

Supplementary information for
“Comment on: ERGC: An efficient referential genome
compression algorithm”

Sebastian Deorowicz, Szymon Grabowski, Idoia Ochoa,
Mikel Hernaez and Tsachy Weissman

Datasets

The 12 FASTA sequences are different assemblies of the *H.sapiens* genome. The following genome assemblies were used:

- hg17 (also referred to as NCBI35.1)

The past version of human genome reference sequence, NCBI Build 35, produced by the International Human Genome Sequencing Consortium (released in 2004). Data access:

```
ftp://ftp.ncbi.nih.gov/genomes//H_sapiens/ARCHIVE/BUILD.35.1/  
Assembled_chromosomes/*.fa.gz
```

- hg18 (also referred to as NCBI36.3)

The past version of human genome reference sequence assembly, NCBI Build 36.3, produced by the International Human Genome Sequencing Consortium (released in 2008). Data access:

```
ftp://ftp.ncbi.nih.gov/genomes//H_sapiens/ARCHIVE/BUILD.36.3/  
Assembled_chromosomes/*.fa.gz
```

- hg19 (also referred to as NCBI37)

The Genome Reference Consortium Human Build 37. Data access:

```
ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/  
Homo_sapiens/GRCh37/Primary_Assembly/assembled_chromosomes/FASTA/
```

- hg38 (also referred to as GRCh38.p3)

The latest version of the Human Reference by Genome Reference Consortium, Human Build 38.2 (released in 2015). Data access:

```
ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes//vertebrates_mammals/  
Homo_sapiens/GRCh38.p2/Primary_Assembly/assembled_chromosomes/FASTA/*.fa.gz
```

- CHM1.1.0

The reference-guided de novo assembly based on human assembly GRCh37, produced by Washington University School of Medicine (released in 2012). Data access:

```
ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes//vertebrates_mammals/  
/Homo_sapiens/CHM1_1.0/Primary_Assembly/assembled_chromosomes/FASTA/*.fa.gz
```

- CHM1.1.1

Reference-guided assembly based on human assembly GRCh37, produced by Washington University School of Medicine (released in 2013). Data access:

```
ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes//vertebrates_mammals/  
Homo_sapiens/CHM1_1.1/Primary_Assembly/assembled_chromosomes/FASTA/*.fa.gz
```

- HuRef

The whole genome shotgun assembly by J. Craig Venter Institute. This assembly represents a composite haploid version of the diploid genome sequence from a single individual, J. Craig Venter (released in 2007). Data access:
ftp://ftp.ncbi.nih.gov/genomes//H_sapiens/ARCHIVE/ANNOTATION_RELEASE.106/Assembled_chromosomes/seq/hs_alt_HuRef*.fa.gz
- WGS

The whole genome shotgun assembly of the human genome, produced by the Celera Genomics (released in 2001). Data access:
http://www.ncbi.nlm.nih.gov/assembly/GCA_000002115.2/
- CSA

The January 2001 compartmental shotgun assembly of the human genome (released in 2004). Data access:
http://www.ncbi.nlm.nih.gov/assembly/GCA_000365445.1/
- YH

The de novo assembled genome of an Asian individual, produced by Beijing Genomics Institute (released in 2010). Data access:
ftp://public.genomics.org.cn/BGI/yanhuang/fa/*fa.gz
- KOREF_20090224

The Korean individual genome sequence, KOREF_20090224 assembly (released in 2009). Data access:
ftp://ftp.kobic.kr/pub/KOBIC-KoreanGenome//KOREF_20090224/fasta/*fa.gz
- KOREF_20090131

The Korean individual genome sequence, KOREF_20090131 assembly (released in 2009). Data access:
ftp://ftp.kobic.kr/pub/KOBIC-KoreanGenome//KOREF_20090131/fasta/*.consensus_fasta

