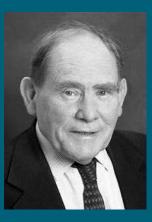
# Human genome, genetic variation, genomic technologies

Dr Bilal Alobaidi Biosciences Institute Faculty of Medical Sciences



"Progress in science depends on new techniques, new discoveries, and new ideas, probably in that order."

- Sydney Brenner, 2002 Nobel Prize Winner



GAAGGGGATAGGACCATAGGGGTCTTGACCAAGCCTGATCTGGTGGACAGAGG TCTGCTGTACTGCTAAGTCCAAAATTAAAGAATAAGATTAAACCAAGAAAAGGAAGCTGAAAACCTGATC TAAACCCTGC GATGACCCAGCACCTGAAAGCCTACTACCAGGAGTGCAGACGGAATA TTCAGCTTTTTACAGGACACCAGTAAGTGCAGCTGGTTCCTGGAGG AGAAG GCCAAATTCTCCGATTAACCAGGCTAGCTATGTGGCTTTCCTGG тсст AGTGCAAG CCCI GTAAACTCTGAGGAGATGAGGAGAGAGGAGCTATGGAGAGTCA GAA TTGTCAGGAGGATGTACTGGCTGAGCACCTCAGAAGGGTCGGC Disease G GTO ACATTTTTTTTGCAAAAATTTTAACAAGGATTCCACCTCTCT TCTTTGGGGGGCAAGCTCAGTCTTCCTTGGCAGCAGTTCAGTCC AGCA AIGAIGUIGG TAGTCCAGTCCTTTCCACAGGCAGAAACCAAGAATGAATCAGC GCTI TATGT TCACAAGTAGTTCTTAGGTGAGTTTCTCTCAATGTCAGTGTTAC GCTG ATGGG'I AACCAAGCCAAGGCTCCGTTGCTCATCTCCAACTGTTTGTGAGGT CTTCAGGGAAACATTCTTTCATGTGTTTGCTTCAGCAAGACATCAT1. LAAAACA G CTTCCAATAAAATTCTTGCTAAGTGACTGGTGCGGCCCTGTATTGACCTG. IAGTTCTGGGAAGAGCTCTGTGCTGGAAGCACTGTCTGGAGTGGCCCCTCCCCAGAGGCAGTGGT GAGAGGCAAAGTCTCCTATGATGACATTGAAGTGGAGCTCTCTGATCCTTCAGAGGTGGAAGAGG 'AACCTGGTGGTAGTCCCCCAGCAATGTGGACATTGCTACCACAGAGGCTCTGAGCATGGCTCAGG CAGAGGTGCTGAAGGCAAGGTCTTGGATGTGATGCGGGAACCTGGTGTATCCTCTGAAGAAGG JAGGCTTATCAGACAGTGCAAGTCTTCTTCAAGGATCACTCATACTTCAGCATTCTTCTGGAA lactititacagacaccagtaagtgcagctggttcctggaggagcagagtgacaccagagag `AAATTCTCCGATTAACCAGGCTAGCTATGTGGCTTTCCTGGTCGCTGTGCAATGATCCTTTAGCT NACTCTGAGGAGATGAGGAGAGAGGAGCTATGGAGAGTCATTCTACTTCAGCAGGAAGACATG TCAGGAGGATGTACTGGCTGAGCACCTCAGAAGGGTCGGCTTCTGGTTTTGTACTGTCCTGTCT "TTTTTTGCAAAATTTTTAACAAGGATTCCACCTCTCTTTTAGCCCACCACCCATCAGAGG TTGGGGGCAAGCTCAGTCTTCCTTGGCAGCAGTTCAGTCCCATAACAACACCAAGTATGACT JTCCAGTCCTTTTCCACAGGCAGAAACCAAGAATGAATCAGCAGCTGAAGTTCAATCTTGATGAA

### Major challenges in medical genetics

- Identifying genetic variation
- Interpreting genetic variation

In research: Large, longterm studies to identify genetic variation that increases/decreases risk of disease, and functional studies to confirm pathogenicity & unravel the mechanism

In clinic: Diagnosis in individual patients, who want reliable and useful answers fast and affordable....



1989 Genetics is an art

#### 2019 Genetics is an industry

- 我我我们

HiSeq 2000

### Our genome: Full of variation

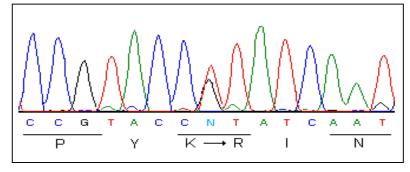
- 6 billion nucleotides per genome
- 2 people vary at 4 million positions
- 1 variation (mutation) can cause a rare disease





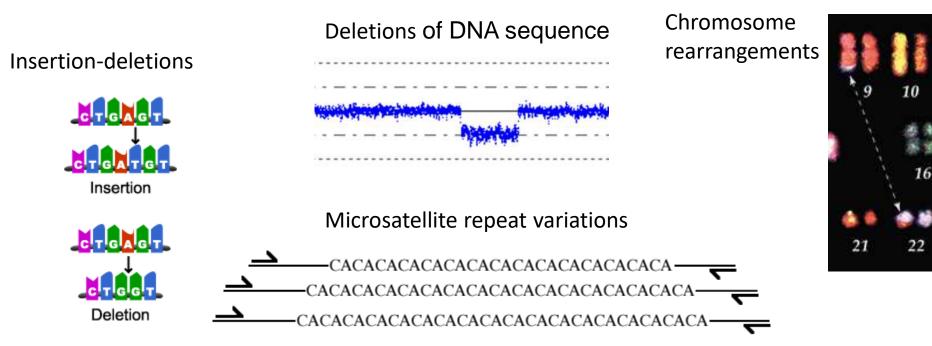
#### Full of variation indeed!

#### Single nucleotide variation



Additional chromosome





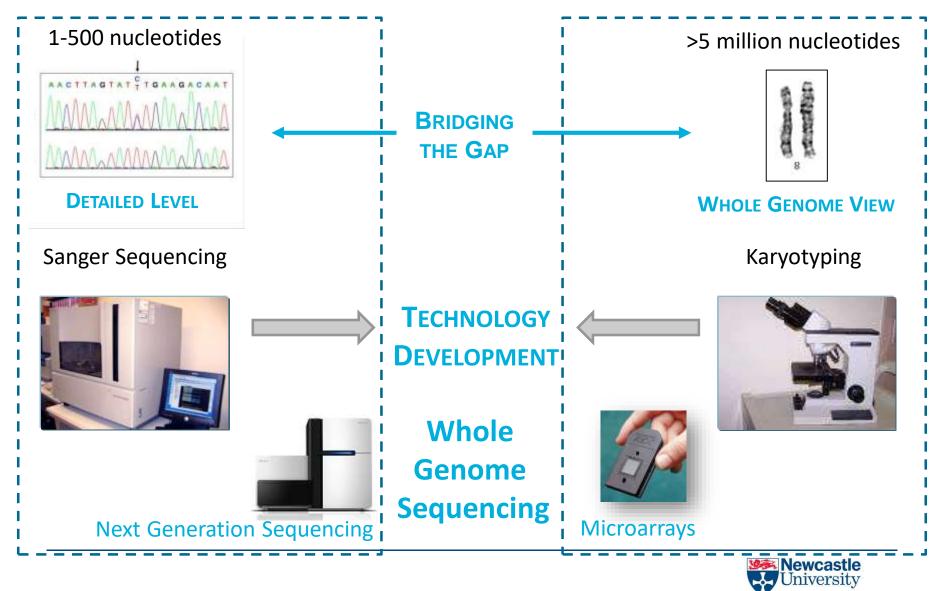


### Variation per genome

- SNVs (single nucleotide variants):
  - ~ 3-3.5 Million SNVs
    - of which vast majority SNPs single nucleotide polymorphisms
    - ~ 500 private/rare coding variants
    - ~50-100 de novo mutations per genome (0-4 coding)
- Indels (insertions/deletions)
  - ~300,000 indels
  - Largest uncertainty, still difficult to detect
- CNVs (copy number variants)
  - 100bp-10Mb: ~1000 per genome
  - >50kb: ~30/genome
  - De novo, >100kb: <1 per genome</p>



Detection of genomic variation at all resolutions From nucleotides to chromosomes!



