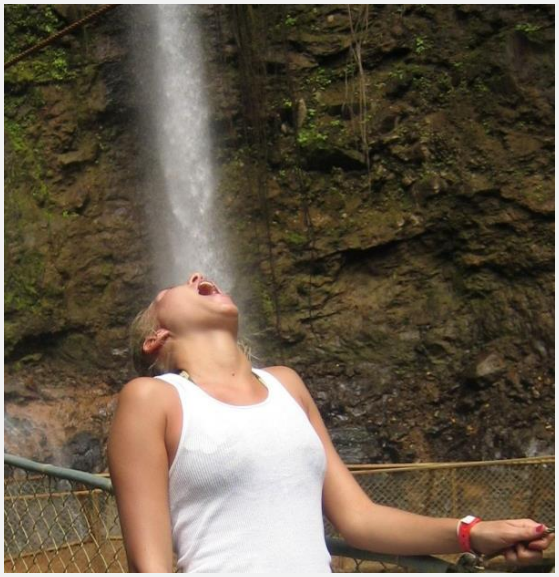
NGS applications and introduction to analysis

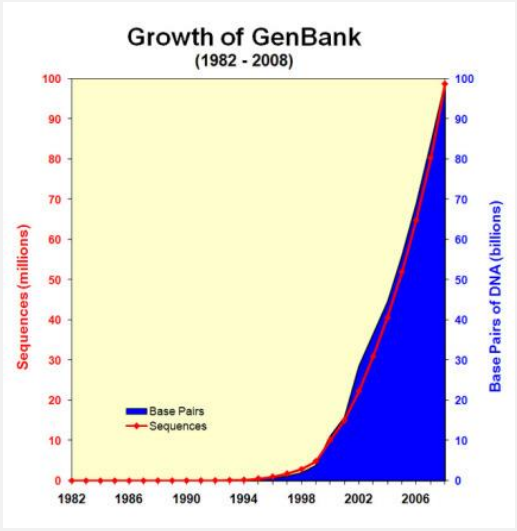
Ester Feldmesser

Introduction to Deep Sequencing Analysis

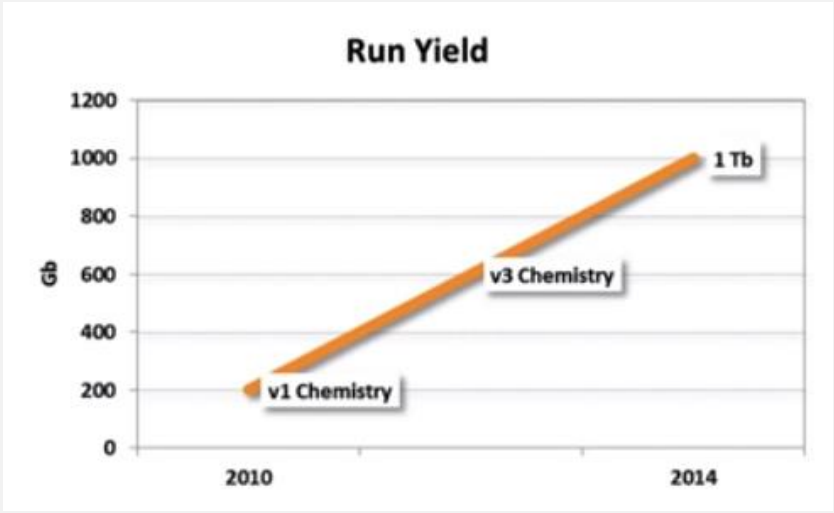
March2016

Sequencing over time



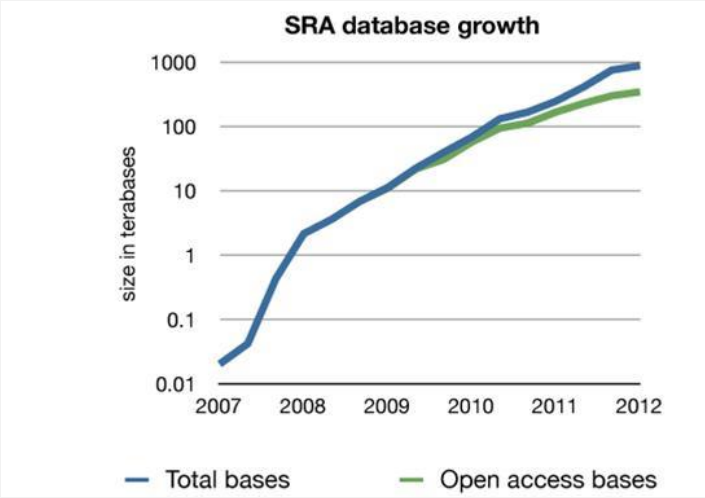


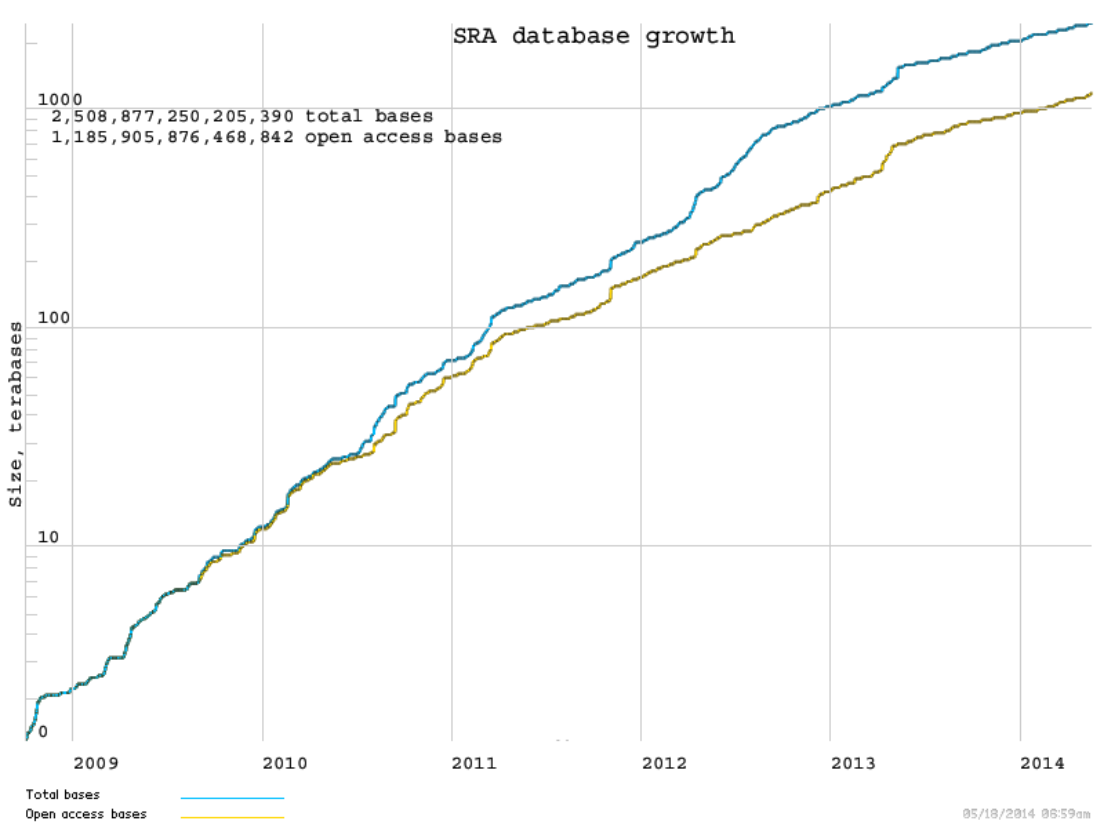
Genbank 2005 – 50 Gb



Illumina's New v4 Reagent Kits for HiSeq 2500

Short Read Archive (SRA) growth





http://www.pharmaphorum.com/articles/data-interpretation-of-ngs-data-in-biomedical-context
http://www.omixon.com/bioinformatics-for-beginners-how-to-get-ngs-data-part-1-short-reads/

Illumina machines yield

HiSeq System Performance Parameters

Read Length	High-Output Run Mode*			Rapid-Run Mode*		
	Dual Flow Cell (HiSeq 2500 only)	Single Flow Cell (HiSeq 1500 or 2500)	Dual Flow Cell Run Time	Dual Flow Cell (HiSeq 2500 only)	Single Flow Cell (HiSeq 1500 or 2500)	Dual Flow Cell Run Time
1 × 36	95–105 Gb	47–52 Gb	2 days	18–22 Gb	9–11 Gb	7 hours
2 × 50	270–300 Gb	135–150 Gb	5.5 days	50–60 Gb	25–30 Gb	16 hours
2 × 100	540–600 Gb	270–300 Gb	11 days	100–120 Gb	50–60 Gb	27 hours
2 × 150	N/A	N/A	N/A	150–180 Gb	75–90 Gb	40 hours
Reads Passing Filter	Up to 3 billion single reads or 6 billion paired-end reads	Up to 1.5 billion single reads or 3 billion paired-end reads		Up to 600 million single reads or 1.2 billion paired-end reads	Up to 300 million single reads or 600 million paired-end reads	

HiSeq 2500 v4 kit

HiSeq System Performance Parameters

High-Output Run Mode			
Read Length	**Dual Flow Cell**	**Single Flow Cell**	**Dual Flow Cell Run Time**
1 × 36	128–144 Gb	64–72 Gb	29 hours
2 × 50	360–400 Gb	180–200 Gb	2.5 days
2 × 100	720–800 Gb	360–400 Gb	5 days
2 × 125*	900 Gb–1 Tb	450–500 Gb	6 days
Reads Passing Filter†	Up to 4 billion	Up to 2 billion	

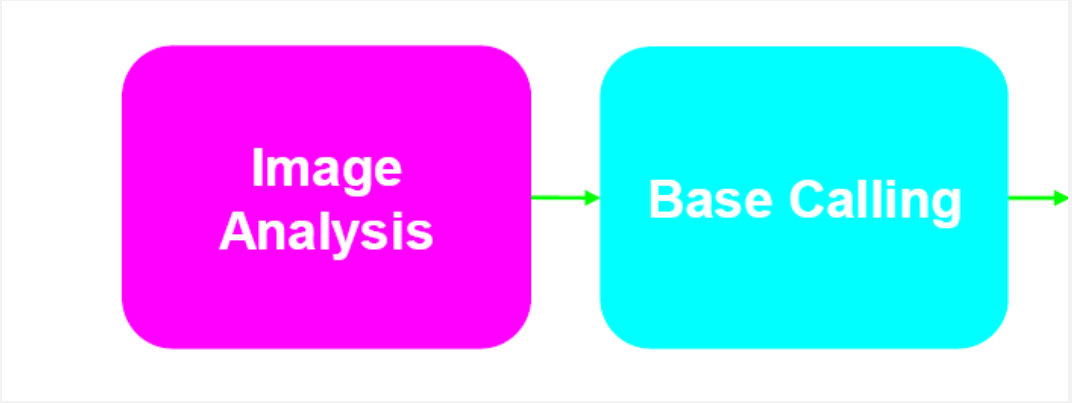
Rapid-Run Mode			
Read Length	**Dual Flow Cell**	**Single Flow Cell**	**Dual Flow Cell Run Time**
1 × 36	18–22 Gb	9–11 Gb	7 hours
2 × 50	50–60 Gb	25–30 Gb	16 hours
2 × 100	100–120 Gb	50–60 Gb	27 hours
2 × 150	150–180 Gb	75–90 Gb	40 hours
2 x 250*	250-300 Gb	125-150 Gb	60 hours
Reads Passing Filter†	Up to 600 million	Up to 300 million	

How can we deal with so much data?

