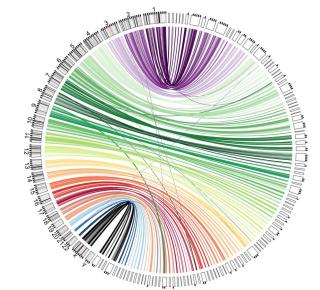


Storskalig analys av strukturella varianter

En överblick över ny teknik och stora genetiska varianter

Jesper Eisfeldt Bioinformatiker

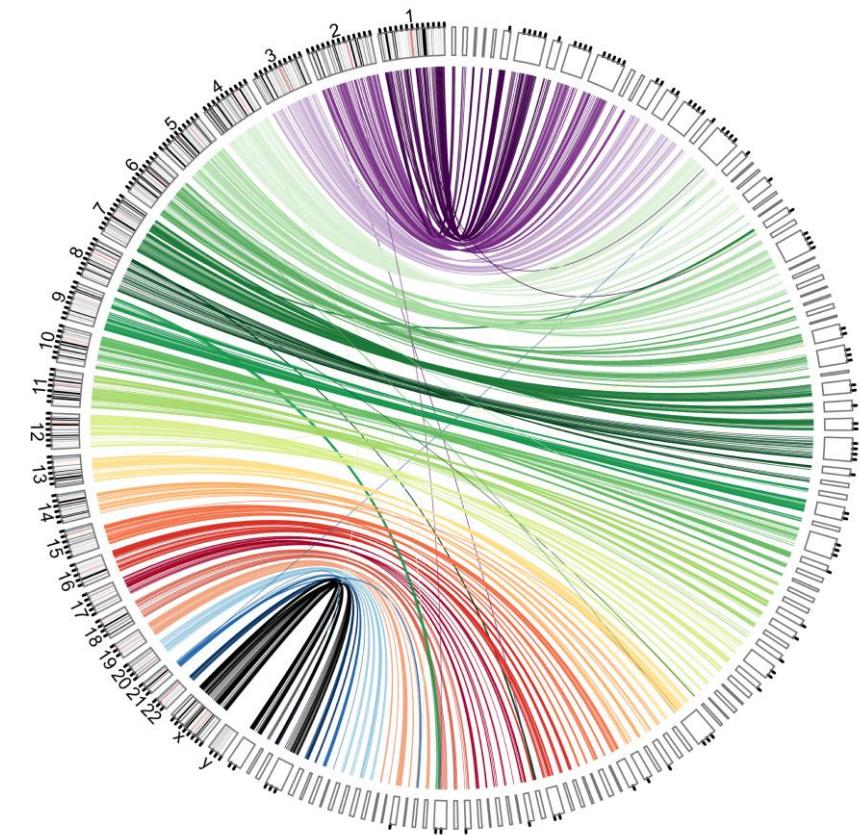


Karolinska
Institutet

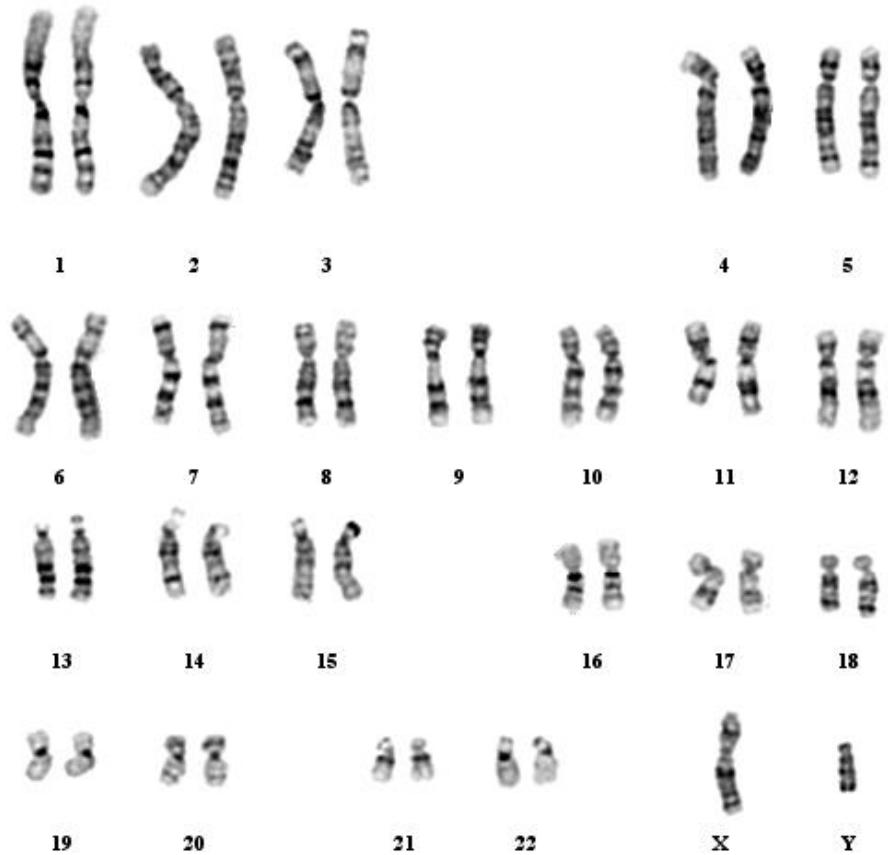
 KAROLINSKA
UNIVERSITETSSJUKHUSET

Storskalig analys av strukturella varianter

- Introduktion
 - Strukturella varianter
 - Metoder
 - Illumina
 - Nanopore
- Projekt
- Sammanfattning



Genetisk variation

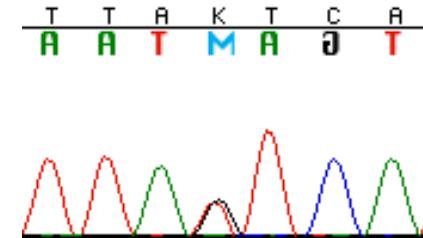


0. Mänskliga genomet

- Baser A,T,C,G
- Ca 3 (*2) miljarder baspar

1. Basparsutbyte (SNVs)

~ 5 M



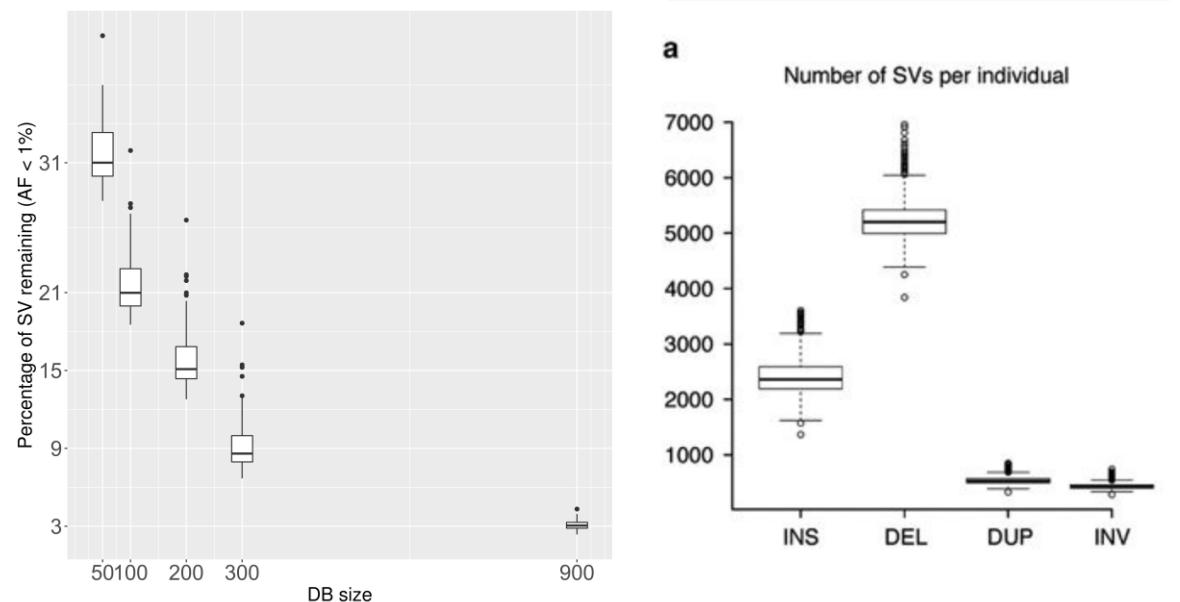
2. Strukturella varianter (> 50 baspar)