# Traditional vs. Next generation sequencing

96 DNA fragments sequenced simultaneously

#### Miniaturization and parallelization

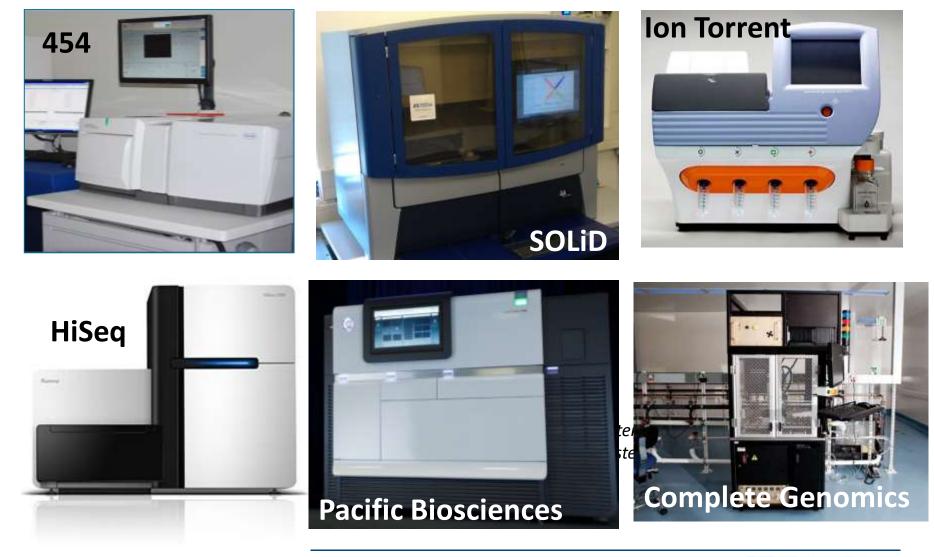
Millions of DNA fragments sequenced simultaneously







### Next generation sequencing equipment





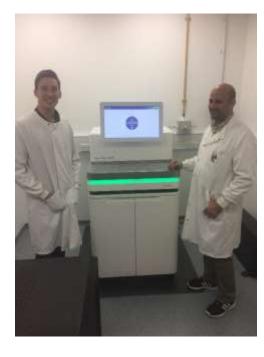
# Genome technology: Big and small scale

Illumina Novaseq





**Oxford Nanopores MinION** 

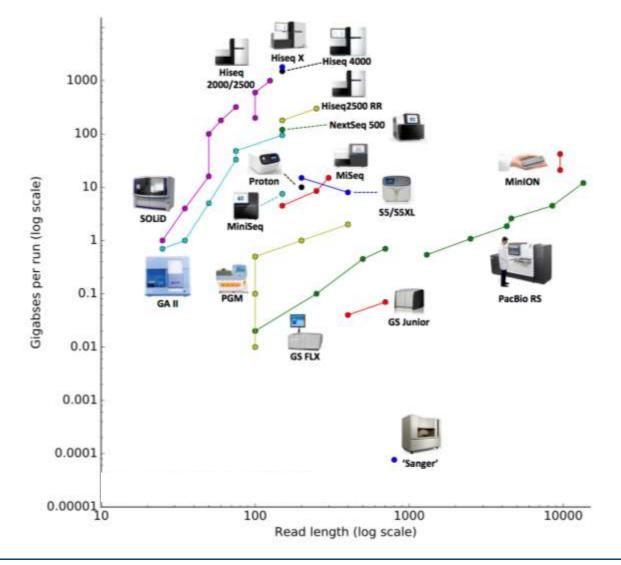






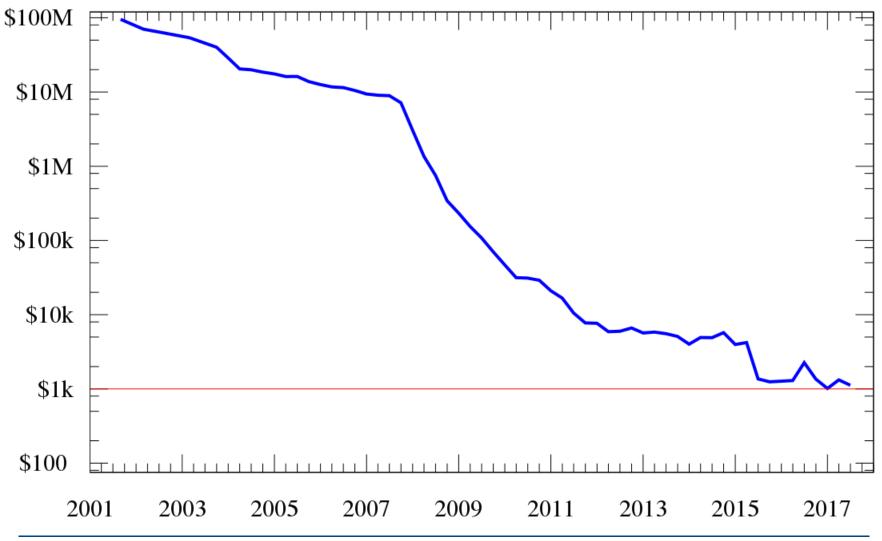


### Reading the DNA: Throughput and length





# Cost of sequencing a human genome (with reasonable quality for variant identification)





\$3,000,000,000 2003 Human Genome Project	M
\$20,000,000 2006 1st individual genome	
\$2,000,000 2007 1st NGS Genome	
\$200,000 2008 1st 30x genome	
\$10,000 2010 1st sub-10K genome	
\$1,000 2014 1st \$1,000 genome	
\$100   2017 1st \$100 genome	

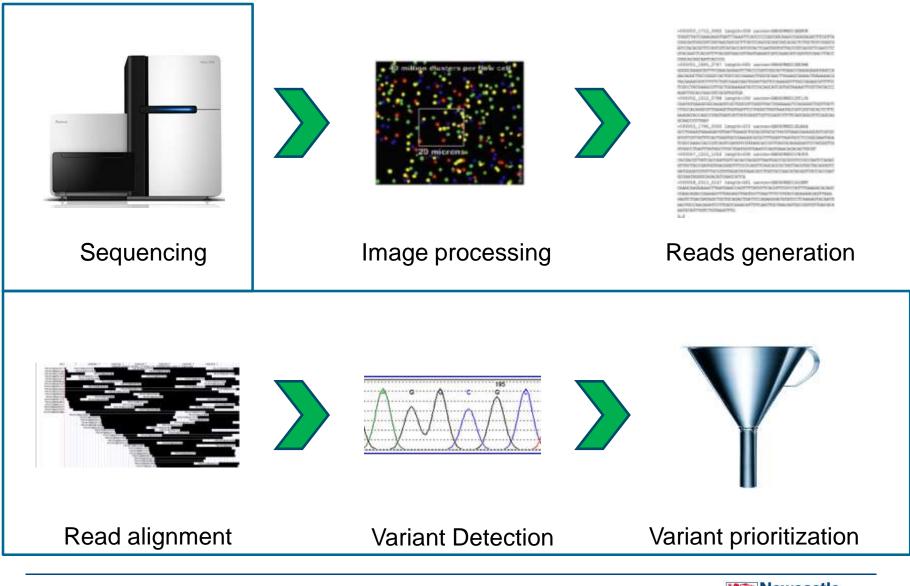
### **Technology choice based on application**

- All types of genetic variation or only single nucleotide variants?
- Gene panel, all genes (the exome) or the entire genome?
- Discovery science or diagnostics?
- Germline variant detection or also somatic (and in what tissue)?

