Supplementary Methods

DNA isolation, KIR genotyping and HLA genotyping

Genomic DNA was isolated from snap frozen, cryosectioned GBM tissue after verification of >70% viable tumor by a neuropathologist, and from blood of patients and healthy controls using DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany), according to the manufacturer's protocol. The GBM patients' and controls' HLA class I (loci A, B, and C) was genotyped by ProImmune (Oxford, U.K.) within two digit resolution using PCR-based sequence-specific oligonucleotide probe hybridization. For KIR genotyping, 2μg of genomic DNA was hybridized to sequence specific oligonucleotides and amplified by polymerase chain reaction using the KIR Typing kit (Miltenyi Biotec, Gladbach Germany), according to the manufacturer's protocol. Presence or absence of the KIR genes was visualized on acrylamide gels by two independent evaluators.

Haplotype Nomenclature

The haplotype nomenclature followed the recommendation of the Allele Frequency Database. The KIR genes present in our cohort were classified into B genotype, if any of *KIR2DL2, KIR2DL5, KIR3DS1, KIR2DS1, KIR2DS2, KIR2DS3* and *KIR2DS5* were present. When none of these KIR genes or KIR2DS4 was present, the genotype was designated as AA. We did not distinguish between AB and BB haplotypes but denoted these as Bx.

MGMT methylation analysis

For the analysis of *MGMT* promoter methylation, 200 ng of tumor DNA was treated with sodium bisulfite using the EZ DNA Methylation-Gold KitTM (Zymo Research, Irvine CA, USA). MGMT promoter methylation was analyzed by methylation-specific PCR, which was carried out for 40 cycles using primers specific for methylated or unmethylated DNA. PCR

products were separated on 2% agarose gels and ethidium bromide-stained bands were recorded. The primer sequences used to detect unmethylated MGMT promoter sequences were 5'-TTTGTGTTTTGATGTTGTTAGGTTTTTGT-3' and 5'-AACTCCACACTC TTCCAAAAACAAAACA-3'. The primer sequences used to detect methylated MGMT promoter sequences were 5'-TTTCGACGTTCGTAGGTTTTCGC-3' and 5'-GCACTCTTCCGAAAACGAAACG-3'.

Flow cytometry

Peripheral blood mononuclear cells were harvested by density gradient centrifugation in Lymphoprep as described above. Thereafter, PBMCs from *n*=56 healthy controls and *n*=21 GBM patients were fixed and stained for intracellular markers prior to staining with antibodies for the surface markers, as listed in Supplementary Table SI. Fluorescence minus one (FMOs) for each florescence channel was used to determine the specific fluorescence of each fluorochrome. For good separation of KIR3DL2 and KIR3DL1, as well as KIR2DL2 and KIR2DL3, two-step staining was performed. NK cells were pre-incubated for 20 min with anti-KIR2DL1 and anti-KIR3DL1 Abs before incubation with anti-KIR3DL2/1 and anti-KIR2DL2/3 Abs. All samples were stained with Live/Dead Near-IR cell Stain Kit (Invitrogen, Waltham, MA USA) for 20 min at 4°C, previous to fixation. Lymphocytes and live cells were gated based on size by forward scatter (FSC) *vs.* granularity by side scatter (SSC) plots and viability using Live/Dead staining. Doublets were excluded based on SSC-A *vs.* SSC-H plot. Flow data acquisition was performed on LSR Fortessa (BD Biosciences, Franklin Lakes, NJ, USA) and the data was analyzed using FlowJo, version 10 (Tree Star Inc., Ashland Or, USA)

TCGA samples and 1000 genome healthy population samples

Glioblastoma (GBM) patient cohort of 300 hybrid-capture whole-exome sequencing data (WES) collected between 1993 and 2013 into The Cancer Genome Atlas (TCGA)

http://cancergenome.nih.gov/ and https://tcga-data.nci.nih.gov/tcga/ were analyzed for KIR2DS4 immunogenotype and their demographics are summarized in Table SII and described in Supplementary Table SIII. We also analyzed the KIR2DS4 gene status in 2504 healthy control samples obtained from 26 populations from Europe, East Asia, sub-Saharan Africa, South Asia and the Americas deposited into The 1000 Genome Project (http://www.1000genomes.org/data) (1) summarized in Table SII and described in Supplementary Table SIV.

