Supplementary information, Data S1 Materials and Methods

Clinical samples

Patients were recruited at the Department of Neurosurgery at Huashan Hospital, an affiliate of Shanghai Medical College, Fudan University from 2010 to 2015. The diagnosis of PA was based on clinical manifestations, imaging, and endocrine laboratory tests according to previously described and currently accepted standard guideline or criteria for each of the PA subtypes[1-8]. Diagnosis was confirmed histologically in blinded fashion by at least two senior pathologists after surgical resection. Invasive adenomas were defined as fulfilling 1 of 2 conditions: (1) Hardy's modified classification grade III, IV and/or stage C, D and E; (2) Knosp classification grade III and IV[9, 10]. In addition, the intra-operative findings and pathological examination were also considered as the evidence of invasive PAs, including invasion of the dura surrounding tumors, the mucosa or bone in sphenoidal sinus. The drug sensitive/resistant of somatostatin analogs such as octreotide were assessed according to 2008 Consensus statement of the Polish Society for Endocrinology[11]. Perioperative blood and tumor were collected and frozen at the time of surgery. Only samples containing more than 90% tumor cells were used in this study. Genomic DNA was extracted from each sample using standard methods. The clinical data for all patients was collected from medical records retrospectively. We included only subjects with no previous malignancies or family history of PA. All patients gave written informed consent, and the ethics committee at Huashan Hospital approved the study. The ACTH-PAs exome sequencing data in our prior study was also included in this study[12].

Exome sequencing

We used the TruSeq™ Exome Enrichment Kit (Illumina) and Agilent SureSelect Human All Exon V5+UTR Kit (Santa Clara, CA) to capture exonic region of tumor/blood samples. The HiSeq 2500 platform (Illumina, Dedham, MA) was used for deep sequencing. Raw reads from each library were trimmed with trimmomatic (http://www.usadellab.org/cms/index.php?page=trimmomatic) and mapped to the reference human genome (UCSC,hg19) using Burrows-Wheeler Aligner (BWA) 0.7.7 with default parameters[13]. PCR duplicates were flagged using Picard (http://picard.sourceforge.net) and the outputs were locally realigned by using the Indel Realignment tool of Genome Analysis Toolkit (GATK) 2.8-1[14]. The outputs were recalibrated using the BaseRecalibrator tool from GATK. Mean sequencing depth was 98x for the tumor samples and 82x for matched normal blood samples, with >10x coverage for 92.2% of target regions in tumor samples and 92.0% of target regions in normal blood samples.

Somatic single-nucleotide variants (SNVs) were detected using MuTect v1.1.7[15] and VarScan2[16], and Somatic indels were identified using VarScan 2 and Genome Analysis Toolkit Somatic indel detector[14] with default parameters. The set of all somatic variants detected by above algorithms were used for the filter process. Filter criteria for high quality variants were as follows: (1) A minimum depth of 5X in both tumors and their normal pairs; (2) Reads depth of variant alleles in tumors should be more than 4X; (3) If the reads depth of variant alleles in tumors are more than 6X, the allelic fraction should be more than 15%, otherwise, the allelic fraction in tumors should be more than 25%.

To reduce sequencing errors and artifacts, we additionally performed deep querying of the putative somatic variants against the entire normal set consisting of all normal blood samples

by using the bam-readcount tool (https://github.com/genome/bam-readcount). The somatic mutations that were detected in more than three reads in 10% or more of germline normal sets represent recurrent sequencing errors or mapping artifacts, and were subsequently removed. Somatic variants were manually inspected for accuracy by using Integrative Genomics Viewer (IGV, Broad Institute, Cambridge, MA, USA).

All high quality somatic variants were annotated based on the information available in catalog of somatic mutations in cancer (COSMIC; http://cancer.sanger.ac.uk/cosmic), dbSNP138 (http://www.ncbi.nlm.nih.gov/snp), the 1,000 Genomes Project (http://www.1000genomes.org/), and EXAC by Annovar[17] (http://exac.broadinstitute.org/). Variants with allele frequencies that were >0.1% of the allelic fraction were removed. For the recurrent ones, we focused on genes with a mutation rate greater than 50 mutations per MB among either all PAs or any subtype of PA. Of 200 somatic mutations that we randomly selected, over 90% were successfully validated using Sanger sequencing. The R package SomaticSignatures was used to identify mutational signatures in our study.