#### And on finances, expertise, bioinformatics capacity, turnaround-time, etc....

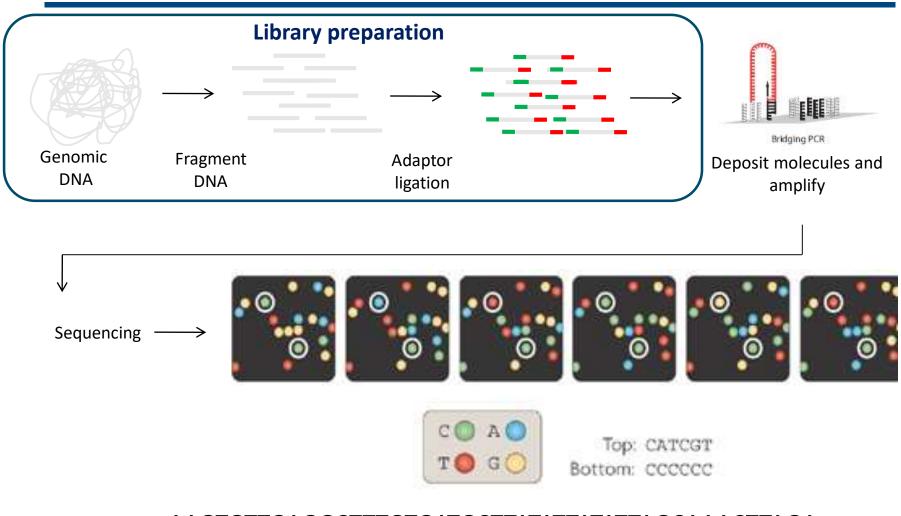


#### Next generation sequencing workflow





#### Next generation sequencing basics

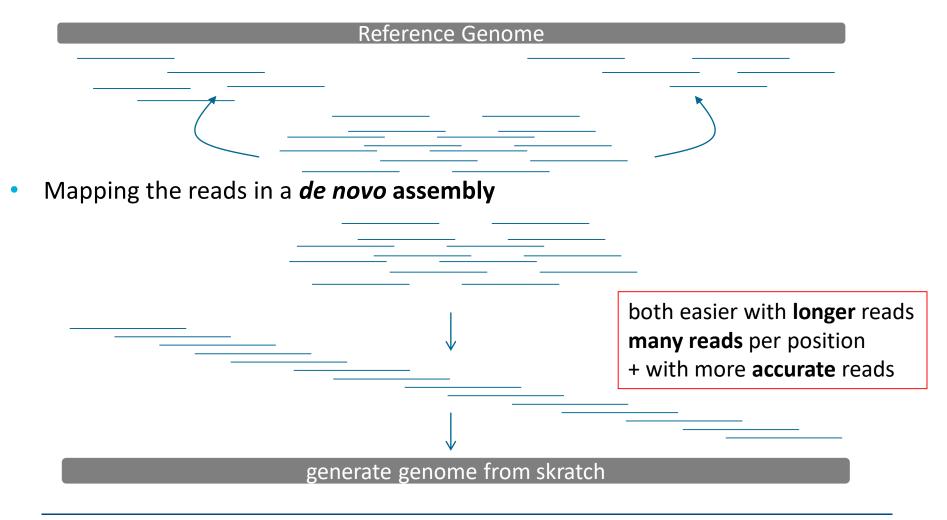


Translate into NGS reads AAGTGTTGAGGCTTTGTGATGCTTATATTATTATTAGCAAACTTAGA ~100 Million Single Reads per Patient!



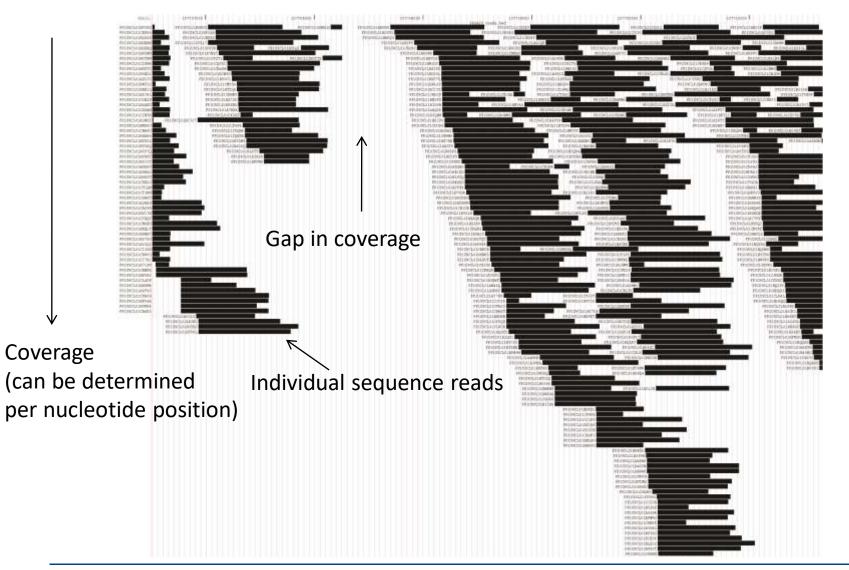
### Mapping sequencing reads

• Mapping the reads to a **reference genome** 



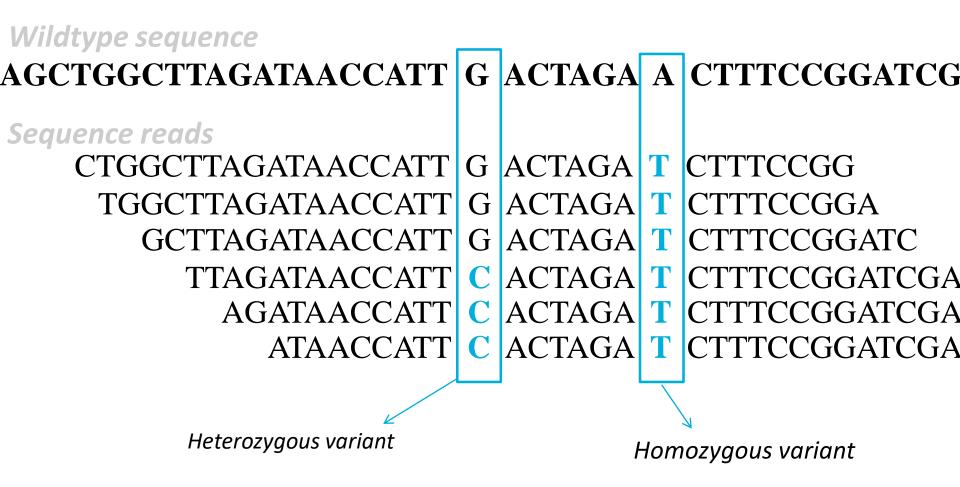


#### Aligning sequencing reads to the reference genome





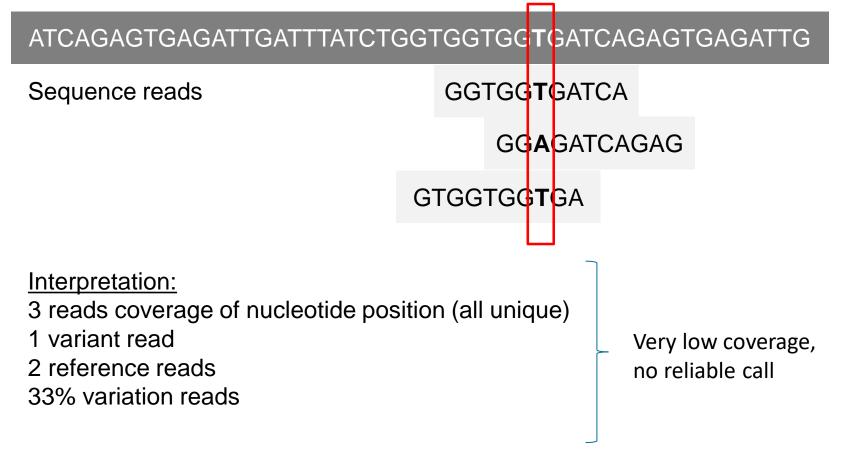
#### Variant detection - Theory





## The importance of coverage

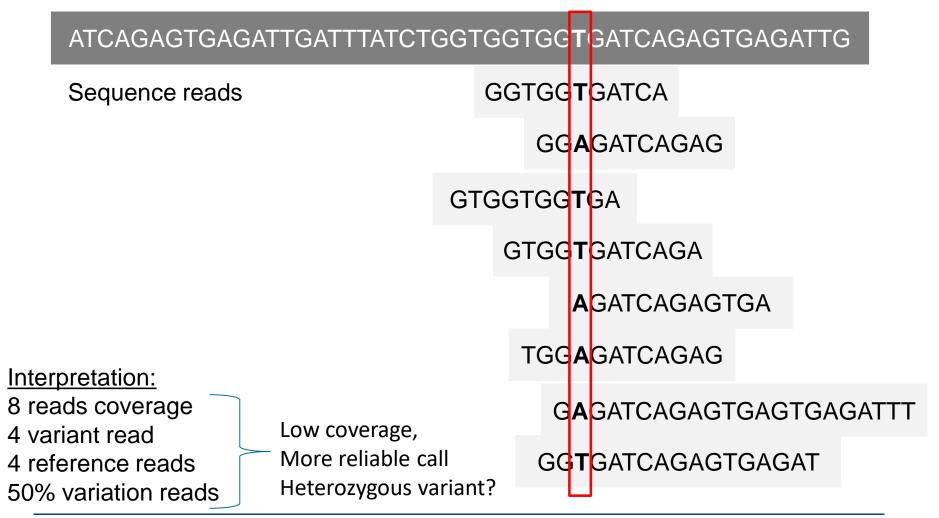
#### Reference genome





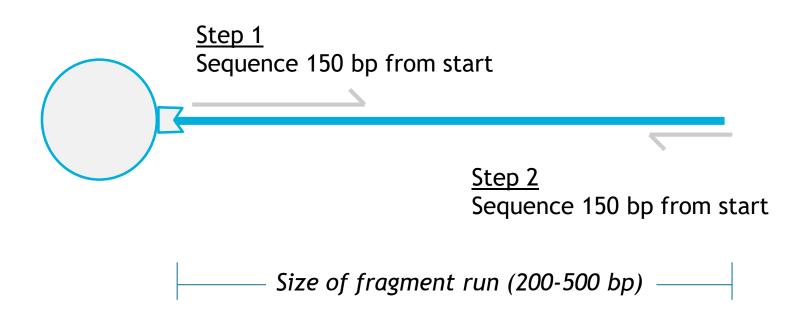
# The importance of coverage

Reference genome





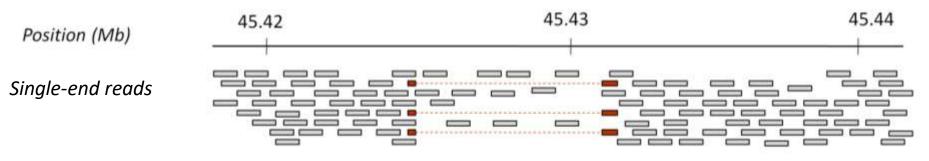
#### Paired-End sequencing



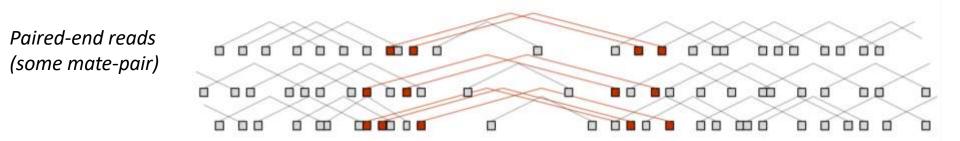
- More <u>data</u> from 1 sequence run (but also takes longer)
- Higher <u>confidence</u> mapping due to relation between sequences
- Useful for studying structural genomic variation