Whole exome sequencing data analysis

Whole exome sequencing of the GBM cohort by TCGA (Study Accession: phs000178.v9.p8) was performed on the Illumina GA2000 and HiSeq platforms. Basic alignment against HG19 reference genome and sequence quality control was done by using the Picard and Firehose pipelines at Broad Institute. The alignment files (BAM files) for primary solid tumor (TP) were downloaded using GeneTorrent client software version 3.8.7 through The Cancer Genomics Hub (CGHub's) (https://cghub.ucsc.edu/) (Access approved on 19th October, 2015). A detailed description of the TCGA data can be found on http://cancergenome.nih.gov/ and https://tcga-data.nci.nih.gov/tcga/. Raw exome sequencing alignment BAM files were analyzed by Hi-performance, high security computer facility (SAFE project) at the IT department, University of Bergen (https://it.uib.no/). Two independent protected servers, *Helena and Concord* ran Linux operation system, with 8 cores, 80 GB of memory and 12 TB of disk space for each.

Variant calling pipeline

The samples selected after passing quality check were further subjected to GATK-3.5 (2), which is currently the most used and standardized tool for analyzing whole exome-sequencing data (Figure S1). GATK-3.5 workflow requires processed and sorted alignment

bam files from Picard 2.0.1(http://picard.sourceforge.net). The TCGA GBM bam files were therefore sorted by SortSam function and then duplicate reads were marked and removed using MarkDuplicates function of Picard. The obtained Picard processed bam files are then used for GATK-3.5 pipeline as per their best practices suggestion. Before variant calling step in GATK-3.5, reads were realigned around indels and quality scores recalibrated using IndelRealigner and BaseRecaliberator, respectively (Full parameters are given in the Supplementary note 1.0). The BaseRecalibarated step (BQSR) corrects for variation in quality for bases (even for most widely dispersed ones). The plots before and after recalibration were (Figure S2) inspected to crosscheck the quality score (which should be closer to their empirical scores) to be used further in a statistically robust manner for downstream GATK steps. These GATK pre-processing steps are important not only to realign indels (especially near the ends) but to also remove artifactual mismatches hampering base quality recalibration and variant detection. It has also been shown to improve the accuracy of several of downstream processing steps as described in best practices (3,4).

Potential Variants were identified on the realigned and recalibrated BAM alignments files using the GATK 3.5 Haplotypecaller in –ERC GVCF mode on each sample separately before applying joint genotyping using GenotypeGVCFs (Full parameters are given in the Supplementary note 1.0.). The GATK authors suggest and strongly recommend calling variants for entire cohorts (joint genotyping) rather than calling variants in each sample and then merging them. This step is critical to remove the strand and allelic biasness (which may be introduced when using call-set for each sample). It removes systematic biases by using evidence from multiple samples to filter out sites and add more valuable information and confidence about the called variation at any site of cohort. The resulting raw variant call-sets are often large and may contain many false positives. GATK provide VariantRecaliberated to filter out noisy variants (VQSR method) to balance sensitivity (discovering all the real

variants) and specificity (tuning filters to reduce false positives) by fitting to "Gaussian mixture model". GATK also provides resource bundle files that includes sites from hapmap, omni, 1000G, dbSNP etc. to evaluate our call sets. SNPs and INDELS are calibrated separately. (Full parameters are given in the Supplementary note 1.0). Variants from this set were mapped by position and alleles to called variants in the full alignments.

The resulting variants are written in VCF (Variant call format) files which is very explicit about the exact type and sequence of variation as well as the genotypes of multiple samples for this variation. We applied hard filters to further remove calls using various variants statistical scores, such as QD (QualByDepth), DP (DepthOfCoverage) and FS (FisherStrand, Phred-scaled pvalue using Fisher's exact test to detect strand bias), using FilterVariant function of GATK. Finally, we extracted different variant from the file and tabulated the GT (genotype), allelic depth (AD) and DP (depth of coverage). We found approx. 25 samples with DP less than 10 in each samples. We manually looked at bam alignment files for the confirmation of reads supports at this site. None of the samples were excluded in our analyses after manually crosschecking the site using IGV genome browser (5,6). The diploid organism, the GT field indicates the two alleles carried by the sample, encoded as 0 for reference allele (REF allele), 1 for first alternative allele (ALT allele), 2 for second allele and etc. It is often single ALT allele the most common case and thus the GT can be either 0/0 (homozygous reference), 0/1 (heterozygous, carrying 1 copy of each of the REF and ALT), and 1/1 (homozygous alternate allele).