- Duplikation
- Deletion
- Translokationer
- Inversioner



Strukturella varianter

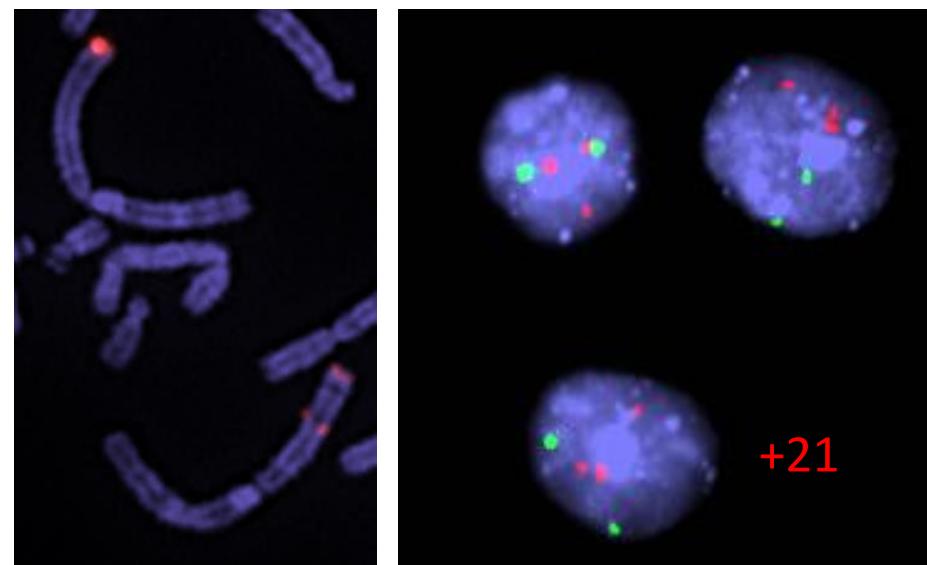
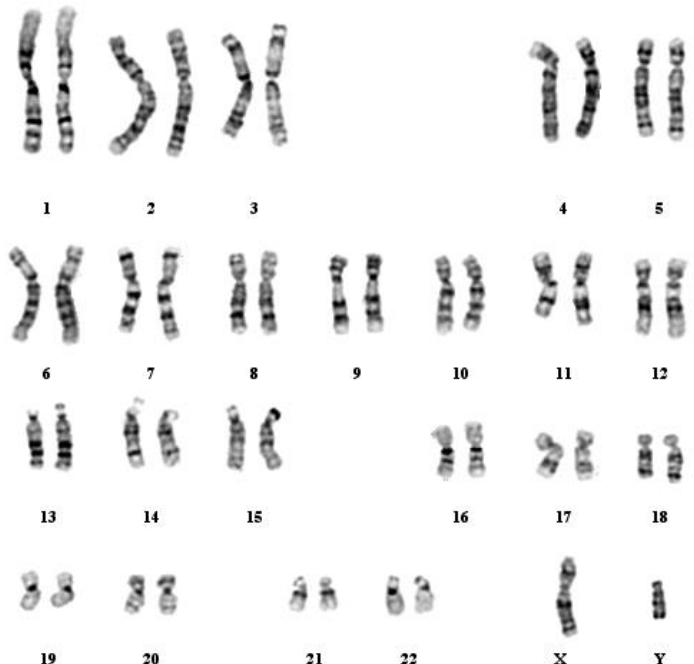
- Ca 5000 -10 000 per individ
- Vissa är sjukdomsorsakande
 - Ovanliga varianter
- Databaser
 - SweGen



Ameur et al 2017

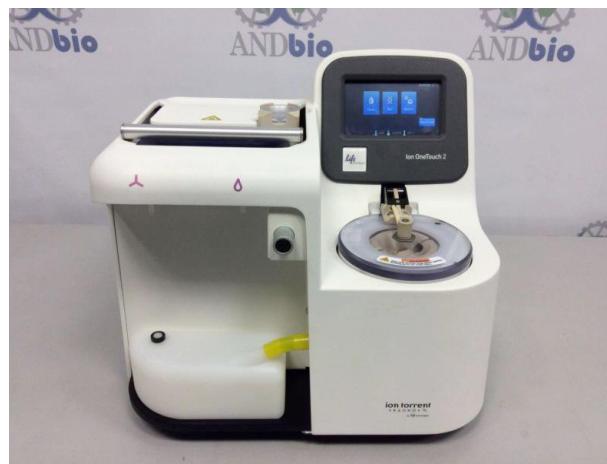
Metoder

- Vanliga metoder
 - Mikroskop (karyotyp)
 - Stora strukturella varianter
 - FISH
 - Förbestämda regioner
 - Och många andra!
- DNA Sekvensering
 - Relativt nytt
 - Att läsa DNA
 - Hela genomet



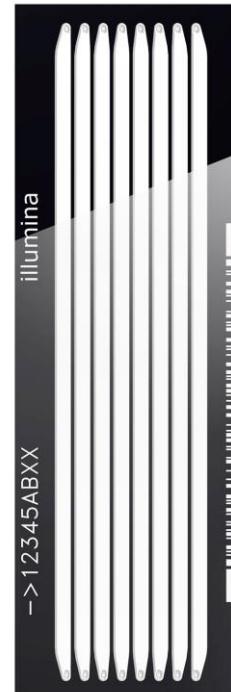
Metoder

- Sekvensering
 - Att läsa DNA
 - Ofta från blodprov
- Många metoder
- För och nackdelar
 - Kostnad
 - Tid
 - Kvalitet
 - Längd på läsningarna



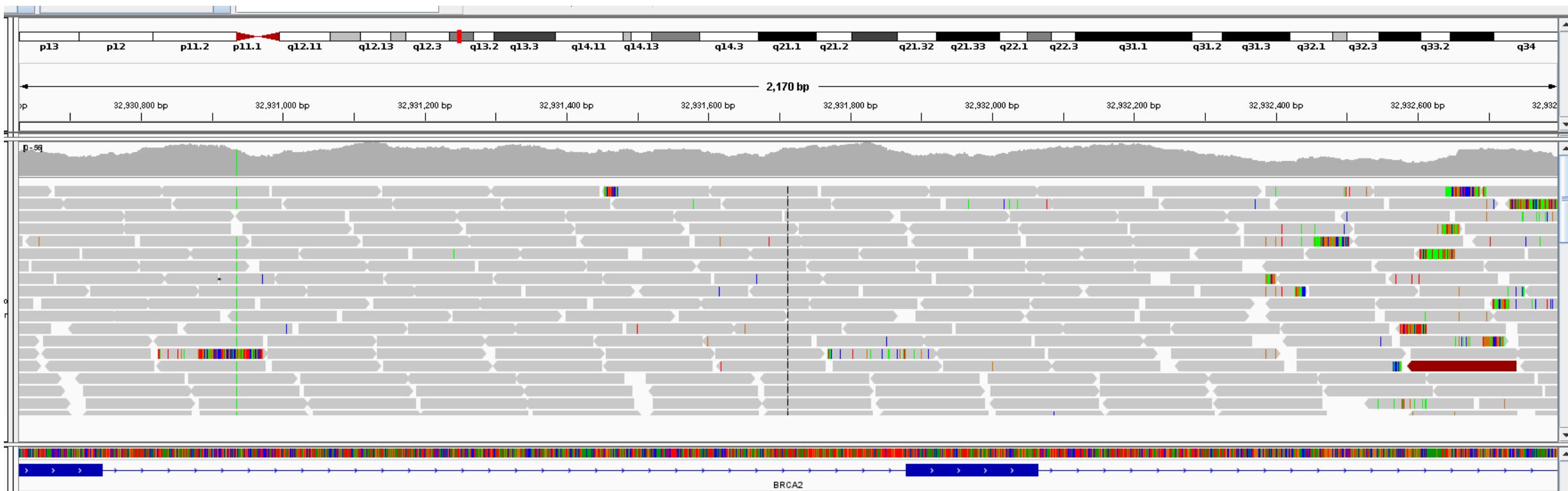
Illumina helgenomsekvensering

- Ledande tekniken
 - I klinik och forskning
- Korta läsningar
 - 150 baspar
- Billigt
 - 10 000 -20 000 kronor
 - 400 miljoner parade läsningar
- Hög kvalitet
 - Mindre än 1 fel per 100 baser



Illumina helgenomsekvensering

- Läs samma genom flera gånger
 - Ofta 30 gånger
 - analysera varianter
 - Hantera fel