Dataset statistics

Table 1: Frequency (in %) of symbols in D1–D5 datasets used in the experiments

Genome	Alph. size	Upper case symbols						Lower case symbols					
		A	C	G	T	N	other	a	c	g	t	n	other
hg18	7	27.4	19.0	19.0	27.4	7.2	10^{-7}	—	—	—	—	—	—
KOREF_20090224	21	26.3	18.2	18.2	26.4	—	0.067	1.0	0.8	0.8	1.0	7.3	0.002
KOREF_20090131	21	26.2	18.1	18.1	26.2	—	0.065	1.1	0.8	0.8	1.1	7.3	0.002
YH	11	27.4	18.9	19.0	27.4	7.2	0.076	—	—	—	—	—	—

Table 2: Frequency (in %) of symbols in 12 *H. sapiens* genomes used in the experiments

Genome	Alph. size	Upper case symbols						Lower case symbols					
		A	C	G	T	N	other	a	c	g	t	n	other
Chromosome 10													
CHM1.1.1	5	28.1	20.0	20.0	28.1	3.9	—	—	—	—	—	—	—
CHM1.1.0	5	27.9	19.9	19.9	27.9	4.5	—	—	—	—	—	—	—
CSA	5	25.8	18.3	18.3	25.9	11.7	—	—	—	—	—	—	—
hg17	5	28.4	20.2	20.2	28.4	2.8	—	—	—	—	—	—	—
hg18	5	28.4	20.2	20.2	28.4	2.8	—	—	—	—	—	—	—
hg19	5	28.3	20.1	20.1	28.3	3.1	—	—	—	—	—	—	—
hg38	12	29.1	20.7	20.7	29.2	0.4	0.00027	—	—	—	—	—	—
HuRef	5	29.0	20.6	20.6	29.1	0.6	—	—	—	—	—	—	—
KOREF_20090131	21	27.0	19.2	19.2	27.1	—	0.073	1.3	1.0	1.0	1.3	2.9	0.002
KOREF_20090224	21	27.1	19.3	19.3	27.2	—	0.075	1.2	0.9	0.9	1.2	2.9	0.002
WGSA	5	28.0	19.8	19.8	28.0	4.4	—	—	—	—	—	—	—
YH	11	28.4	20.2	20.2	28.4	2.8	0.083	—	—	—	—	—	—
Chromosome 20													
CHM1.1.0	5	26.2	20.8	20.8	26.5	5.7	—	—	—	—	—	—	—
CHM1.1.1	5	26.2	20.8	20.8	26.5	5.8	—	—	—	—	—	—	—
CSA	5	26.4	20.9	21.0	26.7	5.0	—	—	—	—	—	—	—
hg17	5	26.5	21.0	21.1	26.8	4.7	—	—	—	—	—	—	—
hg18	5	26.5	21.0	21.1	26.8	4.7	—	—	—	—	—	—	—
hg19	5	26.2	20.8	20.9	26.5	5.6	—	—	—	—	—	—	—
hg38	5	27.7	21.6	21.9	28.0	0.8	—	—	—	—	—	—	—
HuRef	5	27.6	21.8	21.9	28.0	0.7	—	—	—	—	—	—	—
KOREF_20090131	21	26.0	20.6	20.6	26.3	—	0.073	0.5	0.4	0.4	0.5	4.8	0.002
KOREF_20090224	21	26.0	20.6	20.7	26.3	—	0.075	0.4	0.3	0.3	0.4	4.8	0.002
WGSA	5	26.7	21.0	21.1	27.0	4.1	—	—	—	—	—	—	—
YH	11	26.4	21.0	21.0	26.8	4.7	0.083	—	—	—	—	—	—

Software

The following pieces of software were used in the experiments:

- ERGC:

Command-line:

```
./java Utilities ref_seq target_seq output.ergc  
./7z a -t7z archive_name output.ergc -m0=PPMd
```

- GDC:

Version 0.3 (released on Aug 11, 2011) downloaded from:

<http://sun.aei.polsl.pl/REFRESH/index.php?page=projects&project=gdc&subpage=download>

Command-line:

```
./gdc c -rn1 archive_name ref_seq target_seq
```

- iDoComp: Version 1.1 (released on Oct 6, 2014) downloaded from:

<https://github.com/mikelhernaez/iDoComp>

For some experiments (marked by ‘*’) the latest release (version 1.2, released on Oct 11, 2015) downloaded from:

<https://github.com/mikelhernaez/iDoComp>

was used (the former release had some memory allocation bug). The results that did not crash with the previous release are still correct, the memory allocation bug does not affect them (i.e., there is no need to re-run them).

