

Central European Institute of Technology BRNO | CZECH REPUBLIC

Bi7420: Moderní metody pro analýzu genomu

NGS data analysis introduction

Vojta Bystry vojtech.bystry@ceitec.muni.cz

SISNANDBRUD

AS

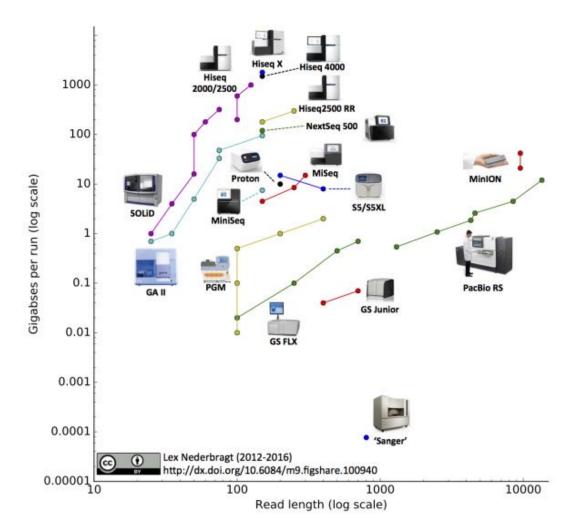
MANA

Plan for Bi7420

- Next generation sequencing methods overview
 - Focus on experiment planning and result interpretation
- 1. Introduction to NGS technology
- 2. miRNA, IncRNA in cancer Marek Mráz
- 3. Basic QC, DNA resequencing
- 4. DNA resequencing, Chip-seq (CLIP-seq)
- 5. Chip-seq (CLIP-seq)
- 6. RNA-seq
- 7. Single-cell RNA-seq, Spatial transcriptomics



- Next generation sequencing
 - New generation sequencing
 - HTP = High throughput
 - Massively parallel sequencing
- Contrast to Sanger sequencing

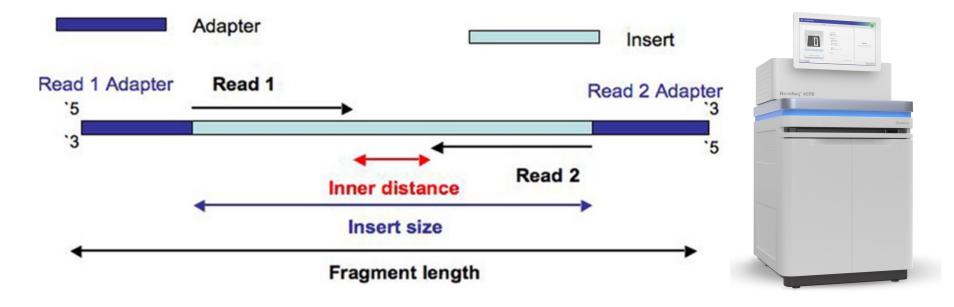




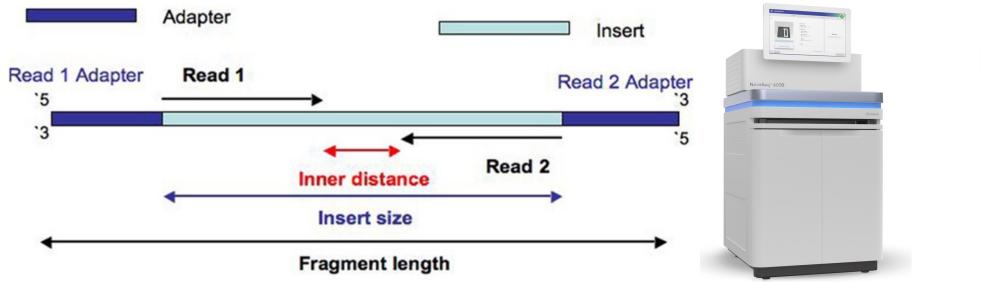
- Illumina sequencing by synthesis short-read sequencing
- Oxford Nanopore Nanopore sequencing
- Pacific Bioscience Single Molecule, Real-Time (SMRT)



- Illumina sequencing by synthesis short-read sequencing
- Oxford Nanopore Nanopore sequencing
- Pacific Bioscience Single Molecule, Real-Time (SMRT)



- Illumina sequencing by synthesis short-read sequencing
- Oxford Nanopore Nanopore sequencing
- Pacific Bioscience Single Molecule, Real-Time (SMRT)