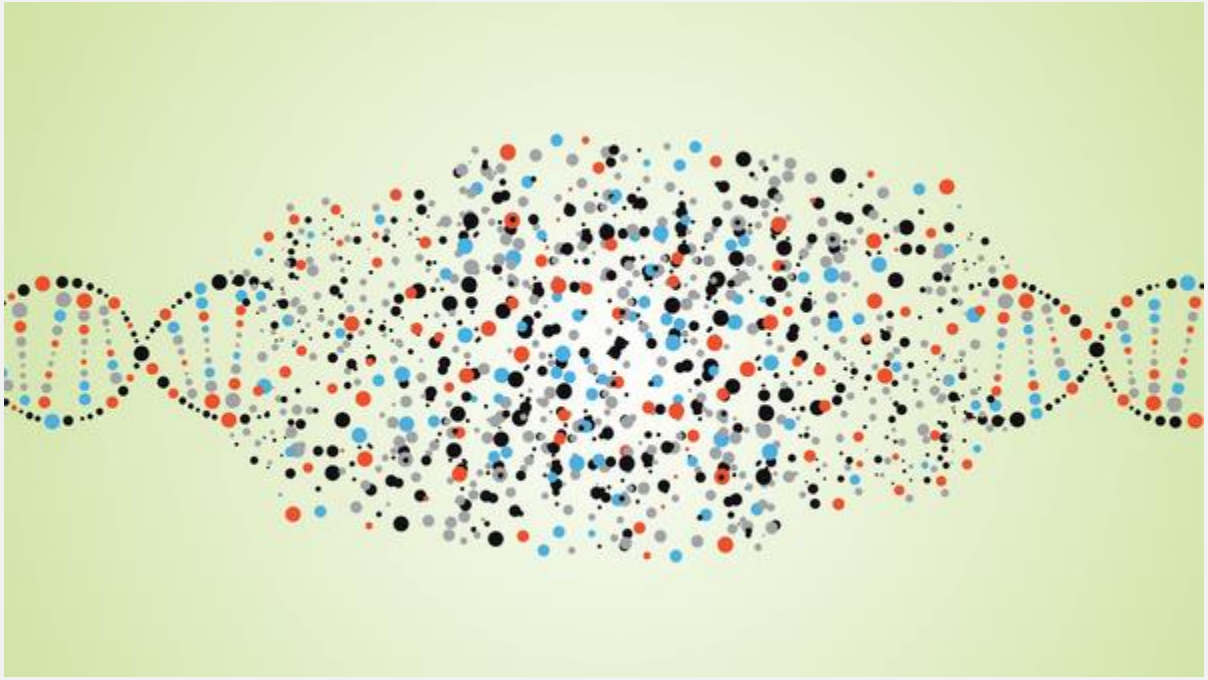
or the story of the Donkey and the Cart

DATA

Illumina Primary Analysis



Next-generation sequencing: The genome jigsaw



To understand why high-throughput gene-sequencing technology often produces frustrating results, says Titus Brown, imagine that 1,000 copies of Charles Dickens' novel A Tale of Two Cities have been shredded in a woodchipper. “Your job is to put them back together into a single book,” he says.

http://www.nature.com/nature/journal/v501/n7466/full/501261a.html

Next-generation sequencing: The genome jigsaw (2)

In one lane we get 150 million sequences, each of length of 50 to 200 bases … and some do not fit correctly.

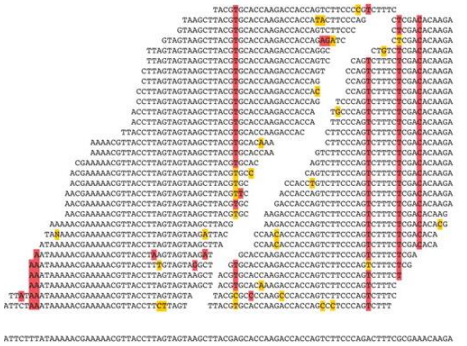
First Question:

Do we have a reference genome?