For the command lines, we followed the instructions on the README file provided with the iDoComp distribution.

Chromosome 10 results

Table 3: Compression results in KB with **CHM1_1.0** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	0.3	0.1	0.3
CHM1_1.1	252.1	305.3	21928.9
CSA	2011.0	2304.7	28203.4
hg17	522.9	618.3	27061.4
hg18	523.0	618.3	27064.1
hg19	520.1	614.1	26992.1
hg38	1026.0	1298.9	22340.1
HuRef	535.3	645.4	30407.6
KOREF_20090131	1888.7	1998.8	27400.4
KOREF_20090224	1709.1	1888.7	27362.7
WGSA	606.6	782.1	23943.4
YH	634.0	730.7	27224.2
Total	10229	11805	288929

Table 4: Compression results in KB with **CHM1_1.1** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	206.5	228.4	21823.8
CHM1_1.1	0.3	0.1	0.2
CSA	2022.3	2315.2	28204.9
hg17	409.4	478.2	27090.3
hg18	409.4	478.2	27092.9
hg19	406.9	474.9	27018.6
hg38	1009.3	1271.5	22477.6
HuRef	528.0	634.4	30408.9
KOREF_20090131	1766.9	1937.1	27403.6
KOREF_20090224	1582.8	1824.3	27346.2
WGSA	645.1	823.3	23894.2
YH	518.0	588.3	27252.5
Total	9505	11054	290014

Table 5: Compression results in KB with **CSA** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	4308.8	5011.4	30746.6
CHM1_1.1	4359.2	5126.5	30903.5
CSA	0.3	0.1	0.2
hg17	4381.7	5219.1	31138.3
hg18	4384.5	5222.0	31141.0
hg19	4369.1	5169.7	31067.1
hg38	4905.0	5967.8	31188.9
HuRef	4068.3	4626.5	30362.9
KOREF_20090131	6126.0	6892.4	31480.3
KOREF_20090224	5945.4	6732.7*	31433.9
WGSA	3659.7	4114.9	29232.2
YH	4549.4	5416.9	31302.8
Total	51057	59500	339998

Table 6: Compression results in KB with **hg17** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	465.7	550.0	22062.5
CHM1_1.1	392.3	456.4	22169.2
CSA	2007.0	2290.1	28204.9
hg17	0.3	0.1	0.3
hg18	7.2	3.1	55.2
hg19	8.7	3.3	21932.0
hg38	644.5	934.4	22303.2
HuRef	516.7	618.2	24730.7
KOREF_20090131	1655.1	946.5	694.6
KOREF_20090224	1470.4	866.4	638.3
WGSA	646.6	823.1	24044.1
YH	331.9	309.3	434.9
Total	8146	7801	167270

Table 7: Compression results in KB with **hg18** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	463.0	546.7	22062.6
CHM1_1.1	389.5	452.7	22169.1
CSA	2007.1	2290.2	28205.0
hg17	3.2	0.1	0.3
hg18	0.3	0.1	0.3
hg19	2.0	0.2	95.5
hg38	641.6	932.2	22303.2
HuRef	514.2	615.1	24730.7
KOREF_20090131	1648.7	935.3	625.4
KOREF_20090224	1463.7	856.1	578.8
WGSA	644.0	819.8	24044.1
YH	328.7	304.7	393.7
Total	8106	7753	145209