Short-read sequencing result

>read_no_1 CGGCCTGGAGGCCCTGCAGAACCTGCTGGGCTACAGGTTCGGCGACGAGGG

>read_no_2 GCAGCGTGAGCGCCATCATGGGCAACCCCCAGGTGAAGGCCCACGGCAAGA

>read_no_3
GGGAGACACCCGCACGTGTGGCCCCGCATGTATGCTGAGCTCTTCCGCGGAT

>read_no_4 TTTGCCCCGCATCGAGCGGGCTGTGCGGGAAATCCTTCTGGCTGTAGGCGA

>read_no_5 CCTGTGGGGCAAGGTGAACCCCGTGGAGATCGGCGCCGAGAGCCTGGCCAG

>read_no_6 GAGGAGGGCCAGGATCCACCAGAGGAAGGGCCTGCTGTGGTTCATCCCCGC

>read_no_7 CTGCACAGCGACTACAACCTGACCTGGTACAGGAACGGCAGCAACATGCCC

>read_no_8 GTGCTGGGCCTGGCCATCAGCCACTTCCTGCTGGAGCAGTTCCCCGACTAC

>read_no_9 AACCTGGGCGAGTACCTGCTGCGGCAAGGGCGAGGAGATGACCGGCGGC

>read_no_10 GTTCCCCGACTACAACGAGGGCGAGCTGAGGAGGCGCGACGAGGGGGCGCCATCGT

>read_no_11 CTTCAGCAAGTTCGGCGACCTGAGCAGCGTGAGCGCCATCATGGGCAACCC

>read_no_12 Accagaggaagggcctgctgtggttcatccccgccgccctggaggacagg

>read_no_13 AAGGGCGAGGAGATGACCGGCGGCAGGAGGAAGGCCAGCCTGCTGGCCGAC

- 10^5 10^10 reads
- 75 300Bp
- · Could be pair-end



NGS experiment workflow



Experimental design

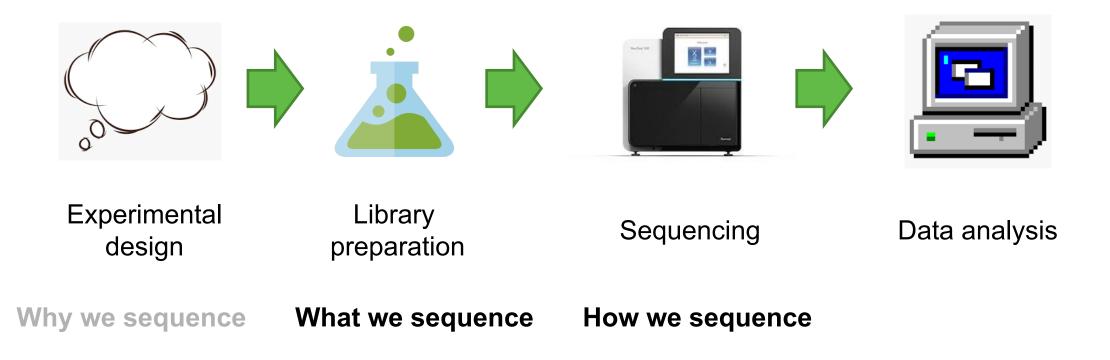
Library preparation

Sequencing

Data analysis

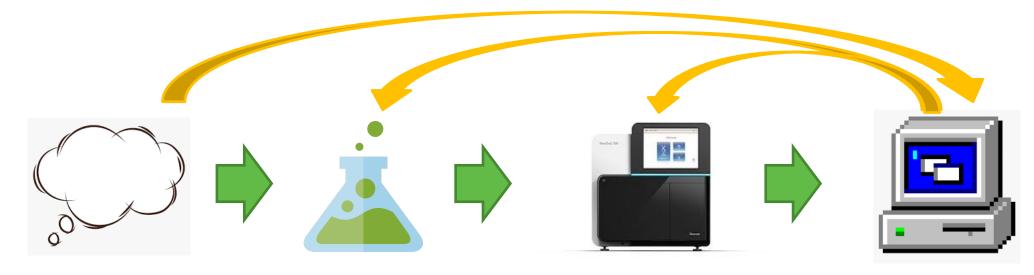


NGS experiment workflow





NGS experiment workflow



Experimental design

Library preparation

Sequencing

Data analysis

Why we sequence

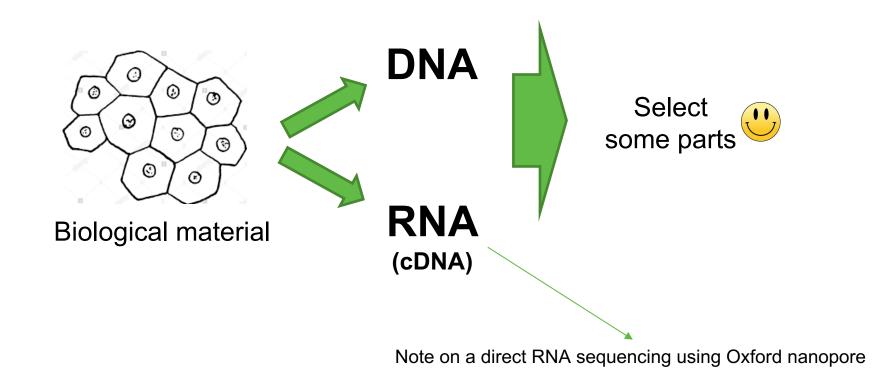
What we sequence

How we sequence

Consultation regarding data analysis is highly advisable.



NGS library preparation - What we sequence







	W The bet The weather we down	
ChIRP-Seq		
GRO-seq BRJC-Seq Bru-Seq BruChase		SHAPE-Seq
-Seq	And an antimetrophysical data and a second second by Budy Budy Strategy Budy Strategy Budy Strategy Budy Strategy Strate	Struc- ture-Seq
Ribo-Seq ARTseq ¹⁴	Sin faithease Researce All and	
RIP-Seq		SPARE
HITS-CLIP CLIP-Seq	ал опадарцирования портороди портород и портород и Портород и портород и пор	
PAR-CLIP	An and a second	PARS dsRNA-Seq
ICLIP	All and the second	Frag-Seq
NET-Seq	teach que arguing teachque arguing tea	
TRAP-Seq	white when a state of the second seco	Cap-Seq
	Targening publicational Adjourney CEPRoot BBA Resentationed place projected will/s-1294-Fang	
CLASH	Change and a final state and a	CIP-TAP
PARE-Seq		
	Papelinkandass (Hele only) (Mar (Hele)) Cappeninkla, frigmen Polyki (Hele (1972)) Repair Repair (Hele) (Hele) (Hele) Brial analism adaption transcription (Hele)	ICE
GMUCT 1.0	Comparison of the second	MeRIP-Seq
GMUCT 2.0	Conservation and the second se	Rearran
TIF-Seq	······································	RAD Paired-end
PEAT	Developt Indeer requesting 185 (sp. 2014) Interpreter and until 185 (sp. 2014) Interp	RAD-Seq
PEAT	anapon efficial (4400) energy products prename information prename info Information prename information pr	

RNA St	ructure			7 -	T		44
SHAPE-Seq	Subsectors 2: Appleand include inclusion for patients section sequencings (2:42PT-long)	_ _ _	• Arrested Table	→ IMPrester	Trans beaughter 10	-b	THE LANGE CONTRACT
Struc- ture-Seq		interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest interest inter		11		→	No series and a series of the
SPARE	Chronicae Deg	tor	* Freedom Providence	Parasida Perandidas Perandidas	+ 1 + 2 Presidence(1) all passe		
PARS dsRNA-Seq		→		III			strad 10646/Cpm
Frag-Seq	_î_		Province and the second		• <u>n</u> •		=
Cap-Seq	Contraction inspectory Figure 2 The second				······································	~	08A
CIP-TAP	CP ToP in standing any real to the second	-	11 - 11 - 11 - 11 - 2 - 11 - 1 - 1 - 11			ta tantan g	- H
ICE	Inside and the second s	****** *	Contral	- Anera Kitan	necolation -> -> -> -> -> -> -> -> -> -> -> -> ->		= ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
MeRIP-Seq	M	M. M.	مر المر المر المر المر المر المر المر ال	fuzzoreciliti	· · ·	-	
Rearra	ngements	and Marke	ers				
	Dances.		the star				