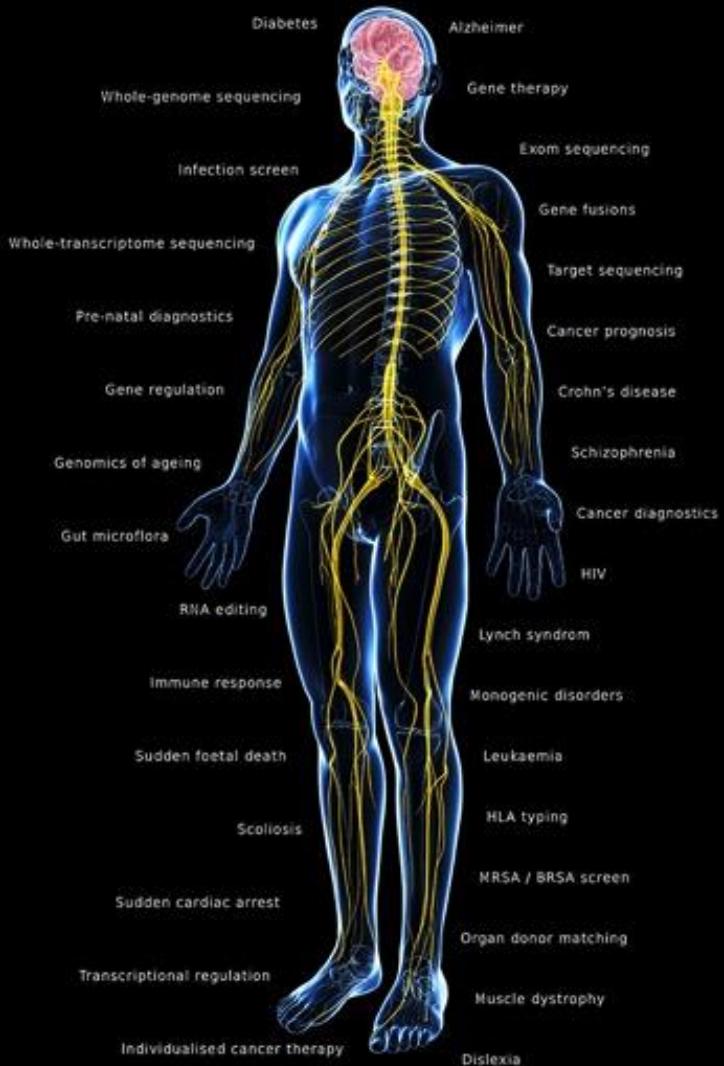


Illumina helgenomsekvensering

- National genomics infrastructure
- Clinical genomics



What we sequenced at SciLifeLab



16S amplicon, *Acinetobacter baumannii*, *Acris lona*, *Acridotheres javanicus*, *Actinobacillus succinogenes*, African swine fever virus, *Agaricomycotina* sp., *Alces alces*, *Alligator mississippiensis*, *Amphibia* *fallax*, *Apis mellifera*, *Apollonia chrysaeus*, *Arabidopsis thaliana*, *Arabis alpina*, *Anaerobizomycetes* Finlay, *Antocephalus gaezetta*, artificial sequences, *Arvicola amphibius*, *Ascaridia galli*, *Aspergillus oryzae*, *Astragalus stephaniae*, Atlantic herring, Atlantic salmon, *Avena sativa*, *Baccharis breviflora*, *Baccharis dracunculifolia*, *Bacteriophages*, *Balaenoptera musculus*, *Balaenoptera physalus*, *Balanus improvisus*, Baltic Sea microorganisms, *Bathymorus* sp., *Bifidobacterium* sp., ***Borrelia burgdorferi***, *Borrelia garinii*, *Bos taurus*, *Bovine viral diarrhoea virus*, *Brachypodium suaveolens*, *Brevisia* sp., *Brettanomyces naardenensis*, *Campbellia elegans*, *Callobiochanus maculatus*, *Candida parapsilosis*, *Candidatus Neopeltischia mikurensis*, *Canis lupus*, *Capreolus capreolus*, *Capella bursa-pastoris*, *Capella grandiflora*, *Capilla orientalis*, *Capilla rubella*, *Ceanothus thyrsiflorus*, *Cervus dama*, *Cervus elaphus*, *Chidia submaculatum*, *Closostachys rosea*, *Clostidium viburnense*, *Coleodictyon anticlavis*, *Colas croceus*, *Collomia heterophylla*, *Coregonus lavaretus*, *Coronavirus*, *Corvus corone*, *Corvus monedula*, *Crassostrea gigas*, *Cryptococcus cinnereus*, *Cryptococcus neoformans*, *Cystomegalovirus*, *Danio rerio*, *Datiscia glomerata*, *Deformed wing virus*, *Dekkera bruxellensis*, *Dicerorhinus sumatrensis*, *Dictyostelium discoideum*, *Diplotheleus gymnosphaericus*, *Diplostomopsis longitubus*, *Drosophila melanogaster*, *Enterobacteriaceae*, *Enterobacter cloacae*, *Enterobacter faecium*, ***Equis caballus***, *Escherichia coli*, *Eurytomostomum macrobaramum*, *Euphorbia lathyris*, *Euphorbia peplus*, *Euplectes afer*, *Euplectes ardens*, *Euplectes auroreus*, *Euplectes macrourus*, *Euplectes orix*, ***Felis catus***, *Ficedula albicollis*, *Ficedula hypoleuca*, ***Fragaria ananassa***, Freshwater microbial communities, *Fucus radicans*, *Fucus vesiculosus*, *Fumaria sp.*, *Geum gallicum*, *Geopsis magnirostris*, *Gordonia murri*, *Globoderma rostafriesii*, *Gnetum gnemon*, *Gnetum luwuense*, *Gnetum montanum*, *Gnetum pendulum*, *Gonyostoma semen*, *Gonzalaguna*, *Hamella maranha*, *Heterobasidion annosum*, ***Hippophae rhamnoides***, ***Homo sapiens***, Human immunodeficiency virus, *Huperzia selago*, *Hymenoscyphus albidus*, *Hymenoscyphus pseudolobulatus*, *Idotea baltica*, ***Influenza A virus***, *Klebsiella pneumoniae*, *Laccaria bicolor*, *Laccobacillus*, *Lepidium campestre*, ***Lepidium sinapigineum***, *Letharia vulpina*, *Littorella saxatilis*, *Lycoperdon cinnabarinum*, *Lynx lynx*, *Malassezia sympodialis*, ***Malus domestica***, *Malus sylvestris*, ***Mammuthus primigenius***, *Marchantia polymorpha*, Marine bacteria whole community, ***Meliagethes aeneus***, Metagenomes, *Methanococcus* sp., *Metschnikowia araliaeana*, *Metschnikowia hawaiiensis*, *Metschnikowia lobiformis*, *Metschnikowia pulcherrima*, *Metschnikowia pulcherrima*, *Neurospora sitophila*, *Neurospora tetrasperma*, *Nora* Virus, *Nochoroacta ornata*, *Nochoroacta perdicaria*, ***Notophthalmus viridescens***, *Nyctereutes procyonoides*, *Ogataea pini*, ***Oryctolagus cuniculus***, *Rana arvalis*, *Oryzopsis latipes*, ***Pacifastacus leniusculus***, *Paenibacillus polymyxia*, ***Panthera leo***, *Panthera pardus*, *Parasite*

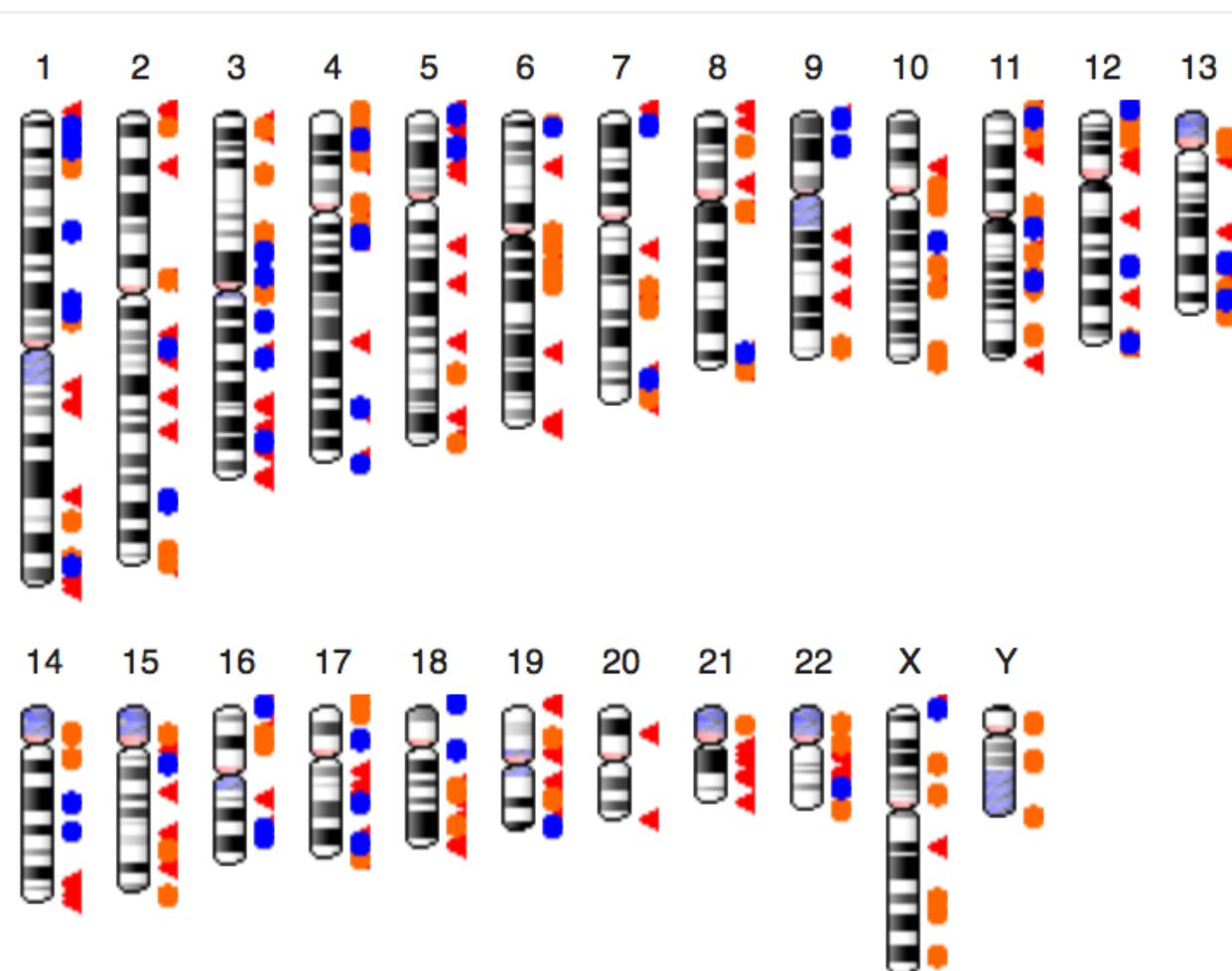
Nanopore helgenomsekvensering

- Långa läsningar
 - Längre än 10 000 baspar
- Många Storlekar
- Relativt dyrt
 - 30 000 kr
 - Mänskligt genom
 - Runt 10 miljoner läsningar
 - stora maskinen
- Sämre kvalitet
 - 10% fel av de lästa baserna