KIR2DS4 deletion analysis of the variant call sets

KIR2DS4*00101, that codes for a full length membrane anchored activating receptor (functional protein (FUNC)) was observed to independently predict favorable prognosis. However, the KIR2DS4 gene also exists as deletion variants (KIR2DS4*0030101-2;

KIR2DS4*004-9) that encode a protein with a 22 bp deletion in chromosome 19q13.4, exon 5 of KIR2DS4 (nucleotide position 55350963, deleted sequence CCCGGAGCTCCTATGACATGTA) (encoding KIR transmembrane and intracellular domains) that causes truncated, soluble KIR2DS4 receptor (non-functional protein (DEL)) to be secreted rather than expressed on the cell surface. The KIR2DS4FUNC and KIR2DS4DEL variants were extracted from the TCGA cohort and tabulated as validation set. Each sample genotype was checked for the 22 bp deletion on position 55350963 in chromosome 19q13.4 and renamed as per the variants type (FUNC or DEL) they constitute. All genotype with 0/0 (homozygous reference) were renamed as DEL/DEL, 1/1 (homozygous alternate) as FUNC/FUNC and 0/1 (heterozygous) as FUNC/DEL. Number of KIR2DS4FUNC and KIR2DS4DEL variants (Figure S3) were calculated for each TCGA sample. The fully analyzed variant calls file (VCF format) for the 2504 samples is available on 1000 Genome database for the healthy control population. To check for specific 22 bp deletion on chromosome 19, the VCF file was downloaded for only chromosome 19 by using NCBI ftp server from 1000 Genome database released on May 2013. Number of KIR2DS4FUNC and KIR2DS4DEL allelic variants were extracted, calculated and used for further analysis.

Clinical data

Clinical information included in Supplementary Table SIII and SIV associated with the TCGA exome sequencing files used in this study was obtained and modified from TCGA open access database. Clinical data comprise different parameters including age, gender, survival, vital status, Karnofsky performance score. For healthy control population, only ethnicity and gender demographics were available and downloaded from 1000 Genome database for further analysis.

Supplementary References

1. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56-65.
2. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297-303.
3. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43(5):491-8.
4. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;11(1110):11 10 1-11 10 33.
5. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29(1):24-6.
6. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14(2):178-92.

Supplementary notes
Software, commands and parameters
1.0. ***Refine TCGA raw exome sequencing data with Picard 2.0.1 and GATK 3.5***
Remove duplicates, re-align around indels, re-calibrate quality scores

```
java -Xmx8g -jar picard.jar MarkDuplicates
        I=sample.sorted.bam
        O=sample.dedup.bam
        METRICS_FILE=sample.metrics
         PG=null

java -Xmx8g -jar picard.jar BuildBamIndex
        I=sample.dedup.bam

java -Xmx8g -jar GenomeAnalysisTK.jar
        -T RealignerTargetCreator -nt $cores
        -R hg19.fasta -I sample.dedup.bam
        -o sample.intervals
         -known Mills_and_1000G_gold_standard.indels.b37.vcf
        -known 1000G_phase1.indels.b37.vcf

java -Xmx8g -jar GenomeAnalysisTK.jar
        -T IndelRealigner
        -R hg19.fasta
        -I sample.dedup.bam
        -o sample.dedup.re.bam
        -targetIntervals sample.intervals
        -known Mills_and_1000G_gold_standard.indels.b37.vcf
        -known 1000G_phase1.indels.b37.vcf

java -Xmx8g -jar GenomeAnalysisTK.jar
        -T BaseRecalibrator
         -nct 4
        -R hg19.fasta
        -I sample.dedup.re.bam
        -o sample.dedup.re.recal_data.table
        -knownSites dbsnp_138.b37.vcf
        -knownSites Mills_and_1000G_gold_standard.indels.b37.vcf
        -knownSites 1000G_phase1.indels.b37.vcf

java -Xmx8g -jar GenomeAnalysisTK.jar
        -T BaseRecalibrator
        -nct 4
        -R hg19.fasta
        -I sample.dedup.re.bam
        -o sample.dedup.re.recal2_data.table
        -knownSites Mills_and_1000G_gold_standard.indels.b37.vcf
        -knownSites 1000G_phase1.indels.b37.vcf
        -knownSites dbsnp_138.b37.vcf
        -BQSR sample.dedup.re.recal_data.table
```

```
java -Xmx8g -jar GenomeAnalysisTK.jar
        -T AnalyzeCovariates
        -R hg19.fasta
        -before sample.dedup.re.recal_data.table
        -after sample.dedup.re.recal2_data.table
         -plots sample.dedup.re.recal_plots.pdf

java -Xmx8g -jar GenomeAnalysisTK.jar
        -T PrintReads
        -nct 4
        -R hg19.fasta
        -I sample.dedup.re.bam
        -BQSR sample.dedup.re.recal_data.table
        -allowPotentiallyMisencodedQuals
        -o sample.dedup.re.recal.bam
```

