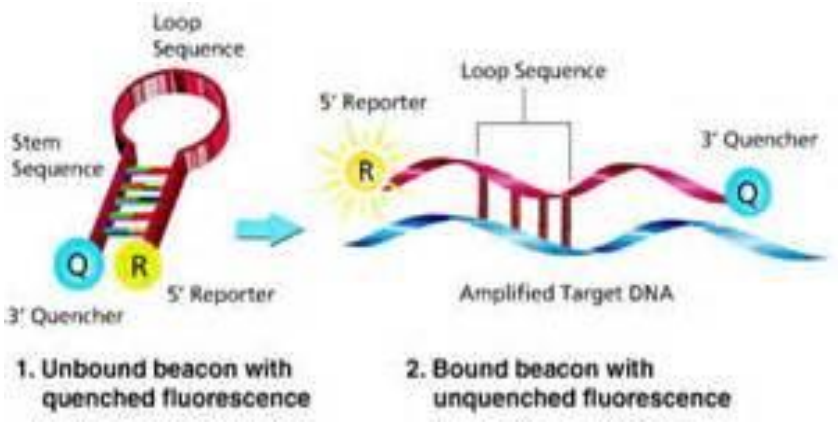


www.neb.com

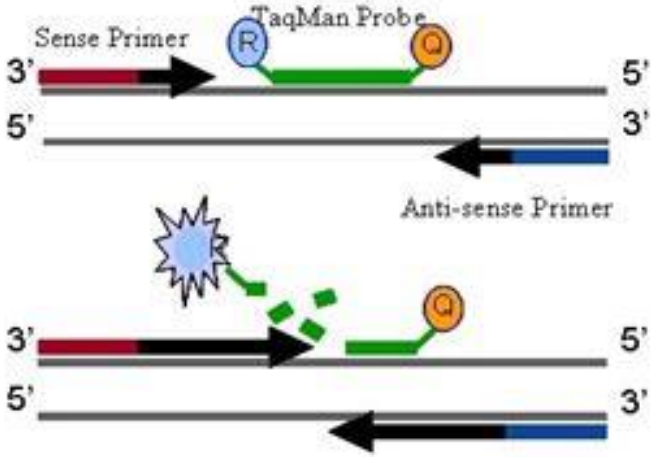
Real-time PCR

- Allows detection and/or quantification of amplified target sequences in real time
- Commonly referred to as quantitative PCR or qPCR, but not technically RT-PCR (i.e., reverse transcriptase)
- Same reagents as traditional PCR, with addition of a fluorescent reporter
  - Either a fluorescent dye that binds to double stranded DNA or a fluorescently labeled oligonucleotide probe specific for the target
- Real-time PCR is performed in a specialized thermal cycler that can visualize the fluorescence during each cycle
- Fluorescence levels are directly proportional to the quantity of target present

Real-time PCR: Detection



http://www.sigmaaldrich.com/technical-documents/articles/biology/molecular-beacons.html



