

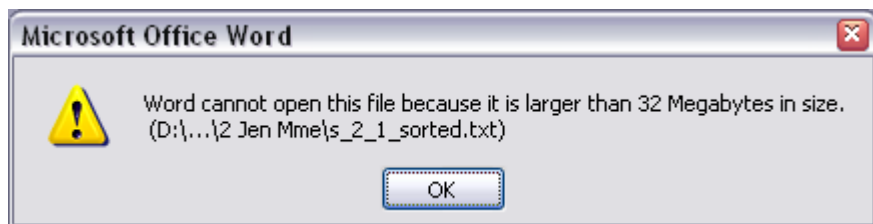
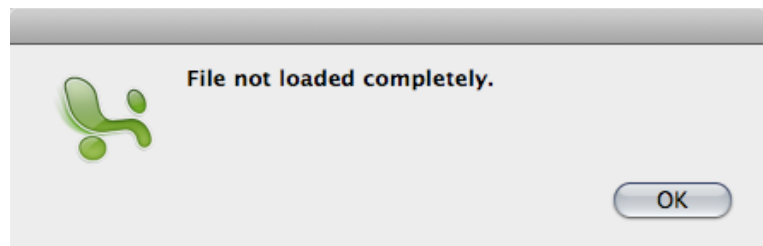
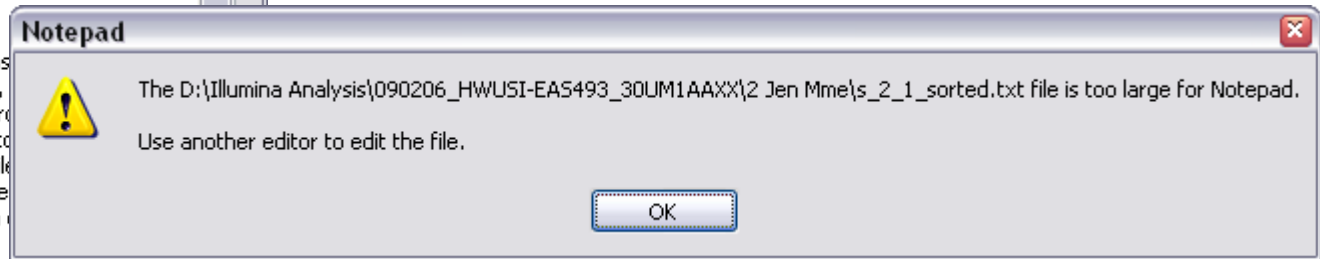
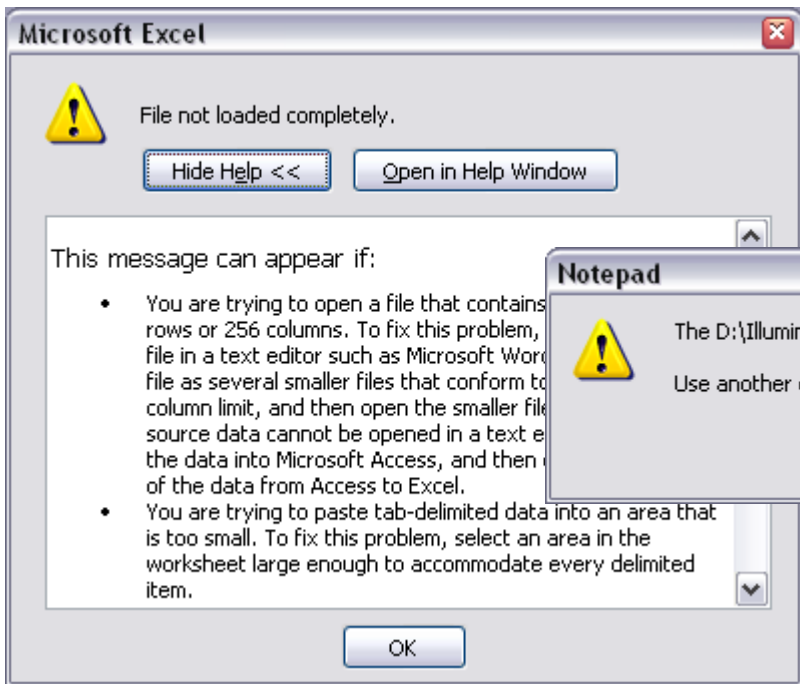
# Managing Large Datasets

Hemant Kelkar – University of North Carolina

Mario Caccamo – BBSRC Genome Analysis Centre, UK

Chair: Simon Andrews, Babraham Institute, UK





# 20 million reads...

HWUSI-EAS493	0004	1	100	10390	4943	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	10615	4706	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	`ffedecee\dddcaececeaeacddd^aeac\dda\d	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	10919	20246	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	10994	19312	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fecffffffffffffcefdffffefffffffeede	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	11170	4957	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffffffffffdf	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	11739	3482	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	dfcfcffffc`fffffffffcefefebfeefafffef\b^ab	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	13214	14296	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffeffffffffffeefccedeefafde]fffe\cedab`	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	13359	20240	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffffffffffffffffffefe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	15526	19623	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	bbb\b` ``}] \x\`bb^Y\Y\YX]bbb`c\TcYUTTQR	phiX	-7	R	32NNNNNNNNN	199
HWUSI-EAS493	0004	1	100	15881	8272	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffffffffff_ BBBB	phiX	-7	R	32NNNNNNNNN	183
HWUSI-EAS493	0004	1	100	17303	3670	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	` ``] ``]Y\]Y\`bbb\cc`^c\__Ubccacc`a\ZZSU`	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	18317	6917	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	2958	15622	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffeSffdfdfedddd	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	3452	18525	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffefffffffffffffefe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	6974	2943	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffddecffffffffffdeeeecceYeecce\`bce`dac	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	7751	7607	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	e`eccccccdc`cc^cabb`bbb`bdd`ccdcad	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	7888	17923	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	100	8806	17983	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffffffffffffffffffefe\	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	10521	14979	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	1112	18316	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffefe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	11694	11459	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	eeaefffefcfeffffdfcdffffdcf	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	12359	17322	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	11483	4504	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffffffffffffffffffec	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	16081	8442	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	16598	19750	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	c`ac`cac`ad`d` `cca`b\bY_`caccS`[Y^`BBBB	phiX	-7	R	32NNNNNNNNN	178
HWUSI-EAS493	0004	1	101	17351	18717	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	efdfdf	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	12427	21277	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	dfffefff`eedecce`ecccccccc	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	13201	8523	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffefd	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	3721	15360	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	3721	17506	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffdfe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	4040	17333	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffefe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	5420	9967	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	16415	4457	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffeff	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	16468	19643	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffeff	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	16611	6714	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fffffefe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	101	7048	17405	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	cbcccdcd`ddd\dd`dca`c^c`^^bV^accdc\ad	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	1937	1127	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffe	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	10	2087	10042	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	ffffffffffffffffffefefUJP]^}[ZXW	phiX	-7	R	32NNNNNNNNN	199
HWUSI-EAS493	0004	1	102	10455	5878	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	fdffff	phiX	-7	R	32NNNNNNNNN	200
HWUSI-EAS493	0004	1	102	11062	13809	0	1	TTAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTT	dcdd^ddc`\`^bb^b\Tbb[ `T[b^]^[[bT__b_Tb]`	phiX	-7	R	32NNNNNNNNN	200

# 20 million reads...