Analys

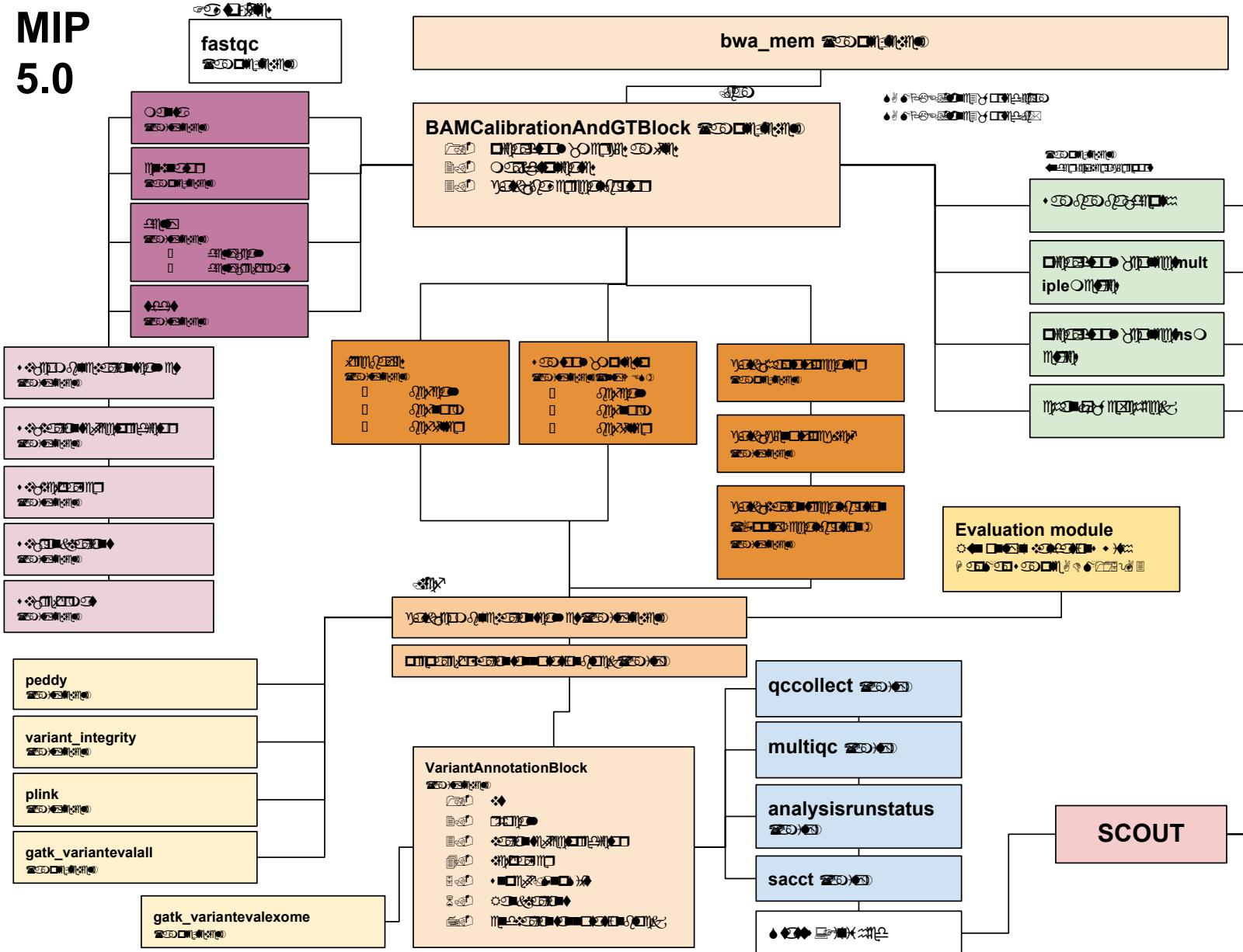
- Jämföra läsningar (sekvenser)
 - Hitta avvikeler
 - Relativt mot referensgenomet
- Referensgenom
 - En model av det mänskliga genomet
 - En stor textfil
 - 3 miljarder tecken
 - 70% från en individ
- Programvara



Ideogram of the latest human assembly, GRCh38.p13

NGS pipeline: MIP

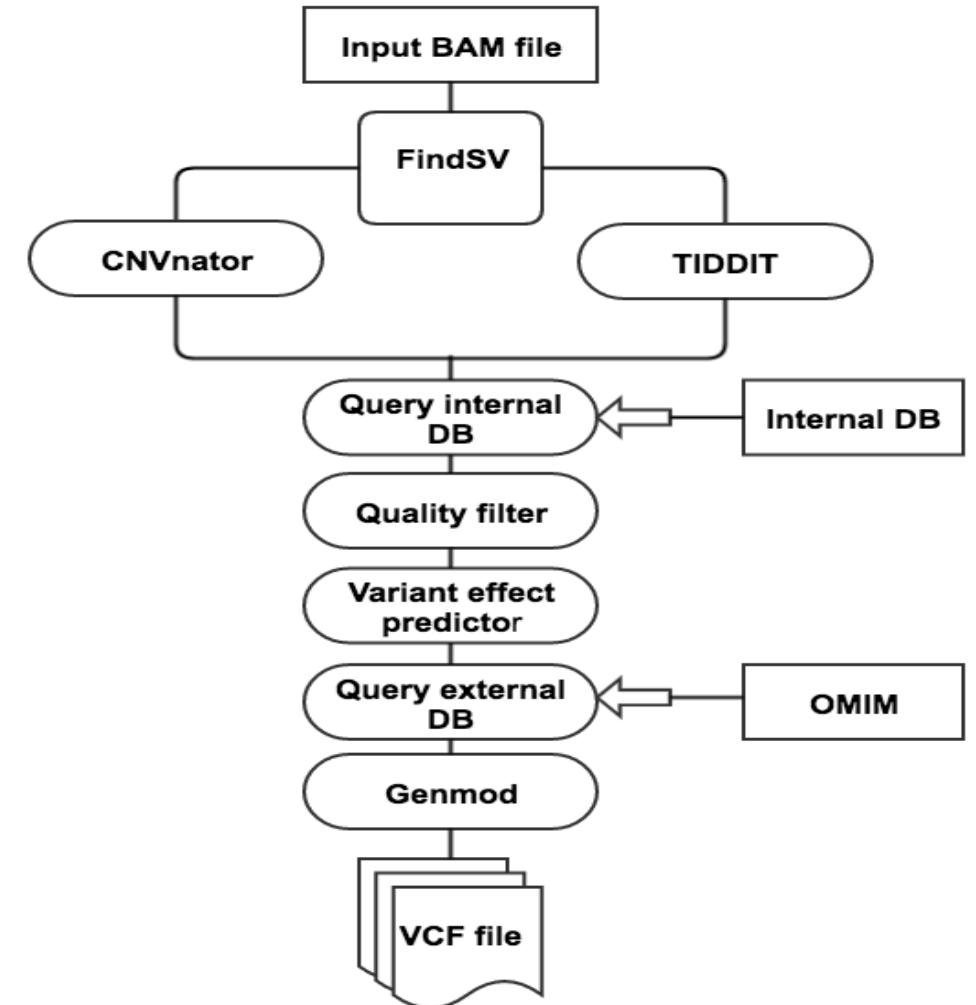
MIP
5.0



Henrik Stranneheim

Leta strukturella varianter i kliniken

- Egen pipeline
 - Hitta stora varianter
- Filtrering
 - Ta bort varianter vanliga varianter
 - Hitta vilka varianter som ligger i gener