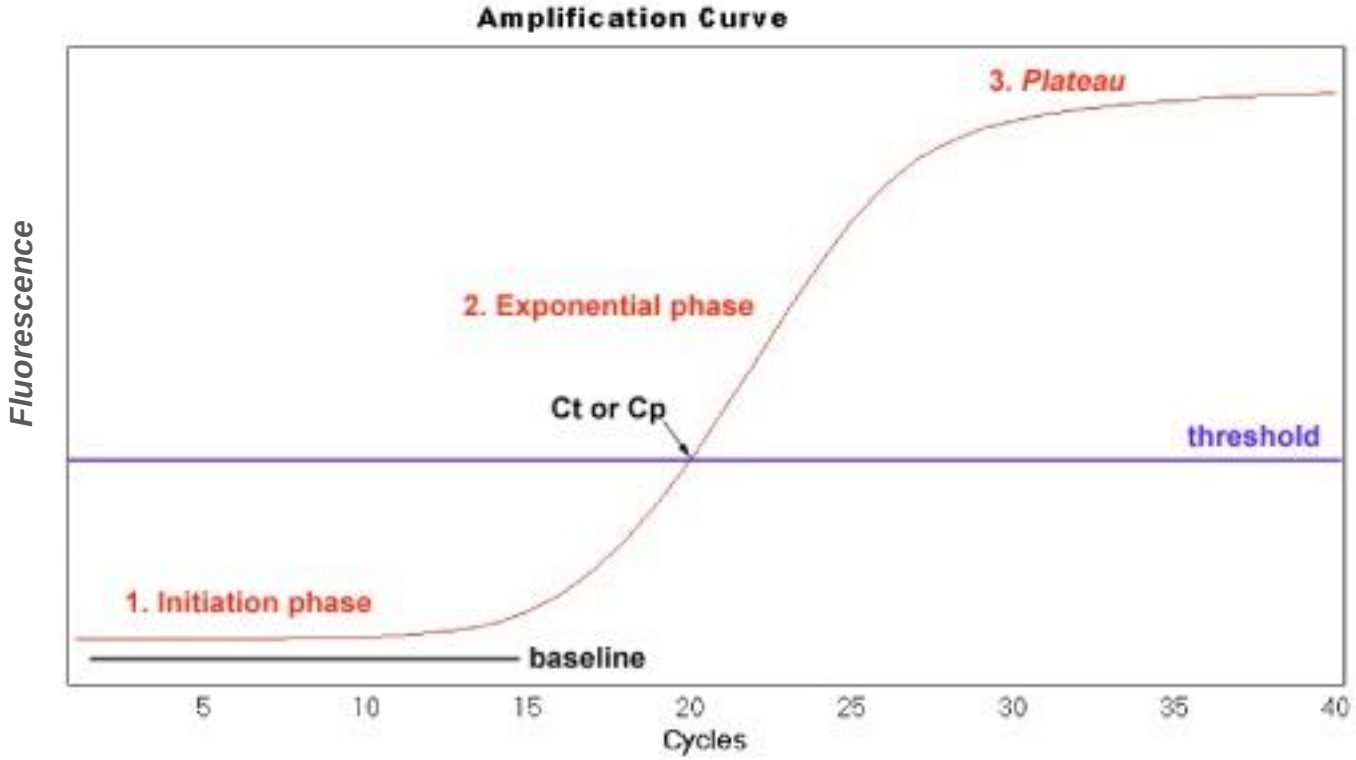
Molecular Beacons Probe

- Short, target-specific oligonuleotide (oligo) with a fluorescent reporter on the 5' end and a quenching molecule on the 3'
- Unbound probe forms a hairpin loop, bringing the quencher close to the fluorophore, thereby inhibiting fluorescence
- Binding to a complementary target sequence separates the quencher and fluorophore resulting in fluorescence

TaqMan Probe

- Short, target-specific oligo with a fluorescent reporter on the 5' end and a quenching molecule on the 3'.
- The proximity of the quencher to the fluorophore inhibits fluorescence.
- Polymerase degrades the probe during replication, separating the fluorescent reporter from the quencher resulting in fluorescence

PCR Amplification Plot



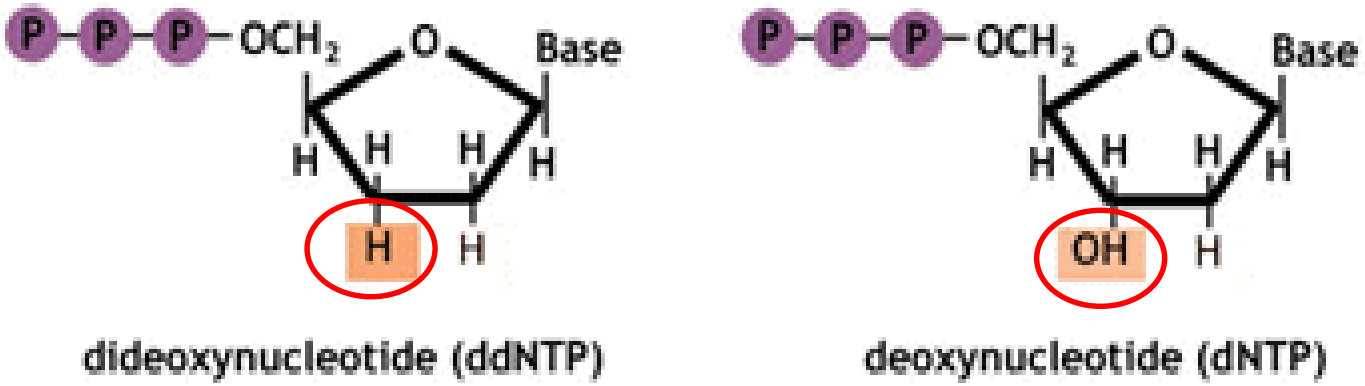
http://www.highveld.com/pcr/real-time-pcr-quantification-analysis.html

Nucleic Acid Sequencing

- The process of determining the specific order of nucleotides within a DNA molecule

- Methods include Sanger sequencing, pyrosequencing, and next generation sequencing (NGS)

