Supplementary Appendix

Table of contents

Supplementary Methods

Patient selection

Patients aged 3-60 years old with IPAH were included. IPAH diagnosis was made by at least 2 experienced PAH experts. Patients with known causes of PAH listed in the classification of the current guideline were excluded, including HHT, drug-induced PAH, associated PAH, and the positive responders in acute pulmonary vasoreactivity testing. Furthermore, all diagnosed IPAH cases were sporadic without first or second-degree relatives who had been diagnosed with PAH or died from a PAH-like syndrome. All patients were of Chinese Han origin and provided written informed consent for DNA analysis in accordance with the institutional review board requirements at the time of collection.

The discovery stage consisted of 251 unrelated IPAH patients and the control group was a random sample of 1,884 subjects who did not have any chronic disease from the Novo-Zhonghua project. All the control samples were also Han Chinese ethnicity as confirmed by identity-by-descent analysis. A total of another incident 80 cases were enrolled in the replication stage. The control exonic sequences of 8,624 East Asian (EAS) were taken from the Genome Aggregation Database (GnomAD) of Broad Institute (http://gnomad.broadinstitute.org/).

DNA sequencing

Genomic DNA samples were either whole-exome sequencing (WES) or whole-genome sequencing (WGS) according to standard protocols. The whole-genome sequencing was performed at WuXi Apptech (Shanghai, China) using a PCR-free library preparation and paired-end (2 x 150 bp) sequencing on the HiSeq X Ten platform (Illumina, Inc., San Diego, CA, USA) to yield a minimum of 35 x coverage. The whole exome sequencing was performed at Novogene

Corporation (Beijing, China) using the Agilent SureSelect Human All Exon kit (50 MB or 60 MB) and sequencing on the Illumina HiSeq instruments to provide a mean coverage of more than 100 × on the target regions of every sample. Prior to sequencing, samples were randomized to minimize batch effects.

The exome data of case and control samples were processed at Novogene Corporation. The quality control procedures were applied to the fastq files to discard any sequence artifacts and contamination, including: (1) read pairs with adapter contamination (>10 nucleotides aligned to the adapter sequences, allowing ≤10% mismatches). (2) read pairs containing > 10% uncertain nucleotides, (3) read has more than 50 percent low quality (Phred quality<5) nucleotides. Then, the fastq files were aligned to the Human Reference Genome (hs37d5) using the Burrows-Wheeler Aligner (version 0.7.8). After the BAM files were sorted and indexed sequentially, Picard (version 1.111) was used to merge the BAM files and to mark PCR and optically duplicated reads for downstream analyses. We used SAMtools (version 1.0) and BCFtools (version 1.0) to simultaneously identify multiallelic single nucleotide variants (SNVs) and small insertions/deletions (InDels) for each subject. The raw calls of SNVs and InDels were further filtered with the following inclusion thresholds: 1) read depth > 4; 2) Root-Mean-Square mapping quality of covering reads > 30; 3) the variant quality score > 20, and merged separately using VCF tools (version 0.1.12b). The variants were annotated with ANNOVAR. The rare copy number variants (CNVs) from the WES data were detected using the SVD-ZRPKM algorithm CoNIFER (version 0.2.2). The homogenous population structure of cases and controls was defined genetically by means of principle-component analysis of common variants that were sequenced.

We used two sequencing platforms (WES and WGS) to genotype the IPAH cases. To reduce systematic errors and alignment artifacts of the two sequencing platform, we restricted our call for

variants to RefSeq coding sequences and called WGS together with WES using very strict criterion. In both WES and WGS, the calling process targeted only regions covered by the Agilent Sure Select Human All Exon kit V6 kit (all exons + 100 bp flanking each exon). We then performed statistical analysis for the matched and mismatched genotype and filtered out variants with a mapping quality < 30, or a genotype quality < 20 or minor-read ratio < 20%.

Gene-based burden analysis

As the IPAH is a rare disease, we focused on rare variants that may have the large effect on the disease. To enrich variants that were likely to alter the function of the proteins, we defined "qualifying variant" as deleterious, rare exonic and splice site variants with a call rate > 90%. Deleterious variants were limited to missense, nonsense, splice sites, frameshift that was not located in the segmental duplication regions or repeat region defined by RepeatMasker. Meanwhile, the deleteriousness score was computationally predicted by the following methods: SIFT, PolyPhen-2, MutationTaster, CADD. Variants were classified as probably deleterious if the variants were predicted by at least 2 software tools to be "damaging" or "probably damaging". The conservation of each variant was scored by GERP++. Rare variants were filtered using a minor allele frequency (MAF) threshold of 1% in 1000 Genomes and 0.1% in ExAC East Asian.