##TIDDTcmd="/proj/sens2017130/nobackup/wharf/jesperei/jesperei-sens2017130/FindSV_johanna/TIDDT-TIDDT-2.0.0/TIDDT.py --sv --bam P2109_123.clean.dedup.recal.bam -									
##VEP=v87 cache=/home/jesperei/.vep/homo_sapiens/87_GRCh37 db=. polyphen=2.2.2 sift=sift5.2.2 COSMIC=78 HGMD-PUBLIC=20162 1000genomes=phase3 regbuild=1.0 ClinVar=201									
##svdbcmdline=/home/jesperei/anaconda2/bin/svdb --merge --overlap 0.9 --vcf P2109_123.clean.dedup.recal_FindSV.vcf									
#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	P2109_123
1	71741295		SV_342_1	N	<DUP>	.	PASS	SVTYPE=DUP;CIPOS=-95,0;CIEND=0,2;END=71741629;SVLEN=335;COVM=75.912;COVA=31.22;COVB=13.84;LFA=	
1	86005620		SV_374_1	N	N[1:146486084[.	PASS	SVTYPE=BND;CIPOS=-244,0;CIEND=0,213;COVM=25.5716494077;COVA=71.9725;COVB=86.03;LFA=42	
1	111957460		SV_419_1	N		.	PASS	SVTYPE=DEL;CIPOS=-47,0;CIEND=0,8;END=111957564;SVLEN=105;COVM=23.76;COVA=29.66;COVB=17.86;LFA=	
1	145382407		SV_471_1	N	N[1:148241134[.	PASS	SVTYPE=BND;CIPOS=-142,0;CIEND=0,129;COVM=28.5746383098;COVA=113.616666667;COVB=56.15;	
2	212161492		SV_3876_1	N		.	PASS	SVTYPE=DEL;CIPOS=-37,0;CIEND=0,8;END=212161600;SVLEN=109;COVM=20.4333333333;COVA=22.27;COVB=2	
3	75276849		SV_8058_1	N	N[3:75278911[.	PASS	SVTYPE=BND;CIPOS=-355,0;CIEND=0,238;COVM=37.16;COVA=55.9;COVB=62.91;LFA=37;LFB=40;LTE=	
3	75569001	CNVnator_del_156		N		.	PASS	END=75579000;SVTYPE=DEL;SVLEN=-10000;IMPRECISE;natorRD=0.548889;natorP1=0.00492305;nato	
3	103164041	SV_8101_1	N		.	PASS	SVTYPE=DEL;CIPOS=-64,0;CIEND=0,1;END=103164407;SVLEN=367;COVM=22.478;COVA=26.14;COVB=26.14;LFA=		
6	57284882	SV_11122_1	N	N[6:57289357[.	PASS	SVTYPE=BND;CIPOS=-290,0;CIEND=0,232;COVM=34.4747826087;COVA=65.1275;COVB=72.05;LFA=64		
6	57423046	SV_10870_1	N	N[6:57429121[.	PASS	SVTYPE=BND;CIPOS=-459,0;CIEND=0,204;COVM=36.1158064516;COVA=97.6933333333;COVB=84.896		
8	84948937	SV_12650_1	N		.	PASS	SVTYPE=DEL;CIPOS=-68,0;CIEND=0,1;END=84949600;SVLEN=664;COVM=18.68125;COVA=32.42;COVB=32.03;L		
9	41346001	CNVnator_del_348		N		.	PASS	END=41855000;SVTYPE=DEL;SVLEN=-509000;IMPRECISE;natorRD=0.42689;natorP1=3.13109e-13;nato	
9	136379164	SV_13046_1	N	<DUP>	.	PASS	SVTYPE=DUP;CIPOS=-72,0;CIEND=0,1;END=136379553;SVLEN=390;COVM=130.03;COVA=33.695;COVB=24.05;L		
10	49865373	SV_14703_1	N		.	PASS	SVTYPE=DEL;CIPOS=-10,0;CIEND=0,0;END=49865532;SVLEN=160;COVM=21.43;COVA=19.75;COVB=18.19;LFA=		
10	59015184	SV_14719_1	N	<DUP>	.	PASS	SVTYPE=DUP;CIPOS=-51,0;CIEND=0,1;END=59015560;SVLEN=377;COVM=68.332;COVA=19.12;COVB=31.88;LFA=		
11	48367480	SV_18771_1	N	N[11:48373693[.	PASS	PASS SVTYPE=BND;CIPOS=-427,0;CIEND=0,100;COVM=37.0284126984;COVA=59.018;COVB=62.71;LFA=40;		
11	114433283	SV_18901_1	N	N[11:131230467[.	PASS	PASS SVTYPE=BND;CIPOS=-362,0;CIEND=0,246;COVM=34.6133650646;COVA=108.37;COVB=84.8025;LFA=9		
12	49708036	SV_24436_1	N]19:272694]N	.	PASS	PASS SVTYPE=BND;CIPOS=0,167;CIEND=-290,0;COVA=47.9266666667;COVB=22.1766666667;LFA=23;LFB=		
13	113990998	SV_25394_1	N		.	PASS	SVTYPE=DEL;CIPOS=-41,0;CIEND=0,1;END=113991106;SVLEN=109;COVM=25.04;COVA=33.39;COVB=24.21;LFA=		
14	107103001	CNVnator_dup_534		N	<DUP>	.	PASS	END=107188000;SVTYPE=DUP;SVLEN=85000;IMPRECISE;natorRD=1.52706;natorP1=0.0416581;nato	
15	21128001	CNVnator_del_543		N		.	PASS	END=21885000;SVTYPE=DEL;SVLEN=-757000;IMPRECISE;natorRD=0.311423;natorP1=2.10532e-13;nato	
15	23496001	CNVnator_del_551		N		.	PASS	END=23565000;SVTYPE=DEL;SVLEN=-69000;IMPRECISE;natorRD=0.160259;natorP1=2.30975e-12;nato	
15	61963001	CNVnator_dup_554		N	<DUP>	.	PASS	END=61989000;SVTYPE=DUP;SVLEN=26000;IMPRECISE;natorRD=1.12968;natorP1=0.00043246;nato	
16	4069109	SV_28814_1	N	N[20:45929626[.	PASS	SVTYPE=BND;CIPOS=-553,0;CIEND=0,3;COVA=32.4885714286;COVB=33.9;LFA=24;LFB=14;LTE=13;E1=0;E2=4		
17	2220001	CNVnator_del_605		N		.	PASS	END=2485000;SVTYPE=DEL;SVLEN=-265000;IMPRECISE;natorRD=0.560547;natorP1=6.01406e-13;natorP2=2	
17	2220316	SV_29905_1	N	<INV>	.	PASS	SVTYPE=INV;CIPOS=-295,0;CIEND=-226,0;END=2617815;SVLEN=397500;COVM=22.3041624748;COVA=32.385;COVB=25.		
17	2618001	CNVnator_del_606		N		.	PASS	END=2648000;SVTYPE=DEL;SVLEN=-30000;IMPRECISE;natorRD=0.547125;natorP1=5.31242e-12;natorP2=1.	
17	21321241	SV_29947_1	N	N[17:21322822[.	PASS	PASS SVTYPE=BND;CIPOS=-394,0;CIEND=0,220;COVM=38.3770588235;COVA=100.97;COVB=88.8133333333		
17	21349689	SV_29955_1	N]17:25331105]N	.	PASS	PASS SVTYPE=BND;CIPOS=0,80;CIEND=-89,0;COVM=39.686481314;COVA=61.765;COVB=84.335;LFA=9;LFB=		
18	50488037	SV_30743_1	N	N[18:50489494[.	PASS	PASS SVTYPE=BND;CIPOS=-356,0;CIEND=0,129;COVM=36.4433333333;COVA=74.894;COVB=63.46;LFA=54;		
20	45048646	SV_32497_1	N		.	PASS	SVTYPE=DEL;CIPOS=-54,0;CIEND=0,1;END=45048749;SVLEN=104;COVM=25.72;COVA=35.92;COVB=21.75;LFA=		
21	11175559	SV_33613_1	N	N[22:20320513[.	PASS	PASS SVTYPE=BND;CIPOS=-340,0;CIEND=0,1;COVA=131.7325;COVB=53.36;LFA=12;LFB=14;LTE=7;E1=0;E2=		
X	61736212	SV_35752_2	N]GL000199.1:162308]N	.	PASS	PASS SVTYPE=BND;CIPOS=-30,0;CIEND=0,10;COVA=1227.405;COVB=1064.11;LFA=146;LFB=445;		
X	61736216	SV_35749_2	N]GL000199.1:59764]N	.	PASS	PASS SVTYPE=BND;CIPOS=-108,0;CIEND=0,6;COVA=1722.725;COVB=1064.11;LFA=154;LFB=445;		
X	61736222	SV_35463_2	N	[GL000226.1:98]N	.	PASS	PASS SVTYPE=BND;CIPOS=-54,0;CIEND=-6,0;COVA=1967.85;COVB=1064.11;LFA=498;LFB=445;L		
X	79238180	SV_34488_1	N		.	PASS	SVTYPE=DEL;CIPOS=-61,0;CIEND=0,1;END=79238302;SVLEN=123;COVM=11.24;COVA=18.16;COVB=10.18;LFA=		
X	84628001	CNVnator_dup_746		N	<DUP>	.	PASS	END=84663000;SVTYPE=DUP;SVLEN=35000;IMPRECISE;natorRD=1.18194;natorP1=0.00746485;nato	
Y	9956728	SV_35758_2	N	N[GL000199.1:5832[.	PASS	PASS SVTYPE=BND;CIPOS=0,265;CIEND=-7,0;COVA=158.2;COVB=509.82;LFA=34;LFB=143;LTE=6;E1=2;E2=		
MT	1	SV_35461_1	N]MT:16541]N	.	PASS	SVTYPE=BND;CIPOS=0,336;CIEND=-489,0;COVM=2314.82837349;COVA=2070.2675;COVB=2259.15833333;LFA=		
GL000231.1	2828	SV_35484_1	N	N[GL000231.1:9189[.	PASS	PASS SVTYPE=BND;CIPOS=-215,0;CIEND=0,124;COVM=112.19515625;COVA=234.393333333;COVB=		

^G Get Help

^O WriteOut

^Y Exit

^R Read File

^W Write To

^Y Prev Page

^N Next Page

^K Cut Text

^U UpCut Text

Leta strukturella varianter i kliniken

Scout Institutes Clinical Genomics AMG-Ashkenazim **Clinical Structural variants** Panels: PU Henrik Stranneheim

SvFilters

Gene Panels PU SVType Nothing selected Region Annotations Nothing selected Functional Annotations Nothing selected Genetic Models Nothing selected