### Detecting structural genomic variation by NGS



Detect CNVs by looking at coverage of sequence reads and at split reads

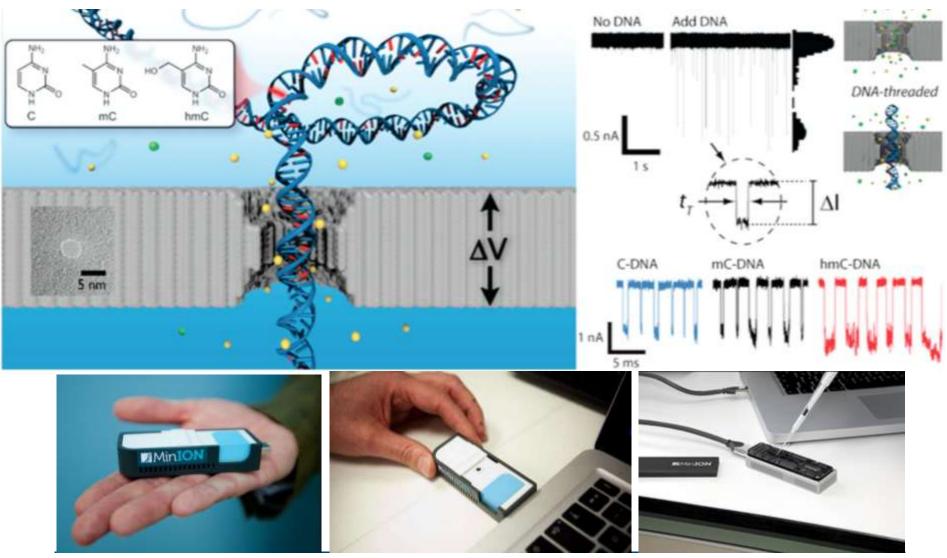


Detects also balanced rearrangements!

Vissers, de Vries & Veltman, JMG 2009



#### Single molecule sequencing without amplification; the future?



Feng et al. Genomics Proteomics Bioinformatics 2015; Wang et al. Frontiers in Genetics 2015

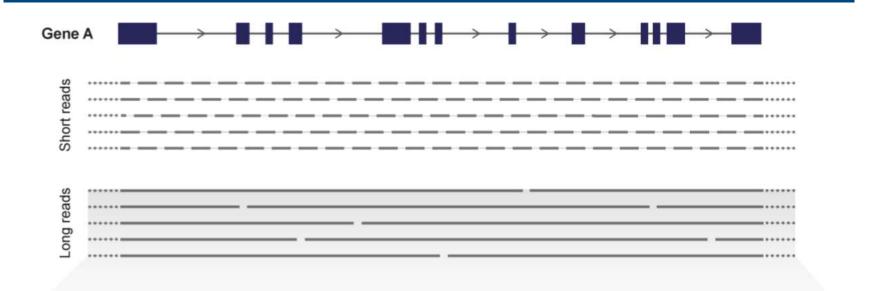


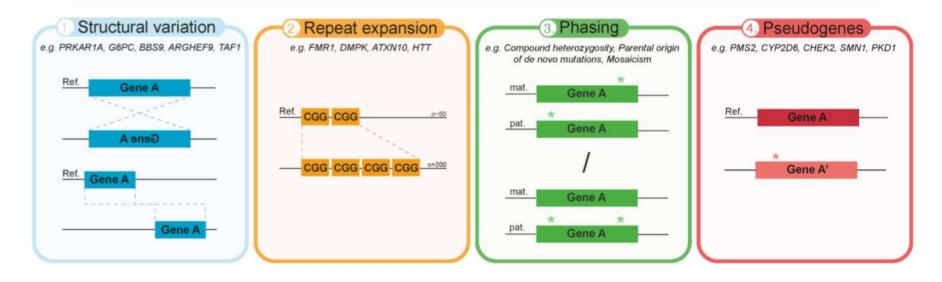
### The promise of single molecule sequencing

- Amplification of DNA prior to sequencing introduces artefacts, DNA needs to be chopped in small fragments, it takes time and is expensive
- Sequencing of one molecule (chromosome) at the time is potentially ideal, especially for analyzing complex genomic regions (e.g. HLA)
- Major advantage: Long sequencing reads
- Major Challenge: Raw sequencing accuracy
- Major companies: Pacific Biosciences & Oxford Nanopores



### Advantages of long read sequencing







Mantere, Kersten & Hoischen. Frontiers in Genetics 2019.

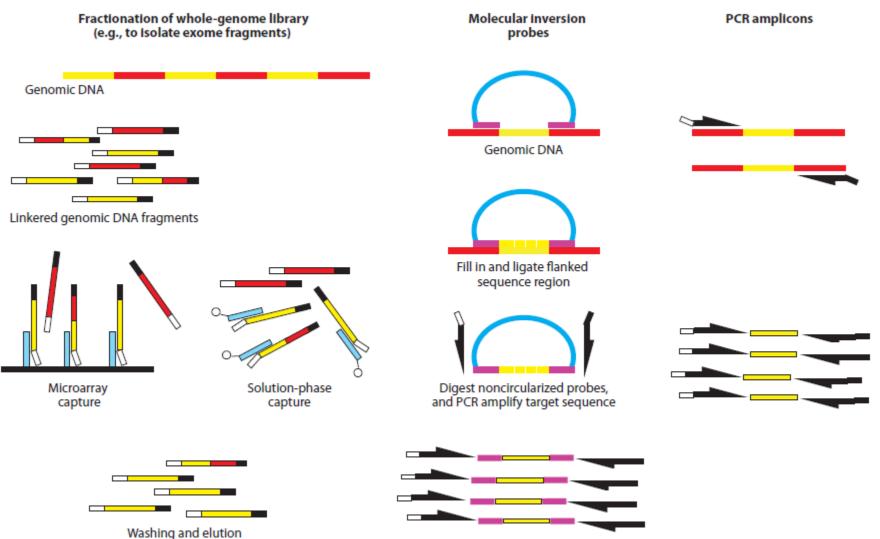
#### Typical questions for which NGS can be used NOW!

- Can we sequence all known disease genes of genetically heterogeneous diseases in parallel (*e.g.* hereditary breast cancer, ataxia, hereditary blindness)?
- Can we sequence entire candidate disease gene loci (*e.g.* from linkage studies/homozygosity mapping)?
- Can we sequence the whole exome (all exons of a genome) to decipher unknown syndromes/diseases?





#### Enriching your DNA to be sequenced





#### • CFTR diagnostics

Enrichment: Amplicon, Molecular inversion probes Sequencing: Ion Torrent, Illumina Miseq & Nextseq

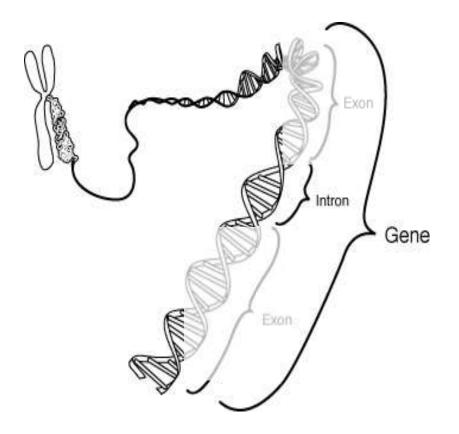
- Sequencing of 100 candidate male infertility genes
   Enrichment: Amplicon, Molecular inversion probe, NimbleGen / Agilent in solution enrichment, Fluidigm, Raindance

   Sequencing: Ion Torrent, Illumina Miseq & Nextseq
- Exome sequencing (diagostics/research)
   Enrichment options: NimbleGen/Agilent/Twist biosciences
   Sequencing: Illumina Hiseq, Novaseq