First step in secondary analysis

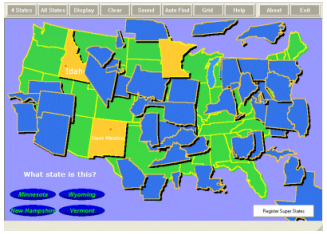
READS

Sequencing

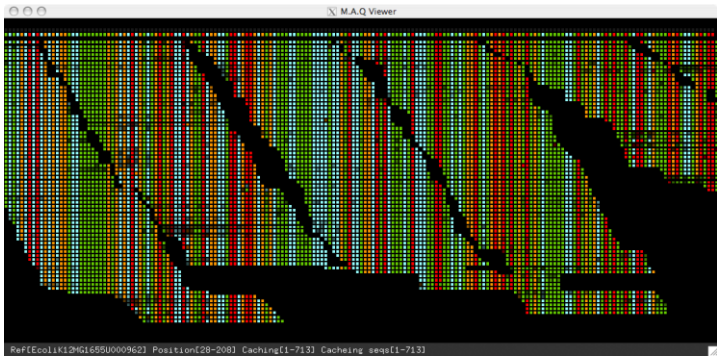


Assembly

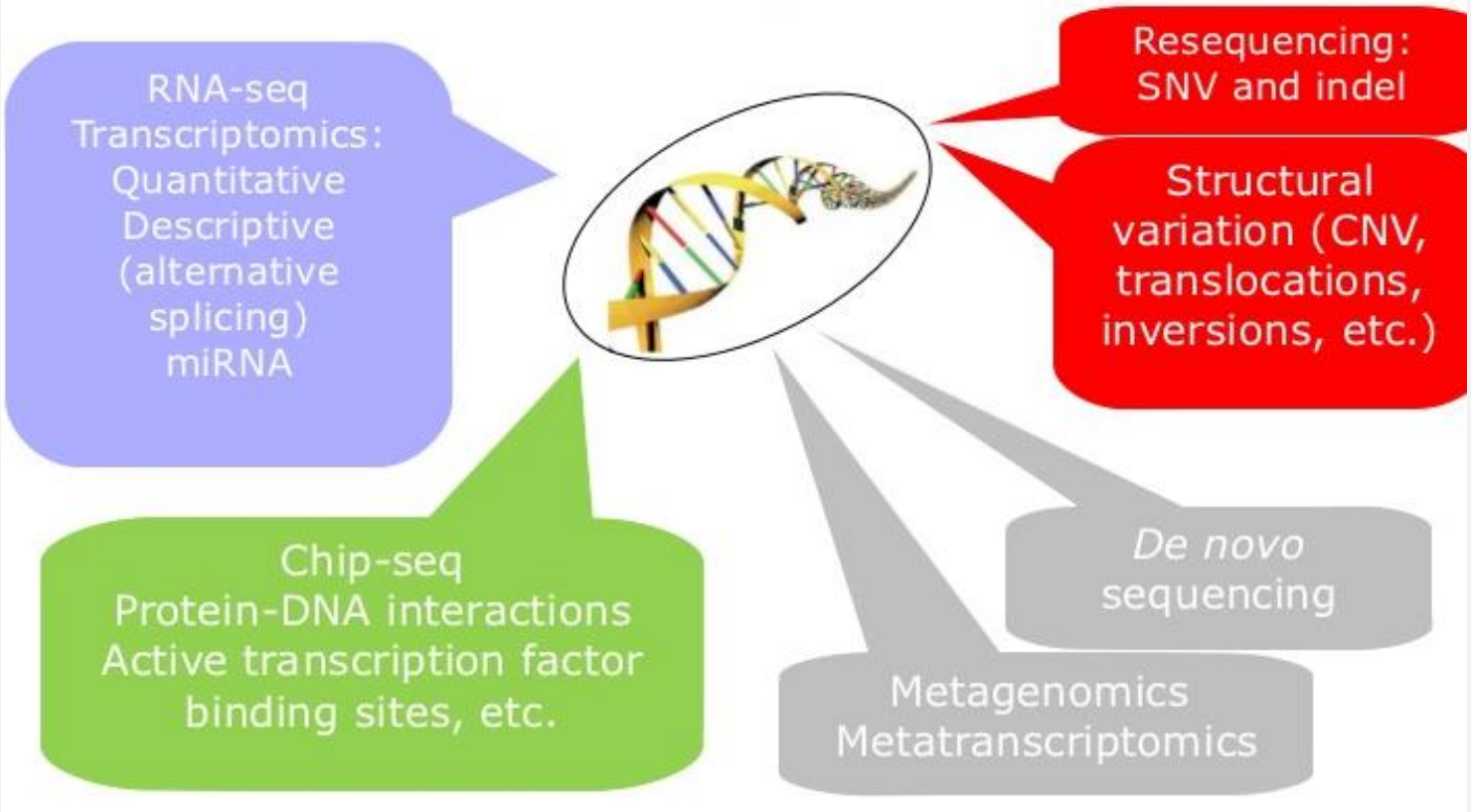
Mapping
Alignment



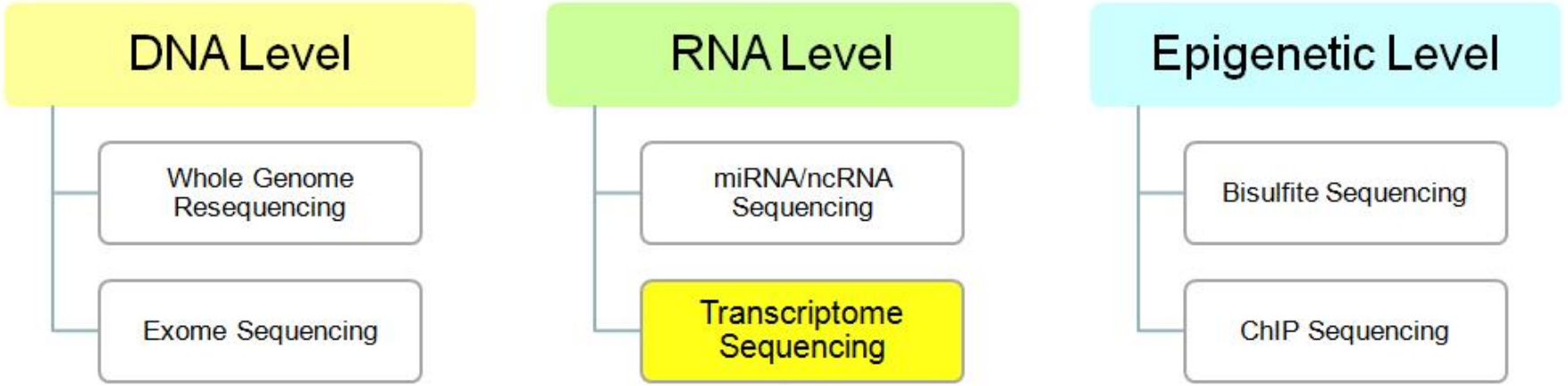
Viewing



Popular NGS applications

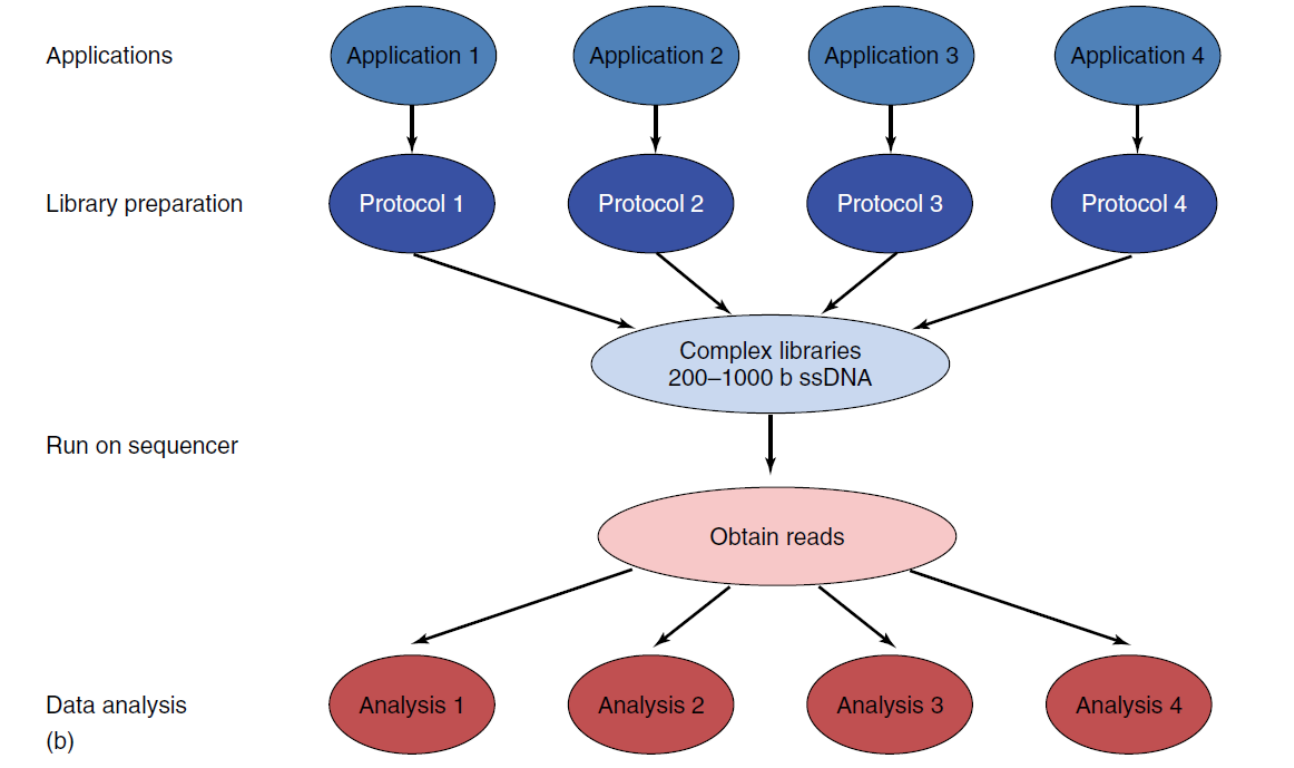


Levels of NGS applications



http://www.bio-miblab.org/research_omic_ngs.php

Protocols and applications



Each application is processed in a different way but all protocols result in short fragments ready to be sequenced. Once reads are obtained, downstream computational analyses are once again application specific.

Horn-Saban, S. and Amann-Zalcenstein, D. 2011. Frontiers in DNA Sequencing: the (R)Evolution of Sequencing Technologies. Encyclopedia of Analytical Chemistry.

Applications: RNA-seq

Process

10 - 100 million reads

↓

Align to reference data

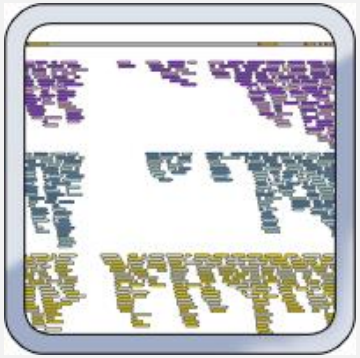
↓

Parse files, reformat data, create reports

↓

Review results, make decisions

mRNA

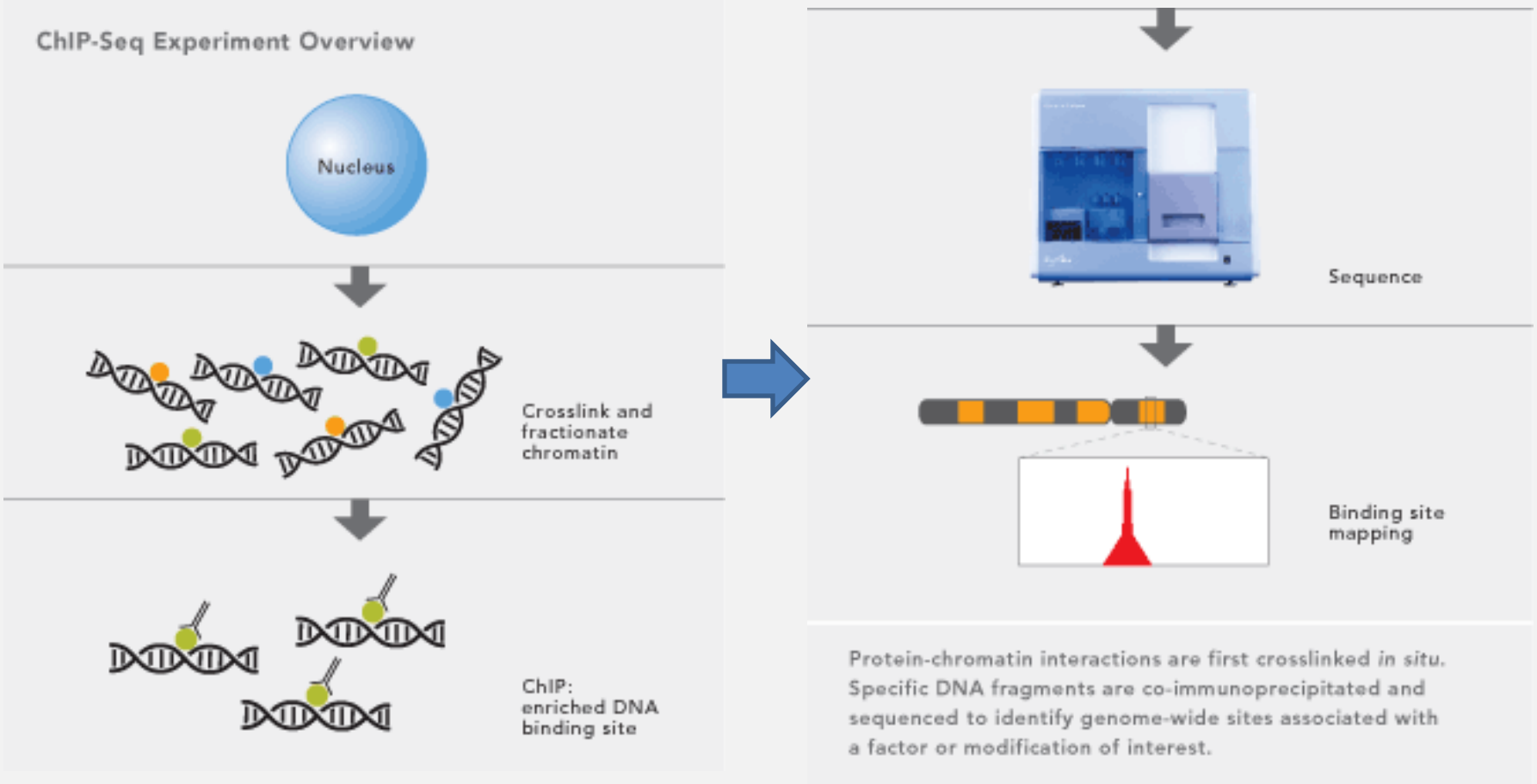


Small RNA

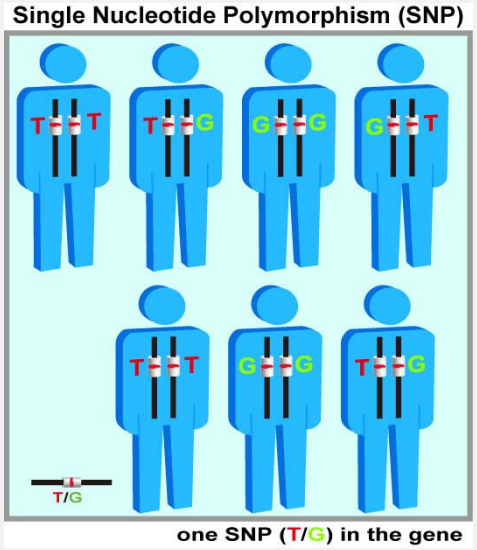
http://www.geospiza.com/finchtalk/uploaded_images/Slide9-733003.png

Applications: DNA (ChIP-Seq)

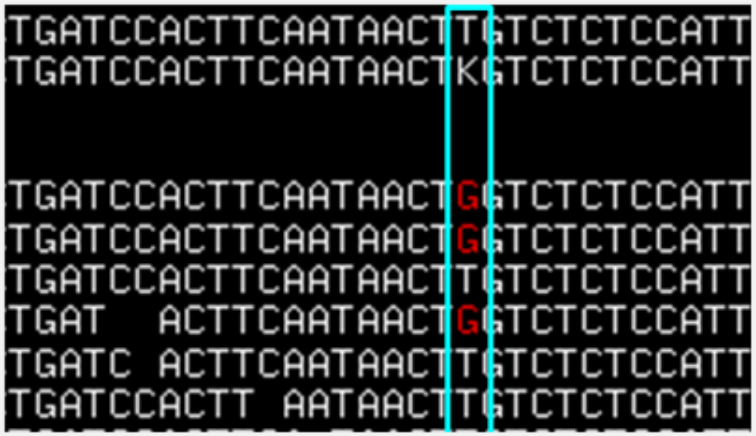
Chromatin ImmunoPrecipitation - Seq:
What DNA sequences are in interaction with a specific protein?

