Best val_loss vs model size (orig + wide)