#### Exome sequencing; Practicing for genomes

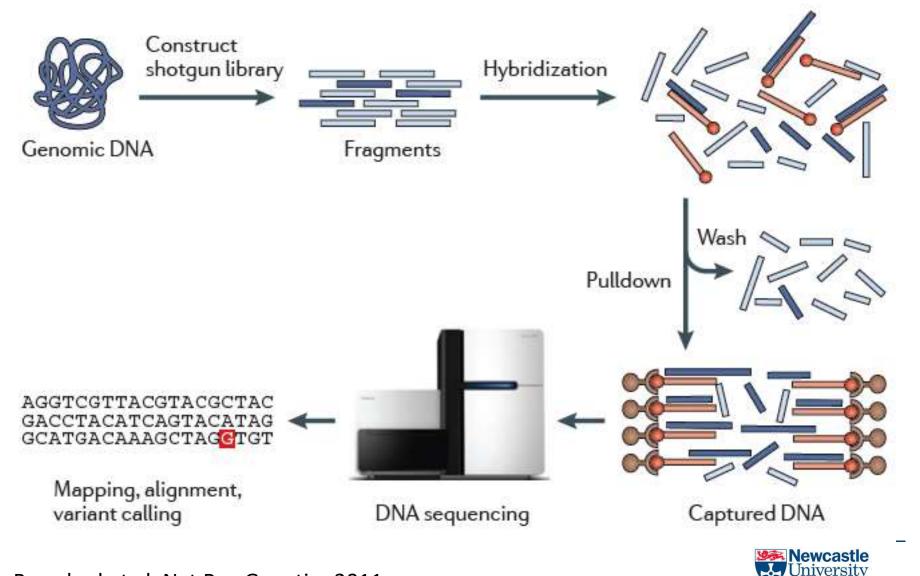
#### 'Exome' (all exons of a genome) ~1% of the human genome