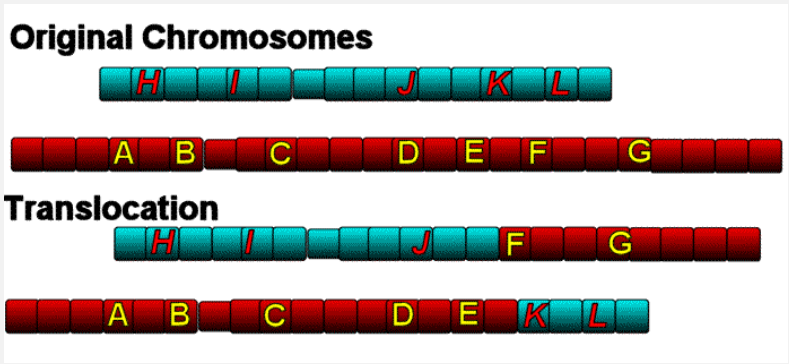


Applications: DNA Variations

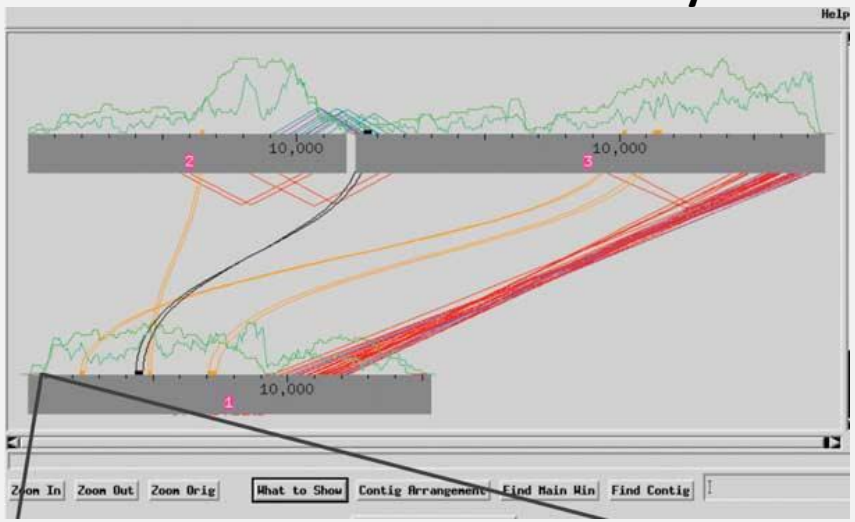


SNP Discovery





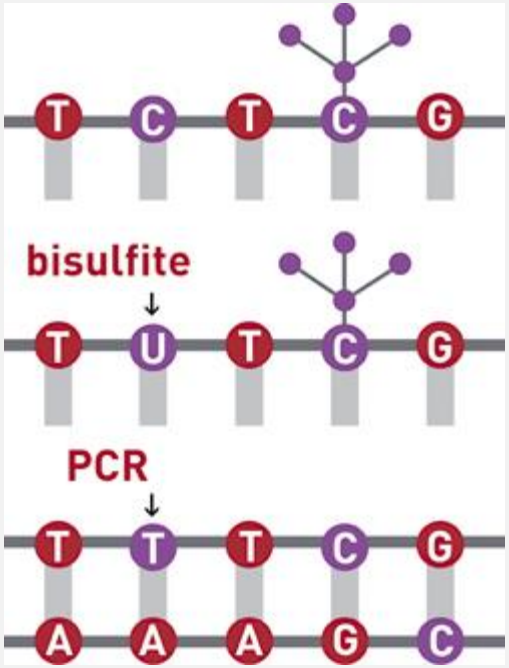
Structural Variant Discovery



Applications: Epigenomics

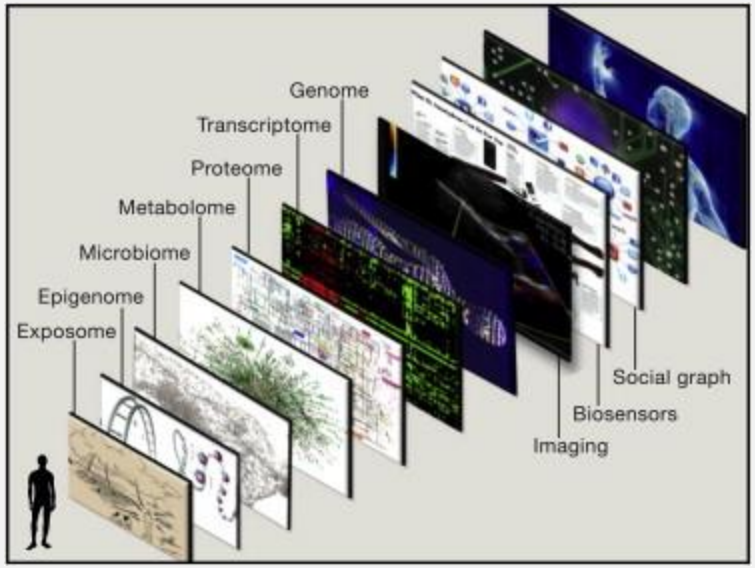
Epigenomics involves the study of stable and long-term changes in the regulation of gene activity and expression that are not dependent on gene.

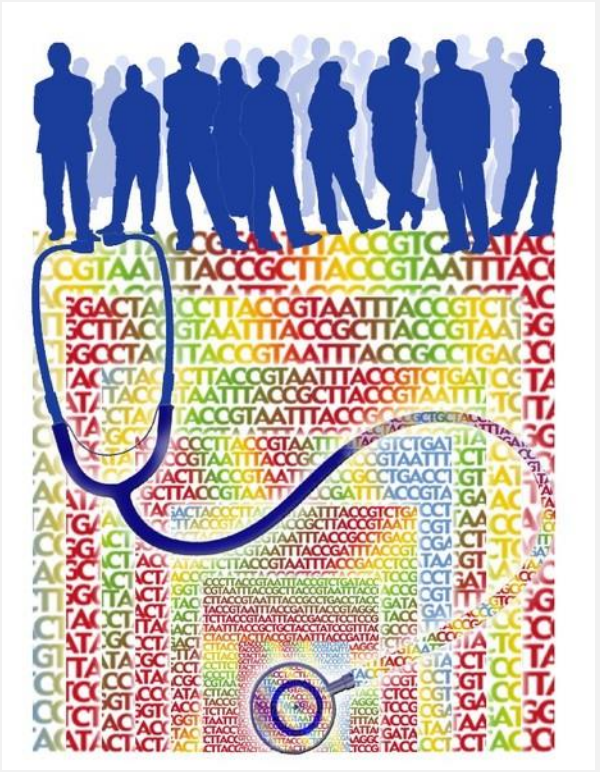
Methylation: Bisulfite conversion



Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected

Personalized medicine using NGS

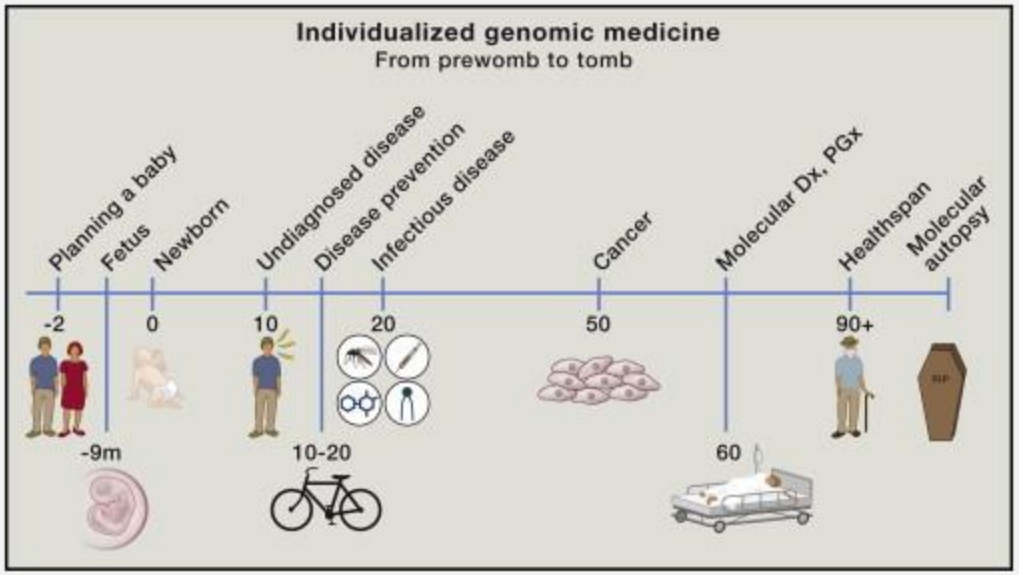




- Diagnosis, prevention (predisposition knowledge) and treatment
- Hereditary rare monogenic diseases
- Metabolic diseases (Obesity, Diabetis, …)
- Infectious diseases (Pathogen tracking)
- Cancer
- Pharmacogenomics (Drug prescription)

http://www.sciencedirect.com/science/article/pii/S0092867414002049#

Individualized Medicine from Prewomb to Tomb



The medical application of genomics is relevant to many points during an individual's lifespan.

- Prior to conception, a couple can have genomic screening for important recessive alleles.
- An expectant mother can have single tube of blood used to accurately assess chromosomal abnormalities of the fetus.

- At birth, sequencing the genome of the newborn can be used to rapidly diagnosis many critical conditions.
- The molecular basis for serious, undiagnosed conditions can often be established by sequencing the individual with parents of siblings.
- Omic information at a young age will be useful by providing susceptibility to various medical conditions that have actionable prevention strategies.