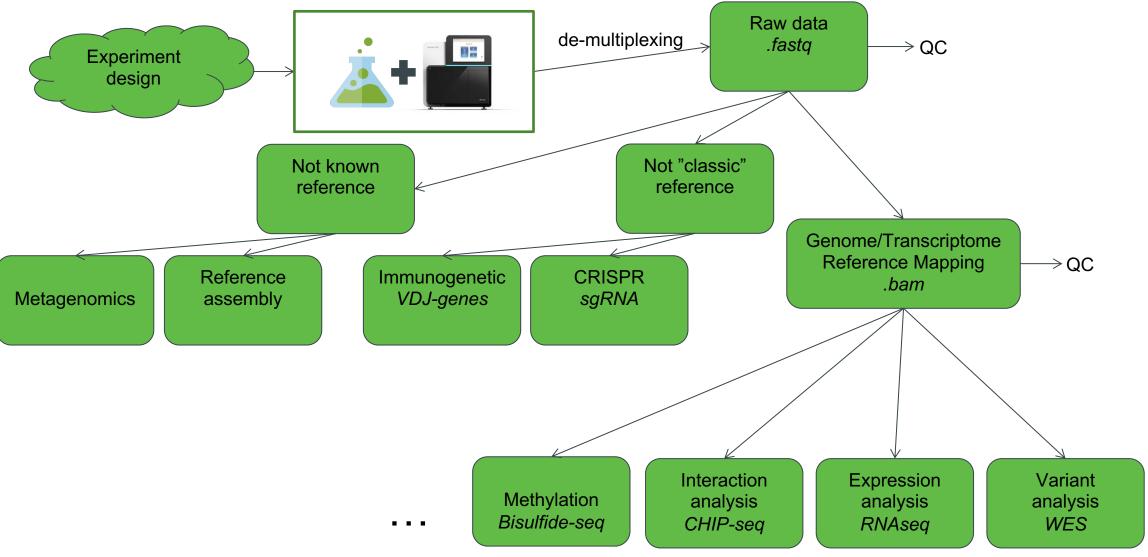
Methylat	*****
BS-Seq WGBS	Authorite and participant (Markan Markan Marka Markan Markan Mark
EpiGnome'' Methyl-Seq	Stability Conversion bindeduct Dim b
RRBS-Seq	Andream sector as the sector a
PBAT	A cla children and a classical
T-WGBS	Topostars had spansfer for the spansfer
axBS-5eq	Second and Analysis and Analysi
TAB-Seq	State state Mill Diff
MeDIP-Seq	M.M. M.M
MethylCap-Seg MBDCap-Seg	

DNA-Protein Interactions

1	DNase-Seq DNasel-Seq	Min m Min - Min - Min
1	MAINE-Seq MNase-Seq	Martin Martin Martin Constant and Antonio Constant
	ChIP-Seq ChIP-exo	MAN - MAN - MAN MAN
	FAIRE-Seq	The second secon