**'All'** coding sequences of a human genome (>200,000 exons), sequenced and analyzed in **one** experiment



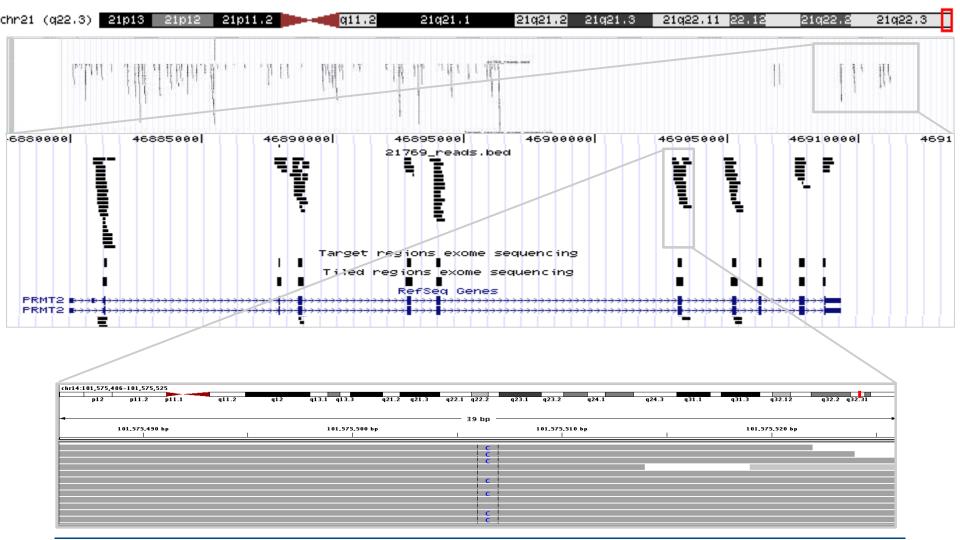
### Exome sequencing workflow



Bamshad et al. Nat Rev Genetics 2011

#### Mapping and annotation of exome sequencing reads

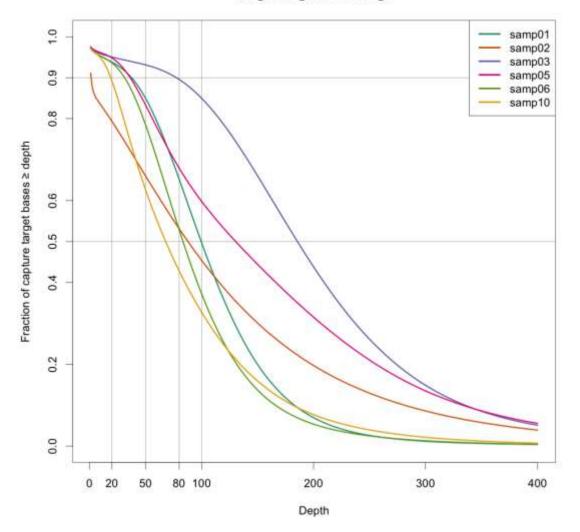
terminal 500kb





#### Enrichment is imperfect, varies per sample

**Target Region Coverage** 



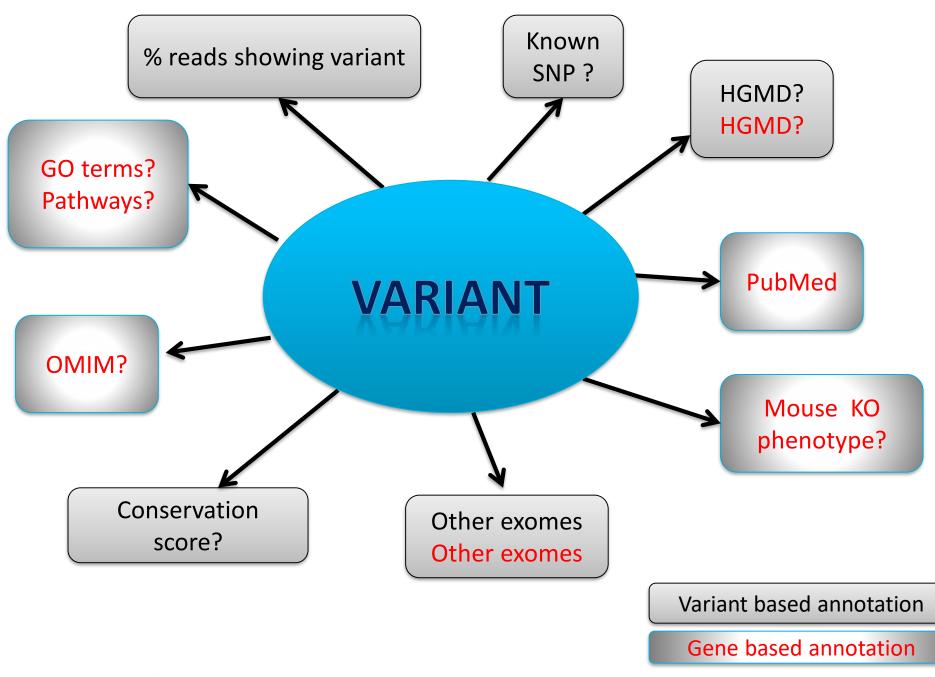


#### Variant calling and variant annotation

4 61	Slarlan Endansi erf	D E			K L H H O	P AD AC AD AE A	F AG AH AI	AJ AK AL AM	AM AO AP AO	AR AS AT	AU AV AW			DJ DK DL DH DH DO DP galaf_ga GOleen Godenn GeneCae OHIH OHIH_E Eulerni
17	559846 559846 G	A	····· , ···	1 10	3 13 ra20568 Suar SH G A			1 1	1 1	1 1 1	PALSE	1 1.24144 Halanar HA HA		25.77
10 abr1	171631 171631 G	A	52	14 45.75 aukalila I	15 16 -= 22727 Sue- SH G A		HOCZL HH_845		15 15 EXOH_F	16 75 114 E.16-		8.68584 3.86475 H/A H/A H/A		l bily://w bily://www.eebi. 25155
13 eke 1 28 eke 1	877664 877664 A	\$	25	11 44 autolila 1 1 25.0065 autolila 1	9 21 98288 5 SH A G 7 19 97485 5 SH T C		HOCZL HH_BIS HOCZL HH_BIS	15837 863446 884542 · 15837 863446 884542 ·	15 15 EXON_F	10 10 105 E.10-1		1.02362 -1.0765 H/A H/A H/A 1.33213 1.23355 H/A H/A H/A		l bily://a bily://aaa.eeki. 26155 bily://a bily://aaa.eeki. 26155
21	112412 112412 C	Ŧ	15	5 38.4645	4 15 ++41285 Same SH C T	i i	HOCZL HH_815	15857 865446 884542 -	15 15 EXON_F	3 31 175 E.S.B		1 8.33481 H/A H/A H/A		B bilg://u bilg://uuu.eebi. 20155
22	114312 114312 G	A	25	11 47.8251 aubalila 8	18 13	7 5 apreak 1 1888Gea 8 1		15837 863446 884542 ·	15 15 INTRON	1 30 105 1751-4	46 PALSE	I -I.514I Halamar HA HA	I I I.61 I I 17.55	i B bily://w bily://www.eebi. 26155
23	114435 114435 G	2	14	8 57.1423 oobolilo 8 26 41.3355 oobolilo 8	1 13 re13313 Suer SH G A 12 24 re13784 Suer SH G C		HOCZL HH_BIS KLHLIZ HH 15	15837 863446 884542 · 5424 885446 884542 ·	15 15 INTRON 12 12 EXON F	1 52 185 1951-5 4 128 222 E+4-1		8 8.58322 Halanar HA HA 8 4844 -4 2029 H/A H/A H/A	1 1 1.51 1 1 17.55 1 1 1.51 1 1 17.55	
25	87684 87684 C	Ť	12	57 45.422	27 23 66565 5 5H C T		KLHL17 HH_158	5425 005030 030350 -	12 12 EXON_F	5 4 117 2.5-4		1 3.82775 H/A H/A H/A	1 1 1.61 1 1 1	lilip://www.grara.org 333451
25 ak-1	833484 833484 G	<u>e</u>	43	22 44.898 aubalila 8	15 16 58257 Sam- 5H G C		PLEKHE PLEKHE HHLUSZ	1686 194240 30034S -	16 16 EXON_F	14 25 215 E.14-3		I I.47882 DAMAG I.81 3.45	74 412 1 5.52 1 1	l Mig://www.granaarda.org MIS3
27	833631 833631 A 338784 338784 T	ŝ	131	138 33.2366 eskelüle 8	41 25 ra23485 Suar SH A G 34 17 ra15842 Suar SH T C		PLEKHP PLEKHP HH_001 ISG1S HH_00S	1686 831748 388345 - 1873 338748 339782 -	15 15 INTRON 2 2 UTR	14 15 43 17514	-24 PALSE PALSE	I -II.3725 Helener HA HA	1 1 5.32 1 1	l bily://www.generaarda.org 04063 I GO:0032 (ISG15-y bily://www.generaarda.org 04063
23	333471 333471 G	Ä	15	5 91.25	5 24 re1921 Suer SH G A		15615 HH_885	1875 558718 555782 -	Z Z EXOM_F	2 245 455 2.2-2		1.00707 -1.0075 TOLERS 8.30 2.37	16 135 1.62 I 6.1 24.31	60:0052 (15015-p kilp://a kilp://aaa.eeki. 3636
38 abr1	333517 333517 A	G		<ol> <li>Hit saladita</li> </ol>	5 13 mil 137 Same SH A G	1 I I I I I I I I I I I I I I I I I I I		1073 330710 333702 -				1.12352 -1.5753 H/A H/A H/A	I I LE2 I E.1 26.31	I. GO:1152  ISC45+p kllp://w kllp://www.eeki. 3535
31 ek-1	367133 367133 T	coco acce	21	21 100 askalila 1		and the second second				7 15 10 1957-0		I -1.3537 Halasar HA HA I I IIITA Malasar HA HA	1 1.61	li Gostill? (planas kilp://w kilp://w Hyselka 37573) li Gostill? (planas kilp://w kilp://w Hyselka 37573)
321	378358 378358 A	G G		55 100 askalila I		1 3 metern Hackage						I -1.5214 Halamar HA HA	1 1.7 1.631	Gottine plana kiljered kiljered Hyselke Braelle Gottine plana kiljered kiljered Hyselke Braelle
34 akr1	972884 972884 T	c		43 35.5556 webstile I	21 21 Pastal Same SH T C	1 I I I I I I I I I I I I I I I I I I I	AGRH HH_111	35331 345366 314355 -	35 35 5A_SITE	24 3 48 19524	HE PALSE	I -1.300 Halance HA HA		E CO:1117  planua kilp://wkilp://wHiganike 375738
35 akr1	572857 572857 T 577865 577865 C	c c		1 75 askalila 1.11334								LIGHS LISZING N/A N/A N/A		E GO:III7  ylarma kily://u kily://u Hyanike 375731 E GO:III7  ylarma kily://u kily://u Hyanike 375731
35 eke1	1185242 1185242 G	Ť		51 100 askalila 100017		2 4 4006						I -1.1411 Halasar HA HA I -1.1424 Halasar HA HA	1.11	Colline (standard bilg://u bilg://u Hyselke 375738 Colline (standard bilg://uuu.standards.arg 254173
38 abr1	1118248 1118248 T	A		12.00.000 askalla I	1 24		EXECUTE TILLE NH 151		S EXON_F		H K PALSE	1.15127 1.12521 TOLER# 1.12 1.24	15 111 1 1.12 1 1	COLUMN (profess http://www.genesards.org 254175
55 akr1	1148434 1148434 A	G		Research Research	28 24 PREERST Same SH A G	1.1.1		15168 1102151 1157318 -		0.001.000500		8.36221 -1.5384 H/A H/A H/A		E CO:1071 Jayringe Milp://www.genenaedu.org 51150
41 -1-1	1221458 1221458 G	A .							24 24 SD_SITE -			I -3.3312 Halasar HA HA		CONTRACTOR AND A CONTRACTOR AND AND A CONTRACTOR AND A
42 skr1	1233858 1233858 G	Å		148 38.5557									7.51	kila://www.araraarda.ara 54973
43 akr1	1244784 1244784 C	G	21	40 - 30 anti-10a - 1	11 17 viel017 Szav SH C G	1 I I I I I I I I I I I I I I I I I I I		13112 123222 1243313 -		4 455 225 254-55	Z G G TRUE	1.31425 -1.3557 H/A H/A H/A	1 1 1 1 1.8	http://www.granarda.org 54575
44	1252454 1252454 C	Ţ		17 Hill scholling 1	18 22 re22164 Szer SH C T	1.1						E -E.7132 Halanay HA HA	1 1 1 12.10	Colling (styralig kilp://www.genenaeda.org 10772
45 -1-1	1252825 1252825 C	G I		d 15.1710 set-file 4					S Z EXON_F			1.33273 1.47525 N/A N/A N/A	1.23 1.2.14	l Go:1146 (deputis klip://www.granaarda.org 18272 1 Go:1114 (defentis klip://w.klip://www.anki. 19755
47	1255417 1255417 T	č	11	1	I CHARTER SHIT C	i i 💼	TASIRS HH_152	3113 1256583 1259787	E EXON_F	E EES 355 E-E-2		1.1171 .41533 To 1 1.12	112 _ 112 _ 1 _ 1 _ 1 _ 1 _ 1 _ 1 _ 1 _	o:1111 (d-Calis klip://w klip://www.aski. 89756
11 abr1	1267996 1267996 T	¢	11	9			DVL1 HH_IH	19895 1268521 1274955 -	15 15 EXON F	1 111 2012	P P TRUE	LITER -LINES HZ		o:1131  piliar kllp://w kllp://www.anki. 1955
Cl aler1	1275228 1275225 GG			45 48 4848 4545				514 3779 6					ntifie	http://www.granearde.org 54587 http://www.granearde.org 54587
51	1919885 1919886 CG			7 77.7 Artelian		The second se		19527		10010	-25 PALS			Control (regular bill)://www.generaedu.org 19557
52 abr1	1421828 1421828 C	т		12 23	1 4721 All 6 4		A1 30	24415 1 7 1			11 P PALSA			http://whitp://www.eeki. 19151
53 abr1	1469196 1469196 A	e.		12 25.887	18 21 +=7599 Sum+SH A G		SSU72			4 15 111 241-0		I -5.113 N/A N/A N/A		1 Costilli (aRMA) kiljs//aua.granaria.org 23181 1 Costilli (aratria kiljs//aua.granaria.org 142578
55	1553652 1553652 C	A	- 1	di dii sekelile i	11 15	27 Studher 1 1111 Gen 1 23	11; HIP2 HIP2 HH III					L. U.S. Nolmary No. 10		1 Gottitti profesa kilp://www.geneseds.org 1426/1
SS abri	1674552 1674552 C	т		7 35.0424 askalite - 0	6 47 xx7487 Szax SH C T	1 1	HADK HH_825			12 151 157 5-12-1	25 P P TRUE	1.74113 -1.4043 N/A N/A N/A		I COMPARENT AND AN AN AN AN AN AN ANY AND AN ANY ANY ANY ANY ANY ANY ANY ANY ANY
57 abri	1675388 1675388 G	T	-	24 55.2592 autolija II	18 28 re4724 Szer SH G		HADK HH_121	27255 1672551 4535765 -	Z H EXON_F	1 51 155 7-11-5	L H K PALSE	8.22855 8.55257 TOLED 8.25 5.25	14 141 I I I I	E GO:1813 (HAD av blig://a blig://aua.anbi. 65228
SI aket	1837363 1837363 G	ŝ		12 S2.0202 askalila 0	11 11 Fall Same SH G		CALMUS HH_111	2461 1056126 0555 -	E E INTRON		TRUSE INC.	I -LEASE Hall HA HA		bilg://a.bilg://aca.asbi. 165688 bilg://aca.granuarda.org 333456
El eler 1	1833648 1833648 G	Â		17 7351 A			11 22 471	17-12 10 13 1511	s atom	E 17 (D. 14)		1.574 TOL 1 6 6		kilp://www.graraarda.org 333456
51 abr1	1843363 1843363 A	G		25 6429 25	15 va21 Sza 1 A							1.101 0711 H7A H7 H.	om th	
62 etc-1	1876875 1876875 A	e .		54 1425 mil	EL Sam I A			51 42 12 5455 -				1 112		hillp://www.grarasrda.org 15452 hillp://www.grarasrda.org 15452
54 skr1	1171153 1171153 C	Å		5 27.2727 askalila	E 21 resizes surrain G A		KIAA1751 NH III	58525 1874612 1925156 -	11 17 EXON F	17 155 165 Ex17-1	21 G C PALSE	1.33213 2.35452 DAMAG	151 114 2.23	Mile://www.erereards.org 15452