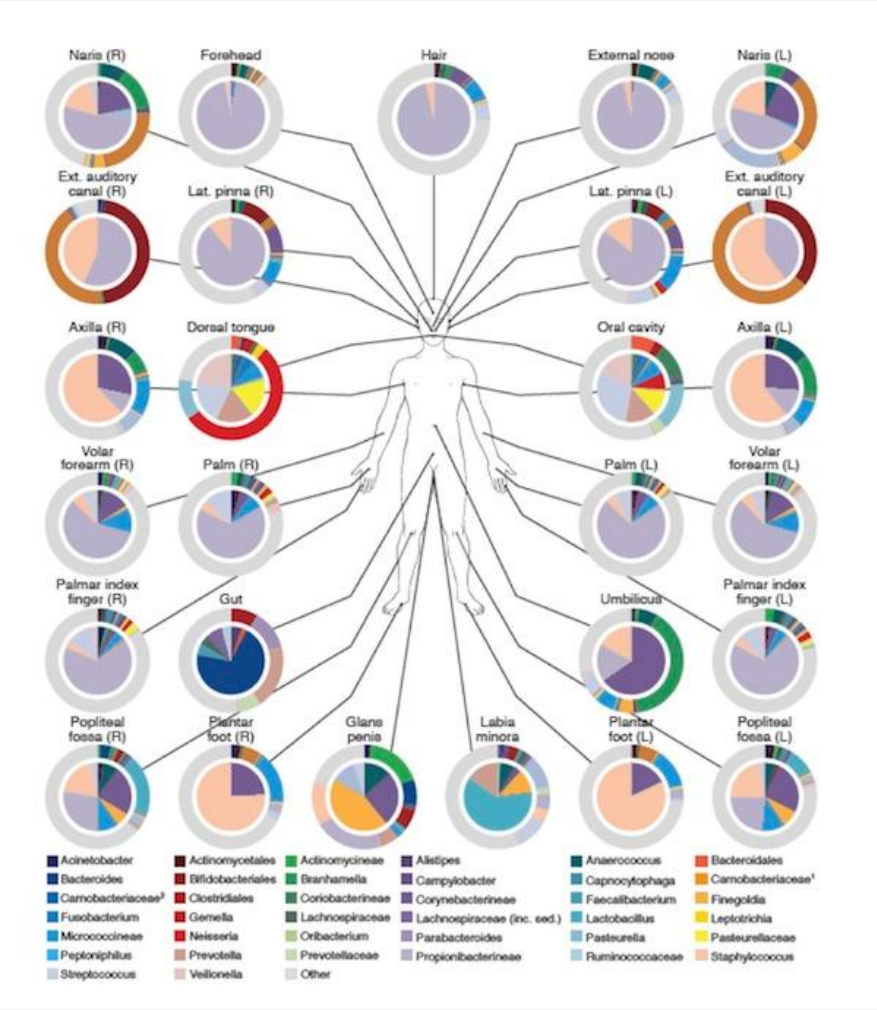
http://www.sciencedirect.com/science/article/pii/S0092867414002049#

Individualized Medicine from Prewomb to Tomb

- Sequencing can be done to define a pathogen for more rapid and accurate approaches to infectious diseases.
- The driver mutations and key biologic underpinning pathways of an individual's cancer can frequently be pinpointed by omics.
- The root causes of common polygenic conditions such as diabetes or coronary heart disease may ultimately be defined at the individual level.
- Specific sequence variants of germline DNA or the gut microbiome have relevance for response to prescription medication (both efficacy and safety). Defining the genomics of healthspan, rather than the traditional focus on diseases, may prove to be especially worthwhile to understand protective alleles and modifier genes.
- For an individual with sudden death, a molecular autopsy via sequencing can be performed, along with family survivors, to determine the cause of death and potentially prevent untimely or avoidable deaths of members of the family and subsequent generations.

http://www.sciencedirect.com/science/article/pii/S0092867414002049#

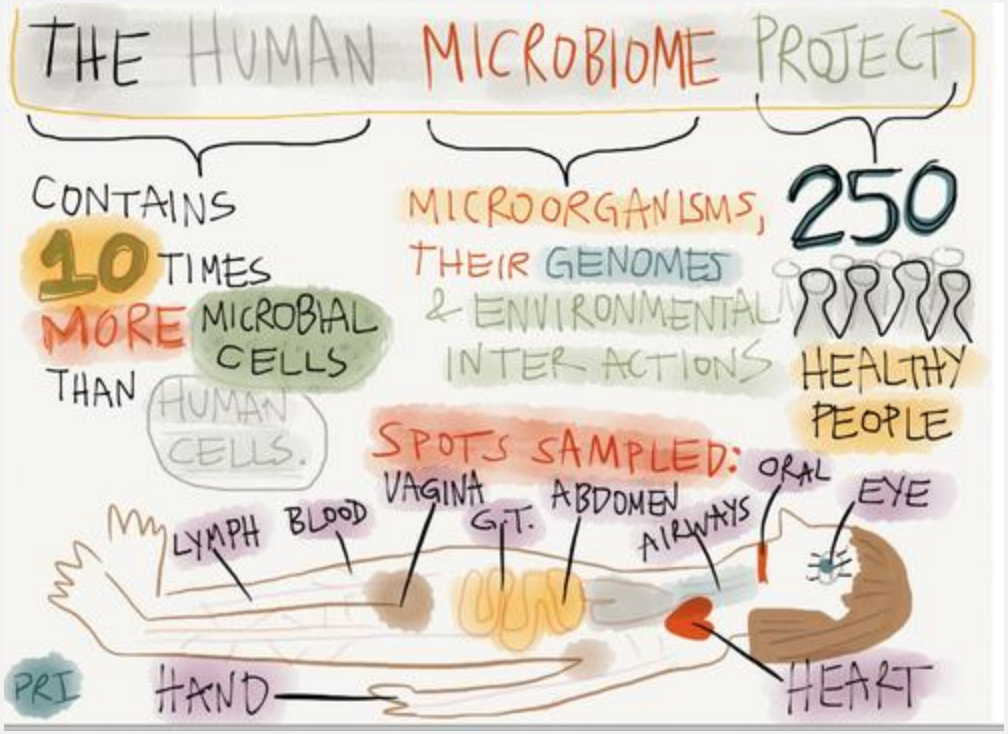
The human microbiome

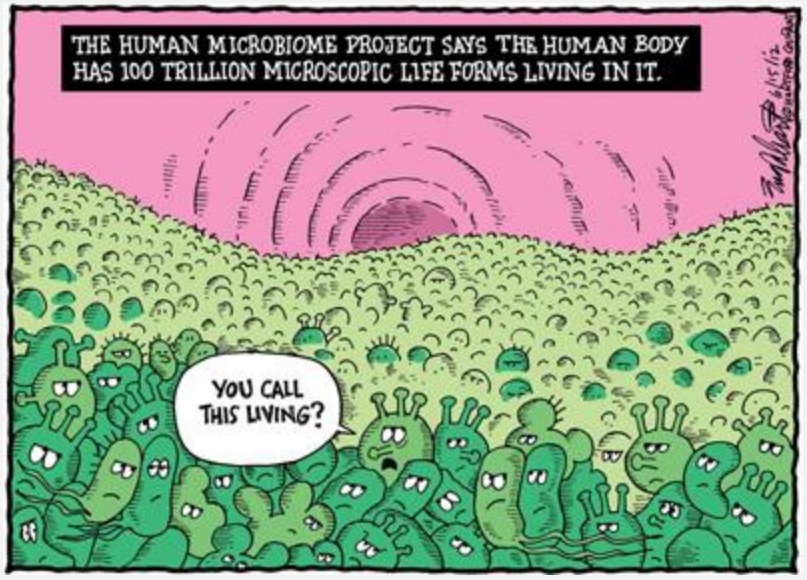


A microbiome is "the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space."
[Wikipedia]

http://www.sciencemag.org/content/326/5960/1694.abstract

The human microbiome

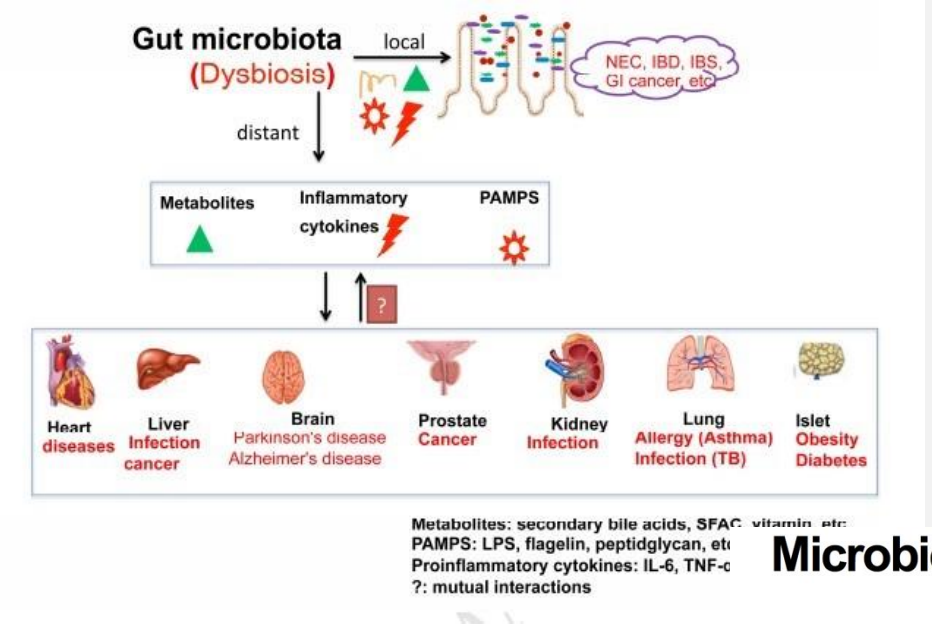




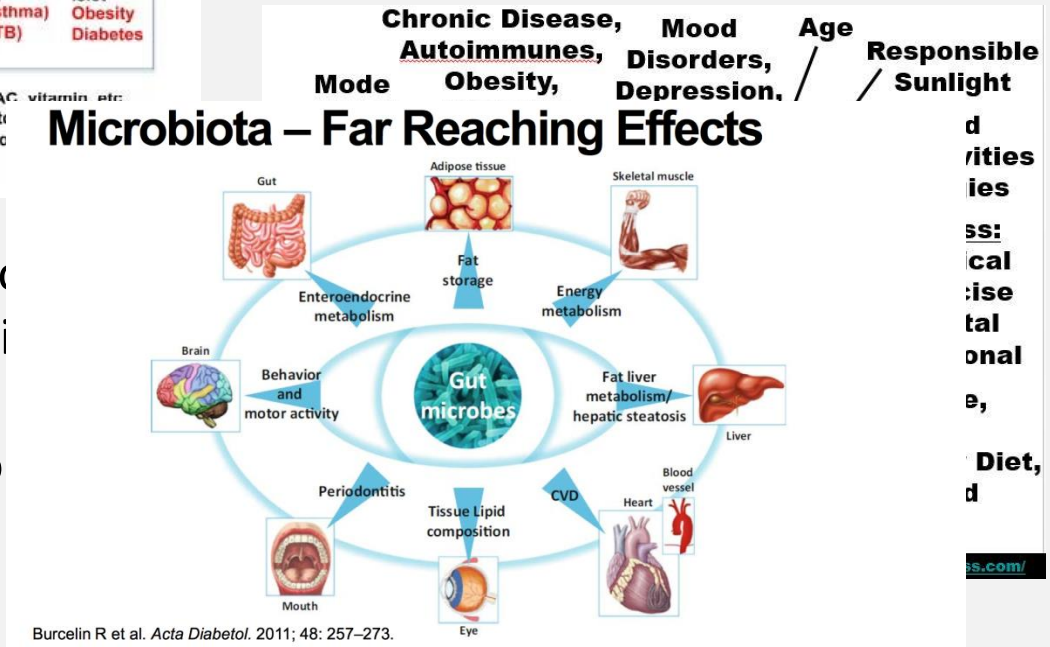
http://www.healthcommentary.org/

http://medicine.emory.edu/research/interest-groups-research-teams/microbiome/index.html