Sanger Sequencing

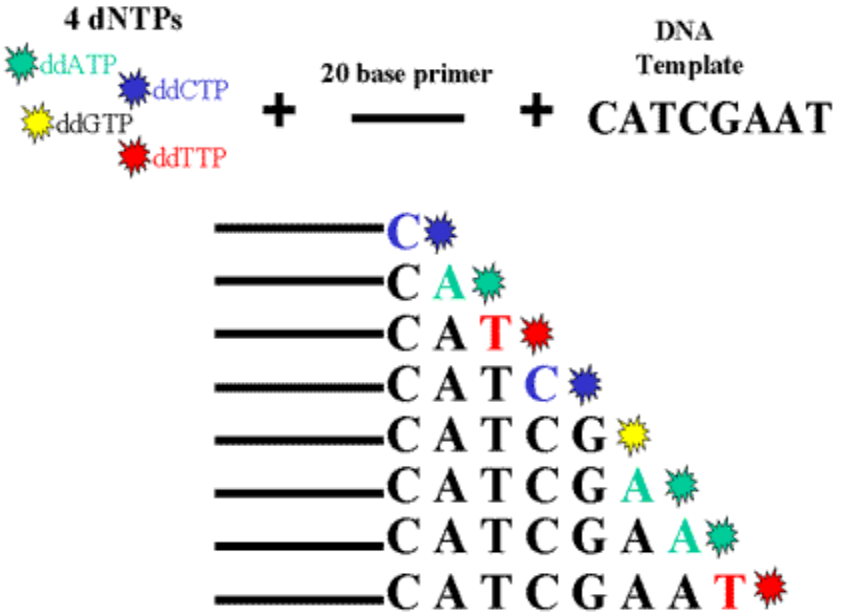


- Dideoxynucleotides (ddNTPs) are nucleotide analogs lacking the 3' hydroxyl group essential for continued chain extension during sequencing reaction
- Incorporation of ddNTPs into a growing DNA sequence results in chain termination, leaving different sized chains

http://www.bioinformatics.nl/webportal/background/techniques.html

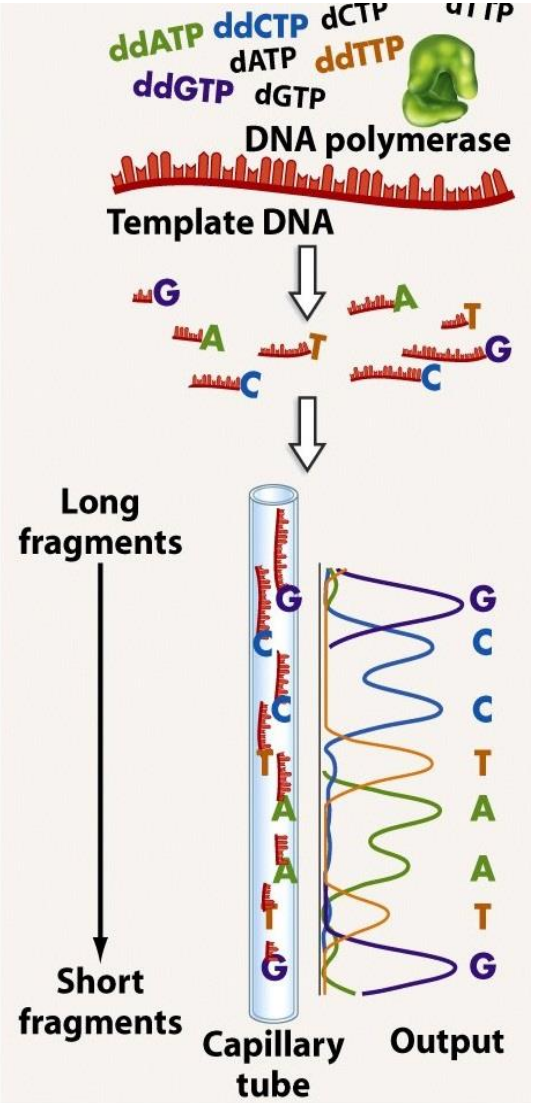
Sanger Sequencing: Detection

- Fluorescently labeled ddNTPs are used in combination with dNTPs ensuring production of DNA segments terminated at every base of the sequence
- Four different fluorescent dyes differentiate between terminal ddATP, ddTTP, ddGTP or ddCTP



http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july1999/images/dnaf7.gif