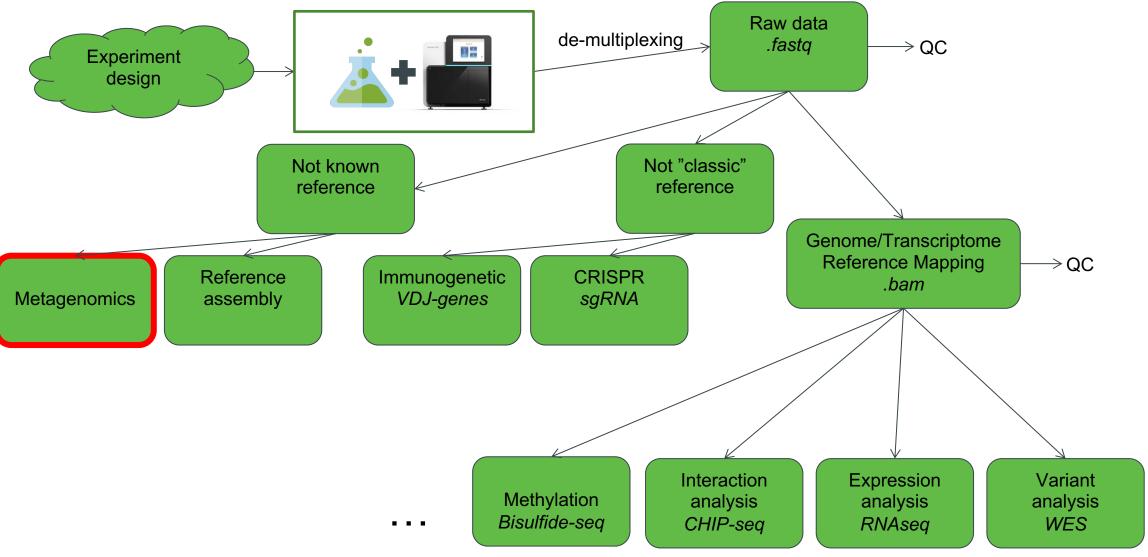
12

NGS data analysis



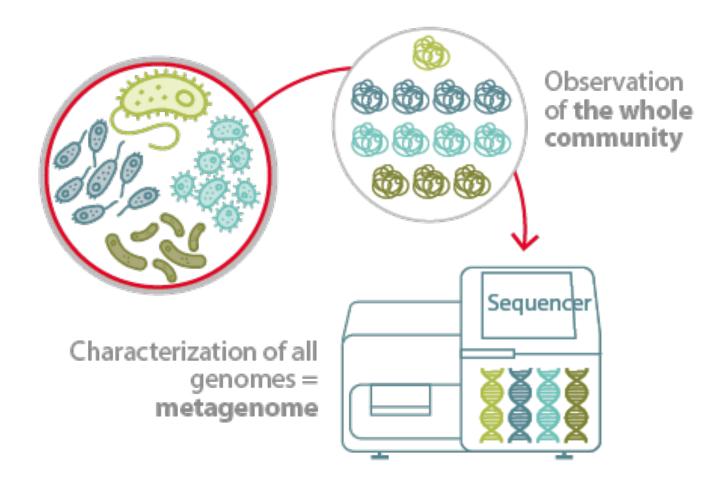


NGS data analysis





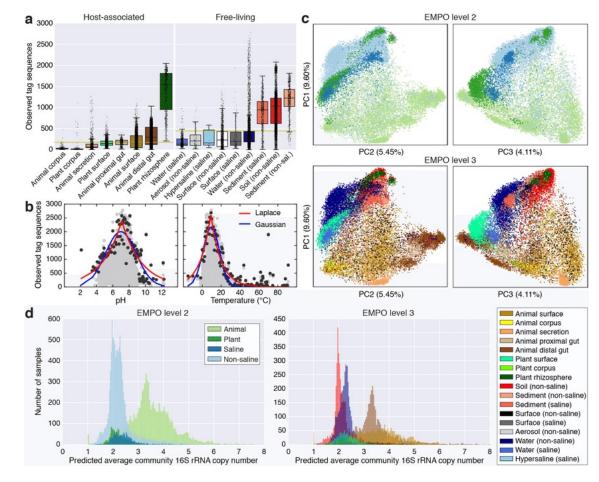
Metagenomics





Metagenomics results

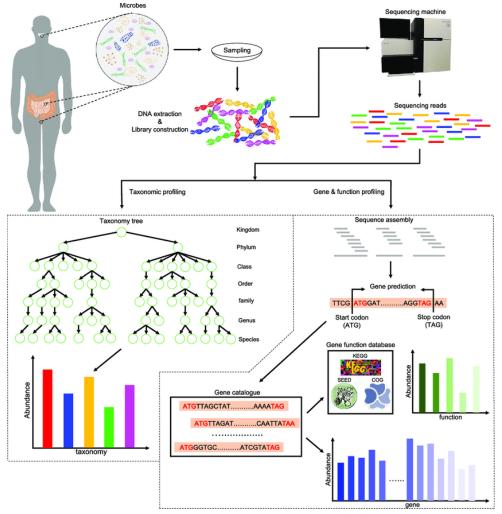
- Environmental statistics about populations
 - alpha, beta, gamma diversity





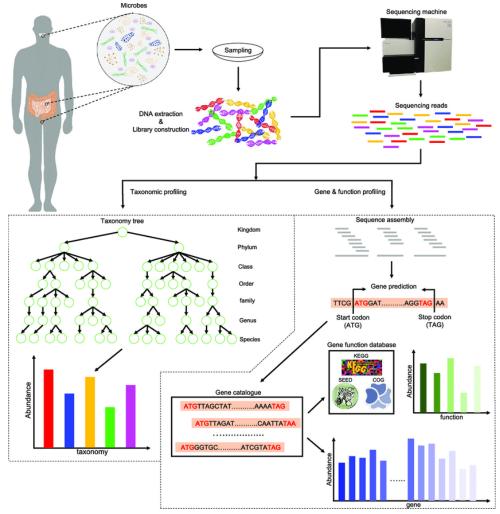
Metagenomics results

- Environmental statistics about populations
 - identify known bacterial species
 - taxonomy profiling
 - eventually functional profiling
 - E.g. antimicrobial resistance genes



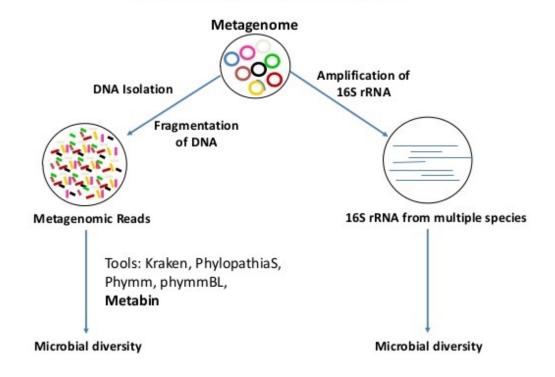
Metagenomics results

- Environmental statistics about populations
 - identify known bacterial species
 - taxonomy profiling
 - eventually functional profiling
 - E.g. antimicrobial resistance genes
- Sequencing techniques
 - 16S rRNA sequencing
 - Shotgun metagenomic sequencing





Metagenomic reads vs 16S rRNA for microbial diversity identification



Factors	16S rRNA sequencing	Shotgun Metagenomic Sequencing
Cost	~\$50 USD	Starting at ~\$150 but price will depend on sequencing depth required
Sample preparation	Similar complexity to shotgun sequencing	Similar complexity to 16S rRNA sequencing
Functional profiling (profile microbial genes)	No (but 'predicted' functional profiling is possible)	Yes (but it only reveals information on functional potential)
Taxonomic resolution: Genus, species, strain?	Bacterial genus (sometimes species); dependent on region(s) targeted	Bacterial species (sometimes strains and single nucleotide variants, if sequencing is deep enough)
Taxonomic coverage	Bacteria and archaea	All taxa, including viruses
Bioinformatics requirements	Beginner to intermediate expertise	Intermediate to advanced expertise
Databases	Established, well-curated	Relatively new, still growing
Sensitivity to host DNA contamination	Low (but PCR success depends on the absence of inhibitors and the presence of a detectable microbiome)	High , varies with sample type (but this can be mitigated by calibrating the sequencing depth)
Bias	Medium to high (retrieved taxonomic composition is dependent on selected primers and targeted variable region)	Lower (while metagenomics is "untargeted", experimental and analytical biases can be introduced at various stages)