We then ran an exome-wide rare variant burden (RVB) analysis by coding individuals based on the presence or absence of rare deleterious variants in each sequenced gene and compared the combined frequency of rare variants in each gene between cases and controls with two-tailed Fisher's exact test in R statistical package. The association P values were corrected by several multiple-testing correction methods including Bonferroni correction and Holm-Bonferroni method. False discovery rates were also calculated by Benjamini-Hochberg procedure and Benjamini-Yekutieli procedure.

Measurement of mature BMP9

The mature BMP9 in blood or conditioned media samples were measured using a Sandwich Enzyme-linked Immunosorbent Assay (ELISA) Development kit (DY3209, R&D Systems, Minneapolis, MN). Baseline venous blood was obtained from subjects at enrollment after fasting overnight (> 12 hours). Written informed consent was obtained from all individuals. All samples were collected into sodium ethylene diamine tetra-acetic acid (EDTA) tubes. After centrifugation at 3000 × g and 4 °C for 15 minutes, the supernatants were frozen in aliquots and stored in cryotubes at -80 °C until assayed. The procedure was completed within 20minutes. According to the manufacturer's protocol. The ELISA recognized the mature peptide using an anti-human mature BMP9 antibody as a capture antibody and a biotinylated polyclonal anti-human mature BMP9 protein as a detection antibody. The inter-assay coefficient of variation was 8%, and the intra-assay coefficient of variation was 5%. All assays were replicated in 3 times.

Construction of *BMP9* wild-type and mutation plasmid

For the six mutant plasmid construction, human wild-type (WT) *BMP9* cDNA (1290bp, NM_016204, Tianyi Huiyuan, China) was subcloned into the *EcoRI/BamHI* site of the pcDNA3.1 plasmid. The six mutant plasmids (V109L, S282fs, R316S, S320C, A353T, and V423M) were constructed based on the WT plasmid using the QuickChange Mutagenesis Kit (Stratagene, La Jolla, CA, USA). All constructs were sequenced to confirm the integrity and the presence of mutations. All plasmids were transfected into cells using Lipofectamine 3000 reagent (Invitrogen). HEK-EBNA cells were seeded in 6-cm dishes and transfected with the plasmids containing either wild-type or six mutants of *BMP9* for 24 h. The supernatant was removed and replaced with CDCHO media (Invitrogen, MA, USA) for 48 h. The conditioned media and cell lysis were collected and aliquoted. After quick frozen in liquid nitrogen, the samples were stored at -80 °C

until used.

Western blotting

The conditioned media or the cell lysis that harvested from HEK-EBNA cells transfected with wild-type or mutant *BMP9* plasmids were separated on non-reducing NuPAGE (Invitrogen) and transferred to polyvinylidene fluoride (PVDF) membranes. Membranes were then blocked with 5% BSA and immunoblotted for the anti-human BMP9 antibody (MAB3209, R&D Systems, Minneapolis, MN) or the BMP9 propeptide antibody (AF3879, R&D Systems, Minneapolis, MN). All blots were re-probed with a monoclonal antibody to HSP70 (bs-0244R, BioSS).

BRE luciferase assay

The BMP9 concentration of human recombinant BMP9, wilt-type conditional medium, and mutant conditional medium were adjusted to 1 pg/ml and 10 pg/ml, respectively. C2C12 cells were seeded in 96-well plate and transfected with 0.1 ug of pGL3(BRE)2-luc, 0.02 ug of pRL-TK-luc and 0.01 ug of pALK1. Four hours later, the cells were washed twice with serum-free DMEM and incubated in serum-free DMEM for 10 h. After treated with human recombinant BMP9, wilt-type conditional medium or mutant conditional medium for 16 h, cells were harvested and assessed the firefly and *Renilla* luciferase activities using the Dual-Glo luciferase assay kit (Promega). Results are calculated as ratios of firefly activity to the *Renilla* activity.

Caspase-Glo 3/7 Activity Assay

PAECs were pre-treated with human recombinant BMP9 (resolved in CDCHO), WT or the mutant condition medium at 5 ng/mL for 16 h. The PAECs were then stimulated with 10 ng/ml tumor necrosis factor-α and 20 μg/ml cycloheximide for 6 h to induce apoptosis. The activity of caspase 3/7 was determined using the Caspase-Glo 3/7 Assay kit (Promega) according to the manufacturer's instruction.