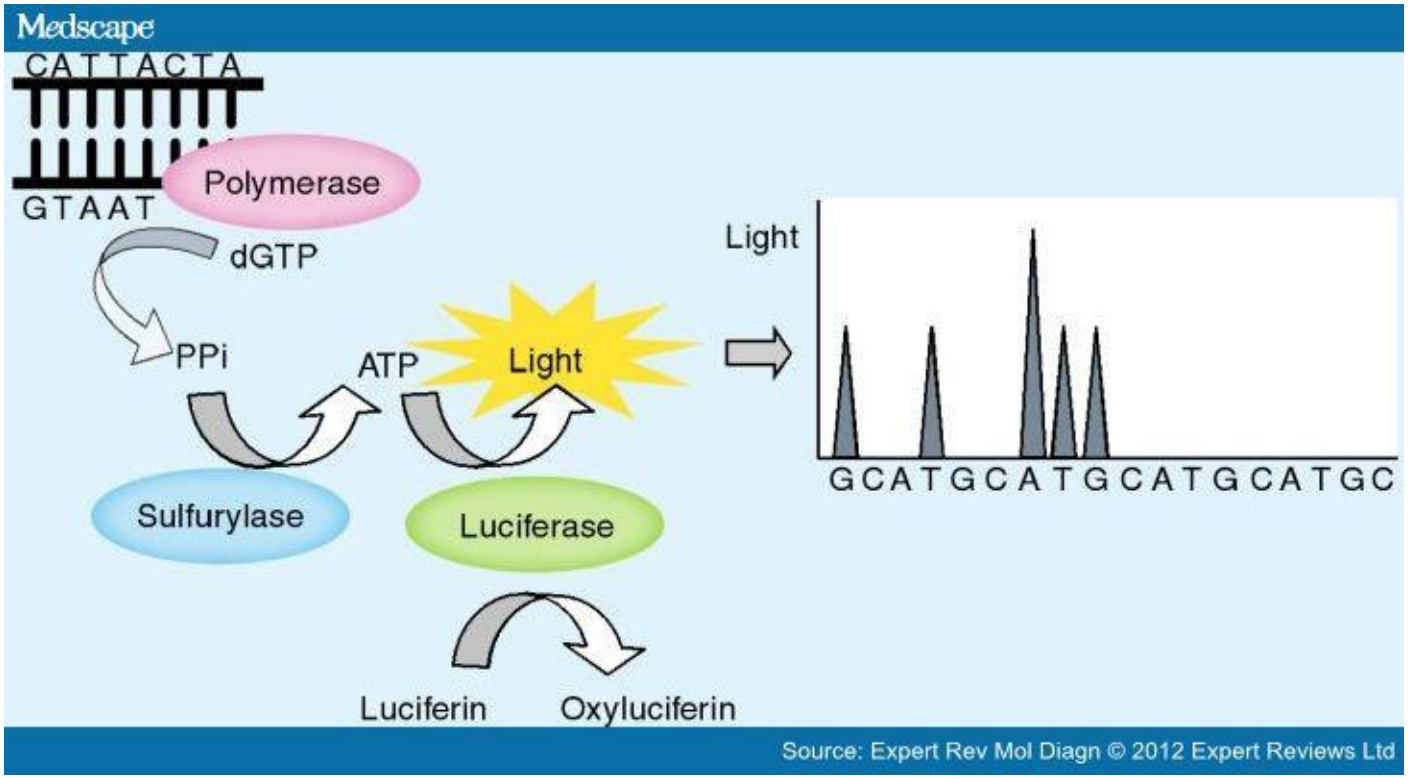
Sanger Sequencing: Detection (2)



http://www.uic.edu/classes/bios/bios100/lectures/sanger03.jpg

Pyrosequencing

A sequencing-by-synthesis procedure that relies on detection of pyrophosphate release following nucleotide addition



http://startravelinternational.com/tag/molecular-diagnosis-of-leukemia-medscape/page/255

Comparison of Sanger and Pyrosequencing

Sanger sequencing

- Long sequence reads (>500 bp)
- Easy to customize
- Ability to detect mixed sequences (LOD ~30%)
- Actual DNA sequence is determined
- Output visualized by a chromatogram—software available to analyze

Pyrosequencing

- Short sequence reads (~100 bp); cannot sequence large segments of DNA
- Dispensation order of nucleotides is customizable
- Throughput may be greater than for Sanger
- Limited ability for mixed sequences
- Actual DNA sequence is determined
- Output visualized by a pyrogram—built-in software to analyze