Table 8: Compression results in KB with **hg19** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	463.1	546.6	22062.7
CHM1_1.1	389.7	452.7	22168.8
CSA	2006.9	2290.4	28205.3
hg17	24.2	2.5	13329.6
hg18	24.2	2.4	12838.8
hg19	0.3	0.1	0.3
hg38	641.6	931.1	22303.2
HuRef	514.3	615.1	24731.2
KOREF_20090131	1649.7	1015.7	13391.9
KOREF_20090224	1464.5	929.5	22384.4
WGSA	644.0	819.8	24044.1
YH	330.6	307.2	22285.7
Total	8153	7913	227746

Table 9: Compression results in KB with **hg38** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	530.4	594.3	5.4
CHM1_1.1	553.7	631.7	26844.5
CSA	2086.7	2355.9	28204.3
hg17	258.5	278.1	26939.9
hg18	258.5	278.0	26942.5
hg19	256.7	275.9	26867.9
hg38	0.3	0.1	0.3
HuRef	630.2	735.9	5356.9
KOREF_20090131	1932.4	1619.1	27358.3
KOREF_20090224	1750.4	1519.9	27309.6
WGSA	503.4	669.5	29.9
YH	567.4	584.9	27126.6
Total	9329	9543	222986

Table 10: Compression results in KB with **HuRef** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	748.6	964.2	26776.6
CHM1_1.1	769.6	988.6	26910.0
CSA	1995.6	2245.5	28206.3
hg17	784.5	1038.6	27111.0
hg18	784.8	1038.8	27113.7
hg19	779.2	1013.8	27039.5
hg38	1340.7	1775.1	22451.6
HuRef	0.3	0.1	0.3
KOREF_20090131	2134.8	2429.2	27447.7
KOREF_20090224	1965.3	2310.0	27401.8
WGSA	475.9	613.6	27610.6
YH	900.4	1164.0	27268.7
Total	12680	15582	295338

Table 11: Compression results in KB with **KOREF_20090131** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	1461.8	1644.9	22104.3
CHM1_1.1	1446.9	1584.7	22162.7
CSA	2617.4	3803.0	28204.6
hg17	1444.6	1418.3	6316.7
hg18	1444.8	1417.0	439.9
hg19	1427.2	1377.8	22117.3
hg38	1986.9	2079.9*	22448.9
HuRef	1369.8	1466.6	24439.5
KOREF_20090131	0.2	0.1	194.5
KOREF_20090224	474.7	288.2	206.7
WGS	1254.9	1733.3	9789.4
YH	1458.5	1432.4	424.4
Total	16388	18246	158849

Table 12: Compression results in KB with **KOREF_20090224** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	1295.5	1521.4	22179.7
CHM1_1.1	1278.3	1456.4	22166.5
CSA	2492.3	3696.8	28204.3
hg17	1270.9	1288.7	6305.4
hg18	1271.1	1287.5	424.6
hg19	1253.8	1248.9	22110.4
hg38	1806.9	1952.3*	22378.5
HuRef	1213.5	1352.3	24509.6
KOREF_20090131	460.0	276.1	251.0
KOREF_20090224	0.2	0.1	145.2
WGS	1118.2	1620.9	9545.6
YH	1275.4	1295.1	414.0
Total	14736	16997	158635

Table 13: Compression results in KB with **WGSA** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	1935.2	2395.9	29983.9
CHM1_1.1	2014.6	2553.7	30128.4
CSA	2657.0	2949.1	28206.4
hg17	2059.0	2701.7	31056.5
hg18	2059.2	2701.9	31059.1
hg19	2050.1	2659.9	30985.0
hg38	2312.0	3074.8	31087.0
HuRef	1597.2	1827.3	29579.1
KOREF_20090131	3450.8	4149.5	31402.0
KOREF_20090224	3280.5	4014.2	31353.9
WGSA	0.3	0.1	0.2
YH	2190.7	2868.7	31222.4
Total	25607	31897	336064

Table 14: Compression results in KB with **YH** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	548.7	630.3	22377.9
CHM1_1.1	477.4	537.4	22192.2
CSA	2078.9	2363.2	28203.6
hg17	310.0	278.7	376.5
hg18	310.1	278.7	375.5
hg19	310.0	278.7	22096.7
hg38	994.0	1223.0	22562.3
HuRef	596.0	698.5	24449.6
KOREF_20090131	1615.1	837.9	595.5
KOREF_20090224	1427.7	757.2	542.6
WGSA	719.5	901.2	9759.0
YH	0.2	0.1	0.2
Total	9388	8785	153532