- Study Examples
 - Assessment of the bacterial microbiome of Amazonian soil



- Study Examples
 - Assessment of the bacterial microbiome of Amazonian soil
 - 16S rRNA sequencing may provide more taxonomic resolution



- Study Examples
 - Assessment of the bacterial microbiome of Amazonian soil
 - 16S rRNA sequencing may provide more taxonomic resolution
 - Changes in microbiome composition and antimicrobial gene carriage following fecal transplant



- Study Examples
 - Assessment of the bacterial microbiome of Amazonian soil
 - 16S rRNA sequencing may provide more taxonomic resolution
 - Changes in microbiome composition and antimicrobial gene carriage following fecal transplant
 - shotgun sequencing to assess both compositional and functional differences



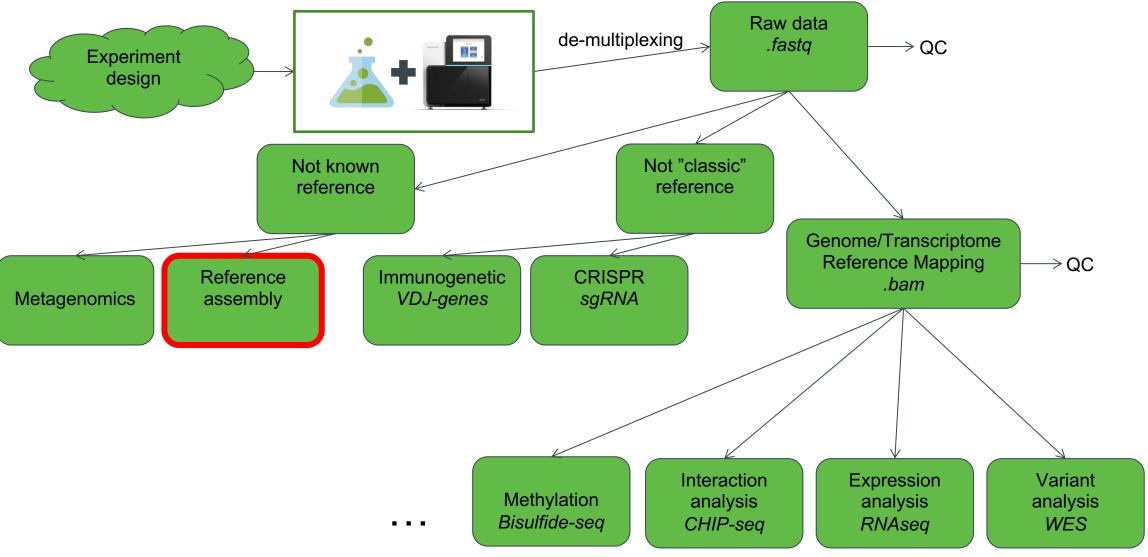
- Study Examples
 - Assessment of the bacterial microbiome of Amazonian soil
 - 16S rRNA sequencing may provide more taxonomic resolution
 - Changes in microbiome composition and antimicrobial gene carriage following fecal transplant
 - shotgun sequencing to assess both compositional and functional differences
 - Daily fluctuations in gut microbiome following 2 week dietary fiber intervention



- Study Examples
 - Assessment of the bacterial microbiome of Amazonian soil
 - 16S rRNA sequencing may provide more taxonomic resolution
 - Changes in microbiome composition and antimicrobial gene carriage following fecal transplant
 - shotgun sequencing to assess both compositional and functional differences
 - Daily fluctuations in gut microbiome following 2 week dietary fiber intervention
 - shotgun sequencing or 16S rRNA
 - assess both compositional and functional differences
 - cheaper and in this case can use 'predicted' functional profiling