Sanger vs. Pyrosequencing

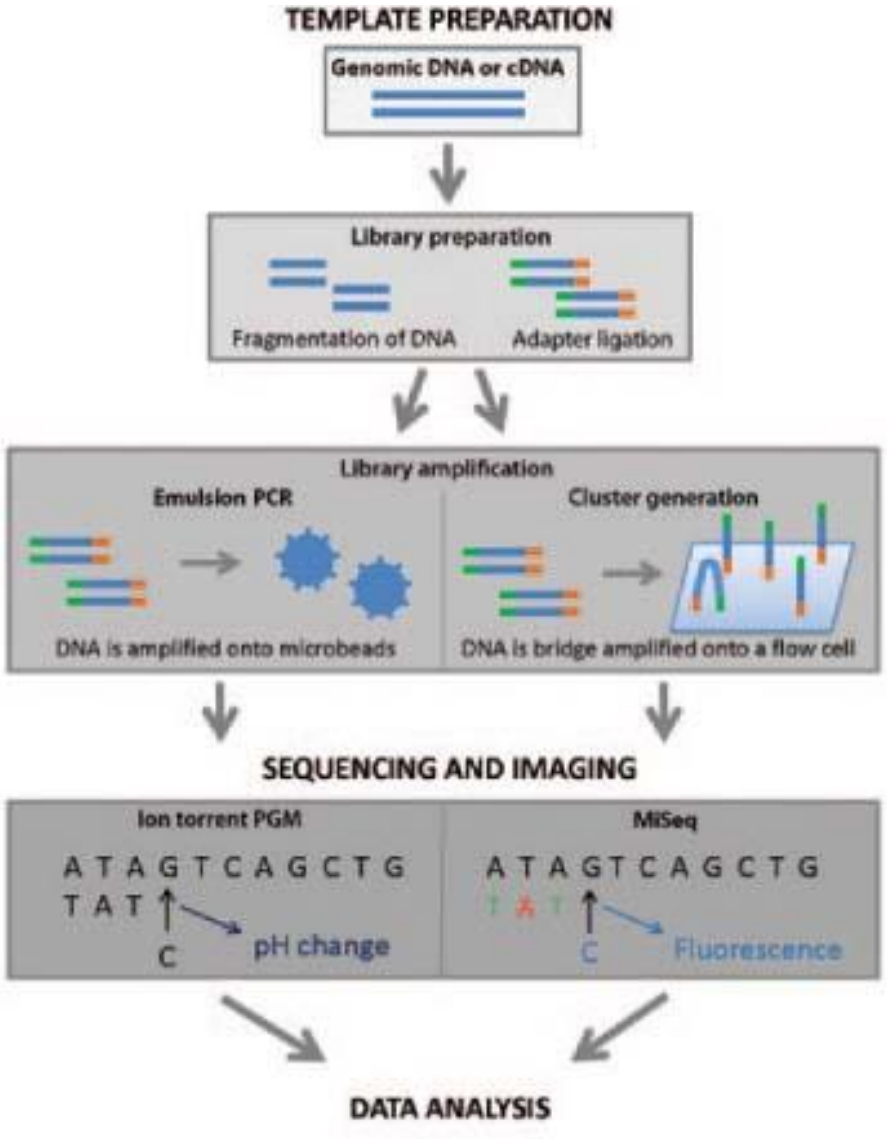
Features	Sanger sequencing	Pyrosequencing
Length of sequence reads	Long; up to ~800 bp	Short; ~60 bp
Customize Assays	Yes	Yes
Ability to detect mixed sequences	Yes; LOD ~30%	No
Ability to find new mutations	Yes	More limited than Sanger
Labor intensive	High	Moderate
Run time per segment (depends on number of loci examined)	1.5 – 2 hours	Minutes to 1.5 hours
Equipment Cost	High; e.g. ABI3500xl ~$180K	Moderate to high ; e.g. Pyromark Q24 ~$35K, Q96 ~$88k
Reagent cost	High	Moderate

Next Generation Sequencing (NGS)

- A collection of techniques/technologies for high-throughput sequencing; also known as massively parallel sequencing
- Includes sequencing of genomes as well as targeted sequencing
- Currently several different technologies and instruments
- Results in large, complex amounts of data and requires intensive bioinformatic analysis
  - Data storage can be problematic

NGS Methodology

- Genomic DNA is fragmented into small segments
- The segments are sequenced in millions of parallel reactions
- The strings of bases are then reassembled (alignment)
  - **reference genome** (used as a scaffold)
  - **de novo** (assembly in the absence of a reference genome)



Applications of NGS

- Provides a high throughput alternative for sequencing
- Applications include basic science as well as clinical diagnostics
  - Molecular typing
    - Pathogen identification and characterization
    - Detection of mutations associated with drug resistance
    - Identification of novel mechanisms of resistance
  - Molecular epidemiology – genotyping
    - Outbreak detection
    - Surveillance