Table S1. Next-generation sequencing data characteristics for IPAH patients.*

Cohort	NGS type	n	Raw data (G)	Q30(%)	Mappable-reads-coverage (x)	Mapping rate (%)	High-quality reads >20x(%)	High-quality reads >10x(%)
Discovery Cohort	WES	215	12.2	91.1	128.5	99.8	98.3	99.4
	WGS	36	115.2	84.7	37.3	99.0	92.4	98.0
Replication Cohort	WES	80	12.3	91.4	123.2	99.9	98.4	99.4

*: NGS denotes next-generation sequencing; WES, whole exome sequencing which was performed in Novogene Corporation. WGS, whole genome sequencing which was performed in WuXi NextCODE Corporation; IPAH, idiopathic pulmonary arterial hypertension.

Table S2. Summary of IPAH cases carrying the rare variants or mutations in the 10 previously reported risk genes.*

Gene	Discovery cohort (*n* = 251)	Replication cohort (*n* = 80)	Combined cohorts (*n* = 331)
BMPR2, no. (%)	49 (19.5)	25 (31.3)	74 (22.4)
ACVRL1, no. (%)	15 (6.0) [†]	6 (7.5)	21 (6.3)
BMPR1B, no. (%)	1 (0.4)	0 (0.0)	1 (0.3)
KCNK3, no. (%)	1 (0.4)	1 (1.3)	2 (0.6)
SMAD9, no. (%)	1 (0.4)	1 (1.3)	2 (0.6)
CAV1, no. (%)	0 (0.0)	1 (1.3)	1 (0.3)
ENG, no. (%)	0 (0.0)	1 (1.3)	1 (0.3)
SMAD1, no. (%)	2 (0.8)	0 (0.0)	2 (0.6)
TBX4, no. (%)	10 (4.0)	2 (2.5)	12 (3.6)
EIF2AK4, no. (%)	4 (1.6)	0 (0.0)	4 (1.2)

BMPR2 & *ACVRL1,* no. (%)	1 (0.4) ‡	0 (0.0)	1 (0.3)
BMPR2 & *BMPR1B,* no. (%)	3 (1.2) §	0 (0.0)	3 (0.9)
BMPR2 & *SMAD9,* no. (%)	2 (0.8) **	0 (0.0)	2 (0.6)
BMPR2 & *TBX4,* no. (%)	1 (0.4) ††	0 (0.0)	1 (0.3)
KCNK3& TBX4, no. (%)	1 (0.4) ‡‡	0 (0.0)	1 (0.3)
Total	91 (36.3)	37 (46.3)	128 (38.7)

*: The number represents IPAH cases carrying the rare variants or mutations in the 10 PAH risk genes. †: One patient carries one truncating mutation (p.R479X) and one rare missense variant (p.V99M) in *ACVRL1*. ‡: One patient carries one splice-site mutation in *BMPR2* (c.853-2A>G) and one rare missense variant in *ACVRL1* (p.R469Q). §: One patient carries one missense mutation in *BMPR2* (p.R491Q) and one rare missense variant in *BMPR1B* (p.F118L), one patient carries one frameshift mutation in *BMPR2* (p.M383fs) and one rare missense variant in *BMPR1B* (p.E246Q), one patient carries one frameshift mutation in *BMPR2* (p.A35fs) and one rare missense variant in *BMPR1B* (p.V140F), **: One patient carries one truncated mutation in *BMPR2* (p.R213X) and one rare variants in *SMAD9* (p.G300E); one patient carries one missense mutation in *BMPR2* (p.C123R) and one rare variants in *SMAD9* (p.P459T) . †:One patient carries one splice-site rare variant in *BMPR2* (c.76+1G>C) and one rare variant in *TBX4* (p.P180S). ‡‡:One patient carries double rare variants in *KCNK3* (p.R390H) and *TBX4* (p.G342C).

1

2

Table S3 Clinical characteristics of the 22 cases with the BMP9 mutations.