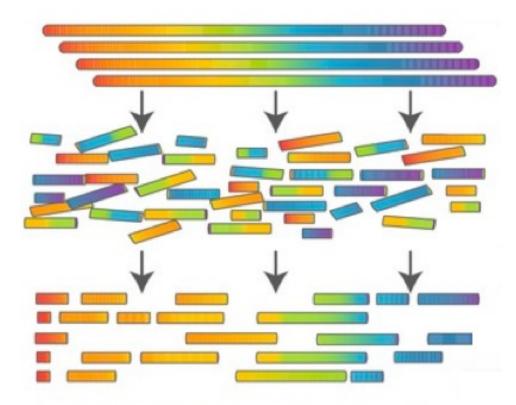


NGS data analysis



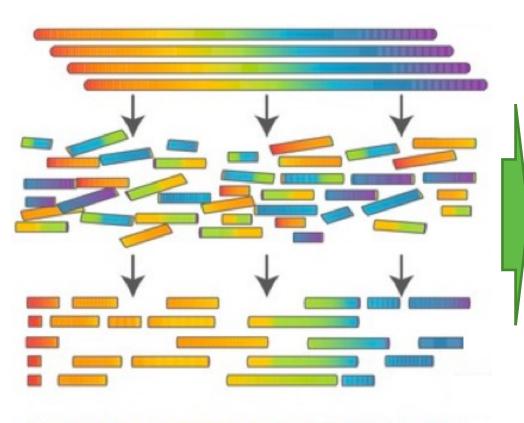


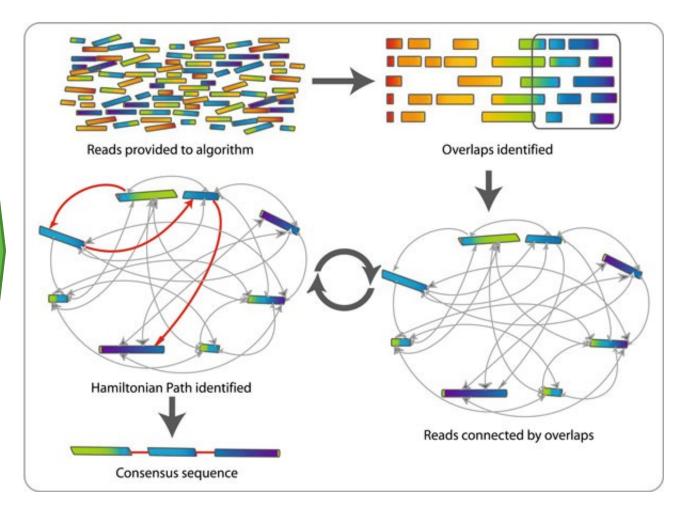
Reference Assembly



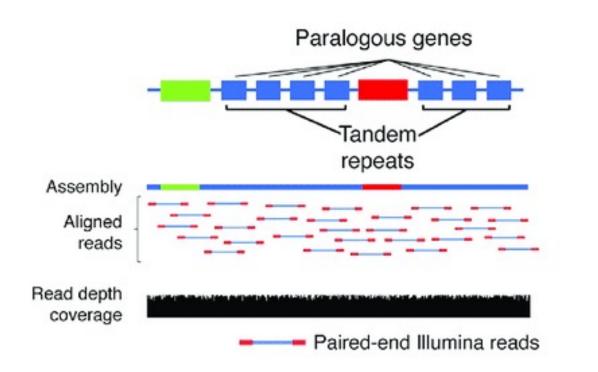


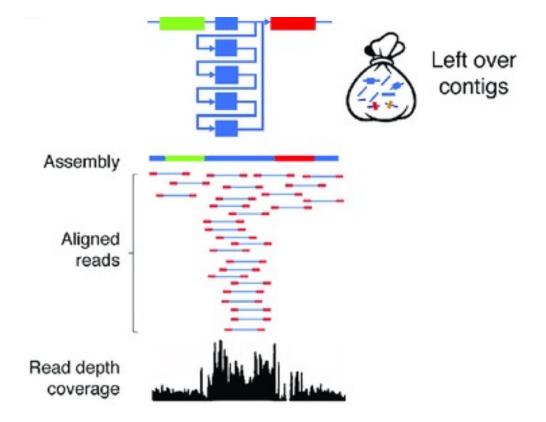
Reference Assembly





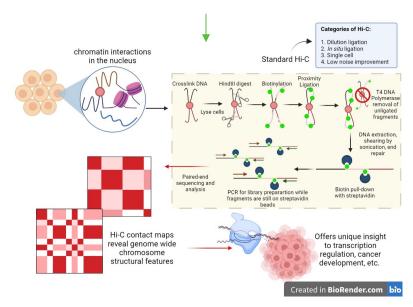
Reference Assembly problematic with short read





Genome Assembly

- Very hard and costly (in eukaryota)
- Multiple sequencing types needed
 - Pair-end short reads
 - Long reads —
 - Mate-pairs (e.g. Hi-C)

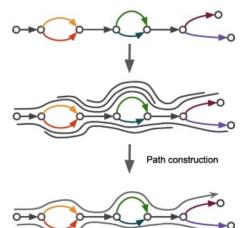




Assembly graph construction

Long reads

Short reads

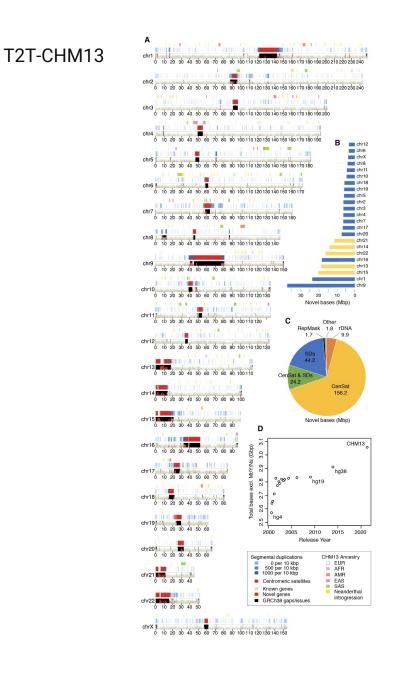






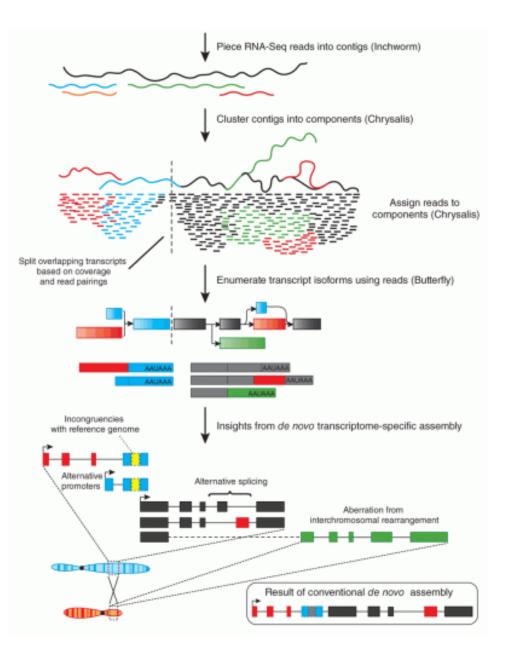
Genome Assembly

- Very hard and costly (in eukaryota)
- Multiple sequencing types needed
 - Pair-end short reads
 - Long reads
 - Mate-pairs (e.g. Hi-C)