NGS Instrumentation

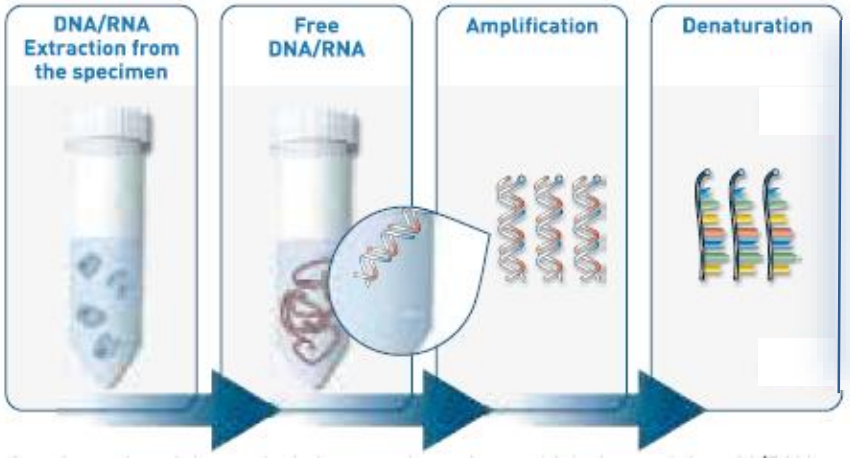
Illumina MiSeq

PacBio RS

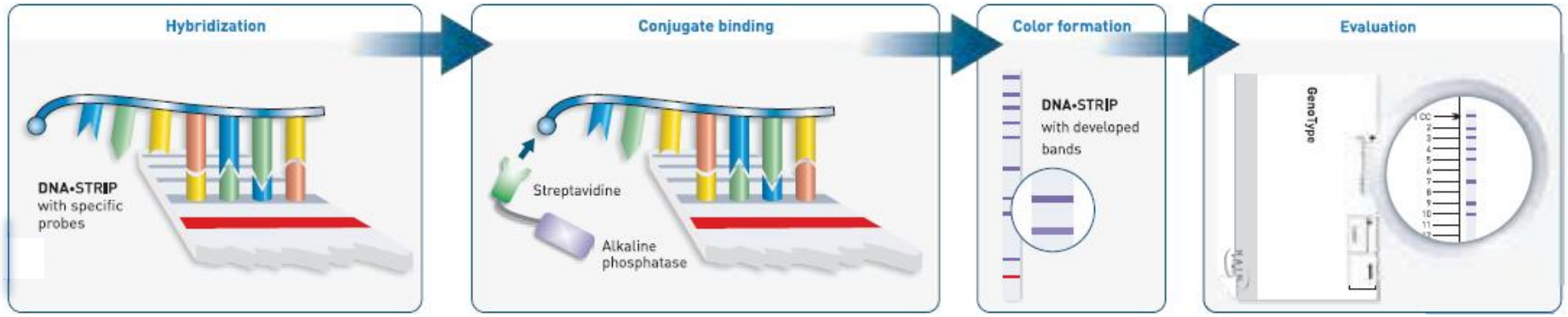
Ion Torrent Personal Genome Machine

Line Probe Assay

Sample preparation



Hybridization and visualization



http://www.hain-lifescience.de/en/technologies/dnastrip.html

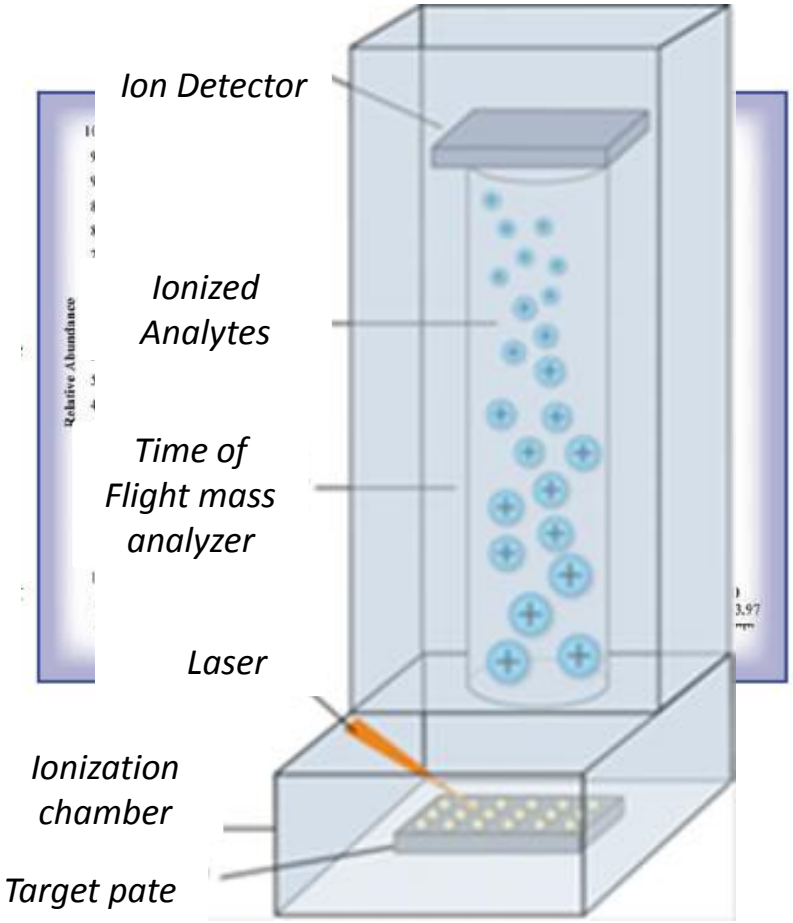
MALDI-TOF MS

- Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry
- Historically used to analyze proteins and other macromolecules
- Increasingly used in clinical realm
- Identifies bacterial species based on protein composition

http://www.bruker.com/

MALDI-TOF MS (2)

- A laser is fired at bacteria combined with matrix/sample crystals
- Energy is absorbed, converted to heat; sample is ionized
- Positive ions accelerate through a vacuum and drift toward the detector
- Speed (time of flight) is proportional to the ion's mass; the different ions are translated to a spectrograph – smaller ions reach detector first
- Spectrograph produced is compared to a database to see which bacterial species matches most closely



