There is a problem with CG chr6 liftover. This output vcf needs to be sorted in order to generate a tbi file. Moreover prior to sorting the header must be removed. All chr6 genomewarp output vcf have “sorted” in the name. Once I move them to DNAnexus I will remove sorted.

* Generate script to sort
  + Hg3-done
  + Hg4-done
  + Hg5-done

HG2 – integration v3.3.2 complete 1/3/17

Initially screwed up the ion combined files, I combined CG not ion. 12/21 I recombined ion and did liftover. Ran integration and chr6 fails, all other chromosomes complete.

Completed integration running chr6 on 12/22/16

HG3 – integration v3.3.2 complete 1/3/17

I screwed up 10X HP1 splie and used HP2 bed. I re-split HP1. For HP1 reran callableloci and sention (12/21). I then reran integration prep (12/22).

* Need to finish callset tables
* Confirm integration script is ok—should be
* Run integration

HG4 – integration v3.3.2 complete 1/3/17

* Chr6 is running now (12/22) if it completes integration will be done

HG5 – integration v3.3.2 complete 1/3/17

* Chr6 integration running now 12/22 to test
* Need to finish callset tables
* Confirm integration script is ok—should be
* Run integration

Combine for Justin and comparison

Highconf vcf and bed

* Hg002- done
* Hg003- done
* Hg004- done
* Hg005- done

Liftover v3.3.2 GRCh37 (hg2-5) highconfidence files –DONE 1/4/17

Justin requested use the phased GRCh37 calls and bed files excluding mendelian errors in the liftover

**HG002**

b38 v.3.3.2 files for comparison

HG002\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC-SOLIDgatkHC\_CHROM1-22\_v.3.3.2\_highconf\_triophased.vcf.gz

HG002\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC-SOLIDgatkHC\_CHROM1-22\_v.3.3.2\_highconf\_noinconsistent.bed

b37 v3.3.2 for liftover (in GRCh37 HG002)

HG002\_GRCh37\_GIAB\_highconf\_CG-IllFB-IllGATKHC-Ion-10X-SOLID\_CHROM1-22\_v.3.3.2\_highconf\_noinconsistent.bed

HG002\_GRCh37\_GIAB\_highconf\_CG-IllFB-IllGATKHC-Ion-10X-SOLID\_CHROM1-22\_v.3.3.2\_highconf\_triophased.vcf.gz

**HG003**

b38 v.3.3.2 files for comparison

HG003\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC\_CHROM1-22\_v.3.3.2\_highconf.vcf.gz

(in GRCh38 HG002) HG003\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC\_CHROM1-22\_v.3.3.2\_highconf\_noinconsistent.bed

b37 v3.3.2 for liftover

HG003\_GRCh37\_GIAB\_highconf\_CG-IllFB-IllGATKHC-Ion-10X\_CHROM1-22\_v.3.3.2\_highconf.vcf.gz

(in GRCh37 HG002) HG003\_GRCh37\_GIAB\_highconf\_CG-IllFB-IllGATKHC-Ion-10X\_CHROM1-22\_v.3.3.2\_highconf\_noinconsistent.bed

**HG004**

b38 v.3.3.2 files for comparison

HG004\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC\_CHROM1-22\_v.3.3.2\_highconf.vcf.gz

(in GRCh38 HG002) HG004\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC\_CHROM1-22\_v.3.3.2\_highconf\_noinconsistent.bed

b37 v3.3.2 for liftover

HG004\_GRCh37\_GIAB\_highconf\_CG-IllFB-IllGATKHC-Ion-10X\_CHROM1-22\_v.3.3.2\_highconf.vcf.gz

(in GRCh37 HG002) HG004\_GRCh37\_GIAB\_highconf\_CG-IllFB-IllGATKHC-Ion-10X\_CHROM1-22\_v.3.3.2\_highconf\_noinconsistent.bed

**HG005**

b38 v.3.3.2 files for comparison

HG005\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC-SOLIDgatkHC\_CHROM1-22\_v.3.3.2\_highconf.vcf.gz

HG005\_GRCh38\_GIAB\_highconf\_CG-Illfb-IllsentieonHC-Ion-10XsentieonHC-SOLIDgatkHC\_CHROM1-22\_v.3.3.2\_highconf.bed

b37 v3.3.2 for liftover

HG005\_GRCh37\_highconf\_CG-IllFB-IllGATKHC-Ion-SOLID\_CHROM1-22\_v.3.3.2\_highconf.vcf.gz

HG005\_GRCh37\_highconf\_CG-IllFB-IllGATKHC-Ion-SOLID\_CHROM1-22\_v.3.3.2\_highconf.bed

* Highconf vcf and beds already downloaded in hg grch38 directories

Comparison of v3.3.2 GRCh38 (hg2-5) highconfidence calls to grch37 high confidence calls – DONE 1/4/17

* Add stats to database

Update database

* Integration summaries -DONE
* Add in HG2-5 callset table example -DONE

Confirm correct in callset tables

* SV bed file
* HG #
* CG #
* Ion #

Notes:

* Because of screw up in misidentifiying files I confirmed that combined files for liftover -ion, solid, and cg vcf - were correct. Also confirmed all 10x bam were generated using the correct combined bam and bed.
* HG2 and HG4 had been run on 12/21 however today (1/3/17) following phasing Justing noticed an error in the HG2-5 callset tables. Specifically use of the beds (0/1s). New callset tables were prepared and all integrations rerun