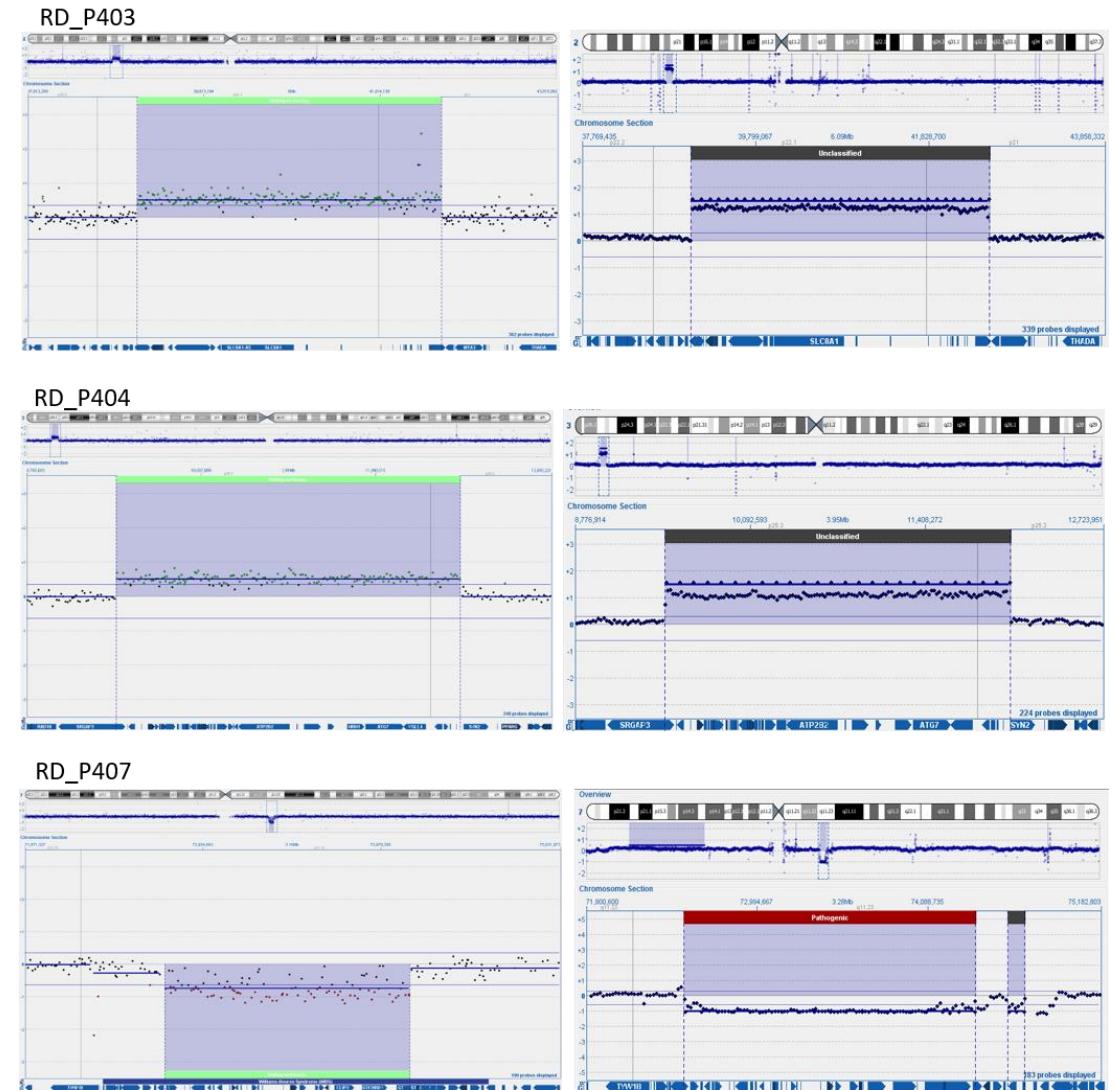
HGNC Symbols/ids (case sensitive) Length Length shorter than Decipher CLNSIG

gnomAD ClinGen NGI obs SweGen obs Chromosome Start position End position

Filter variants Clinical filter **Filter and export**

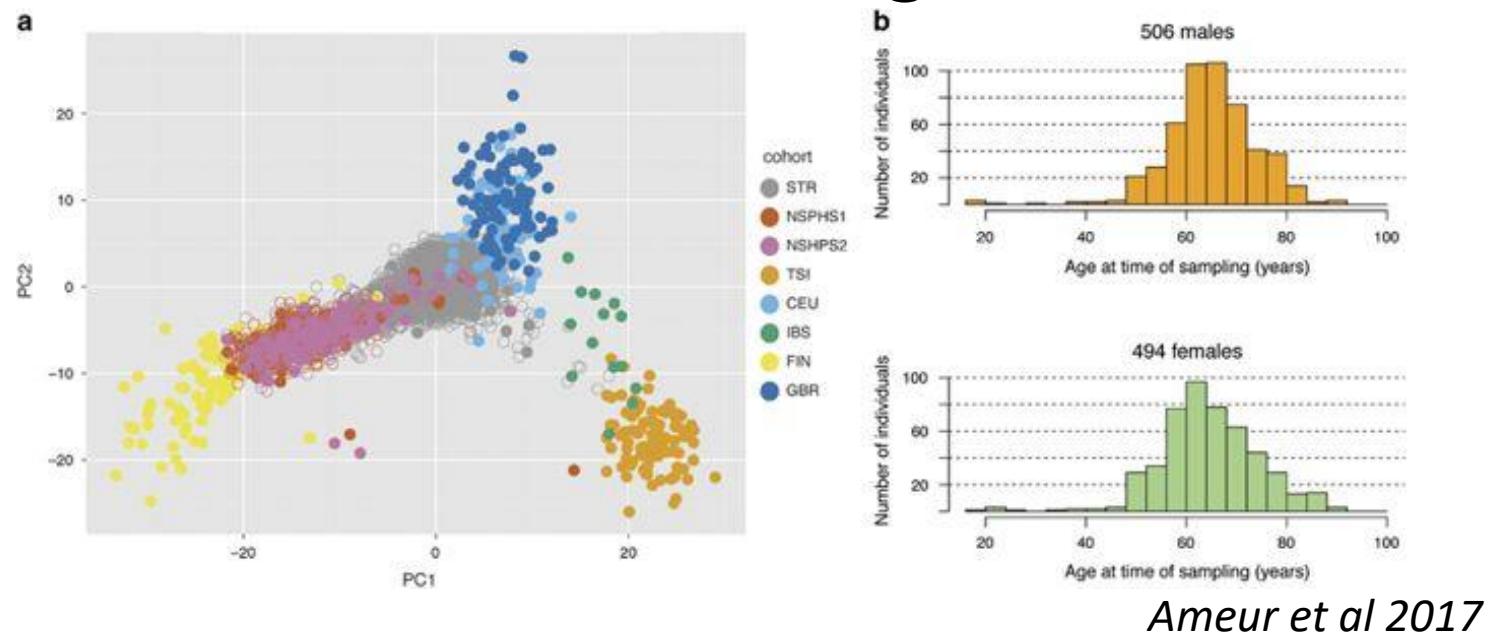
Save filter Load filter Delete filter

Rank	Score	Type	Chr	Start loc	Stop loc	Length	Region	Function	Frequency	Gene(s)
6	0	DEL	1	2193484	2193554	70	ncRNA_exonic	non_coding_transcript_variant	~	SKI
25	0	DEL	1	6858652	6858832	180	ncRNA_exonic	non_coding_transcript_variant	~	CAMTA1
26	0	DEL	1	6997498	6997616	118	intronic	intron_variant	~	CAMTA1
27	0	DEL	1	6997622	6997676	54	intronic	intron_variant	~	CAMTA1
28	0	BND	1-19	7043791	34363372	inf	intronic	intron_variant	~	CAMTA1
29	0	INS	1	7044150	7044354	204	intronic	intron_variant	~	CAMTA1
30	0	BND	1-19	7044518	32575950	inf	intronic	intron_variant	~	CAMTA1
31	0	BND	1-2	7078199	134420863	inf	intronic	intron_variant	~	CAMTA1

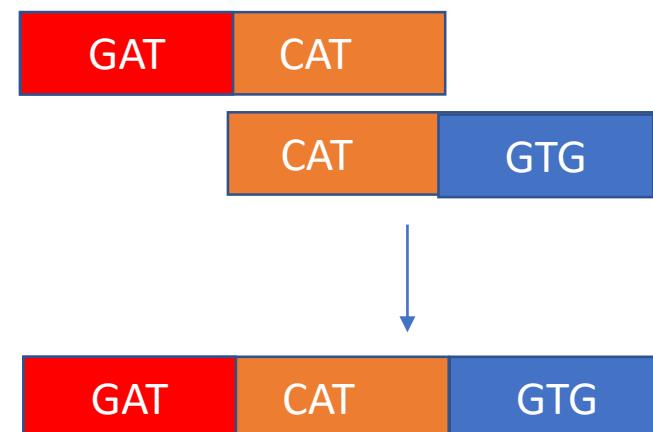


Hitta sekvenser som saknas i referens genomet

- 1000 Svenskar
 - Tvillingregistret
 - Sekvenserade 2017
- Pussla ihop genom
 - Överlappande läsningar
- Jämför mot Mänskliga referensgenomet
 - Analysera de bitar som inte matchar