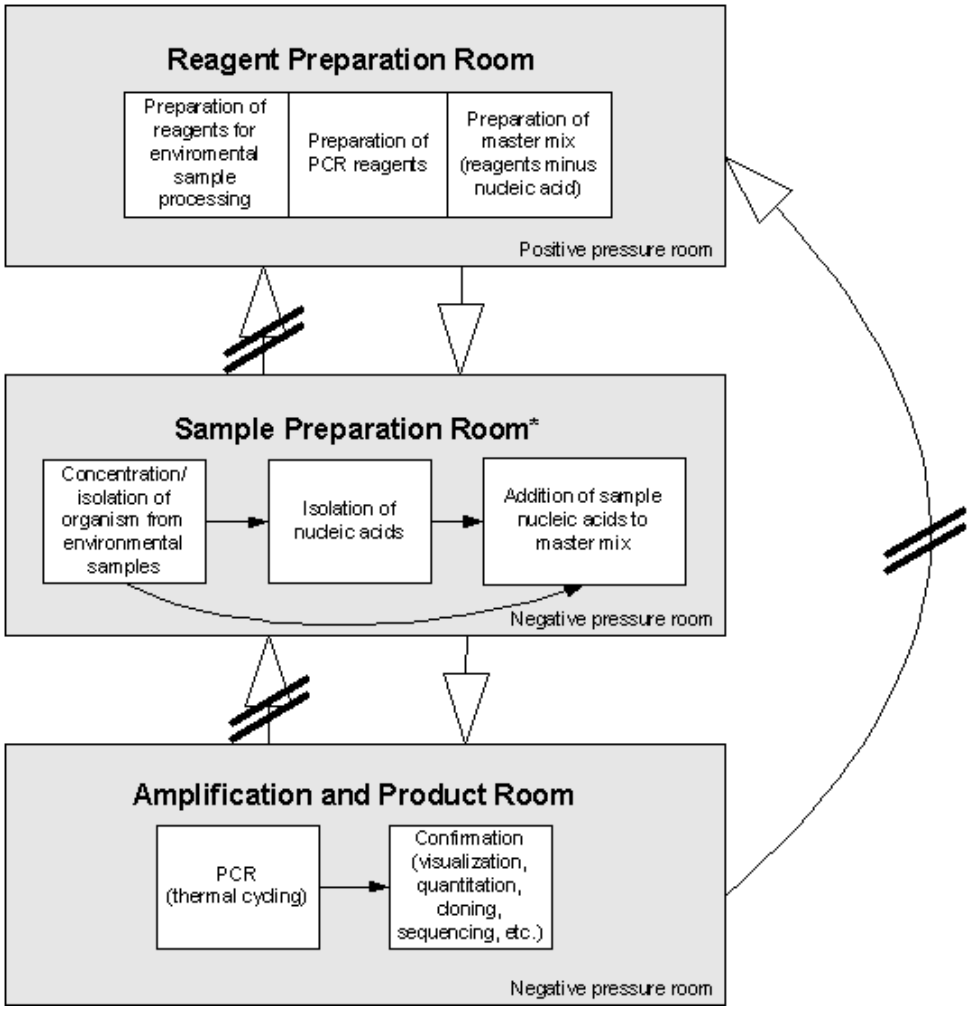
MOLECULAR BIOLOGY 101

WORKFLOW, CONTAMINATION, SAFETY, AND REGULATION

Workflow Considerations for Implementing Molecular Testing

- Need distinct areas/rooms
  - Reagent preparation
  - Specimen preparation/nucleic acid extraction
  - Amplification
  - Product analysis
- Must have unidirectional workflow
  - Individuals always work from “clean” to “dirty” and never in the opposite direction
  - Minimizes risk of amplicon cross-contamination
  - Dedicated equipment including refrigerators, freezers, pipettes, aerosol resistant tips, lab coats, pens, notebooks in each area
  - Gloves must be worn and changed often

Workflow Through Laboratory Rooms



**Sample/negative control preparation and positive control preparation should occur in physically separated areas (hoods) or different rooms.*

http://www.epa.gov/microbes/qa_qc_pcr10_04.pdf

What is PCR Contamination?

- The unintentional presence of nucleic acid target
- Since PCR is highly sensitive, the presence of even 1 copy of a contaminant can compromise the reaction causing false-positive PCR results
- Types of contamination
  - sample to sample (template) contamination
  - carryover of DNA (amplicon) from a previous amplification of the same target

Potential Sources of Contamination

- Template or amplicon transfer
  - Aerosol generation
  - Tube breakage
  - Inadequate plate sealing
  - Inadequate cleaning procedures
  - Pipette or reagent contamination
- Noncompliance with unidirectional workflow
  - Movement of individuals from dirty to clean areas
  - Transfer of reagents, paper, racks, lab coats, etc., from dirty to clean areas

Detection of Contamination

- A no template control (NTC) should be included with each run to check for contamination of reagents or cross-contamination between samples
- If NTC is positive, action must be taken to identify the potential source of contamination
  1. Consider reagents and equipment as possible source of contamination
  2. Environmental wipe testing of molecular areas may be performed to detect both template and amplicon contamination
  3. Clean suspect areas (5- 10% freshly made bleach, followed by ethanol, has been recommended [1, 2])
  4. Change practices as necessary to reduce chances of repeat contamination event
  5. Rerun assay—results are not valid

http://www.biofiredx.com/media/Advisory-Notice-Contamination-Prevention-0113.pdf

CLSI Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine; Approved Guideline, Second Edition. CLSI document MM09-A2, 2014