ES abri	1946897 1946897 T	č		E HER and all a		i i	GADRD HH_III	11425 1948528 1952852 -	S SEXON_F	4 84 224 Ea4-P	H G G TRUE	1.55245 1.24267 N/A N/A N/A	1 1 1 2.30 1 1	CO:1117  squapli klip://w klip://w  Graves 2565
55 abr1	2998868 2998868 T	ç		24 HII askalila II	10 25 rs24545 Szar SH T C		PEXIS PEXIS HH_SS	7778 2325484 2333878 -	E E EXOH_F	5 510 07 Ev3-0		I -7.752 H/A H/A H/A	1 1 1.62 1 1 1	GO:1116 pealein klip://w klip://w Zellweys 5192
E7 akr1	2998872 2998872 G	A A		13 SE.S217 askalila II	10 25 re11450 SamrSN G A								1 1.62 1 1.62	E Gostilitä jaratein kilja://a kilja://a Zellarga – 5192 E Gostilitä jaranga kilja://a kilja://aaa.eeki, – 55223
63	2442425 2442425 T	ĉ		et statut and a			AND NH 11	10161 2423035 2447035 -	15 15 EXOLU	1 11 21 <u>5 6</u> 41-1	IK K TRUE	1.3 3 -1 17 H/A H/A H/A		GO:   15  ang: kl p://w kl p://www.enki. 55223
28 abr1	2518162 2518162 T	¢	45	41-20.000 minibit 1	I 23 PRESERTS Same SH T C	wild		5117				I - C Halanar HA HA		http://www.granarda.org 127281
74 ek-1	2518368 2518368 G	A		8 55.557 autorite 8.88564	E 27 PRE7812 Same SH G A		1a.el 10 52	(ES) II			R1 PALS	I L H Halamar HA HA		60:1115 jperiod bills://www.generards.org 127281 50:1115 jperiod bills://www.generards.org 73258
73	2525475 2525475 C	A		1 72.7273 minibility 1.111113	I 13 PHENES SAME SHIC A		HEA 133				21		5.23	Gottini protest kilet/fuuu.geerearde.org 7325
74 abr1	2551125 2551125 A	G		23 (1.333) (addition 1	23 28 ve18737 Szev SH A G		HHELI (133	124011 2511041 2554041 -	24 25 EXON_F	5 161 162 245-0	A A TRUE	1.31551 -1.5531 N/A N/A N/A	1 1 1 5.23 1 1	1 CO:1115 protest http://www.granarda.org 79251
75	2328843 2328843 G	A		48 44.5552 ++4+654 8	1 17 +=37352 Sum+SH G A	1.1	ACTRT2 MH_III	1422 2327386 2323327 -	1 1 EXON_F				55 1514 1.64 1 1	l bily://a bily://aaa.eeki. 148625   Go:1145  ee:ilien bily://a bily://aaa.eeki. 69976
75	1231581 1231581 C	č		20 SULT235 extention II									113 331 1.67	Gotteld parties bly://a bly://aaa.eski. 53376   Gotteld parties bly://a bly://aaa.eski. 53376
28 abr1	3373587 3373587 C	Ť	- 11	d 40.55566 1	25 19 Pazielise SzarrSH C T	i i	ARNGETIS HH_111			7 16 151 547-6	I H Y PALSE	1.14252 2.15586 TOLER# 1.4 1.47	11 221 1 1 1	1 GO:1135 (regulat bilg://www.generarda.org 27257
75 ek-1	3384346 3384346 T	ç		01 97.591 askellar 1		1.1.1				11 11 152 5-11-	IIS S TRUE	I SUIL N/A N/A N/A		GO:1135 projekt http://www.granaarda.org 27257
II shell	3384833 3384833 G	н 6		55 100 askalila 0 0334	41 21 Fe10737 SteerSH G R								445 446	E GO:1135 (regelat bilg://www.genesards.org 27257 E GO:1111 (kistori: bilg://w bilg://www.eebi, 1355
12 abri	1411517 1411517 T	ĉ		5 THE ALLONG D		110	HEGE HELE	010101000000000000000000000000000000000	d be dis	5 1515 1511 1V55-5	Sec. 198	I -I.SISS Nalanay NA NA		CONTRACTOR (Contractor Magel/Contractor March 1959
13 abr1	3537336 3537336 T			51-50.000 astallar 1		: : (A/)		ne cour	a peais	sease (	COUSIAC	I -1.7117 H/A H/A H/A		kily://a kily://aaa.eeki. 49856
II shell	9598692 9598692 G	Ť		S2 45.0704 askalila		· · · · · ·	TP23 TP23 HH H4	12348 2337138 2338211 43242 3332816 3648322 -	12 12 EXOLUS	1 10 10 10 10 10		1.50250 -0.743 TOLERF 1.22 3.12	11 121	61155
15	3634575 3634575 C	Ť		3 33.3333 askalila I	I 15 Part2146 Same SH C T	i i				7 23 13 547-23	S A A TRUE	I 1.21515 N/A N/A N/A		11.1 COLUMN Incloane Milpillia Milpillia Phenesk - 2464
17 abr1	3634614 3634614 T	5		45 23.4448		1.1				7 12 13 5-7-2	S H H TRUE	1.12111 -1.7521 N/A N/A N/A		11.1 GO:1011 (velvane klip://w klip://w PHenenk 2161
11 111	3633263 3633263 G	R G		42 bit.254 substitut							SEP P TRUE	8.78873 -8.7254 N/A N/A N/A 8 -4.7427 N/A N/A N/A	1.7	E GOSTIEL Josef Same Milys/Vac Milys/Vac PHenrok 2461 Milys/Vacan-generatedu.org 140021
10 .4+1	SESSIES SESSIES C	Ğ		11 11.5552	12 15 re1818 Szer SH C G	i i				1 10 10 54-0	IST Q E FALSE	1 -1.1717 TOLERF 1.1 4.12	23 13 1 1.23	http://www.grana.edu.org_10021
31 abr1	3668513 3668516 CT	таті сттеті				11 Caprocks 1713312, 113			12 12 INTRON	1 1212 1211 1751-2		E E.5335 Halanar HA HA	1 1 1 1.21 1 1	http://www.graracela.org_10020
12 akr1	5724565 5724565 T	c a	40	C HE saladita E							ALSE FALSE	I -1.1133 Halanay HA HA		http://www.granarda.org 5751 http://www.granarda.org 5751
14	5757465 5757465 T	ĉ	51	12 (2.3333	18 28 re64197 Same SH T C							I -2.5073 Halanar HA HA		http://www.granarda.org 3/31
85 abri	5742875 5742875 G	c		47 55.425	19 25 +=12746 Szar SH G C	1 I I I I I I I I I I I I I I I I I I I			22 24 INTRON			I -1.2283 Halanay HA HA		http://www.genenards.org 5751
35	9745595 9745595 T	2	51	12 38 autorida - 8	19 22	: :		45153 3718585 3763657 · 28143 3763785 3731853 ·	22 21 EXON_F 7 7 INTRON	1 148 155 E.8-7 4 1521 1525 1954-2		1 8.72343 H/A H/A H/A 8 8.42858 Helener HA HA		l
10 11-1	9776886 9776886 C	e e	14	11 78.5714	1 25 125161 5 SH C G	: :	D770 HH_884	28143 3763785 3791853 -	7 7 INTRON 7 7 EXON F	4 1521 1525 IV54-2 5 77 171 E-5-7		8 8.42858 Halasar HA HA 8.89245 8.85441 TOLER6 1 5.15	121 111	l
33	5758182 5758182 A	Ğ	55	14 48 autolila - B	14 28 S2858 Sum- SH A G	i i	D77D HH_884	28145 5765785 5751855 -	7 7 EXOH_F	7 172 255 E.7-6	2 P P TRUE	1 -1.1122 H/A H/A H/A		B. GO:BBBE Japapian kilp://u. kilp://uuu.anki. 1677
111 .4.1	8787458 8787458 G	5	28	21 111	15 15 42748 Sum-SH G C	1 1	C14+6174 HH_287	11155 9795569 9886717 ·	4 4 EXON_F	5 461 485 E.S-2	7 T R PALSE	1.11717 1.14351 DAMAG 1 4.32	51 127 1 1 1	1 bilp://www.graraarda.org 333441
181 44-1	4672577 4672577 G	A G	41	40 97.564 aukalila - 0	28 28 -=24285 5 SH G A 24 19 -=22238 5 SH A G	: :	AJAP1 AJAP1 NM_818 Kchad; Kchad; NM 172		6 5 EXON_F 15 14 EXON_F	2 751 111 E.2.4 15 11 31 E.15-		8.75228 2.77558 TOLER& 8.25 5.45 8.25772 -1.5254 H/A H/A H/A	56 1584	l
183	6281957 6281957 G	č	13	11 57.8947 aukalila 8	7 21 ral4611 Samr SH G C		RHF207 HH_207	15171 6488776 6289946	18 17 EXON_F	10 75 172 E.10-		1 1.75177	68 265 1 1	l lip://www.granarda.org 30551
104	6447275 6447275 T	ę	55	15 57.5758 I	15 2851786 Sam-SH T C		THPRSP THPRSP HH_881	2005 6446750 644042 -	S SEXON_F	4 77 168 E.4-7	7 a a TRUE	I -2.5372 H/A H/A H/A	1 1 1 40 1 1	8 GO:8887 forll our bilg://a bilg://aua.oubi. 8748
185 abr1	6451198 6451198 G 6582194 6582194 C	A T	17	9 17.6471	2 18 28 22 re15568 Suer SH C T	1 Barleia 85285_a	I PLEKNC PLEKNC HM_191 PLEKNGS HM_191	59948 6448799 6582656 · 59948 6448799 6582656 ·	22 22 EXOH_F	24 667 762 E.244	34 S F PALSE PALSE	1 2.38979 DAMAG 8 4.92 8 -8.3479 Halamar HA HA	37 331 4.03	l
187	ESERES7 ESERES7 C	÷		14 400 establis 0	27 25 ra75488 Saur SH C T		HOLS HH_824	53310 6440733 6582656 ·	12 12 SA_SITE	12 5 18 IV512-	S FALSE	8.81575 -1.5454 Halanar HA HA		B GOLDES (regulat bilg://w.bilg://w.spisaliw_S7445
111 .4-1	6557848 6557848 A	G		40 400 autolita - 0	25 28 ra1864 Suar SH A G	i i	TASIRI TASIRI HH_177	24984 6598824 6562484 -	4 4 EXON_F	5 541 752 E.S.2	ZEK E PALSE	8 8.48546 TOLER# 8.55 2.55	56 912 1 2.95 1	E GO:EEF7 [Geprate Allp://w Allp://www.anki. EEF55
183	6628977 6628977 C	Ţ	25	42 41.3735	11 18		DHAJCHI HH_BH		15 15 EXON_F	12 35 78 E.12-1	33 D H PALSE	1.37631 2.6886 TOLER6 1.35 5.83	25 557	bilg://www.graraurda.org 55795
111 -1-1	6627222 6627222 A	e c	54	52 55.2365 aukalila 8	51 2420045 55H A G 17 1020045 55H G C	1 1		67798 6646846 6684559 ·	16 16 EXON_F 16 16 EXON F	10 10 117 E.10- 1 25 131 E.1-2		1.55781 -1.7785 H/A H/A H/A 1 1.51283 TOLER6 1.77 5.88	24 14	http://www.granearda.org 55735 http://www.granearda.org 55735
1121	7646175 7646175 C	Ť	25	14 53.8462 aubalila 8	18 22 17375 Sam- SH C T		CANTA1 HH_815	384381 6767371 7752351 -	25 25 EXOH_F	1 175 1847 E.S.4	75 H H TRUE	1.35863 -1.5428 H/A H/A H/A		8 GO:8845 (regulat http://w.http://www.eehi. 29261
115	ZB46544 ZB46544 G	A		97 98.9796 eebelile B	48 13 re12128 Same SH G A			384384 6767374 7752354 ·	23 23 EXON_F	3 545 1847 E.S.S		8.33781 -8.5361 H/A H/A H/A	1 1 1.31 1 1 1	60:0045 (regulat bilg://a bilg://aaa.oobi. 29264
114 445	7714558 7714558 C	T C	28	9 45 ocholilo 8 79 58.9677 ocholilo 8	7 27 ra23814 Suar SH C T 54 22 ra28713 Suar SH A G	: :		384384 6767374 7752354 - 384384 6767374 7752354 -	25 25 INTRON 25 25 INTRON	11 54618 54686 19511 13 33 362 19513		S -1.8332 Hul anne HA HA S -3.5612 Hul anne HA HA	5.65	
447		2			1 14 1107 7 PM T C						AL P TOUR			