Chromosome 20 results

Table 15: Compression results in KB with **CHM1_1.0** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	0.3	0.1	0.3
CHM1_1.1	66.1	75.4	0.2
CSA	966.3	1066.7	14089.1
hg17	201.5	228.4	14028.3
hg18	201.5	228.4	14028.3
hg19	201.6	228.5	14028.4
hg38	750.3	1196.4	14619.1
HuRef	241.2	289.8	13708.5
KOREF_20090131	559.0	590.6	14167.0
KOREF_20090224	489.6	554.6	14148.5
WGSA	310.2	386.4	8489.2
YH	254.8	286.2	14104.7
Total	4242	5132	135412

Table 16: Compression results in KB with **CHM1_1.1** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	87.0	92.0	640.3
CHM1_1.1	0.3	0.1	0.2
CSA	965.7	1066.3	14088.6
hg19	187.8	211.9	8137.4
hg18	187.8	211.9	8137.4
hg19	187.9	212.0	8137.5
hg38	777.7	1183.7	8690.3
HuRef	249.7	300.2	13717.8
KOREF_20090131	552.7	584.6	8259.3
KOREF_20090224	482.9	548.6	8236.0
WGSA	330.8	406.7	8390.0
YH	241.8	269.9	13871.6
Total	4252	5088	100306

Table 17: Compression results in KB with **CSA** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	1131.0	1267.4	13240.9
CHM1_1.1	1110.4	1249.8	13266.9
CSA	0.3	0.1	0.2
hg19	1131.0	1261.8	13808.1
hg18	1131.0	1261.8	13808.1
hg19	1131.2	1261.8	13808.1
hg38	1683.8	2372.6	14361.3
HuRef	1114.8	1235.9	13718.8
KOREF_20090131	1600.2	1714.5	13443.9
KOREF_20090224	1524.2	1663.9	13424.0
WGS	985.2	1066.5	9467.0
YH	1197.5	1339.1	13885.5
Total	13741	15695	146233

Table 18: Compression results in KB with **hg19** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	181.4	201.4	8068.6
CHM1_1.1	141.6	167.0	351.6
CSA	938.3	1031.0	14088.6
hg19	0.3	0.1	0.3
hg18	0.3	0.1	0.3
hg19	16.4	0.4	0.5
hg38	584.7	1002.6	8440.7
HuRef	225.6	269.0	8147.9
KOREF_20090131	490.7	367.3	269.3
KOREF_20090224	420.3	336.7	243.0
WGS	311.4	382.9	8403.1
YH	145.5	134.4	179.0
Total	3457	3893	48193

Table 19: Compression results in KB with **hg18** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1.1.0	181.4	201.4	8068.6
CHM1.1.1	141.6	167.0	351.6
CSA	938.3	1031.0	14088.6
hg19	0.3	0.1	0.3
hg18	0.3	0.1	0.3
hg19	16.4	0.4	0.5
hg38	584.7	1002.6	8440.7
HuRef	225.6	269.0	8147.9
KOREF_20090131	490.7	367.3	269.3
KOREF_20090224	420.3	336.7	243.0
WGS	311.4	382.9	8403.1
YH	145.5	134.4	179.0
Total	3457	3893	48193

Table 20: Compression results in KB with **hg19** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1.1.0	181.4	201.4	8068.6
CHM1.1.1	141.6	166.9	347.2
CSA	938.3	1031.2	14088.6
hg19	16.3	0.2	0.4
hg18	16.3	0.2	0.5
hg19	0.3	0.1	0.3
hg38	584.3	1002.2	8440.7
HuRef	225.6	269.0	8147.9
KOREF_20090131	490.8	380.3	310.0
KOREF_20090224	420.3	351.3	317.4
WGS	311.3	382.1	8403.1
YH	145.6	134.5	169.4
Total	3742	3919	48294