Safety

- DNA extractions of primary specimens and culture material should include an inactivation step (kill step)
  - To ensure safety of all staff, all protocols used to kill or deplete organisms must be validated in individual laboratories in accordance with laboratory and institution guidelines
  - Routine viability monitoring of inactivation procedures is necessary
- All work should be performed by properly trained and competency-assessed staff at the appropriate biosafety level for your facility

Considerations for Molecular Testing

- Confer with your institution's Compliance Director or Quality Assurance Officer before bringing on new technologies
- All newly implemented assays require some level of validation, even those that are FDA approved
- Consult the CLIA and FDA websites for detailed information
- Please see references and resources at the end of the module

Regulatory Classification of Assays

- Molecular assays for TB that are cleared, approved, or authorized by FDA
  - Hologic Amplified MTD—for detection of MTBC
  - Cepheid GeneXpert MTB/RIF—for detection of MTBC and mutations associated with rifampin resistance
  - Hologic Accuprobe—for ID of MTBC from culture
- Use of Laboratory Developed Tests (LDTs), or assays labelled as Research Use Only (RUO) or Investigational Use Only (IUO) require proper validation in compliance with the regulations of your accrediting body (e.g. CLIA, CAP)

Why the Increased Interest in Molecular Diagnostics for TB?

- Rapid detection of *M. tuberculosis* complex
- Rapid detection of mutations associated with rifampin and isoniazid resistance; rapid detection of mutations associated with resistance to second-line drugs
- Simultaneous detection of *M. tuberculosis* complex and mutations associated with drug resistance (e.g. GeneXpert)
- Lower cost and greater variety of platforms/reagents available
- More rapid outbreak detection

Use of Molecular Assays in the TB Laboratory

Assay Purpose	Specimen Types	Importance	Platforms Available
Direct detection of MTBC by NAAT	• AFB smear positive and smear-negative clinical specimens	• Patient isolation and initiation of therapy	• Cepheid GeneXpert • Hologic MTD • LDTs
Identification of Mycobacteria	• AFB-positive cultures • Clinical specimens (dependent on assay)	• Rapid, accurate and reproducible identification of mycobacteria • Initiation of therapy	• Line Probe Assays • DNA sequencing (16S, rpoB, hsp65) • MALDI-TOF • Other LDTs
Detection of drug mutations	• Clinical specimens • MTBC-positive cultures	• Rapid results for prompt patient therapy decisions	• Cepheid GeneXpert • Line Probe Assays • Molecular beacons • Pyrosequencing • Sequencing • Real-Time PCR

MOLECULAR BIOLOGY 101

REFERENCES AND RESOURCES

CLIA Program
Regulatory Information Resources

- CMS CLIA website (program info, statistics, etc.) http://www.cms.hhs.gov/clia/
- CLIA and LDT FAQs http://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/LDT-and-CLIA_FAQs.pdf
- CLIA Subpart K Part 1 (verification, QA) http://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/apcsubk1.pdf
- CLIA Subpart K Part 2 (standards, mycobacteriology) http://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/apcsubk2.pdf
- CLIA Subpart M (personnel and staffing) www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/apcsubm.pdf
- CDC CLIA website (CLIAC, regulations, etc.) http://wwwn.cdc.gov/clia/
- FDA CLIA website (complexity categorizations, waiver) http://www.fda.gov/cdrh/CLIA/index.html

FDA Regulatory Resources

- FDA (laboratory develop tests) http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm407296.htm
- FDA (in-vitro diagnostics) http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/default.htm

Other Resources

- Burd, Eileen. Validation of Laboratory-Developed Molecular Assays for Infectious Diseases. 2010. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2901657
- Update on FDA Reclassification, Gitterman, Steve. APHL 2013 TB Conference http://www.aphl.org/conferences/proceedings/Documents/2013/2013-National-TB-Conference/ID_2013August12_23_Gitterman.pdf

CLSI References

- MM03 A2 Molecular methods for infectious diseases
- MM06-A2 Quantitative methods for infectious diseases
- MM09-A2 NAA sequencing methods—covers Sanger, Massively parallel, NGS, validation, QA.
- MM19-A Establishing Molecular testing in clinical labs
- MM22 -A Microarrays for infectious diseases http://clsi.org/standards/

Other Resources

- Pyrosequencing video: http://www.youtube.com/watch?v=jylCHBxTKkw-
- CDC Advanced Molecular Detection (AMD) for TB site http://www.cdc.gov/amd/project-summaries/detecting-tuberculosis-accurately.html
- Heat-kill/viability study http://www.ncbi.nlm.nih.gov/pubmed/15667662