#### What do we know about this position in the genome?



#### Quality

Short read sequencing reliable for most applications Average sequencing coverage reliable for detection of point mutations Long read sequencing better for structural variation, repeat expansions, but high error rate for single nucleotide variation detection

#### Throughput/Speed

Short read technology allows exome and genome sequencing in days Thousands genomes can be sequenced on individual systems annually Long read genomes still take more time and have less throughput

#### Costs

Short read technology: 100x coverage exome < €300, 30x coverage genome < €1.200. Long-read genome 30x coverage ~ €10.000 Prices genomes below €1.000 in 1-2 year, below €200 in 5 years?



#### Genome sequencing: All variation in one experiment!

#### Important:

**DNA** from

blood/saliva

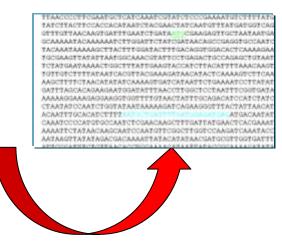
- Accuracy
- Completeness

(all genome, all variations)

- Speed
- Price



# Genome with 'all' variation





# Why perform whole genome sequencing?

If you consider genetics may play a role in your patient: Why not read the entire book? Why settle for studying what we now know? **We still live in the dark ages of genetics!** 

Key advantages of genome sequencing:CompletenessAll variationSimplicityOne test

Price is dropping, quality will continue to improve, no enrichment, better for structural variation



#### Genome sequencing centers established around the world



Transformative Genomics: England Begins Daunting Task of Sequencing 100,000 Genomes by 2017







"Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier."

- President Barack Obama, State of the Union Address, January 20, 2015



### The UK 100,000 Genomes Project (and beyond)









100,000 genomes

70,000 patients and family members

**21** Petabytes of data. 1 Petabyte of music would take 2,000 years to play on an MP3 player.

 13 Genomic Medicine Centres, and
 85 NHS Trusts within them are involved in recruiting participants

**1,500** NHS staff (doctors, nurses, pathologists, laboratory staff, genetic counsellors)

2,500 researchers and trainees from around the world



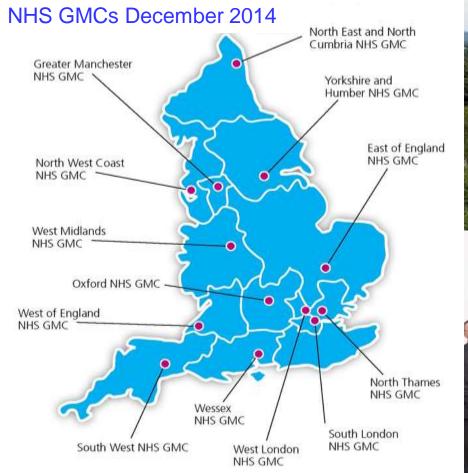


#### Illumina Partnership July 2014





#### Data Centre November 2014







# What are we telling participants?

- Information about a patient's main condition
- Information about additional 'serious and actionable' conditions (optional)
- Carrier status for non affected parents of children with rare disease (optional)



Main findings All participants agree to receive results about the main condition for which they were referred

Additional findings Participants can opt in to receive feedback on a selection of known genetic alterations of high

clinical significance

Carrier status Eligible adults can opt in to find out their carrier status for certain genetic diseases

Image courtesy of Health Education England



### A new national diagnostic service

- Increasing move towards exome and genome sequencing; further use of patient/parent trios
- Genomics England will provide data infrastructure for a new Genomics Laboratory Service for the NHS in England
- Central test request system with shared 'genomic test directory'
- Central WGS pipeline (lab and bioinformatics). Reporting by hubs
- Central shared genomic knowledgebase for NHS laboratories
- 7 genomics laboratory hubs for performing other/additional genetic tests, interpretation and reporting









#### Newcastle Fertility Centre @ Life



#### Genetic Medicine



**Northern Genetics Service** 