Identification of SCNAs

SCNAs analysis was performed by using exome SCNA with default parameters[18], which achieves high levels of sensitivity and specificity to predict copy number alterations in WES tumor-normal paired samples. Recurrent SCNA regions were identified by GISTIC2.0[19] with join segment size=4 and confidence level=0.9, Significance threshold for q-values=0.25, Maximum number of segments allowed for a sample=4200, amplifications_threshold =0.1, deletions_threshold=0.1, which uses amplitude of the SCNA aberration and frequency of its

occurrence across samples to identify the boundaries of aberrations. Genomic disruption level was calculated by the sum of copy number aberration (not equal to 2) from exome CNV results, and then divide by the sum of all segments region for each sample.

Pathway Enrichment Analysis

For pathway enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.jp/) database was used and a hypergeometric test was carried out. For each tumor subtype, all of the pathways contain N genes (including n mutated genes) and a pathway contains M genes (including k mutated genes, k>=2), the probability p of observing at least k mutated genes in this pathway by chance is:

$$p = 1 - \sum_{i=0}^{k-1} \frac{C_n^i C_{N-n}^{M-i}}{C_N^M}$$

Two pathway enrichment analysis approaches were undertaken. First, based on our mutation data combining previous whole exome sequence data[20-23], the mutated genes were divided into seven groups (GH, GT, TSH, ACTH, PRL, Nonfunctioning and Plurihormonal). Second, all mutated genes were divided into two groups (invasive and non-invasive samples). In each approach, the enrichment analysis by hypergeometric test was calculated. If mutated genes in each group were significantly enriched (P<0.05) in some pathways, the group name and pathways were connected. Furthermore, the relationship between two subtypes was quantified and then the results were visualized in Cytoscape 3.4.0(http://www.cyto scape.org/).

Drug target analysis

Candidate cancer genes were identified using the Cosmic V70 Cancer Gene census (http://cancer.sanger.ac.uk/cosmic) and 624 Cancer Associated Genes[24]. All such cancer genes mutated in PAs in this study were then mapped to the KEGG database (http://www.genome.jp/kegg/pathway.html). Seven major pathways including more than 2 mutant cancer genes were identified: cAMP signaling, cell cycle, PI3K-Akt signaling, immune response signaling (including NF-kB signaling and Epstein Barr Virus infection-related pathways), MAPK signaling, endocrine signaling (including GnRH signaling, thyroid hormone synthesis, diabetes signaling, and carcinoid in neuroendocrine tumor related pathways) and Rap1 signaling pathways. Genes in the different pathways were displayed in Figure 1D according to KEGG pathway database. The FDA-approved drugs or drugs screened in phase I to phase III clinical trials that target one of the genes in these pathways were based on Drugbank database V4.5 (http://www.drugbank.ca/).

Statistical analysis

All the statistical tests were carried out in R (https://www.r-project.org/). We estimated that 125 samples would be needed to detect a gene mutated at a frequency of ~6% mutation rate among 125 samples with a power of 80% and α=5%. We also estimated that 20 samples would be needed to detect a gene mutated in ~30% of samples. Power analysis estimates were conducted using a binomial power model [25]. Mann-Whitney U-test was used for comparisons of tumor size. Fisher's exact test was used to compare of invasiveness and drug resistance among WT and mutant tumors in a group.

Additional acknowledgements

We thank and acknowledge all of the participants in the project. We gratefully thank Professor Lu Xie, Professor Chao-Wei Fu and Dr. Dai-Zhan Zhou for helpful discussions. We thank Professor Rong Xie, Dr. Ming Shen, Dr. Zhao Ye, Dr. Jian-Ping Song, Dr. Wen-Qiang He, Dr. Yi-Chao Zhang, Dr. Ni-Dan Qiao, Dr. Zhi-Feng Shi, Mrs. Ye Wang and Mrs. Qiu-Wei Song for the sample and clinical data collection. We appreciate Professor Yi-Ming Li, Professor Zhao-Yun Zhang, Professor Hong-Ying Ye, Professor Bin Lu, Professor Hong Chen, Professor Yin Wang, Professor Feng Tang, Professor Dao-Ying Geng, Professor Zhen-Wei Yao, Professor Wei-Jun Tang and Professor Xing-Dang Liu for the assistance in clinical data evaluation. We also thank Professor Yun Lu, Dr. Hong Peng, Dr. Xue-Min Jian, Mrs. Xiao-Lan Qin and Ms. Jie Zheng for their technical assistance.