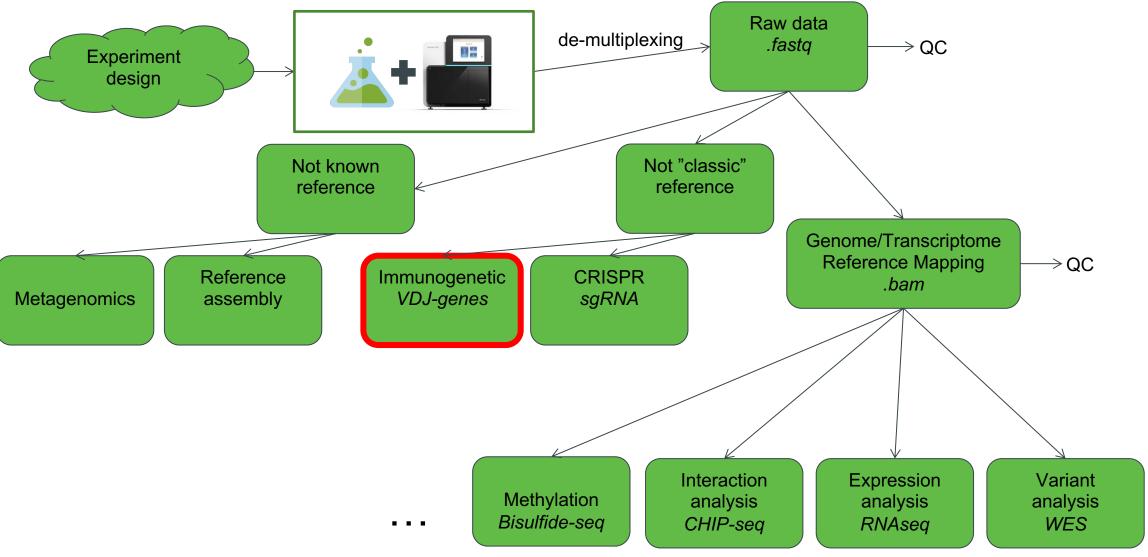
Transcriptome Assembly

- Assemble RNA fragments
 - Similar reference helpful
- Genome guided assembly
 - Good for poorly annotated organisms with known genomic reference





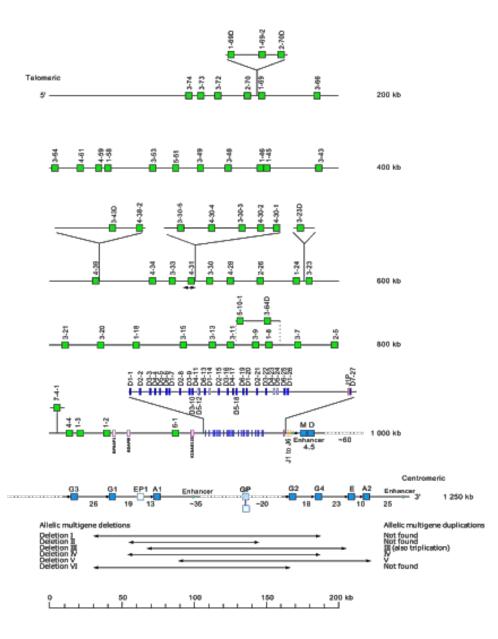
NGS data analysis





Immunogenetic

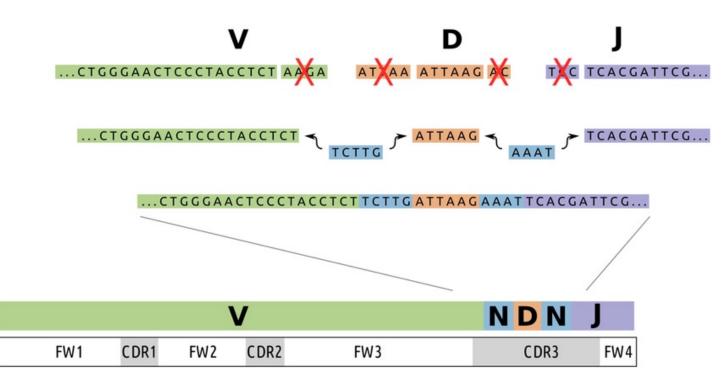
- T-cell receptor , Immunoglobulin (B-cell)
- Gene rearrangement during cell maturation
 - VDJ recombination





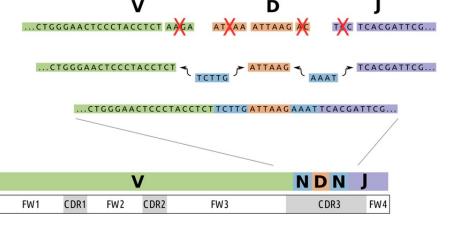
Immunogenetic

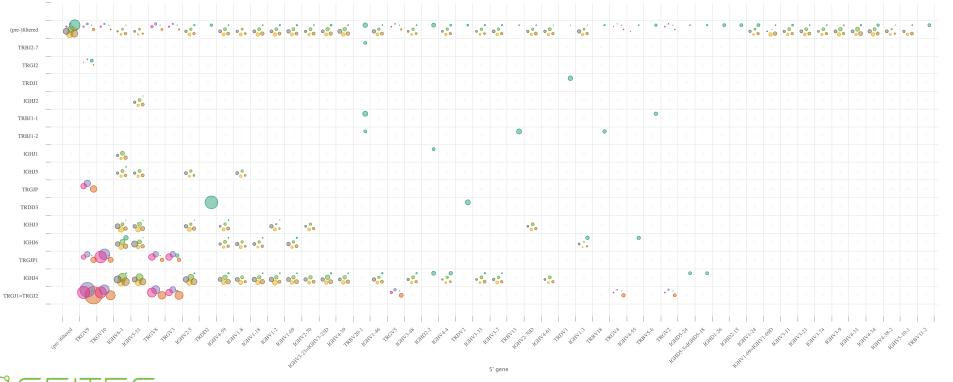
- T-cell receptor, Immunoglobulin (B-cell)
- Gene rearrangement during cell maturation
 - VDJ recombination