Variant calling whole genome with GATK 3.5

```
java -Xmx8g -jar GenomeAnalysisTK.jar
        -T HaplotypeCaller
        -R hg19.fasta
        -I sample.dedup.re.recal.bam
        -o sample.raw.snps.indels.hc.g.vcf
        -ERC GVCF
        -variant_index_type LINEAR
        -variant_index_parameter 128000
        --dbsnp dbsnp_138.b37.vcf

java -Xmx8g -jar GenomeAnalysisTK.jar
        -T GenotypeGVCFs
        -R hg19.fasta
        -V sample_1.raw.snps.indels.hc.g.vcf
        -V sample_2.raw.snps.indels.hc.g.vcf
        -V sample_3.raw.snps.indels.hc.g.vcf
        -V sample_n.raw.snps.indels.hc.g.vcf
        -o ggvcf_variants.vcf
```

Variant recalibration and filtration with GATK 3.5

```
java –Xmx8g -jar GenomeAnalysisTK.jar
        -T VariantRecalibrator
        -R hg19.fasta
        -input ggvcf_variants_sample.vcf
        -resource:hapmap,known=false,training=true,truth=true,prior=15.0
        hapmap_3.3.b37.vcf -resource:omni,known=false,training=true,truth=false,prior=12.0
        1000G_omni2.5.b37.vcf -
        resource:1000G,known=false,training=true,truth=false,prior=10.0
        1000G_phase1.snps.high_confidence.b37.vcf -
```

```
        resource:dbsnp,known=true,training=false,truth=false,prior=6.0 dbsnp_138.b37.vcf -
        an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR
        -mode SNP
        -recalFile ggvcf_variants_sample.snp.vr.recal
        -tranchesFile ggvcf_variants_sample.snp.vr.tranches
        -rscriptFile ggvcf_variants_sample.snp.vr.plots.R

java –Xmx8g -jar GenomeAnalysisTK.jar
        -T ApplyRecalibration
        -R hg19.fasta
        -input ggvcf_variants_sample.vcf
        --ts_filter_level 99.0
        -tranchesFile ggvcf_variants_sample.snp.vr.tranches
        -recalFile ggvcf_variants_sample.snp.vr.recal
        -mode SNP
        -o ggvcf_variants_sample.snp.vcf

java-Xmx8g -jar GenomeAnalysisTK.jar
        -T VariantRecalibrator
        -R hg19.fasta
        -input ggvcf_variants_sample.snp.vcf
        --maxGaussians 4 -resource:mills,known=false,training=true,truth=true,prior=12.0
        Mills_and_1000G_gold_standard.indels.b37.vcf -
        resource:dbsnp,known=true,training=false,truth=false,prior=2.0 dbsnp_138.b37.vcf -
        an QD -an DP -an FS -an SOR -an ReadPosRankSum -an MQRankSum
        -mode INDEL
        -recalFile ggvcf_variants_sample.snp.indel.vr.recal
        -tranchesFile ggvcf_variants_sample.snp.indel.vr.tranches
        -rscriptFile ggvcf_variants_sample.snp.indel.vr.plots.R

java -Xmx8g -jar GenomeAnalysisTK.jar
        -T ApplyRecalibration
        -R hg19.fasta
        -input ggvcf_variants_sample.snp.vcf
        --ts_filter_level 99.0
        -tranchesFile ggvcf_variants_sample.snp.indel.vr.tranches
        -recalFile ggvcf_variants_sample.snp.indel.vr.recal
        -mode INDEL
        -o ggvcf_variants_sample.snp.indel.vcf

java –Xmx8g -jar GenomeAnalysisTK.jar
        -T VariantFiltration
        -R hg19.fasta
        -V ggvcf_variants_sample.snp.indel.vcf
        -filter \"QD < 2.0 || FS > 60.0 || SOR > 4.0 || MQ < 40.0 || MQRankSum < -12.5 ||
        ReadPosRankSum < -8.0\"
        -filterName \"snp_filter\"
        -o ggvcf_variants_sample.snp.indel.filt_snp.vcf

java -Xmx8g -jar GenomeAnalysisTK.jar
```

```
-T VariantFiltration
-R hg19.fasta
-V ggvcf_variants_sample.snp.indel.vcf
-filter \"QD < 2.0 || FS > 200.0 || SOR > 10.0 || ReadPosRankSum < -20.0\"
 -filterName \"indel_filter\"
-o ggvcf_variants_sample.snp.indel.filt_indel.vcf
```