Cohort*	Case†	Gender	Age at diagnosis (years)	Mutation	Protein alternation	BMP9 Level (pg/mL)	mPAP (mm Hg)	CI (l/min/m^2)	PVR (Wood units)	Survival, months‡
1	WSY[a]	F	6	c.844_845del	p.S282fs	10.6	85	4	36.1	108, Dead
1	CHG[a]	M	29	c.G948T	p.R316S	N/A	59	2.2	13.5	93, Dead
1	QXF[a]	F	27	c.C405A	p.F135L	10.9	39	1.7	10.9	112
1	MSH[a]	F	40	c.A958T	p.S320C	N/A	43	3.1	7.4	140, Dead
1	CY[a]	M	32	c.G1057A	p.A353T	11.1	57	2.4	10	2, Dead
1	RHJ[a]	M	32	c.G1267A	p.V423M	5.2	72	1.9	20.7	55, Dead
1	WPY[a]	F	22	c.G1263A	p.M421I	N/A	50	2.2	12.8	69
1	SHM[b]	F	37	c.T1016G	p.I339S	7.1	51	1.5	18.9	52
1	LSH[b]	F	44	c.G329A	p.R110Q	19.8	50	2.2	11.3	48
1	ZP[b]	F	40	c.A776G	p.N259S	26.6	101	2.4	21	41
1	MCY[b]	F	37	c.C328T	p.R110W	13.6	45	1.8	10.6	38
1	FYH[b]	F	49	c.G964A	p.G322R	22.7	71	2	16.4	25, Dead
1	WMY[b]	F	18	c.G325T	p.V109L	37.2	76	2.9	14.7	35
1	FYY[b]	F	24	c.T590C	p.I197T	13.4	71	2	17.6	22
1	LJM[b]	M	10	c.T614A	p.L205X	21.8	74	2.9	13.4	27
1	GJN[&b]	F	3	c.G1267A	p.V423M	8.9	N/A	N/A	N/A	24, Dead

1	LX[b]	F	28	c.T455C	p.L152P	12.5	47	1.7	11.8	22
2	YLL[b]	F	30	c.G1120C	p.A374P	30.9	64	2.5	13.9	21
2	CYF[b]	F	28	c.1062delC	p.Y354X	17.5	47	2.3	10.6	20
2	NPP[b]	F	24	c.G3A	p.M1I	26.2	55	2.5	12.3	19
2	WLL[b]	F	29	c.881delA	p.E294fs	6.3	88	2.5	18.6	17
2	WBM[b]	F	26	c.C232T	p.Q78X	12.1	47	3.3	5.4	9

*: Cohort: 1 denotes the discovery cohort, 2 denotes the replication cohort. †: "a" denotes patients from Shanghai, and "b" denotes patients from Beijing. &: Patient GJN is a pediatric patient. PAH was diagnosed by two PAH experts based on echocardiography. ‡: The survival time from diagnosis to death. mPAP denotes mean pulmonary artery pressure; PVR, pulmonary vascular resistance; CI, cardiac index; N/A, not available. M, male; F, female.

Table S4. The 21 distinctive *BMP9* mutations found in IPAH patients.

Chromosome position	Coding DNA change*	Exonic Function	Protein change†	SIFT, Polyphen2, MutationTaster, CADD‡	gerp++gt2§	Allele frequency in 1000G_Chinese** (*n* = 301)	Allele frequency in ExAC_EAS†† (*n* = 4327)	Allele frequency in Novo-Zhonghua (*n* = 1884)
Chr10: 48416691	c.G3A	missense	p.M1I	D, P, D, 22.8	5.7	Absent	Absent	Absent
Chr10: 48416462	c.C232T	stopgain	p.Q78X	NA, NA, NA, 11.1	5.4	Absent	Absent	Absent
Chr10: 48416369	c.G325T	missense	p.V109L	D, P, D, 28.1	5.4	Absent	Absent	Absent
Chr10: 48416366	c.C328T	missense	p.R110W	D, D, D, 35.0	4.4	Absent	Absent	Absent
Chr10: 48416365	c.G329A	missense	p.R110Q	D, D, D, 35.0	5.4	Absent	Absent	Absent
Chr10: 48414463	c.C405A	missense	p.F135L	D, D, D, 25.6	2.7	Absent	Absent	Absent
Chr10: 48414413	c.T455C	missense	p.L152P	D, D,D, 26.0	5.6	Absent	Absent	Absent
Chr10: 48414278	c.T590C	missense	p.I197T	D, P, D, 24.4	5.7	Absent	Absent	Absent
Chr10: 48414254	c.T614A	stopgain	p.L205X	NA, NA, D, 40.0	5.7	Absent	Absent	Absent

Chr10: 48414092	c.A776G	missense	p.N259S	D, P, D, 23.8	5.3	Absent	Absent	Absent
Chr10: 48414022	c.844_845del	frameshift deletion	p.S282fs	NA, NA, NA, NA	NA	Absent	Absent	Absent
Chr10: 48413986	c.881delA	frameshift deletion	p.E294fs	NA, NA, NA, NA	NA	Absent	Absent	Absent
Chr10: 48413920	c.G948T	missense	p.R316S	T, D, D, 12.5	NA	Absent	Absent	Absent
Chr10: 48413910	c.A958T	missense	p.S320C	D, D, D, 23.3	5.5	Absent	Absent	Absent
Chr10: 48413904	c.G964A	missense	p.G322R	D, D, N, 23.1	4.6	Absent	0.0001	0.0002
Chr10: 48413852	c.T1016G	missense	p.I339S	D, D, D, 24.3	5.3	Absent	Absent	Absent
Chr10: 48413811	c.G1057A	missense	p.A353T	D, P, D, 24.0	4.7	Absent	Absent	Absent
Chr10: 48413805	c.1062delC	stopgain	p.Y354X	NA, NA, NA, NA	NA	Absent	Absent	Absent
Chr10: 48413748	c.G1120C	missense	p.A374P	D, D, D, 26.9	5.6	Absent	Absent	Absent
Chr10: 48413605	c.G1263A	missense	p.M421I	D, D, D, 28.1	5.5	Absent	Absent	Absent
Chr10: 48413601	c.G1267A	missense	p.V423M	D, D, D, 32.0	5.5	Absent	Absent	Absent