Microbiome: A popular subject



- Inflammatory diseases
- Musculoskeletal conditions
- Cardio-metabolic and chroni
- Cancer
- Depression and anxiety diso



The ISME Journal
Multidisciplinary Journal of Microbial Ecology

Short Communication

Subject Category: Microbe-microbe and microbe-host interactions

The ISME Journal (2013) **7,** 880–884; doi:10.1038/ismej.2012.153; published online 13 December 2012

An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice

Open

Na Fei[1] and Liping Zhao[1,2]

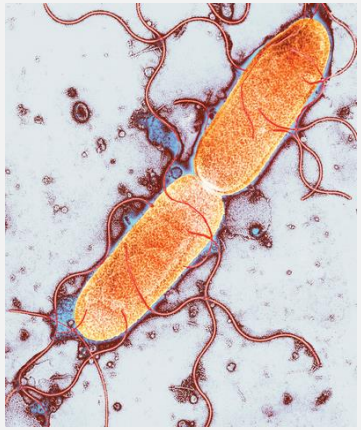


Table 1. Changes of endotoxin load, inflammation indicators, metabolic phenotypes and the gut microbiota during weight loss of a morbidly obese volunteer

Figures and tables index

Measurements	*Day 0*	*9 Weeks*	*23 Weeks*	*Reference range*
Body weight (kg)	174.8	144.8	123.5	—
BMI (kgm−2)	58.78	48.66	41.50	18–23

The endotoxin-producing *Enterobacter* decreased in relative abundance from 35% of the volunteer's gut bacteria to non-detectable, during which time the volunteer lost 51.4 kg of 174.8 kg initial weight and recovered from hyperglycemia and hypertension after 23 weeks on a diet of whole grains, traditional Chinese medicinal foods and prebiotics.

http://www.nature.com/nature/journal/v508/n7496_supp/full/508S61a.html

Microbiome: A complicated relationship status (2)

Research Article

A gut microbiota-targeted dietary intervention for amelioration of chronic inflammation underlying metabolic syndrome

OPEN

Shuiming Xiao[1], Na Fei[1], Xiaoyan Pang[1], Jian Shen[2], Linghua Wang[1], Baorang Zhang[1], Menghui Zhang[1], Xiaojun Zhang[1], Chenhong Zhang[1], Min Li[1], Lifeng Sun[1], Zhengsheng Xue[1], Jingjing Wang[1], Jie Feng[1], Feiyan Yan[1], Naisi Zhao[1], Jiaqi Liu[1], Wenmin Long[1] and Liping Zhao[1,2,*]

Article first published online: 21 OCT 2013

DOI: 10.1111/1574-6941.12228

Issue

FEMS Microbiology Ecology

Volume 87, Issue 2, pages 357–367, February 2014

After 9 weeks, the nearly 100 study participants had

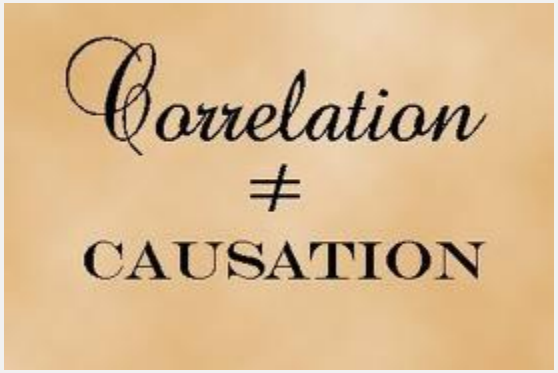
- improved markers of metabolic health and
- lower levels of potentially harmful bacteria, including *Enterobacter*,
- but they only achieved a modest weight loss of about 6 kg on average

Obesity and the gut microbiome: Striving for causality

System	Phenotype	Microbes implicated in obese state	Proposed mechanism	Caveats
mm	Increased adiposity	Whole microbiome	↓Angptl4 and ↑energy storage	Mixed strain background of knockout mice
mm	Increased weight gain	Whole microbiome	↓AMPK activity and ↓energy expenditure	
mm	Microbiota associated with obesity—Lepob/ob	↑Firmicutes: Bacteroidetes		Leptin's newly recognized role in gut immunity
hs	Microbiota associated with obesity/weight loss	↑Firmicutes: Bacteroidetes	Increased energy harvest	small sample size (n<15/group)
mm	Increased adiposity	↑Firmicutes: Bacteroidetes	Increased energy harvest	Leptin's newly recognized role in gut immunity
mm	No difference in obesity	↑Firmicutes: Bacteroidetes		Mouse strain choice
mm	Microbiota changes related to diet, not obesity	↑Firmicutes: Bacteroidetes		Use of purified diets versus natural ingredient diet
mm	Increased obesity	↑Firmicutes: Bacteroidetes		
mm	Microbiota associated with obesity	↑Firmicutes: Bacteroidetes	SCFA production and fecal energy uncoupled from obesity	
hs	Microbiota associated with obesity, obesity	↑Bacteroidetes	↑SCFA (propionate) production	
hs	Microbiota associated with obesity	Reduced diversity ↓Bacteroidetes, ↑Actinobacteria		
hs	Microbiota associated weight loss/diet	No difference in Bacteroidetes, ↑Firmicutes		Relatively small sample size (n<30/group)
mm	Increased obesity	Whole microbiome	Immune-mediated dysbiosis	
mm	Increased obesity—Leprdb/db, liver damage	Whole microbiome	Immune-mediated dysbiosis	
mm	Increased epidymal fat pad weight	Bacteroides thetaiotamicron + Methanobrevibacter smithii	↑SCFA production	
mm	Increased adiposity/weight gain	Bacteroides thetaiotamicron+Methanobrevibacter smithii	GPR41-mediated SCFA sensing	Mixed strain background of knockout mice
mm	Increased insulin resistance	Whole microbiome	Immune-mediated dysbiosis	
dm	Increased insulin resistance	Acetobacter pomorum		
mm	Increased insulin resistance	Whole microbiome	Metabolic endotoxemia	
mm	Increased insulin resistance	↓Firmicutes: Bacteroidetes	Metabolic endotoxemia	
hs	Microbiota changes after gastric bypass	**↑Firmicutes, ↓Gammaproteobacteria**		small sample size (n<15/group)
rn	Microbiota changes after gastric bypass	**↑Firmicutes and Bacteroidetes, ↓Proteobacteria**		
hs	Microbiota associated with overweight/weight gain	↑Bacteroides, Staphylococcus aureus		Relatively small sample size (n<30/group)
hs	Microbiota associated with overweight—children	↑Staphylococcus aureus, ↓Bifidobacteria		Relatively small sample size (n<30/group)
hs	Microbiota changes associated with weight loss—adolescents	↓Bacteroides, Lactobacilli		Relatively small sample size (n<30/group)
hs	Microbiota associated with weight loss—adolescents	↑Clostridium histolyticum, Eubacterium rectale–Clostridium coccoides, ↓Bacteroides–Prevotella		Relatively small sample size (n<30/group)
hs	Hepatic Steatosis Severity	Whole microbiome	Small intestinal bacterial overgrowth	
hs/mm	Increased adiposity	↑Erysipelotrichi, ↑Bacilli, ↓Bacteroidetes		
hs	Normalized BMI	Correlation with Functional Modules consisting of ATPase complex and ectosine biosynthesis	Increased energy harvest	Relatively small sample size (n<30/group)

Harley and Karp, http://www.sciencedirect.com/science/article/pii/S2212877812000051#

Microbiome: cause?



- Diet is a major factor in obesity
- Diet also shapes the microbiota.
- Changes in the levels of gut microbes produced by healthy or unhealthy diets are broadly similar to the differences seen in lean versus obese individuals.

“So you already have this diet-to-microbiota relationship that's difficult to disentangle from the microbiota-to-obesity relationship,” says Eric Martens, a microbiologist at the University of Michigan.