Immunogenetic

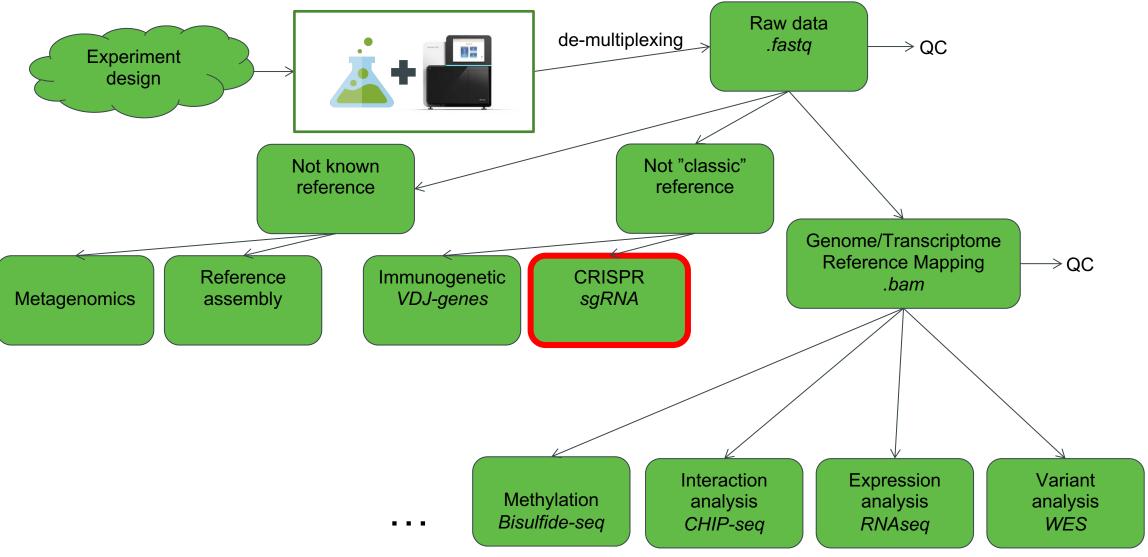
- Different cell populations
 - Clonal studies
 - Repertoire usage
- Main usage blood malignancies (leukemias)





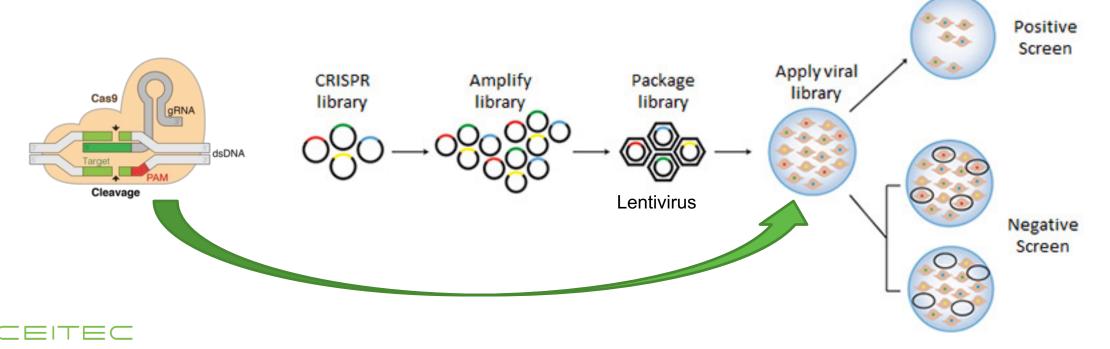


NGS data analysis

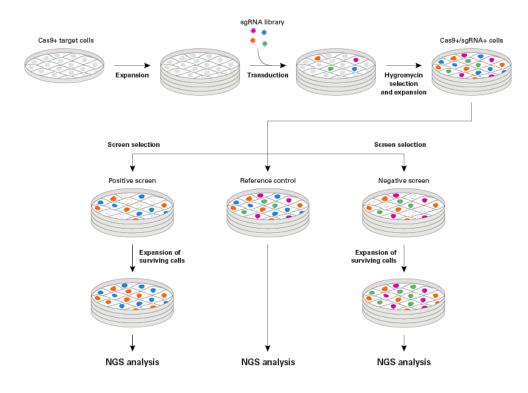




- Cas9 (CRISPR associated protein 9) is a protein which plays a vital role in the immunological defense of certain bacteria against DNA viruses
- sgRNA libraries
 - Each sgRNA knockout specific gene
 - 76,000 guide RNAs (sgRNAs) with four highly active guides per gene, targeting about 19,000 genes as well as non-targeting sgRNA controls

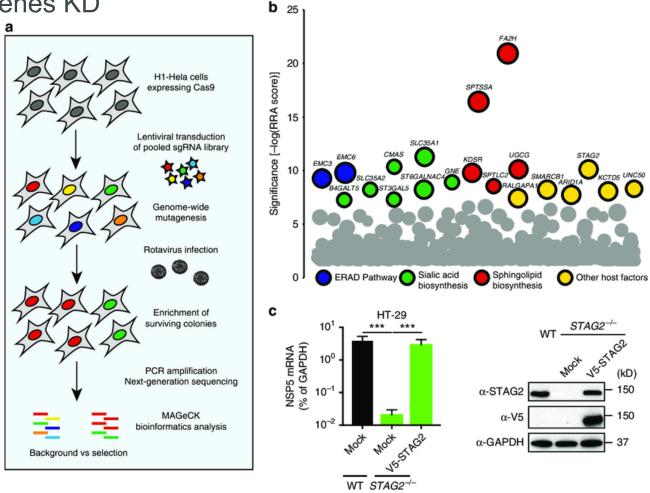


- Screen selection + expansion/enrichment of surviving cells
- NGS sequencing



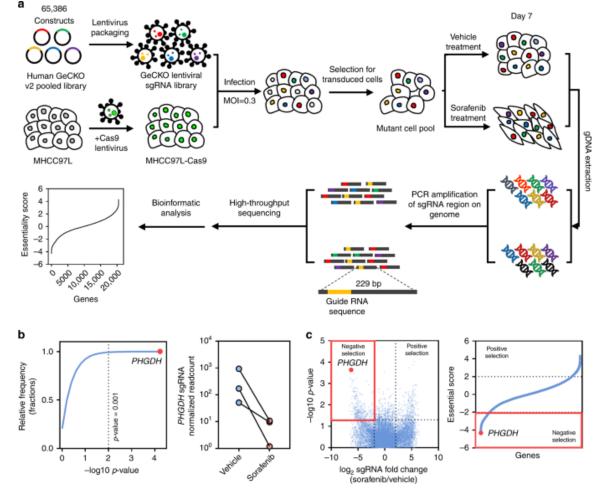


- NGS data analysis
 - Counting cells with different genes KD
 - Counting sgRNA fragments
 - Compare conditions









Wei, L., Lee, D., Law, CT. *et al.* Genome-wide CRISPR/Cas9 library screening identified PHGDH as a critical driver for Sorafenib resistance in HCC. *Nat Commun* **10**, 4681 (2019). https://doi.org/10.1038/s41467-019-12606-7



NGS data analysis