*: Abbreviations are in accord with nomenclature guidelines as recommended by the Human Genome Variation Society (http://varnomen.hgvs.org). The letter c. is used to indicate coding DNA, where nucleotide 1 is the A of the ATG translation initiation codon.

†: p. is used to indicate the change at the protein level.
‡: A CADD score of >15 indicates deleteriousness for the variant; T, tolerate (not considered deleterious by SIFT); P, possibly deleterious; D, deleterious; NA, not available.
§: The larger the score, the more conserved the site.
**: The minor allele frequency in the Chinese cohort of the 1000 Genomes dataset (www.1000genomes.org). This cohort includes 2 populations from China: CHS (105 individuals from China South) and CHB (103 individuals from Beijing).
††: The minor allele frequency in the database of East Asia in Exome Aggregation Consortium (ExAC).

Table S5. Clinical characteristics of the patients with or without *BMP9* mutations.*

	***BMP9* mutation Carriers n = 22**	**IPAH patients without *BMP9* mutation n = 309**
Female, no. (%)	18 (81.8)	240 (77.7)
Age at diagnosis (yr)	28.7 ± 11.7	28.0 ± 11.0
Right Atrium Pressure (mm Hg)	9.8 ± 6.6	8.8 ± 6.3
Mean Pulmonary Artery Pressure (mm Hg)	62.1 ± 16.2	62.6 ± 15.1
Pulmonary Artery Wedge Pressure (mm Hg)	9.2 ± 3.2	9.2 ± 3.7
Cardiac Index (L/min/m^2)	2.4 ± 0.6	2.5 ± 0.9
Pulmonary Vascular Resistance (Wood Units)	14.9 ± 6.4	15.5 ± 7.8
Mixed Venous Oxygen Saturation (%)	61.3 ± 8.0	64.0 ± 10.5

*: Plus-minus values are means ±SD. There were no significant differences between the patients with or without BMP9 mutations in any of the characteristics listed.

Table S6. The demographic characteristics of healthy controls, IPAH patients with or without *BMP9* mutations.

Demographic characteristics	NC	IPAH without *BMP9* mutation	IPAH with *BMP9* mutation
N	87	38	19
Age, years	31.9±7.0	28.0±11.8	29.1±12.5
Female, (%)	62 (71.3%)	31 (81.6%)	16 (84.2%)

Table S7. The frequencies of rare variants of *ATP13A3, AQP1, and SOX17* in the combined cohort.

Gene	Number of cases*	Frequency of cases†	Number of controls*	Frequency of controls†	*P*‡	Adj. *P*§
AQP1	8	0.024	78	0.007	0.006	1
ATP13A3	7	0.021	117	0.011	0.107	1
SOX17	4	0.012	31	0.003	0.025	1

*: The numbers denote subjects carry rare deleterious variants in the cases or controls.
†: The prevalence of the mutation carrier in the cohort.
‡: The raw *P* value without correction for comparing the alleles between case and control by two-tailed Fisher's exact test.
§: The adjusted *P* value after Bonferroni correction for comparing the alleles between case and control.

Figure S1. The workflow of exome-wide rare variant burden analysis.