Table 21: Compression results in KB with **hg38** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1.1.0	147.0	169.3	6283.1
CHM1.1.1	141.1	167.3	6201.0
CSA	925.2	1012.2	14089.9
hg19	11.9	2.7	6603.9
hg18	12.0	2.7	6603.9
hg19	11.9	2.6	6604.0
hg38	0.3	0.1	0.3
HuRef	172.4	214.1	7130.6
KOREF_20090131	491.1	395.8	5760.6
KOREF_20090224	420.7	365.8	5745.9
WGS	263.3	330.1	7288.6
YH	146.2	135.7	5904.3
Total	2743	2798	78216

Table 22: Compression results in KB with **HuRef** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1.1.0	271.1	320.2	8034.0
CHM1.1.1	257.4	314.6	2363.3
CSA	964.4	1051.6	14085.7
hg19	275.1	324.5	8168.3
hg18	275.1	324.5	8168.3
hg19	275.1	324.6	8168.4
hg38	822.1	1043.6	8762.9
HuRef	0.3	0.1	0.3
KOREF_20090131	676.4	732.0	8303.0
KOREF_20090224	605.9	690.1	8283.0
WGS	284.9	343.9	12904.0
YH	328.9	391.9	8202.4
Total	5037	5862	95444

Table 23: Compression results in KB with **KOREF_20090131** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	466.7	524.8	14053.3
CHM1_1.1	436.9	488.3	8057.8
CSA	1170.8	1462.2	14089.3
hg17	442.8	411.5	202.0
hg18	442.8	411.5	202.0
hg19	442.9	422.6	7951.5
hg38	1087.4	1399.2	14346.5
HuRef	517.4	576.7	8159.7
KOREF_20090131	0.2	0.1	75.4
KOREF_20090224	165.9	104.9	78.4
WGS	541.2	752.9*	8408.7
YH	454.4	429.1	202.1
Total	6169	6984	75827

Table 24: Compression results in KB with **KOREF_20090224** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	400.0	481.3	14053.3
CHM1_1.1	370.8	445.3	413.3
CSA	1113.5	1421.6	14088.8
hg17	374.5	369.3	200.5
hg18	374.5	369.3	200.5
hg19	374.6	380.9	7950.8
hg38	993.9	1349.3	14347.0
HuRef	451.9	533.5	8160.8
KOREF_20090131	166.7	100.8	98.0
KOREF_20090224	0.2	0.1	54.2
WGS	482.5	706.6*	8405.9
YH	382.7	384.5	199.9
Total	5486	6543	68173

Table 25: Compression results in KB with **WGSA** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	762.7	895.9	9197.7
CHM1_1.1	762.5	899.1	10145.9
CSA	1259.6	1366.2	14088.2
hg17	787.1	920.5	13764.1
hg18	787.2	920.5	13764.1
hg19	787.2	920.6	13764.4
hg38	1300.3	1969.6	14691.9
HuRef	721.4	825.5	13758.9
KOREF_20090131	1230.4	1402.1	10314.0
KOREF_20090224	1156.3	1346.8	10293.1
WGSA	0.3	0.1	0.2
YH	847.9	1003.3	13840.8
Total	10403	12470	137623

Table 26: Compression results in KB with **YH** as the reference sequence

Target sequence	GDC	iDoComp	ERGC
CHM1_1.0	228.1	246.5	14053.3
CHM1_1.1	189.6	213.6	397.8
CSA	978.4	1077.2	14082.8
hg17	137.6	123.1	162.2
hg18	137.6	123.1	162.3
hg19	137.8	123.4	181.0
hg38	727.9	1108.0	8532.3
HuRef	274.0	316.9	8157.9
KOREF_20090131	486.2	339.8	265.6
KOREF_20090224	414.3	309.1	237.6
WGSA	351.7	425.1	3718.4
YH	0.2	0.1	0.2
Total	4063	4406	49951