Reference

1 Bancos I, Natt N, Murad MH, Montori VM. Evidence-based endocrinology: illustrating its principles in the management of patients with pituitary incidentalomas. *Best Pract Res Clin Endocrinol Metab* 2012; **26**:9-19.

2 Beck-Peccoz P, Lania A, Beckers A, Chatterjee K, Wemeau JL. 2013 European thyroid association guidelines for the diagnosis and treatment of thyrotropin-secreting pituitary tumors. *Eur Thyroid J* 2013; **2**:76-82.

3 Freda PU, Beckers AM, Katznelson L *et al.* Pituitary incidentaloma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2011; **96**:894-904.

4 Katznelson L, Laws ER, Jr., Melmed S *et al.* Acromegaly: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2014; **99**:3933-3951.

5 Mayson SE, Snyder PJ. Silent (clinically nonfunctioning) pituitary adenomas. *J Neurooncol* 2014; **117**:429-436.

6 Melmed S, Casanueva FF, Hoffman AR *et al.* Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2011; **96**:273-288.

7 Nieman LK, Biller BM, Findling JW *et al.* Treatment of Cushing's Syndrome: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2015; **100**:2807-2831.

8 Ntali G, Capatina C, Grossman A, Karavitaki N. Clinical review: Functioning gonadotroph adenomas. *J Clin Endocrinol Metab* 2014; **99**:4423-4433.

9 Knosp E, Steiner E, Kitz K, Matula C. Pituitary adenomas with invasion of the cavernous sinus space: a magnetic resonance imaging classification compared with surgical findings. *Neurosurgery* 1993; **33**:610-617; discussion 617-618.

10 Wilson CB. A decade of pituitary microsurgery. The Herbert Olivecrona lecture. *J Neurosurg* 1984; **61**:814-833.

11 Bolanowski M, Bar-Andziak E, Kos-Kudla B *et al.* Consensus statement of the Polish Society for Endocrinology: presurgical somatostatin analogs therapy in acromegaly. *Neuro Endocrinol Lett* 2008; **29**:59-62.

12 Ma ZY, Song ZJ, Chen JH *et al.* Recurrent gain-of-function USP8 mutations in Cushing's disease. *Cell Res* 2015; **25**:306-317.

13 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**:1754-1760.

14 McKenna A, Hanna M, Banks E *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**:1297-1303.

15 Cibulskis K, Lawrence MS, Carter SL *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**:213-219.

16 Koboldt DC, Zhang Q, Larson DE *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; **22**:568-576.

17 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**:e164.
18 Amarasinghe KC, Li J, Hunter SM *et al.* Inferring copy number and genotype in tumour exome data. *BMC Genomics* 2014; **15**:732.
19 Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011; **12**:R41.
20 Newey PJ, Nesbit MA, Rimmer AJ *et al.* Whole-exome sequencing studies of nonfunctioning pituitary adenomas. *J Clin Endocrinol Metab* 2013; **98**:E796-800.
21 Reincke M, Sbiera S, Hayakawa A *et al.* Mutations in the deubiquitinase gene USP8 cause Cushing's disease. *Nat Genet* 2015; **47**:31-38.
22 Ronchi CL, Peverelli E, Herterich S *et al.* Landscape of somatic mutations in sporadic GH-secreting pituitary adenomas. *Eur J Endocrinol* 2016; **174**:363-372.
23 Valimaki N, Demir H, Pitkanen E *et al.* Whole-Genome Sequencing of Growth Hormone (GH)-Secreting Pituitary Adenomas. *J Clin Endocrinol Metab* 2015; **100**:3918-3927.
24 Lu C, Xie M, Wendl MC *et al.* Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun* 2015; **6**:10086.
25 Lawrence MS, Stojanov P, Mermel CH *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014; **505**:495-501.