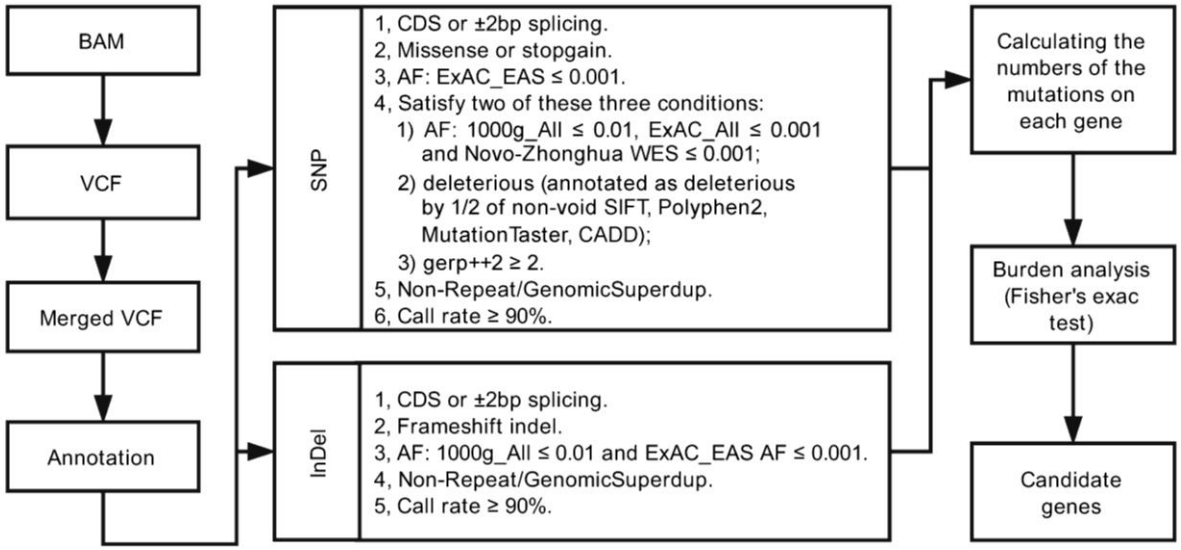
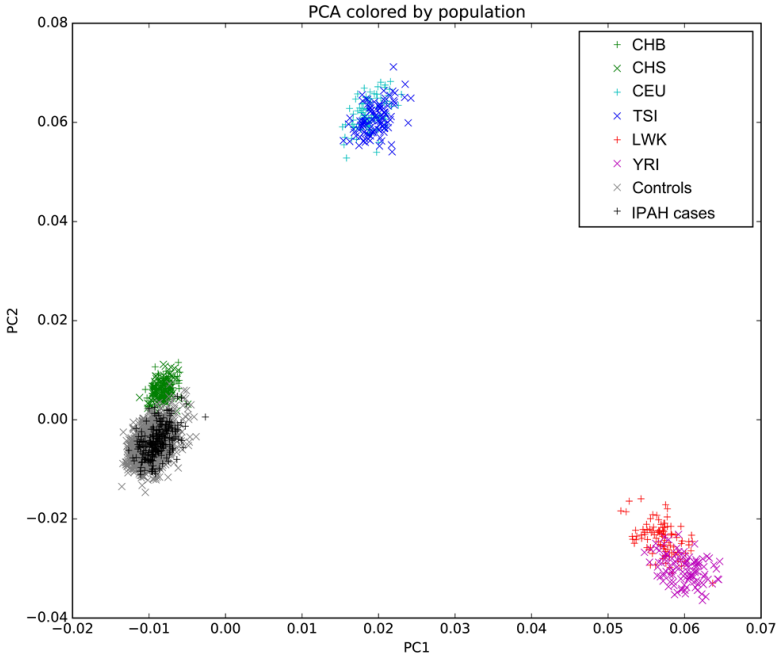
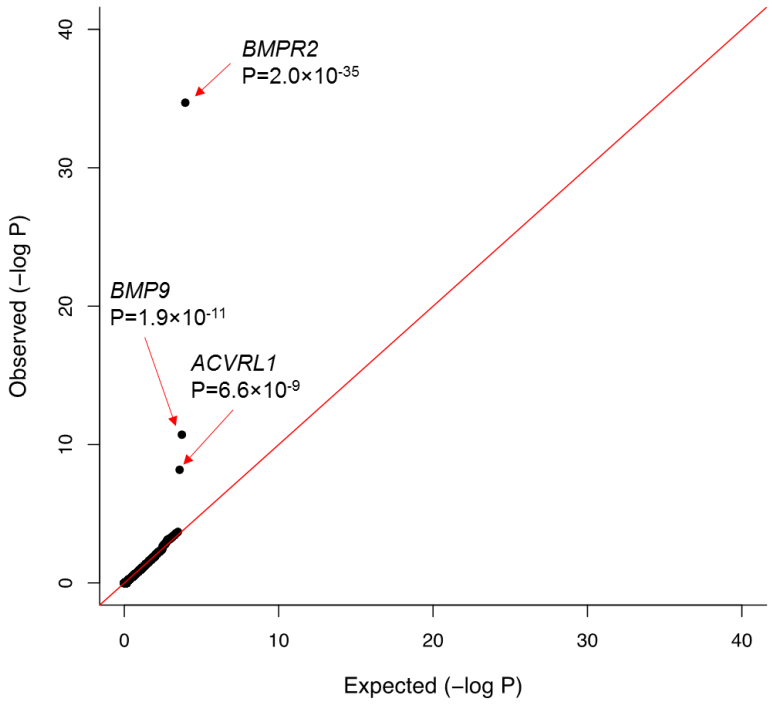


Figure S2. Principal component analysis for ancestry of cases and controls in the discovery cohort.