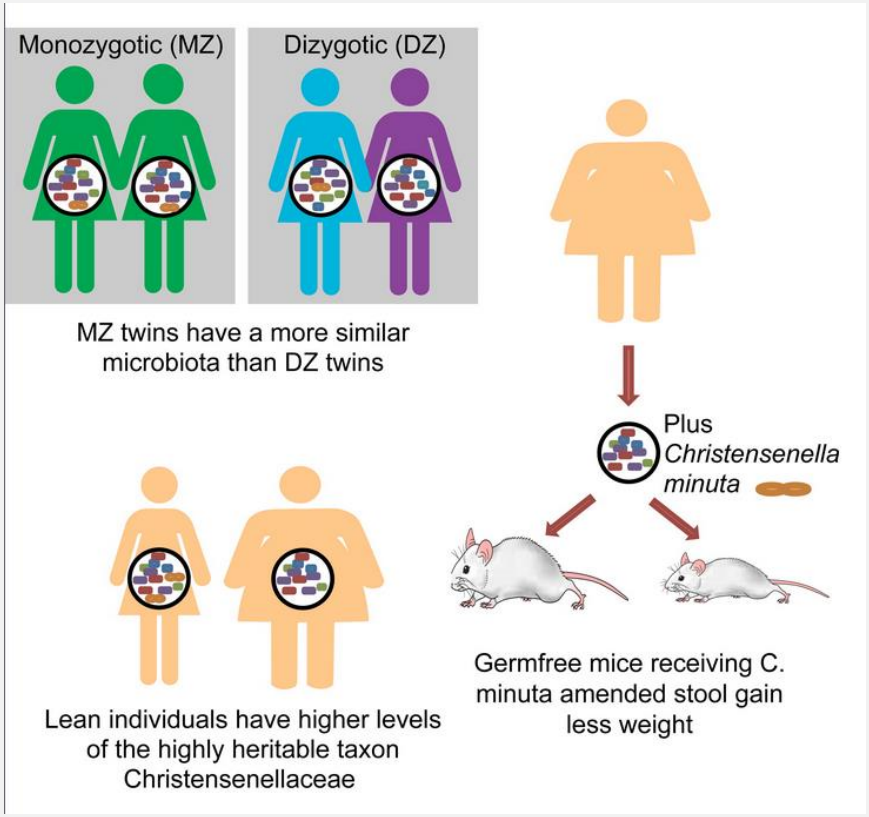
Cell

Volume 159, Issue 4, 6 November 2014, Pages 789–799

Article

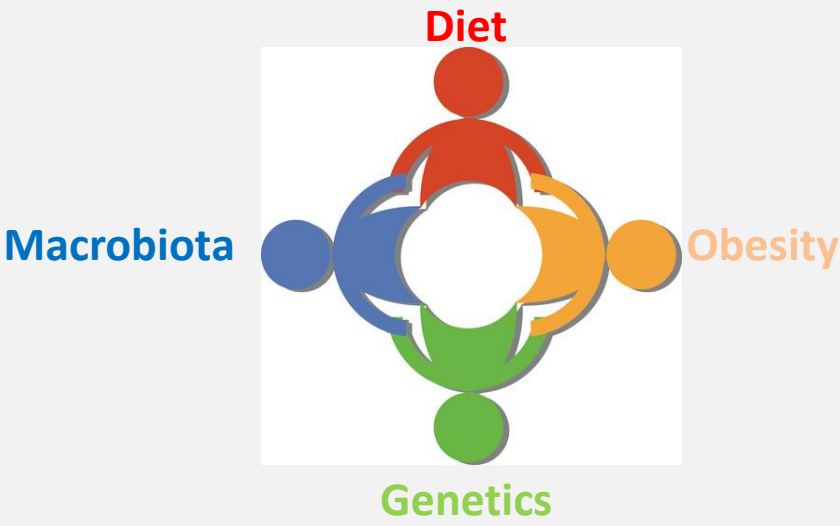
Human Genetics Shape the Gut Microbiome

Julia K. Goodrich[1, 2], Jillian L. Waters[1, 2], Angela C. Poole[1, 2], Jessica L. Sutter[1, 2], Omry Koren[1, 2, 7], Ran Blekhman[1, 8], Michelle Beaumont[3], William Van Treuren[4], Rob Knight[4, 5, 6], Jordana T. Bell[3], Timothy D. Spector[3], Andrew G. Clark[1], Ruth E. Ley[1, 2]



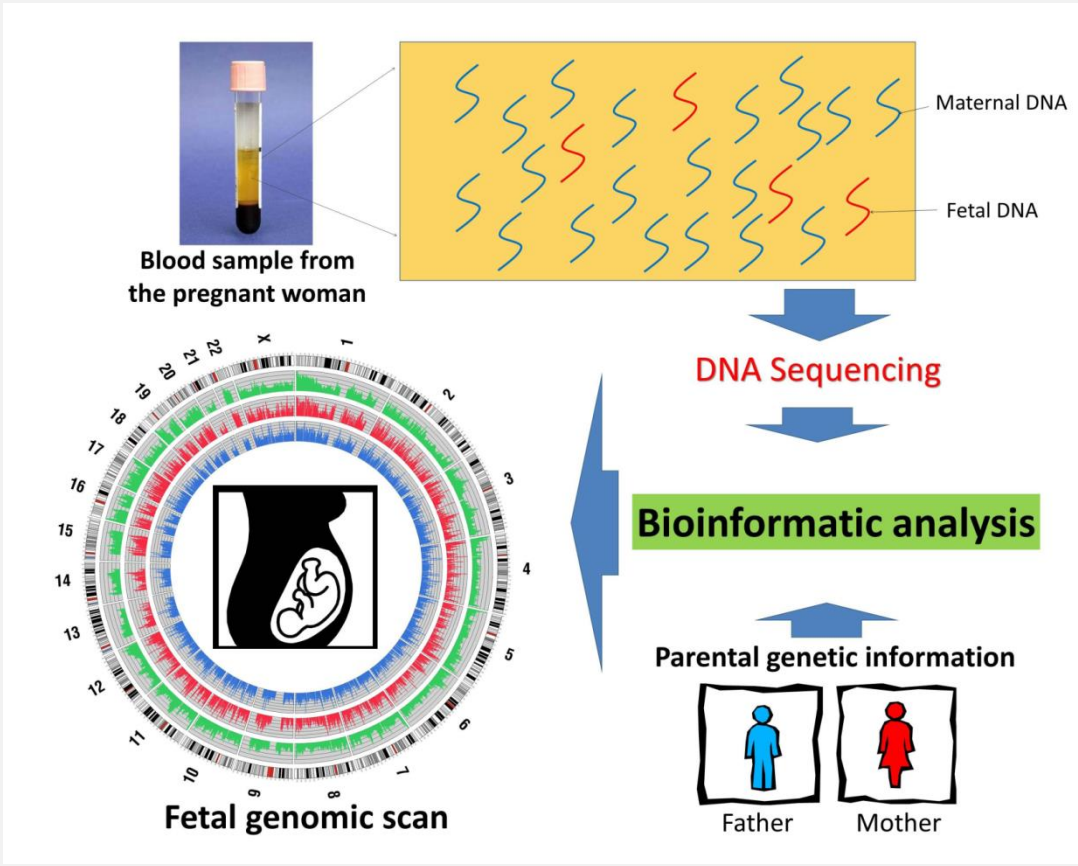
Highlights

- Host genetics influence abundances of health-associated gut bacteria
- Many heritable taxa co-occur
- The most heritable, Christensenellaceae, associates with a lean BMI
- Heritable taxa reduce weight gains in germ-free transplant experiments



Noninvasive Prenatal Diagnosis

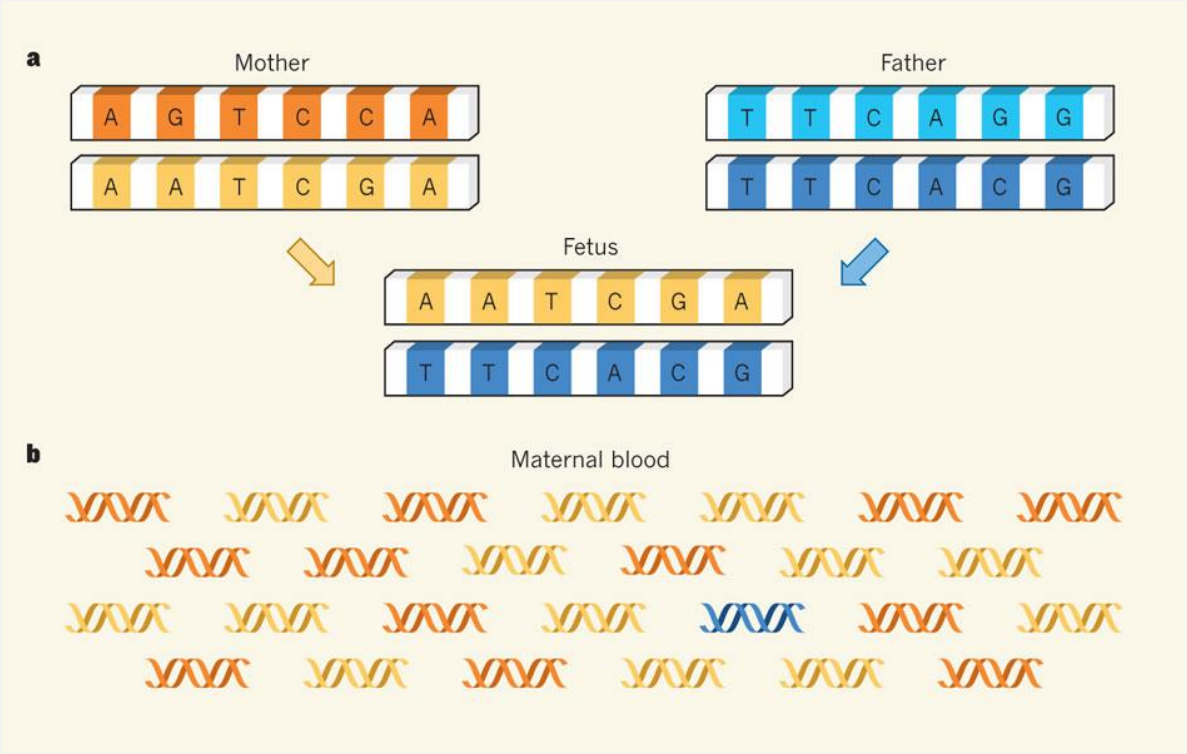
During pregnancy, a median of 10% of the DNA in the plasma of pregnant women is fetally derived.



http://www.smartplanet.com/blog/rethinking-healthcare/babys-genome-found-in-moms-blood/

Noninvasive Prenatal Diagnosis

During pregnancy, a median of 10% of the DNA in the plasma of pregnant women is fetally derived.



Noninvasive Prenatal Diagnosis

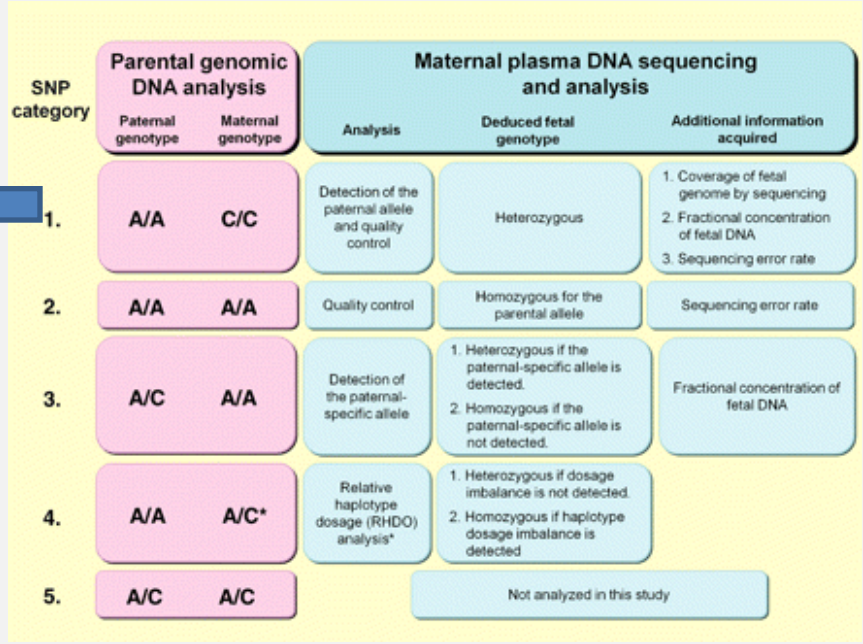
Diagnosis of β-thalassemia in a fetus by Lo et al in 2010

- β-thalassemia is an autosomal recessive disease that causes severe anemia
- Mutations in the HBB gene (subunit of hemoglobin) cause the disease
- To inherit the disease, the fetus must carry mutations from both parents
- The father was a carrier of the –CTTT, 4–base pair (bp) deletion of codons 41/42
- The pregnant mother was a carrier of the A→G mutation at nucleotide −28 of the HBB gene
- The pregnant mother DNA was sequenced to a coverage of x65

Sci Transl Med 2, 61ra91 (2010), http://stm.sciencemag.org/content/scitransmed/2/61/61ra91.full

Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus

- SNP calling was performed for both parents using a SNP array

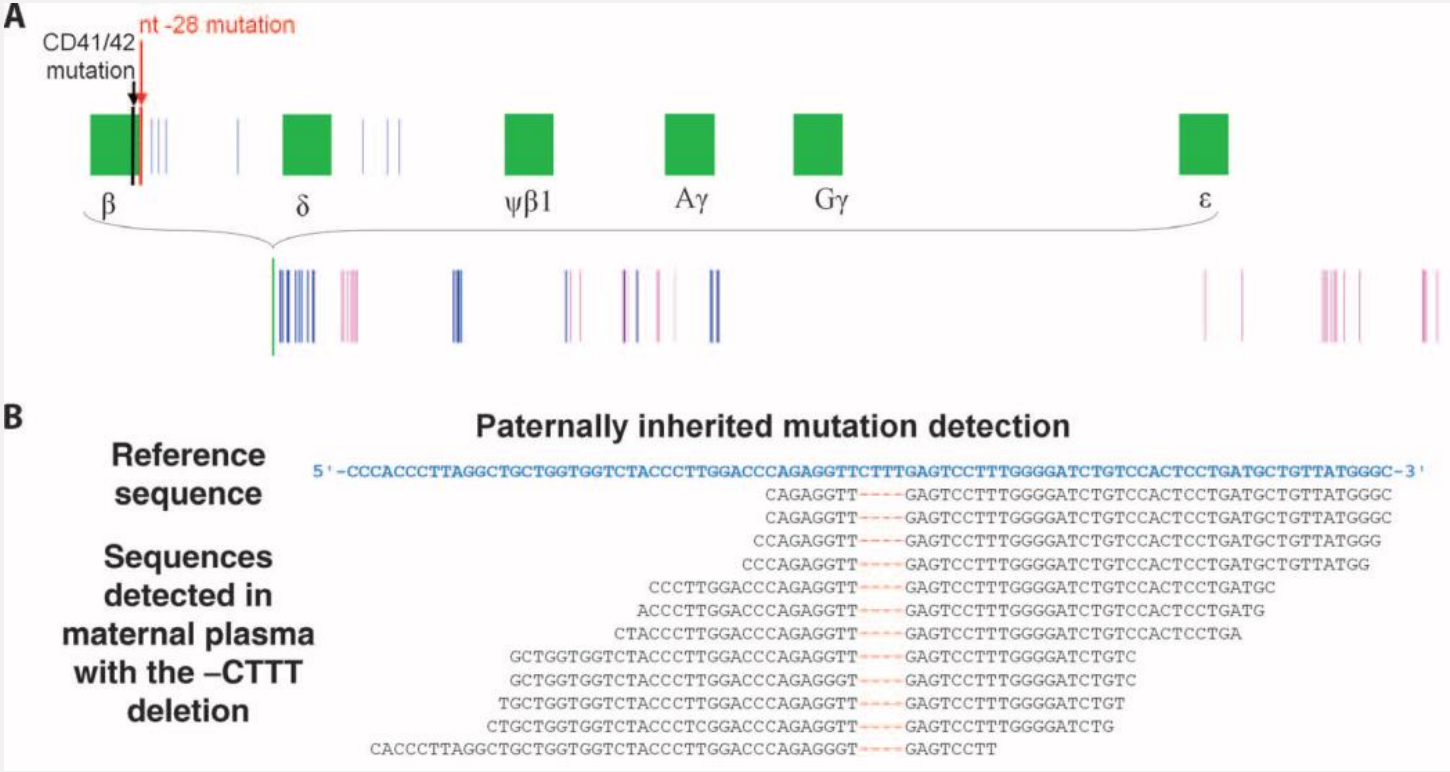


SNP category	Parental genomic DNA analysis: Paternal genotype	Parental genomic DNA analysis: Maternal genotype	Maternal plasma DNA sequencing and analysis: Analysis	Maternal plasma DNA sequencing and analysis: Deduced fetal genotype	Maternal plasma DNA sequencing and analysis: Additional information acquired
1.	A/A	C/C	Detection of the paternal allele and quality control	Heterozygous	1. Coverage of fetal genome by sequencing 2. Fractional concentration of fetal DNA 3. Sequencing error rate
2.	A/A	A/A	Quality control	Homozygous for the parental allele	Sequencing error rate
3.	A/C	A/A	Detection of the paternal-specific allele	1. Heterozygous if the paternal-specific allele is detected. 2. Homozygous if the paternal-specific allele is not detected.	Fractional concentration of fetal DNA
4.	A/A	A/C*	Relative haplotype dosage (RHDO) analysis*	1. Heterozygous if dosage imbalance is not detected. 2. Homozygous if haplotype dosage imbalance is detected	
5.	A/C	A/C	Not analyzed in this study		

- Maternal and fetal genomes were found at constant proportions
- Knowing the expected proportion it is possible to calculate the expected proportion for each case of inheritance

Sci Transl Med 2, 61ra91 (2010)

Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus



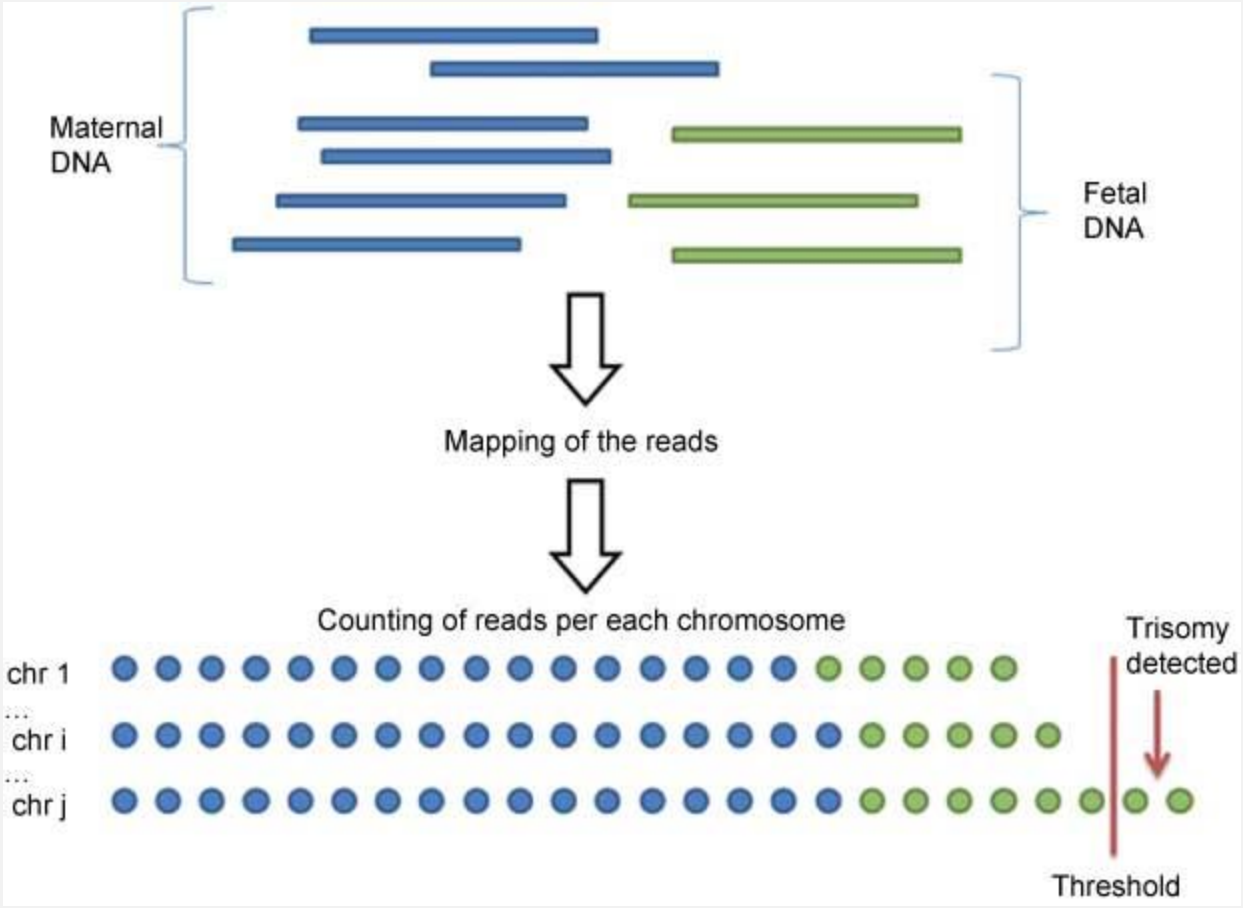
The fetus was found to carry only the paternal mutation

Sci Transl Med 2, 61ra91 (2010)

Screening for trisomy

phg foundation making science work for health

Non-invasive prenatal diagnosis using fetal DNA in maternal plasma: From dream to reality



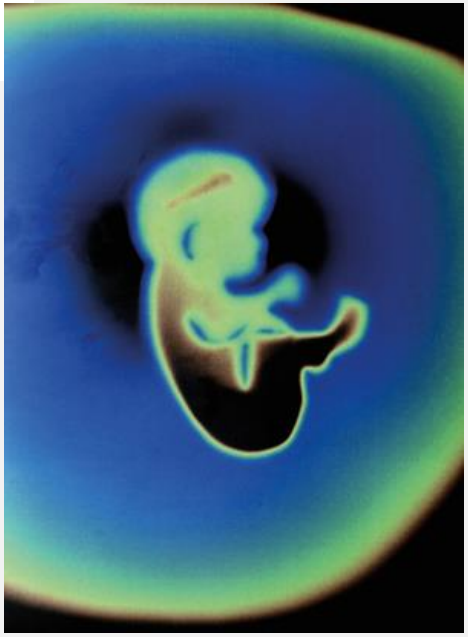
NATURE REVIEWS GENETICS | RESEARCH HIGHLIGHT

GENETIC TESTING

cfDNA screening for trisomy 21 tested in unselected pregnancies

Orli G. Bahcall

Nature Reviews Genetics **16**, 316–317 (2015) | doi:10.1038/nrg3953



A new prospective study provides further evaluation of the potential use of cell-free DNA (cfDNA) screening for trisomy 21 (Down syndrome) as a primary screening method among unselected pregnancies in the general population.

- Low false discovery rate relative to standard screening
- Positive predictive value of 80.9%
- False positive rate for cfDNA screening was considered low at 0.06%, in comparison to 5.4% for standard screening.
- Limitation: the 'no-call' rate, as no cfDNA result was returned for ~3% of women

Bianchi, D. W. DNA Sequencing versus standard prenatal aneuploidy screening. *N. Engl. J. Med.* **370**, 799–808 (2014)
Norton, M. E. Cell-free DNA analysis for noninvasive examination of trisomy. *N. Engl. J. Med.* **372**, 1589–1597 (2015)

We can deal with so much data!

or the story of the Donkey and the Cart

Thank you

for your attention !