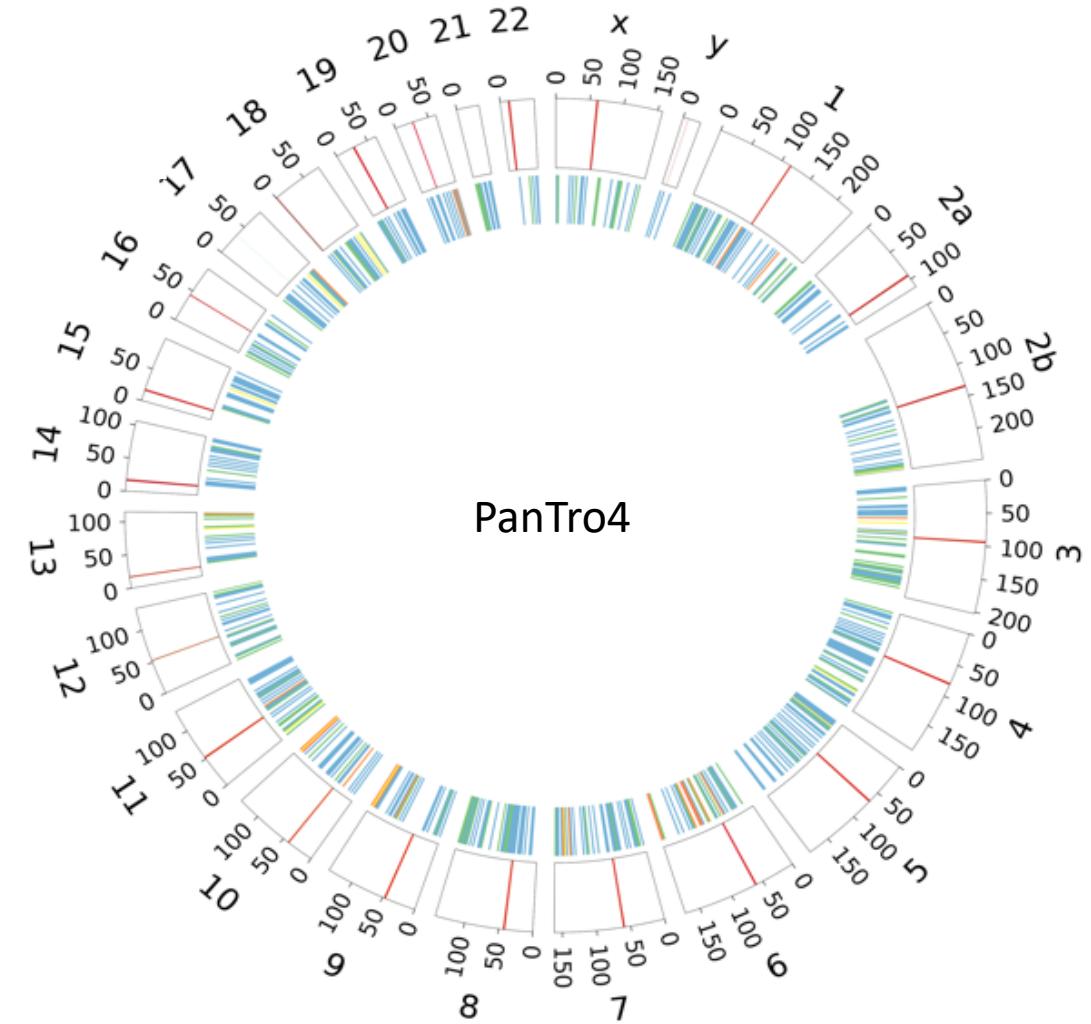


Ameur et al 2017



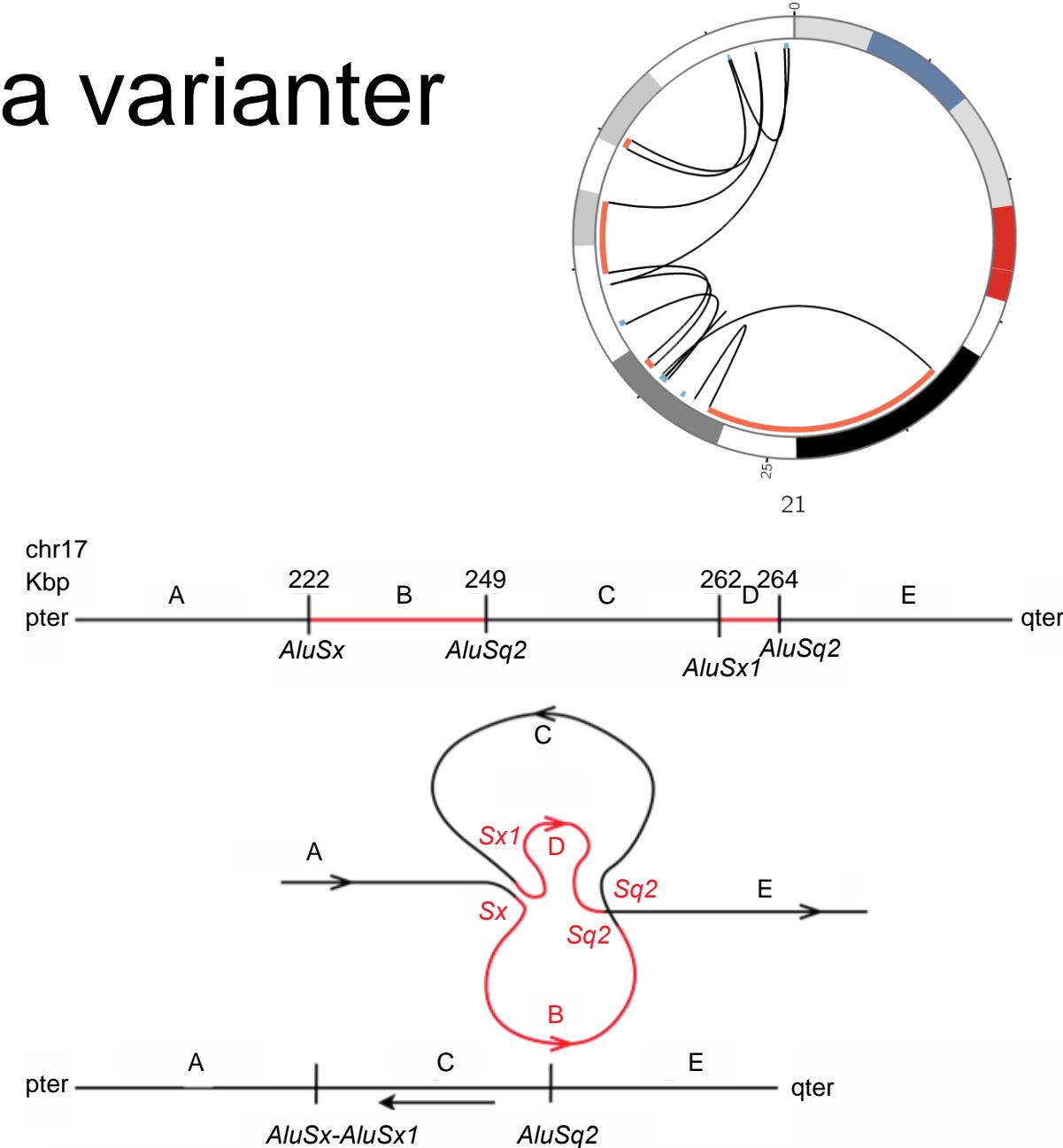
Hitta sekvenser som saknas i referens genomet