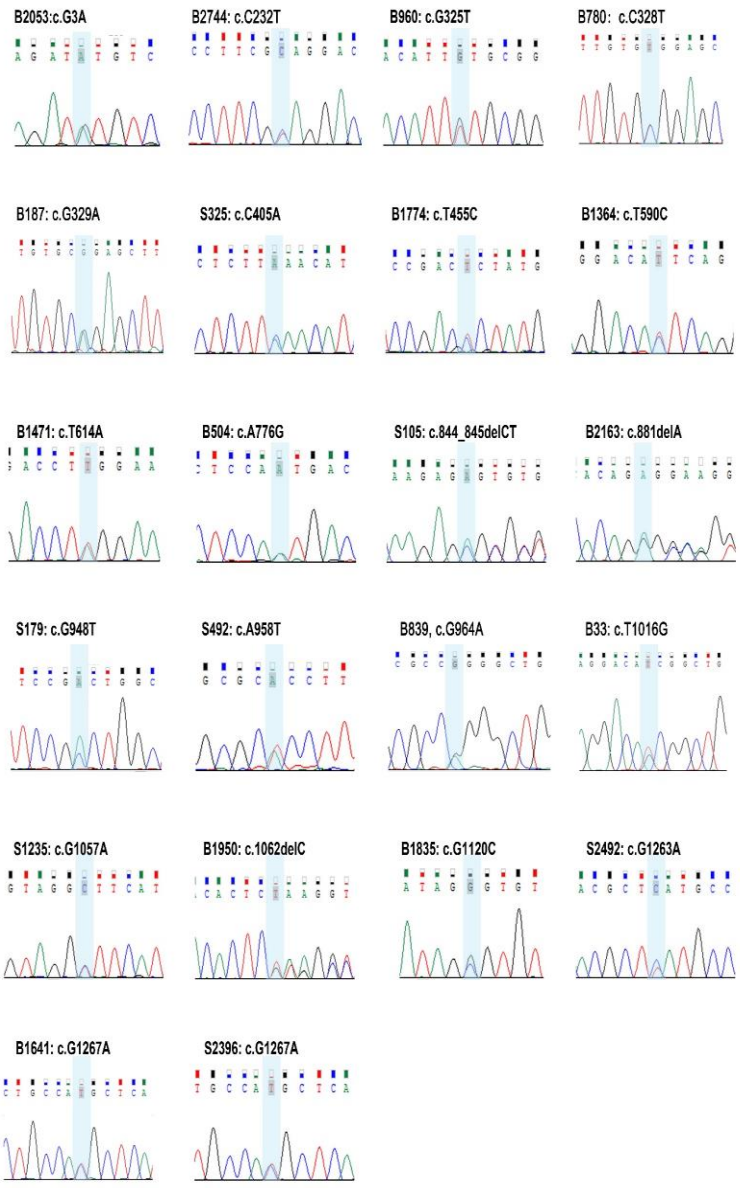
Principal component analysis was performed by projecting exome-sequencing samples onto 1000 genome samples using Plink. Results demonstrate the presence of 3 groups, corresponding to East Asian descent (in green), Caucasian descent (in blue) and African descent (in red). Both the cases and controls from the discovery cohort were clustered together with East Asian descent. The CHB denotes Han Chinese in Bejing, China. (N=103). CHS denotes Southern Han Chinese (N=105). CEU denotes Utah Residents with Northern and Western European Ancestry (N=99). TSI denotes Toscani in Italia (N=107). LWK denotes Luhya in Webuye, Kenya (N=99). YRI denotes Yoruba in Ibadan, Nigeria (N=108). Cases denote 251 patients with idiopathic pulmonary arterial hypertension. Controls denote 1,884 normal controls from Novo-Zhonghua project.

Figure S3. Quantile-quantile plot of discovery cohort for the dominant coding model.



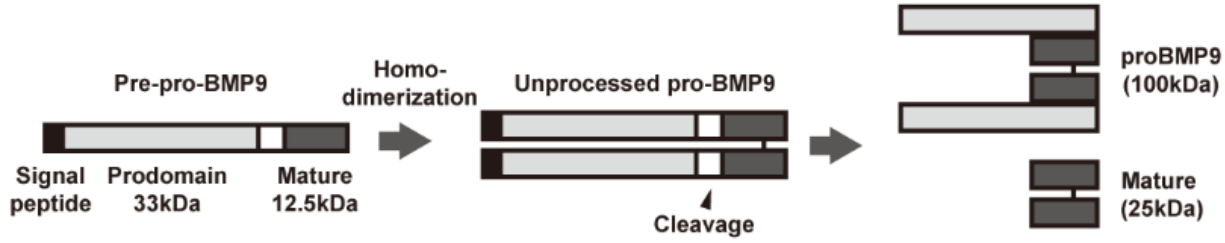
Results are shown for the burden analysis of rare variants in protein-coding genes in the discovery stage. This test included a total of 13,318 genes that had more than one case carrier with the rare deleterious variant under the dominant model. The distribution of observed P values for each gene was compared to the distribution of expected P values. The 3 genes passed correction for multiple tests were labeled.

Figure S4. Verification of the *BMP9* mutations in IPAH cases by Sanger sequencing.



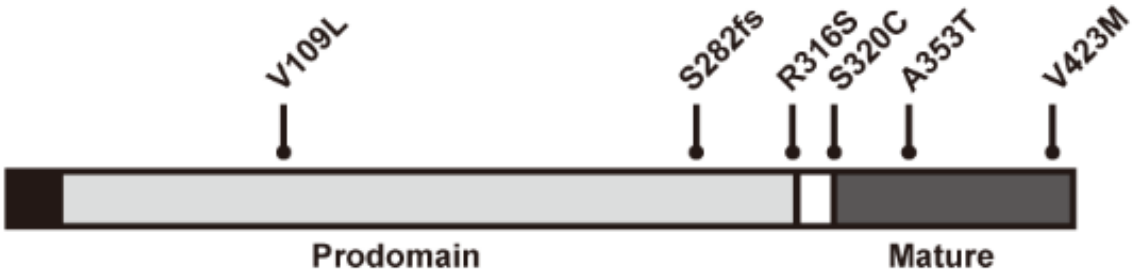
The genetic mutations were highlighted by blue color**.**

Figure S5 The schematic diagram for BMP9 biosynthesis and processing.



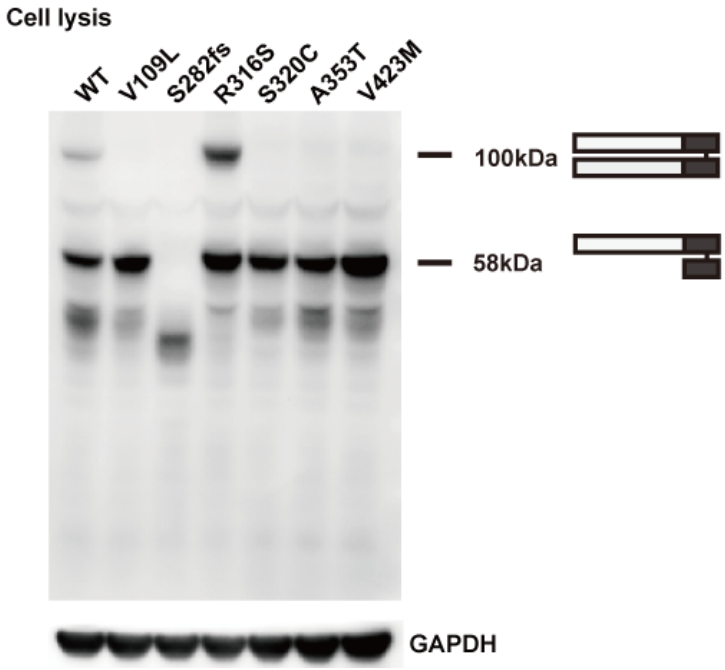
The pre-pro-BMP9 precursor contains a signal peptide (amino acids 1-22), a prodomain (amino acids 23-319), and the mature BMP9 (amino acids 320-429). The proBMP9 and mature BMP9 are both the active forms.

Figure S6. The locations of 6 *BMP9* mutations for functional study.



Six *BMP9* mutations were selected for functional study, including 1 truncating mutation in the prodomain (S282fs), 2 missense mutation in the prodomain (V109L), 2 missense mutations in the cleavage site (R316S, S320C), and 2 missense mutation in the mature domain (A353T, V423M).

Figure S7. The expression pattern pro-BMP9 for wild-type *BMP9* and 6 mutants in HEK cells.



Protein blot analysis of HEK cells transfected with plasmids for either the wild-type *BMP9* or the six *BMP9* mutants using an antibody to the pro-BMP9. Wild-type BMP9 and mutant protein fractioned as 2 bands corresponding to the proBMP9 (100kDa), partially processed proBMP9 (58kDa). GAPDH served as an internal control.