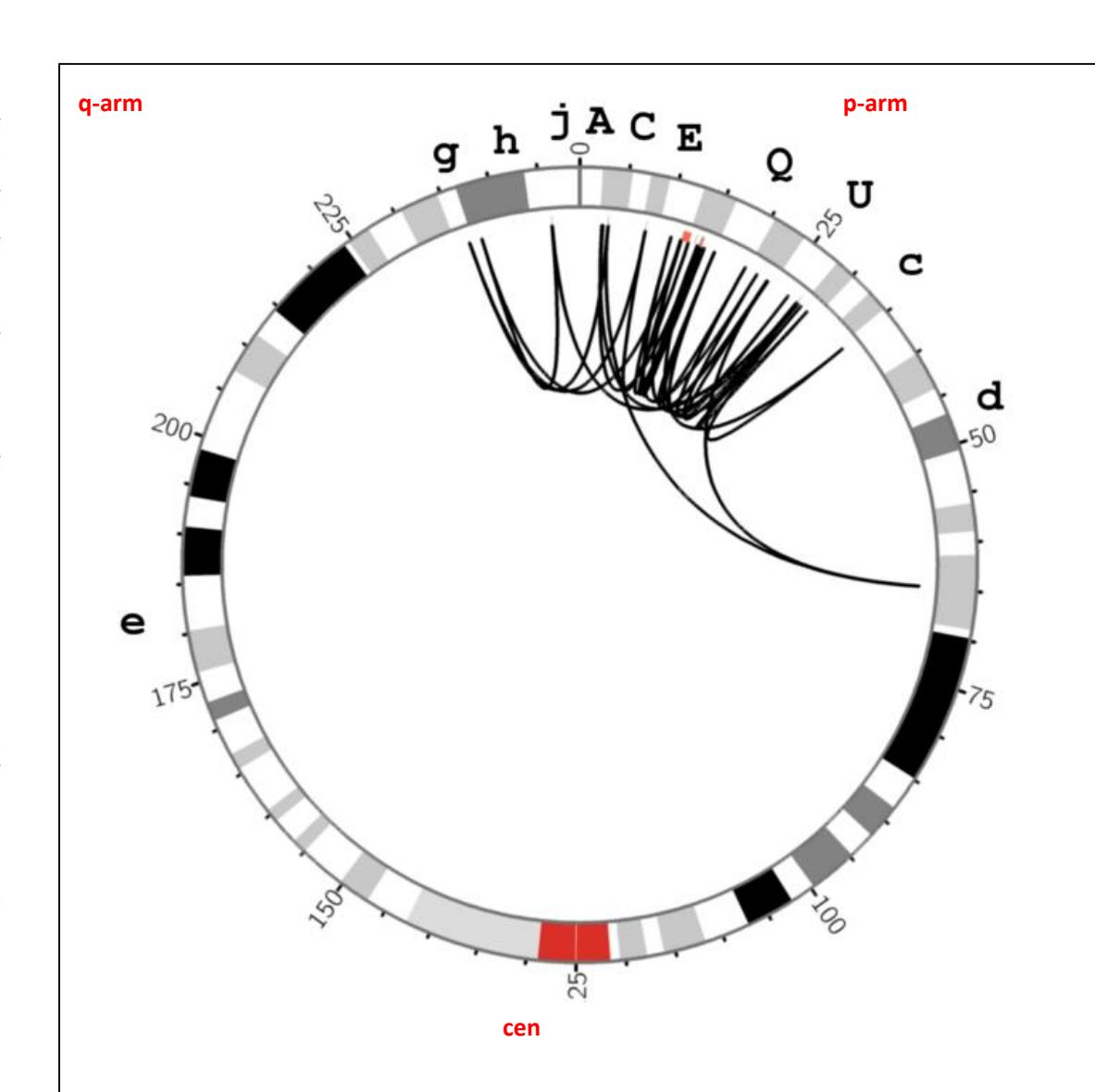
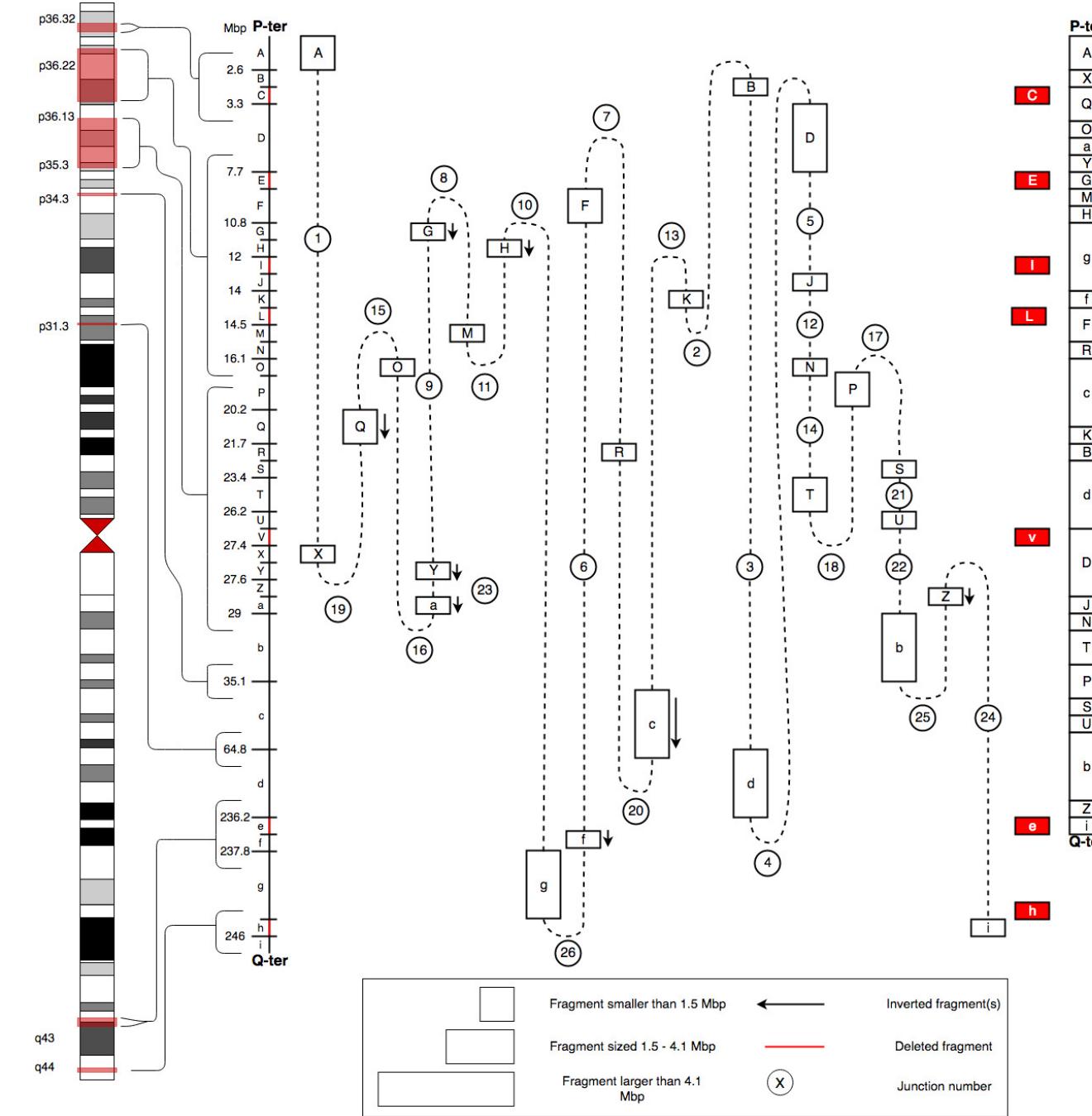
- Sekvens som saknas i referensgenomet
 - Vi hittade 46 Miljoner baser
 - Som kromosom 19
 - Ca 1 miljon baser per individ
- Totalt 60 000 nya sekvenser
 - 20 000 matchar chimpansgenomet



Lösa komplexa strukturella varianter

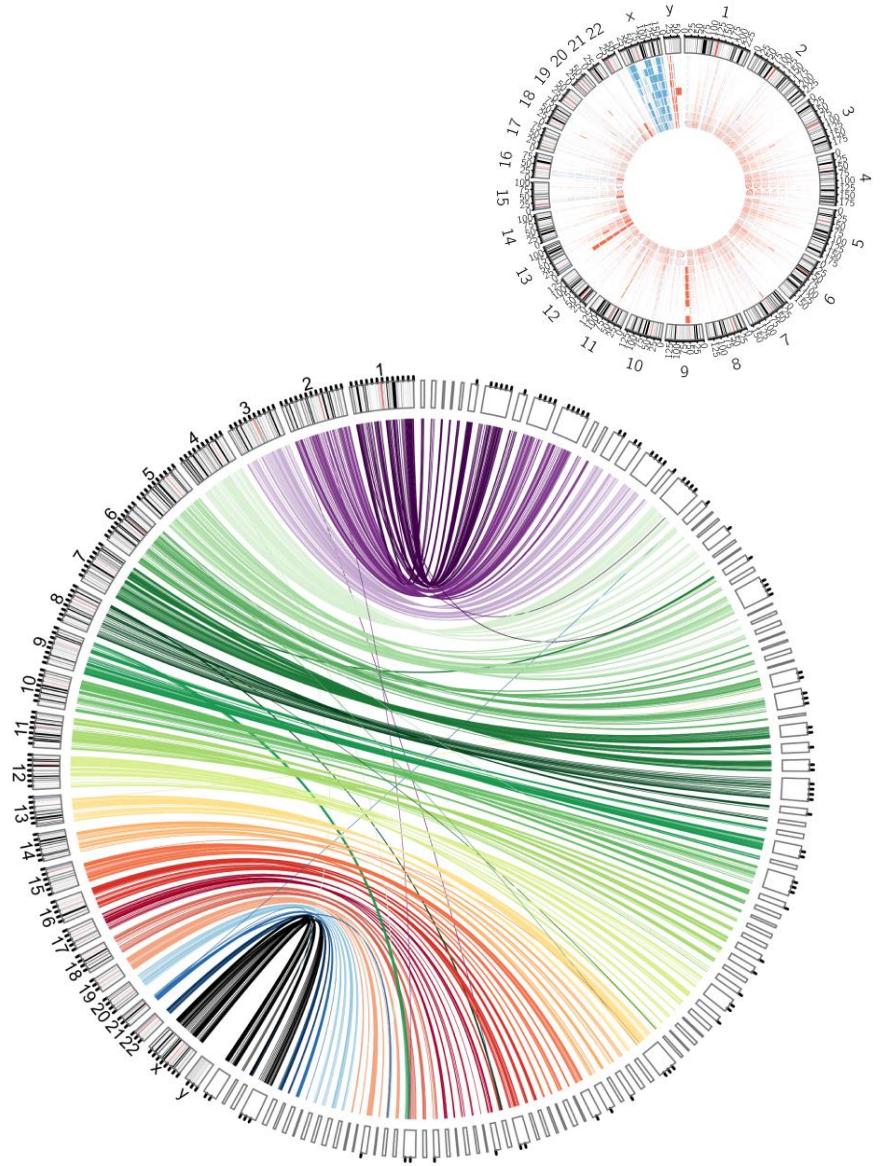
- Komplex strukturell avvikelse
 - Flera DNA fragment huller om buller
- Förstå genetiska mekanismer
- Hitta trasiga gener





Sammanfattning

- En stor variation av stora varianter
- Många sekvenseringstekniker
 - För och nackdelar
- Kliniskt relevant information
 - Sjukdomsorsakande varianter
 - Bättre verktyg



Tack!

